ma
the
ma
tisch

cen
trum

amsterdam   1974

GENETICS OF REPRODUCING AUTOMATA *)

Paul M.B. Vitányi

Summary

Sexual reproduction is modeled and investigated in
the formal framework of John von Neumann's theory of
self-reproducing cellular automata. It is argued that
the transition from asexual to sexual reproduction ne-
cessitates a change in number and structure of the ge-
netic types involved. It is shown that the recombination
of the parents' characteristics in the offspring close-
ly conforms to nature. Similarities with biological sys-
tems is discussed and e.g. a concrete hypothesis on a
X-Y mechanism for the physiology of sexual processes is
presented.

1. Introduction

Abstract automata are information processing dis-
crete parameter systems and may be viewed as mathemati-
cal models for natural automata (e.g. biological organ-
isms, solar systems) and artificial automata (e.g. com-
puters, slot machines).

To study formally the notion of machine self-repro-
duction von Neumann in about 1953 introduced cellular
automata.[5] In general terms, a cellular automaton con-
sists of a finite aggregate of interacting automata and
is said to reproduce if it constructs a replica of it-
self. This process clearly constitutes asexual reproduc-
tion: the offspring is an exact copy of a single parent.
When we model and investigate sexual reproduction in
this framework, the transition from asexual to sexual
reproduction necessitates a change in number and struc-
ture of the genetic tapes involved, Vitányi.[4] As the
terminology in use is apt to create confusion we wish
to clarify some matters at the outset:
- One cellular automaton consists of many interacting
  automata called cells.
- The self-reproduction of a cellular automaton should
  be taken as a model for the reproduction of a single
  natural cell rather than as a model for the repro-
  duction of a multicellular organism.
- We sometimes use "automaton" for "cellular automaton"

---
*)
  This paper will be published in the proceedings of
  the Conference on Biologically Motivated Automata
  Theory, June 19-21, 1974, MITRE Corporation, McLean,
  Virginia (USA).

when no confusion can result.
- The terminologies "machines" and "automata" are used
  interchangeably.

2. On Methodology

Biological methodology usually consists of descrip-
tion and classification according to actual observations.
There are important exceptions to this: The Darwinian
theory of evolution and the Mendelian theory of heredity
are examples of hypothetico deductive thinking in biol-
ogy.[1] Von Neumann gave a formal vehicle, i.e. cellular
automata theory, in which to express a notion of self-
reproduction, i.e. asexual reproduction. We shall embed
a model of sexual reproduction in this framework thus
supplying together a hypothesis on the physiological
mechanism of the sexual genetic processes and a formal
framework in which to express alternative hypotheses;
both of which are lacking notwithstanding recent remark-
able advances in biochemistry. It is contended that even
in an abstract system of selfreproduction bearing almost
no direct relationship to biology interesting theorems
about the logical requirements and limitations of ma-
chines - and biological organisms - may be obtained. For
a further discussion along these lines see e.g.[4].

We may point out that one of the biological hypoth-
eses concerning the physiology of sexual processes
(both organically and with respect to behavior), impli-
cit in our model, is that each of the two constituent
sets of the double chromosome set has the potential of
causing the organism to grow into any one of both sexes
and behave accordingly. That sexual behavior can be ge-
netically determined is apparent from the experimental
evidence concerning the mating behavior of certain
birds.

The development of the organism is governed by one
of the two constituent sets of the double chromosome
set, say the dominant one, which always contains the Y
chromosome if present. In actual biological fact the
dominant chromosome set may not be physically separated
from the other one but be present as such in some other
way by e.g. activating and suppressor mechanisms between
genes or blocks of genes.

1

Which sex is selected depends on the dominant set containing an X or Y chromosome in the sense that the X chromosome determines one part of the chromosome set to become activated resulting in growth and behavior of the "female" type and the Y chromosome does the same thing to another part of the chromosome set resulting in growth and behavior of the "male" type. Sex linked inheritance would then be that certain genes or blocks of genes are (de)activated by (de)activated genes residing in sex oriented parts of the chromosome set. Alternative models where the presence of one X and no Y instead of 2 Xs acts exactly like the presence of a Y are obtained by slight changes. Hypotheses as described above are subject to experimental methods and hence can be verified or refuted on such a basis. The Mendelian theory of heredity and the Darwinian theory of evolution are also present in our model, as well as an indication why all organisms which have evolved use the same genetic code. This latter remark may be clarified as follows. A chromosome contains the building plan of the organism to be constructed. To carry out this construction we need a complicated read-out, interpreting and executing mechanism. Hence two species using different codes use also different code processors and so can have no progeny although they may seem similar. Whenever a primitive reproducing species arises it starts competing with species using different codes. By an evolutionary argument it is clear that if a species using a certain code starts to evolve into a higher species better equipped for the struggle for life this will cause an over-all advantage to all species using this code, thus eventually eliminating possible competitors at an early stage of evolutionary history.

3. Asexually Reproducing Automata

The notions in this section will be sketchily treated; for further information we refer to[4] or Codd[2] or the references contained therein.

Let us consider machines composed from a suitable collection of elementary parts. We may choose these components to be self-reproductive and computation-universal in their environments and so remove the problem or consider it at a descriptive level. We may also use components which are very simple (e.g. not self-reproducing, not computation-universal, having few different states) but aggregates of which can be self-reproducing.

To exclude trivial cases of selfreproduction such as crystallizationlike processes we require our automata to do something meaningful. They must be capable of simulating a *Universal Turing machine* i.e. a machine which

can execute any algorithm if supplied with an appropriate program. Our cellular automata exist in a *cellular space* which may be visualized as an infinite chessboard with each square or *cell* capable of assuming any one out of a finite number of states (e.g. 8). All cells change their states simultaneously in discrete time according to their own state and the states of the four nondiagonal neighbors. An assignment of states to a set of cells in the cellular space is called a *configuration*. Configurations consisting of cells in state 0 or 1 do not change in time without being influenced by cells in other states. A *cellular automaton* consists of a finite configuration in the cellular space, embodying the logical structure of an information processing device, which reads and writes information, coded in 0s and 1s, from an attached linear array of cells by means of a constructing arm. Such a constructing arm is an array of cells in state 1 which can be extended to and retracted from any location in the cellular space if the central configuration feeds it the appropriate series of signals coded in propagating sequences "0s", $s \in \{4,...,n\}$ which travel along the array such that s leads and 0 trails. It has been shown that an appropriate cellular automaton can simulate a Universal Turing machine, using a binary coded array of cells as its tape, and can construct any configuration of 0s and 1s anywhere in the cellular space.[5,2] Since the basic configuration of a cellular automaton is such a (0,1) configuration and is completed and starts to function if some activating signals are injected at an appropriate entry point there is a cellular automaton which can construct every such cellular automaton if supplied with an appropriate *description* on its tape. Hence this automaton is a *Universal Computer Constructor* (UCC). If the tape of the UCC contains its own description and if, moreover, the UCC copies its tape at the appropriate location of the constructed machine *self-reproduction* has been attained. Since the offspring is an exact copy of a single parent we have here an asexually reproducing automaton with a *genetic tape* containing, as it were, a blue print of itself. In modeling sexual reproduction we shall consider two parents pro automaton since more parents only complicate the picture and do not contribute to the advantages of the model. According to what is known "nature" seems of a similar opinion.

4. Sexually Reproducing Automata

Mainly, the sexually reproducing automaton consists of the *body* B, two *genetic tapes* T_1 and T_2 containing the encoded construction and behavioral algorithms, *construc-*

ting arms C (both M- and F-type) and D (only F-type) to execute these algorithms, and reading-writing constructing arms R_1, R_2, R_3 and R_4 to read from and write on T_1 and T_2 (fig. 1). Furthermore there need be some auxiliary tapes and reading-writing constructing arms which are not of interest here. The two specialized types or sexes of machines (M- and F-type) result from our aim to simplify the individual automata by a delegation of tasks that have to be performed, e.g. *searching* for the other automaton (M-type), *contributing genetic material* (M- and F-type), *constructing the offspring* (F-type).
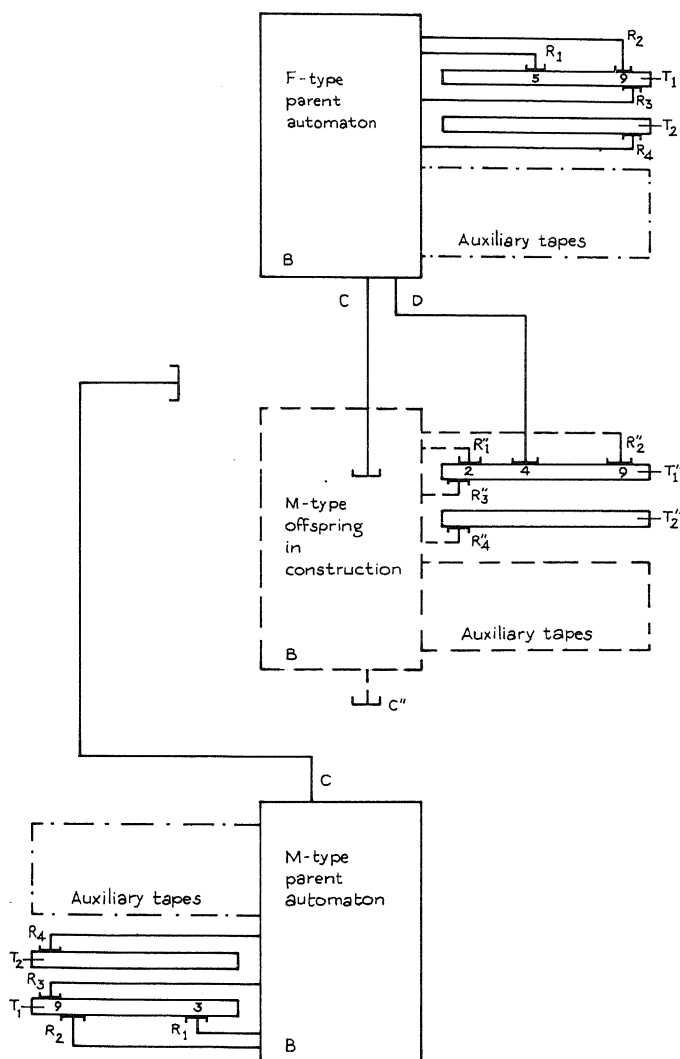
Prior to the construction of the offspring we need its genetic material (if it is redundant also a clearly determined part of it constituting a complete description), according to which the new automaton is to be constructed, since we have to know in advance which characteristics of what parent will be incorporated in the offspring. More specifically, we want the description of the offspring to be unambigously extractable from its total genetic material. Because each automaton has two parents and due to the above (and additional considerations set forth in [4]) every automaton posesses two genetic tapes each of which is *complete* in the sense that it contains all algorithms involved; and *similar* with respect to structure, instruction sequences and the diverse algorithms. By similar we mean here that although instructions in identical positions on the tapes may be different an interchanging of (sequences of) them will not render the algorithms involved incoherent or meaningless.

The recombination of the parents' characteristics in the offspring is due to the processes used to convey the genetic material from the parents to the offspring. By means of a random *copying procedure* each parent maps its two genetic tapes onto one *initial tape image*: M and F produce T_1' and T_2', respectively. Subsequently, by means of constructing arm C, M places each *dominant* (marked by an additional 1) instruction or characteristic that has a *recessive* (marked by an additional 0) counterpart on T_1' while the other is placed on T_2'; if both are of the same kind the distribution is random. The *definitive tape images* T_1'' and T_2'' result from this mixing phase and it is from T_1'' that the offspring is constructed by the F-type parent. F activates the offspring by injecting the activating signals and separates its constructing arm C in the process.

M-algorithm. In the course of its reproductive behavior M executes the following algorithm.

m1. M searches by means of its constructing arm C the cellular space until it discovers and recognizes a *fertilization prone* F automaton (i.e. an F which has constructed an initial tape image T_2').

m2. M constructs the initial tape image T_1' at the appropriate location.

m3. M compares words in identical positions on T_1' and T_2' and places all dominant words that have a recessive counterpart on T_1' and the recessive ones on T_2'; if both are of the same kind then the distribution is random. T_1'' and T_2'' result from this process.

m4. M retracts arm C, changes the search parameters and starts again at m1.

Fig. 1. Sexually reproducing cellular automata.

3

F-algorithm

f1. With arm C F constructs the initial tape image T_2' at a location computed from some parameters.

f2. F is fertilization prone and checks after a certain time interval whether some M has performed m2 and m3.

f3. If m2 and m3 have been performed F extends arm D towards the first word on T_1'' and arm C towards the location where the central configuration (body B) of the offspring will be constructed. If m2 and m3 have not been performed F erases T_2'' retracts C and starts again at f1.

f4. F changes its input from R_1 to D and constructs the central configuration of the offspring according to the building plan contained by T_1''.

f5. The last executed instruction on T_1'' changes the input from D to R_1; D is retracted and through C the activating signals are injected in the offspring.

f6. C is retracted, F changes its parameters and starts again at f1.

For details concerning the algorithms executed by M and F, the *behavioral algorithms*, we refer to [4].

Obviously, the F automaton has to start its construction job with a fixed instruction on T_1'', e.g. the first one. Since each tape contains all algorithms the first instruction must select the tape section containing the description of the specific part of the offspring's sexual type, i.e. the part of the configuration (body B) that is different for M and F. After its construction has been completed the offspring's reading arm R_1'' starts reading the second instruction on T_1'' directing it to the behavioral algorithm suited to its sex. Hence we require, in contrast with the asexual case, a tape *partitioned* into behavioral and construction sections. Note that there is a marked difference between the construction sections from which the "physical layout" of the automaton is constructed and the more algorithmic sections which govern the behavior to be performed, these latter sections are read, interpreted, and executed by the configurations constructed according to the former sections. In asexual reproduction no distinction is made between different tape sections, as the problem of different sexes and behavior does not arise, i.e. the automaton computes a location, proceeds to execute the construction sequence, copies the tape and activates the offspring. These four different actions are accomplished by using different interpreting sections of the body B in sequence: the behavior is built into the automaton more or less as hardware subroutines in an electronic computer. The complicated nature of sexu-

al reproduction, however, necessitates special behavioral and construction algorithms and hence tape sections, thus accentuating differences and similarities between construction and behavior as embedded in a cellular space.

A genetic tape is composed of 9 sections numbered 1-9. Each section contains a sequence of binary coded instructions either embodying a behavioral algorithm or a construction algorithm. Sections 1 and 2 determine whether the automaton is M- or F-type; these sections play the role of Y and X chromosomes in biology. *Section 1.* A *jump*, i.e. transfer of the head of the reading arm to a designated instruction word on the tape, to section 4 if the tape determines an M-automaton, to section 6 if it determines an F-automaton. *Section 2.* A jump to section 3 if the tape determines an M-automaton, to section 5 if it determines an F-automaton. *Section 3.* A subprogram that embodies the behavioral algorithm of M. Note that when the genetic tape determines an M-automaton section 2 may consist of the empty instructions since section 3 follows immediately. *Section 4.* The instruction sequence for the construction of the specific reproducing part of the M-automaton. The last instruction is a jump to section 7. *Sections 5-6.* As sections 3-4 with F substituted for M. *Section 7.* The instruction sequence for the construction of the identical part of the M and F automaton. *Section 8.* The instruction sequence for the construction of the individual part of the automaton. The last instruction on 8 gives back the control to section 5 of the constructing F-type parents' T_1. *Section 9.* Instructions for the individual nonreproductive behavior of the automaton. These instructions may be read, interpreted and executed by the part of the automaton that is specified in section 8.

When the *instruction code* or the *partitioning* of the genetic tapes is different for two automata, we may talk about different *species* of automata. Usually, within a species secs. 1-7 of T_1 and T_2 will be identical. Secs. 1,2 of T_2 will always specify jumps to secs. 6 and 5, respectively: i.e. T_2 is always X-type. In an M-automaton T_2 is X-type while T_1 is Y-type, i.e. secs. 1, 2 specify jumps to secs. 4 and 3. Because an automaton is constructed according to its T_1 genetic tape, this tape controls the sex of the automaton. If we attach a dominant marker to the first two sections of a Y-type genetic tape, and a recessive marker to the first two sections of an X-type genetic tape, then because of the processes used the offspring is M- or F-type on a fifty-fifty basis. Note that every genetic tape carries the potential for the development of both an F-type and an M-type automaton. Which one is realized depends on the instruc-

tions at the commencement of the tape and the interpreting apparatus of the F-type parent. Such a mechanism may take many forms as indeed it does in "nature".

5. Automata Genetics

(i) The two parents use different (binary) coding for identical instructions. The mixing phase m3 will scramble the instructions on the offspring's genetic tape in such a way that the constructing automaton will construct a meaningless configuration, if any. Thus two automata using different instruction codes can have no progeny, notwithstanding that they may consist of correct compatible configurations and use a compatible fertilization technique. No fertility among seemingly compatible, but genetically different species. (Note the exception as specified in (ii).)

(ii) The parents use the same instruction code, but the partitioning of their respective genetic tapes differs with respect to positioning and/or lengths of secs. 1-7. This will result in meaningless parts of the offspring's genetic tapes, viz. the part after the first difference, causing faulty parts of the offspring's configuration and/or senseless behavior after activation. In case one of the parents contributed exclusively dominant and the other exclusively recessive characteristics the offspring will be well formed but sterile owing to the random copying procedure which will create a totally garbled initial tape image.

(iii) (i)-(ii) do not occur but secs. 8 and 9 differ with respect to length and/or positioning. The offspring will be well formed with respect to reproduction, but the individual characteristics severely disturbed.

(iv) (i)-(iii) do not occur, but the initial tape images T_1' and T_2' differ with respect to the instructions in one or more of the sections 3-9. The difference occurs in: *section 3*. The reproductive M algorithm is deranged i.e. the M-type offspring is sterile due to behavioral defects. The trait is dormant in F-type offspring but will exhibit itself in the offspring's M-type progeny. *Section 4*. The same as in sec. 3 but with regard to the reproductive part of the M-type body: organic sterility. *Secs. 5-6*. As in secs. 3-4 with F substituted for M and vice versa. *Section 7*. Disfigurement in the reproductive part of the offspring's body both for M- and F-types. We are reminded of sterile hybrids. *Secs. 8-9*. We assume that this usually holds in a species of automata; it is meant to convey *individual traits* to the automata, respectively *physical* (qua in-

terpreting configuration) and *behavioral* (qua performed nonreproductive algorithm). It implies the existence of a population of genetically different individuals of sexually reproducing automata for which notions like "genetic pool", "evolution", "adaptability", "evolutionary variability", etc. are appropriate.

Mutations can be brought into the model in an obvious way by suitable changes in the genetic tapes resulting in, for instance, the consequences mentioned above, viz. a change in one of the genetic tapes of the parents gives via the copying procedure the effects as treated. A beneficial change in the progeny can be brought about by tentative small changes in the genetic tapes of a population of automata such that the cumulative effect of a set of these changes incorporated in one automaton in the course of the sexual reproductive processes promotes its viability. Assume a certain redundancy in tape structure to the effect that the change of one word on a genetic tape need not have fatal consequences. We may than observe a transition of *varieties* of automata, i.e. classes of automata differing in important respects qua secs. 1-7 but not qua instruction code and which are still reconcilable with respect to reproduction, into different species of automata using an identical instruction code but not reconcilable with respect to reproduction. We obtain a universe populated with different species of automata using the same instruction code (assuming that they all stem from the same stock). *Sex linked inheritance* is introduced easily by e.g. enlarging secs. 3-6 with nonreproductive parts. Genetically induced "transsexuality" occurs when the dominant and recessive markers of sections 1 or 2 are changed to their opposites, e.g. the offspring will consist of an M-type configuration executing the F-algorithm or vice versa. For further discussion along these lines and comparison with natural systems we refer to[4], where e.g. an extension of the model to cover multicellular organisms is treated. This extended model has genotypically identical cells, cell differentiation etc. and is consistent with the "axioms" of development presented by Apter.[1] An interesting question arising from our model seems to be which chromosomes (i.e. tape sections) are necessary and/or sufficient and what variations in the build-up of the genetic tapes are possible.

6. Conclusion

- We have modeled sexual reproduction in the formal framework of a cellular space and have obtained several of the familiar properties of heredity, not by treating

them as a priori given but by deriving them indirectly from certain logical assumptions.

- Our method may prove a useful tool with which to model and investigate hypotheses about mechanisms for sexual reproduction especially with respect to the genetic aspects.

- We have proposed a specific X-Y mechanism for the physiology of sexual reproduction in which sex determination is governed by a chromosome region of a different order of magnitude than those that are responsible for Mendelian characteristics and where the chromosome regions that are involved are equivalent to a large number of Mendelian genes. Such a theory is suggested by experimental evidence.[3]

- Our model is a first formal (nondescriptive) model for sexual reproduction, may have bearing on behavioral genetics and (the relevant essentials being translated into the appropriate biochemical terms) is of the kind that can be experimentally verified or refuted in whole or in parts.

References

1 M.J. Apter, Cybernetics and Development, Pergamon, New York, (1966).

2 E.F. Codd, Cellular Automata, ACM Monograph Series, Academic, New York (1968).

3 U. Mittwoch, Sex, growth, and chromosomes, New Scientist Sci. J. 51, 126-128 (1971).

4 P.M.B. Vitányi, Sexually reproducing cellular automata, Math. Biosci. 18, 23-54 (1973).

5 J. von Neumann, The Theory of Self-Reproducing Automata, (edited and completed by A.W. Burks), Univ. of Illinois Press, Urbana, Ill. (1966).