with the compliments of the author
28 November 1973, Edward F. Moore

MACHINE MODELS OF SELF-REPRODUCTION

BY

EDWARD F. MOORE

Reprinted from
Mathematical Problems
in the Biological Sciences
Proceedings of Symposia in Applied Mathematics
Volume 14, 1962

Reprinted 1965 with revisions.

[49] Edward F. Moore, "The firing squad synchronization problem", published in the book *Sequential Machines, Selected Papers.* Reading, *Massachusetts: Addison-Wesley Publishing* Company, March 1964, pp. 213-214.

[50] Edward F. Moore, "Machine models of self-reproduction", *Proceedings of Symposia in Applied Mathematics, Vol. 14 (Dec. 1962), pp. 17-33.* (Published by the American Mathematical Society, Providence, Rhode Island.)

[51] Edward F. Moore, "Mathematics in the biological sciences" *Scientific American*, Vol. 211, No. 3 (Sept. 1964), pp. 148-164.

Note: [51] is a popularized version of [50]. Except for the colored illustrations in [51], [50] is more complete than [51].

[52] John Myhill, "The converse of Moore's Garden-of-Eden theorem", *Proceedings of the American Mathematical Society, Vol. 14, No. 4 (Aug. 1963), pp. 685-686.*

Note: The converse of Theorem 2 (page 26) of [50] is proved in [52].

[53] John Myhill, "The abstract theory of self-reproduction", published in the book *Views on General Systems Theory.* New York: *John Wiley & Sons, 1964, pp. 106-118.*

[54] Howard H. Pattee, "On the origin of macromolecular sequences", *Biophysical Journal*, Vol. 1, No. 8 (November 1961), pp. 683-710.

[55] L. S. Penrose, "On living matter and self-replication", published in the book *The Scientist Speculates.* New York: Basic *Books, Inc., 1962, pp.* 258-271.

[56] Walter R. Stahl, Robert W. Coffin, and Harry E. Goheen, "Simulation of biological cells by systems composed of string-processing finite automata", *Proceedings of the Spring Joint Computer Conference, 1964, pp. 89-102.*

MACHINE MODELS OF SELF-REPRODUCTION

BY

EDWARD F. MOORE

Introduction. The ability of living organisms to reproduce themselves has long been considered to be one of their most characteristic features. Von Neumann was the first to treat in any detail the problem of how to make machines reproduce themselves in a purely mechanistic fashion as a way of throwing light on some fundamental problems of biology and as a problem (of intrinsic interest aside from biology) concerning the capabilities and limitations of machines. Over a period of years von Neumann gave several individual lectures and several series of lectures on *The general and logical theory of automata*, which partly dealt with self-reproducing machines. One of the early individual lectures was reprinted in **[17]**, but some of the later lectures varied in content and in approach. I heard only one series of the later lectures, but by reading and by hearsay I am acquainted with the ideas of some of the others. At first he considered a kinematic model of self-reproduction in which he considered three-dimensional physical objects, which were the parts out of which the self-reproducing machine was to be built. The machine was assumed to be in a stockroom in which unlimited supplies of these parts were floating around (much like particles of food in a nutrient medium), and it found the parts which it needed and then assembled them to make a copy of itself.

In his later lectures on this subject, von Neumann constructed a different model somewhat more abstract and less analogous to the real world. He considered a universe which is a 2-dimensional Euclidean space subdivided into square cells of equal size, like the squares of graph paper or of a checkerboard. I will call such a space a *tessellation*, but will somewhat modify the dictionary definition of tessellation. Located in each of the cells of this tessellation there is to be one copy of a finite-state machine. Each cell-machine is to be deterministic and synchronous; that is, at each integer-valued time $T > 0$, the state of each cell-machine is to depend only on its own state at time $T - 1$ and on the states of its neighboring cell-machines at time $T - 1$. All of the cell-machines are to be exactly alike as to their list of states and the rule determining their transitions, but different cell-machines are permitted to be in different states. The list of the possible states of the cell-machines must include a special state called the quiescent state, and all except a finite number of cell-machines will be in the quiescent state. The quiescent state is to have the property that if any cell-machine and all of its neighbors are in the quiescent state at time $T - 1$, then the cell-machine will be in the quiescent state at time T.

The entire system consisting of the underlying tessellation space, the set of allowable states, the quiescent state, the rules for transitions between states, etc., will be called a *tessellation structure*. This tessellation structure differs from the tessellation space in much the same way that a group (including the group product and the inverse) differs from the set of elements of the group.

A finite block of cells of a tessellation will be called an *array*. A *configuration* or a state of such an array is a function which associates a state with each cell. That is, a configuration of an array specifies what state each cell of the array is in at one particular time T. An example of such a configuration is given in Fig. 1, where there are seven cells in the array, and the state of each cell-machine is indicated by a symbol written in the cell.

Von Neumann considered the problem of choosing a tessellation structure such that its cell-machine has only a small number of states, and then making up a

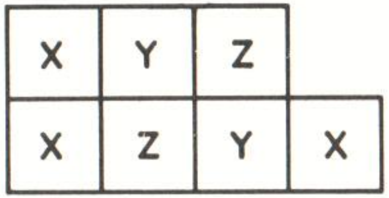

FIGURE 1. Example of a tessellation, showing a configuration defined for only seven cells.

configuration of these which would have the property of being able to reproduce itself. The details of constructing this would be somewhat like writing a digital computer program and would avoid all the problems of motion, assembly, and geometry which would be present in a kinematic model. Von Neumann solved most of the details of the problem of how a self-reproducing machine could be constructed both by a kinematic model and by a tessellation model. Although his work on this was not quite finished at the time of his death, publication of it is planned [**16**].

Kinematic models of self-reproduction. Although von Neumann's early discussions of self-reproduction were in terms of a kinematic model which is made of parts which are in motion, he later changed to a tessellation model which is less vivid and dramatic, less realistic, and less difficult to deal with mathematically. However, there have been some quite interesting recent examples of kinematic models which have been actually built and demonstrated.

One way of performing the necessary spatial motion by simply controlled electrical means is to use model railroads. The starting, stopping, and reversing of trains under electrical control is already easy without having to design and develop new manipulating devices. Also, the rearranging of parts can be done by having cars switched onto sidetracks. The necessary logical control and programming can be built in the form of relay circuits mounted aboard the individual cars. The self-reproducing machine is to be a train made up of a sequence of different kinds of cars, and the cars are considered to be the elementary parts or raw materials.

The actual construction of a model railroad layout and cars which perform this was done by Jacobson [**6**]. He described several different models of varying degrees of generality and actually constructed the simplest of these. This model was partly trivial in certain respects and was the least general, with regard to self-reproducing ability, of the models which he described.

However, other kinematic models somewhat more general and less trivial than Jacobson's, but based on techniques quite different from model railroading, have been constructed by Penrose. The simplest one of these models (illustrated in Fig. 1 of [**19**] and of [**20**] and in Fig. 2 of [**21**]) will be called Penrose's basic model, since his other models resemble this but have various modifications and other features.

Penrose's basic model makes use of only two kinds of units, called A units and B units. Each of these is a single rigid piece of solid material, cut out in a shape having certain hooks and interlocks. It is possible for a pair of these units to hook together in two different ways to make an AB machine or a BA machine. If a number of the unconnected units are placed in "random order" in a rectangular box of the right size containing an AB machine, and the box is shaken regularly, more copies of the AB machine will be constructed by means of mechanical forces, friction, and gravity alone without any electricity, magnetism, or chemical reactions being involved. If the unconnected units are similarly shaken with a BA machine present, more copies of the BA machine will be constructed. Hence either the AB or the BA machine can appropriately be called a seed. However, if the unconnected units are similarly shaken with neither kind of seed present, no "spontaneous generation" occurs and no seeds are formed; but by shaking the box with unusual force, "spontaneous generation" can occur.

No picture of Penrose's basic model is included with this paper, since if the reader attempts the problem of how to design the shapes of the units A and B so as to have the specified properties, the difficulties he will encounter in his attempt will cause him to more readily appreciate the ingenuity of Penrose's very simple solution to this problem.

After constructing an exact copy of Penrose's basic model, I have found that it not only operates mechanically with reasonable satisfaction but is very useful in suggesting to audiences some of the problems and possibilities of self-reproducing machines.

There is one misunderstanding of the biological interpretation of Penrose's basic model which I should clear up. When this is shown to engineers, mathematicians, or physical scientists, many of them assume that the two kinds of units correspond to male and female individuals. This is not at all the reasonable analogy. One of the AB or BA machines corresponds to an individual, and the units correspond to distinct molecules of the chemical compounds of which chromosomes are made. The fact that there are two kinds of units may mislead persons who have forgotten their biology. In this case the reproduction is asexual and the individual consists of only one chromosome made up of two molecules. A more general and less trivial example of reproduction would have many units arranged in a long chain, more like a real chromosome. Penrose does in fact give such more general models in his papers. Still more lifelike models which would have several chromosomes and also have additional units arranged in structures analogous to cytoplasm and cell walls have not yet been described in print in detail by anyone.

Another kinematic model of self-reproduction (involving electromagnets and electrets floating in a liquid) has been proposed by Morowitz [12], but has apparently not been designed in detail or built, so it is not known how well such a machine would work.

Tessellation models of self-reproduction. Before the general definitions and descriptions of tessellation structures are given in further detail, the reader should be cautioned that there are three different kinds of things which can be called machines. The tessellation itself can be considered to be a machine, although in a tessellation model of self-reproduction the tessellation is more naturally considered to be the environment or the universe (including the supplies of parts or raw materials) in which the self-reproduction takes place. A configuration (which is restricted to an array of finite size) can be considered to be a machine, and in fact it is a machine of this kind which can be shown to reproduce itself. A cell can be considered to be a machine, since it has a list of states and transitions, but it corresponds in the tessellation models of self-reproduction to one of the elementary parts out of which the machine is built. Because of the fact that these three kinds of things can be called machines, it is necessary to be careful to avoid any confusion between them. An example of a mistake which arises in this way is in the paper by Rosen [25], where it is asserted that there is a paradox involved in the existence of a self-reproducing machine. This alleged paradox can best be explained as originating from his failure to maintain the distinction between the tessellation and the configuration, although he does not actually use these words.

Tessellations of N-dimensional Euclidean space, for any positive integer N, could be defined in a manner analogous to the case for $N = 2$; and the constructions and the results would be very similar. In fact, $N = 3$ is the case of real biological interest; and von Neumann stated in his Vanuxem lectures (Princeton University, March 2–5, 1955) that he had at first thought that it might be necessary to use a 3-dimensional model, since the wiring diagram of the machine might turn out to be non-planar. However, he said he had succeeded in devising a method for wires to cross one another which could be used with what I have been calling a 2-dimensional tessellation structure. It certainly is simpler to work with a 2-dimensional space, since pictures can be drawn on paper to represent configurations. Shannon has indicated to me (unpublished personal communication) a scheme for a limited kind of self-reproduction which could be carried out with a 1-dimensional tessellation structure, but the 2-dimensional model seems to be of greater interest. Except where otherwise indicated, this paper will deal with tessellations which are subdivisions of Euclidean 2-space into square cells.

A *tessellation structure* should perhaps be formally defined as a quintuple (N,T,S,q_0,f), where N is a positive integer, T (the tessellation itself) is a subdivision of Euclidean N-space into cells which are N-cubes of unit dimensions and whose centers have integer coordinates, S is a finite set whose elements are called states, q_0 is a distinguished member of S called the quiescent state, and f is a function which maps the set of all states of the 3^N cells which are neighbors of any cell x at time $T - 1$ into states of x at time T. However, this would be unnecessarily formal and complicated for the present discussion and will be skipped with only this brief mention.

In this paper, the *neighbors* of a given cell will be taken to be all of those cells (including the cell itself) which have each of their coordinates differing by at most

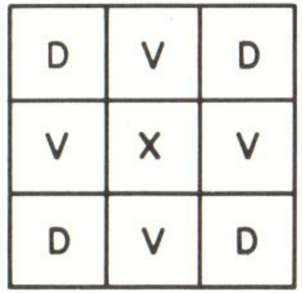

FIGURE 2. The nine cells which are considered neighbors of the cell marked X.

1 from the coordinates of the given cell. In Fig. 2, all nine of the cells are neighbors of the cell marked X. This definition permits the set of neighbors of all the cells in a rectangular array to be rectangular, and hence it is easier to compute the numbers of cells in certain arrays. The exact definition of neighbor used is not too important. Von Neumann's construction in [16] considers the five cells labeled V and X in Fig. 2 to be the ones of interest and permits the state of a cell to depend on previous states of only these. However, his definition can be included within mine, since I can merely specify the functional dependence of the next state to be such that it does not actually depend on the state of the cells labelled D.

The functional correspondence f, which specifies how each cell has its state, for all time $T > 0$, depend on the states of the neighboring cells at time $T - 1$, is to be the same for all cells of the plane. This is a homogeneity requirement, specifying that the physical laws of this universe are to be the same in all parts of it. The functional dependence f is also required to be such that a cell whose neighbors at time $T - 1$ are all in the quiescent state will be itself in the quiescent state at time T. In Fig. 3, the quiescent state is indicated by 0's.

In order that the states of the entire tessellation will be constructively attainable, we will require that the states at time $T = 0$ be such that all but a finite number of cells are in the quiescent state. It then readily follows that at each time $T > 0$

all but a finite number of the cells will be in the quiescent state, although it is possible for the number of non-quiescent cells to increase with increasing T.

We will say that a configuration c is a *copy* of a configuration c', if there is a translation of the plane which maps the array of c onto the array of c', and each cell has the same state as the cell into which it is mapped.

We will say that a configuration c^* *contains n copies of* a configuration c, if there exist n disjoint subsets of the array of c^*, and each of these subsets is a copy of c.

This definition is illustrated in Fig. 3, which shows a configuration which contains only three copies of the configuration of Fig. 1, although, if the word

X	Y	Z	O	X	Y	Z	O
X	Z	Y	X	X	Z	Y	X
O	O	O	O	O	O	O	O
X	Y	Z	X	Y	Z	O	O
X	Z	Y	X	Z	Y	X	O

FIGURE 3. This configuration contains three copies (but not four copies) of the configuration of Fig. 1.

"disjoint" were changed to "distinct" in the above sentence, it would contain four copies.

A configuration c will be said to be *capable of reproducing n offspring by time T* if, starting from the initial conditions of the entire tessellation at time $t = 0$ such that the set of all non-quiescent cells of the tessellation is an array whose configuration is a copy of c, there is a time $T' < T$ such that at time T' the set of all non-quiescent cells will then be an array whose configuration contains at least n copies of c.

A configuration will be said to be a *self-reproducing configuration* if for each positive integer n, there exists a T such that c is capable of reproducing n offspring by time T.

The above definition does not rule out trivial examples of self-reproducing configurations. Consider the tessellation structure such that there are only two states, X and 0, and having the transition function f such that each cell will be in state X at time T if at least one of its neighbors was in state X at time $T - 1$. Then the configuration consisting of one cell in state X will be a self-reproducing

configuration. This is more nearly a model of crystal growth than of self-reproduction.

To obtain a kind of self-reproduction which would certainly be considered non-trivial, von Neumann required that each configuration contain a universal Turing machine. It might be thought desirable to be able to avoid triviality by being able to have a self-reproducing configuration which would contain a copy of any given configuration, but Theorem 2 will show that this is impossible.

It is usually assumed that in reproduction of actual organisms, the population can grow exponentially with time; for instance, the size of the population might double once each generation. That this cannot occur in a tessellation universe is indicated in the following theorem.

THEOREM 1. *If a self-reproducing configuration is capable of reproducing $f(T)$ offspring by time T, then there exists a positive real number k such that $f(T) \leq kT^2$.*

PROOF. Let c be the self-reproducing configuration. Let the smallest square array large enough for a configuration containing a copy of c be of size $D \times D$. Then at each time T, the total number of non-quiescent cells is at most $(2T + D)^2$. If r is the number of cells in the array of c, then $f(T) \leq (2T + D)^2/r$, from which the conclusion of the theorem follows.

This Malthusian sort of argument, that the number of offspring cannot go up faster than the square of the time, since there would not be room for them, depends on the fact that there is a finite velocity of propagation of the non-quiescent region, since each unit of time permits this region to grow only one cell in each direction. This corresponds roughly to a physical limitation such as a finite velocity of light.

This argument also depends on the dimension of the space, since for an N-dimensional tessellation structure the corresponding bound would be kT^N.

At various points in the preceding discussion, the special nature of the time $T = 0$ has been alluded to. It will be assumed that the cells of a tessellation structure will be governed by the transition rules of the tessellation structure for all $T > 0$, but for $T = 0$ we can arbitrarily specify the initial conditions. That is, to test whether a given configuration is self-reproducing we set up a copy of it, surrounded by quiescent cells, at $T = 0$. This is the only time at which we permit such an arbitrary configuration to be specified.

It happens that under certain reasonable assumptions about the tessellation structure there will always exist configurations which cannot occur except at time $T = 0$. That is, these configurations are not only unstable, but they are non-constructible in the sense that there is no configuration at time $T - 1$ which will give rise to the given configuration at time T by means of the function f which defines the rules for the transition from one state to another. Such a configuration will be called a *Garden-of-Eden* configuration. This term, from the Biblical account in the second and third chapters of Genesis, was suggested by John W. Tukey.

Since a Garden-of-Eden configuration cannot be produced by any other configuration, no self-reproducing configuration can contain a copy of a Garden-of-Eden configuration. Hence an investigation of the conditions under which they can occur throws light on the limitations of the ability of machines to reproduce themselves.

The conditions under which these Garden-of-Eden configurations can occur involve the ability to perform erasing. After a blackboard has been erased, it is no longer possible to tell what had been written there; and by analogy with this, the term erasing is used by the designers of memory units in digital computers.

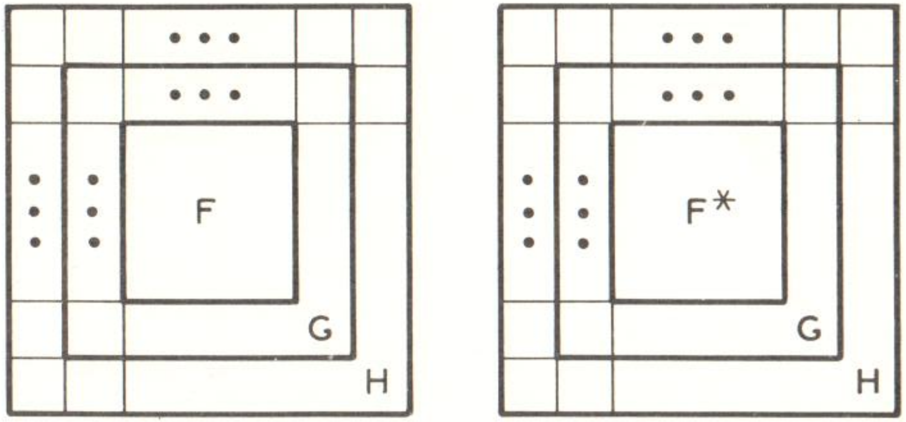

FIGURE 4. The descriptions of the configurations of two different arrays of the same size, as occurring at time T in the definition of an erasable configuration.

Erasing is an irreversible process whereby a given action produces a state from which it is impossible to determine the preceding states from which it could have arisen.

However, if we have a tessellation structure for which the transition function f is such that the state of each cell is merely the previous state of the cell just to the left of this, we would not want to consider that erasing was taking place. Within a fixed array of cells we might not be able to reconstruct the past state, but the information as to the past state has merely been shifted off to the right and not destroyed. Thus, it is necessary for us to be careful in the formal definition of erasing that we watch the cells which are neighbors of the configuration to be sure that the information is not carried away to them. In addition to this, it is necessary to specify what happens on the cells which are neighbors of these to be sure that new information is not shifted in from outside. Thus the formal definition will be somewhat involved.

Consider a pair of configurations of two arrays of the same size, as shown in Fig. 4. Without loss of generality these arrays may be taken to be square. We consider in each case an inner array, whose configuration is specified as being F and F^* respectively, at time T. Next we consider the hollow square which consists of the set of all the neighbors of the previous array which are not members

of that array. The configuration of this array is required to be G at time T in each case; that is, one configuration is a copy of the other.

Then we consider the similarly extended boundary array around these, which we require to have the configuration H in each case; that is, the two configurations are required to be copies of each other. We require that $F \neq F^*$; that is, the innermost arrays must have different configurations, but they must agree along the two layers of boundary at time T.

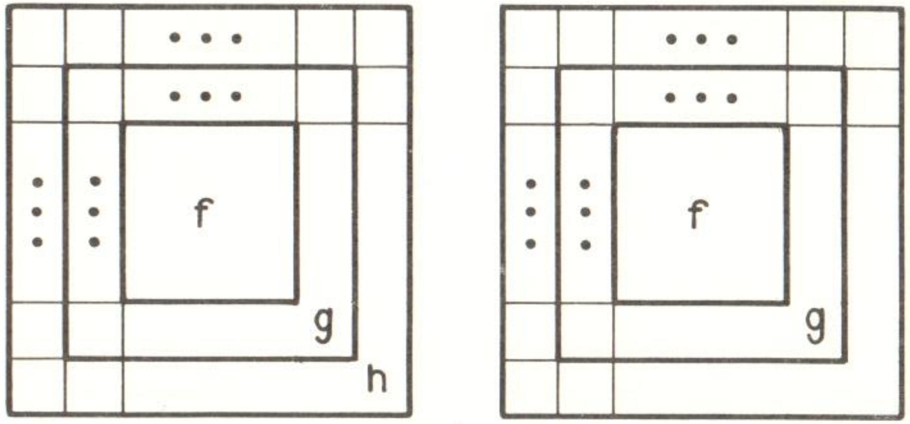

FIGURE 5. The descriptions of the configurations of two identical arrays, as occurring at time $T + 1$ in the definition of an erasable configuration.

If we have such a pair of configurations, and if the configurations which follow them (shown in Fig. 5) at time $T + 1$ are copies of each other as far as the configuration f of the inner array and the configuration g of the intermediate array is concerned (note that the configuration h of the outer array cannot generally even be specified, since the state of the cells of this array can depend on the previous states of cells outside all of the arrays which have been considered), the pair of configurations will be said to be *mutually erasable*.

It should be noted that the relation of being mutually erasable is transitive and symmetric, hence configurations can be put into equivalence classes, each of whose members are either copies of each other or are mutually erasable.

A configuration c will be said to be an *erasable configuration* if there exists another configuration c' such that they are mutually erasable. An erasable configuration was called a "configuration which can forget" in [**11**], but the term "erasable" is not only less anthropomorphic but more in accordance with the terminology used in electrical engineering.

It should be noticed that if c is an erasable configuration and c' is a configuration of a rectangular array such that c' contains a copy of c, then c' is erasable. This is what permits us to consider an erasable configuration to be associated with a square array without loss of generality.

Not all tessellation structures are such that erasable configurations of their cells can be defined. If the function f defining the transitions and the states has the property that such irreversible transitions cannot take place, then only a restricted class of methods of construction and computation can take place.

We now proceed to state the main result of this paper.

THEOREM 2. *For a tessellation structure for which there exist erasable configurations, there exist Garden-of-Eden configurations.*

First I will give a sketch of the intuitive idea of the proof, and then a detailed proof will be given of the inequality from which this theorem follows. Let n be

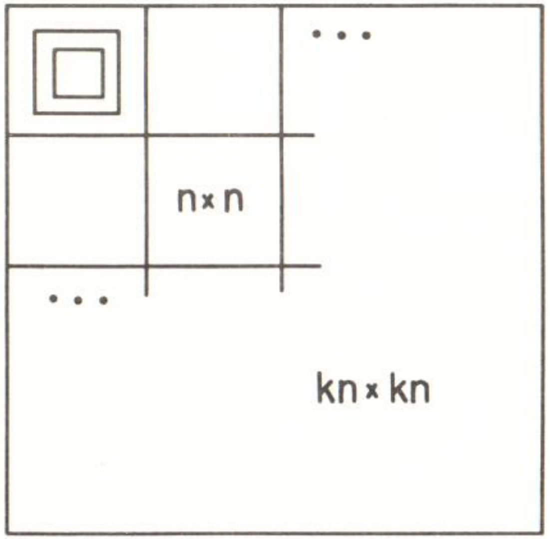

FIGURE 6. An array of size $kn \times kn$, for use in proving Theorem 2.

an integer such that there is an array of size $n \times n$ which has an erasable configuration. Then we consider, for an integer k to be chosen later in the proof, an array of size $kn \times kn$, as shown in Fig. 6.

Each of the k^2 arrays of size $n \times n$, as shown in Fig. 6, is the proper size to contain a copy of an erasable configuration; and k is to be chosen large enough that, averaging over all configurations of the array of size $kn \times kn$, there will often be many such erasable configurations.

If we consider the array of size $(kn - 2) \times (kn - 2)$ into which this array is mapped at time $T + 1$, where T is the time of the original array (again since the cells along the border are not necessarily defined as to what state they will take on), we can compute how many possible configurations or states it can have as being $A^{(kn-2)^2}$, compared with $A^{(kn)^2}$ for the array of Fig. 6, where A is the number of states of each cell.

If we indicate the states of these two arrays in Fig. 7 and indicate which states are mapped into which, we will note that whenever we start with a state containing one erasable configuration, then this state and the other one which is mutually erasable with it will both map into one state at time $T + 1$. Wherever we have a state containing two copies of the $n \times n$ erasable configuration, we will have four states at time T mapping into one state at time $T + 1$. In general, wherever we have a state containing s copies of the erasable configuration, we have 2^s states at time T mapping into one state at time $T + 1$. Then we need only show that the loss in number of states due to erasure is more than the loss due to the difference between $A^{(kn-2)^2}$ and $A^{(kn)^2}$.

Let us consider the logarithms of the numbers of states rather than the numbers of states themselves. The logarithm of the ratio which indicates how many states are lost because of this boundary layer increases approximately linearly with

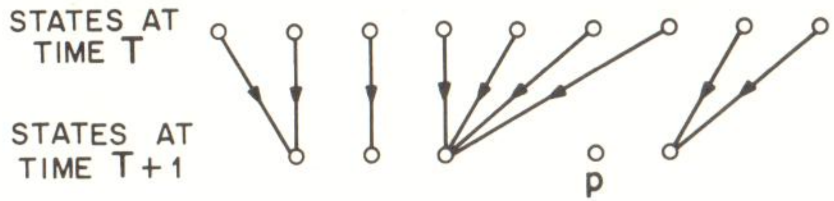

FIGURE 7. Diagram showing the states of the array at time T and the states into which they make transitions at time $T + 1$.

k. Also, if we consider the logarithm indicating the number of states lost due to erasure, this increases approximately proportionally with the area of the array. This is because this logarithm indicating the states lost due to erasure increases with the number of erasable configurations present, except that appropriate averaging must be done over all states of the array. Since the loss due to erasure goes up approximately with the square of k, then by choosing k large enough it can be shown that more states are lost that way than by the shrinkage at the boundary layer. Hence there must exist a state p (as shown in Fig. 7) of the $(kn - 2) \times (kn - 2)$ array which cannot be reached from any of the states of the array at time T. This state p is the desired Garden-of-Eden configuration.

The detailed proof follows below. Consider the equivalence relation R defined between $n \times n$ configurations which holds between any two configurations if they are either mutually erasable or are copies of each other. Since, by the hypothesis of the theorem, there is a pair of mutually erasable configurations of size $n \times n$, the relation R divides the set of A^{n^2} configurations of this size into at most $A^{n^2} - 1$ equivalence classes. Then consider two configurations c and c' of the $kn \times kn$ array. The configuration c will be said to have the relation R^* to the configuration c' if each of the k^2 subconfigurations of size $n \times n$ of configuration c has the relation R to the subconfiguration in the corresponding location in configuration c. Then R^* is an equivalence relation, and the number of equivalence classes of

R^* is at most $(A^{n^2} - 1)^{k^2}$. Since any two configurations which are equivalent under relation R^* lead to the same configuration at time $T + 1$, in order to prove the theorem, we need only prove that there exists a positive integer k sufficiently large that

$$(1) \qquad (A^{n^2} - 1)^{k^2} < A^{(kn-2)^2}.$$

To prove this, we note that since $A > 1$ and $n > 1$, $A^{n^2} > A^{n^2} - 1 > 0$, and so $(A^{n^2}/(A^{n^2} - 1)) > 1$. Thus we can choose a positive integer k such that

$$k > 4n/\log_A(A^{n^2}/(A^{n^2} - 1)),$$

so that

$$\log_A (A^{n^2}/(A^{n^2} - 1)) > 4n/k,$$

and thus

$$\log_A ((A^{n^2} - 1)/A^{n^2}) < -\frac{4n}{k} + \frac{4}{k^2}.$$

By raising A to these powers,

$$(A^{n^2} - 1)/A^{n^2} < A^{(-4n/k+4/k^2)},$$

and hence

$$A^{n^2} - 1 < A^{(n^2-4n/k+4/k^2)}.$$

By raising these to the k^2 power we obtain

$$(A^{n^2} - 1)^{k^2} < A^{(k^2n^2-4kn+4)},$$

which is easily seen to be equivalent to (1), completing the proof of the existence of Garden-of-Eden configurations.

Such a Garden-of-Eden configuration corresponds to a machine which cannot arise as the result of any past state of its universe, but can occur only at time $T = 0$. This also corresponds to a machine which cannot be built out of the available parts, but whose physical structure can be described as an arrangement of those parts.

There are a number of assumptions about the nature of the tessellation structure, some of which were hidden along the way in definitions. Listing them all together below, it was assumed that

(1) The universe is homogeneous.
(2) Space and time take on only discrete integer values.
(3) Only local action can occur at any one time.
(4) The universe is in Euclidean N-space.
(5) The laws of behavior of the universe are deterministic.
(6) Erasing is possible.

From these six assumptions (and perhaps other hidden ones which I have not noticed) it can be concluded that Garden-of-Eden configurations exist. Each of these assumptions can be examined in detail, since there might be reason to wish to prove a result similar to Theorem 2 in some mathematical model of the physical

universe in which not all of these assumptions hold. In particular, if it could be shown to hold in some universe of modern theoretical physics, it might have some cosmological interest in indicating that certain states of the physical universe are describable but not attainable.

The assumption (1) that the universe is homogeneous is used in that the possibility of erasure in one place then gives the possibility of erasure in many places.

The assumption (2) that space and time are quantized is used in the method of proof given above, but methods might be possible which would depend on integration in a continuous universe rather than counting in a discrete one.

The assumption (3) that the behavior at any location can depend on the immediate past of only the immediate neighborhood is used to confine the effects of any information coming in from outside an array to a thin boundary layer, and it corresponds to an assumption that information cannot be transmitted faster than the velocity of light.

The assumption (4) that the universe is in Euclidean N-space only requires that we have a space in which regions can be found which are large enough that the volume of the interior can be made arbitrarily many times as large as the volume of a thin layer at the boundary.

The assumption (5) that the laws of the behavior of the universe are deterministic prevent states from having more than one possible successor. If probabilistic transitions were permitted, they would cause the states to branch apart; and it might be necessary to have some way of insuring that this branching apart did not have more effect than the erasing in order to prove the theorem.

The assumption (6) is vital to the proof, and one of the tessellation structures already defined in this paper will serve as a counter-example showing that the conclusion of the theorem does not hold when the other five assumptions hold but this does not.

Assumption (6) violates Newtonian mechanics but not quantum mechanics. However, assumption (5) violates quantum mechanics but not Newtonian mechanics.

Further problems. There are many problems which can be formulated about tessellation structures, tessellation models of self-reproduction, and kinematic models of self-reproduction. A few are indicated below.

What can be done to make the statement that one machine is more general or less trivial than another in its self-reproducing behavior more precise? There does not seem to be a clear-cut line of demarcation between the trivial and the non-trivial models. Could this relation of being less trivial in this respect be formalized as a partial ordering between machines?

In the attempt to show how life could have arisen on the earth **[18]** by the chance interaction of non-living materials, it would be desirable to have some way of computing how likely this would be to occur. If there were some way of seeing how complicated an assemblage of parts (in particular, molecules) must be in order to have the self-reproducing property (and further, the property of being capable of undergoing evolution to produce much more complicated descendants), this would certainly be of interest. An attempt in this direction, in characterizing complexity in terms of bits of information, was made by Jacobson [6] for some of the models he described.

In a tessellation structure in which Garden-of-Eden configurations exist, how large can the smallest array which has such a Garden-of-Eden configuration be? Although the method of proof of Theorem 2 given in this paper gives a very large size for an array having a Garden-of-Eden configuration, an array of 5×5 size or smaller is large enough for each tessellation structure which I have examined in detail.

How simple a tessellation structure (in terms of number of states or some other measure of complexity) can permit a non-trivial self-reproducing configuration?

Can a tessellation structure have a self-reproducing configuration without having an erasable configuration?

Among all tessellation structures having n states, obtain upper and lower bounds (or possibly exact values) on how many of them have erasable configurations, and how many of them have self-reproducing configurations. In particular, prove that as n increases, the fraction of tessellation structures which have erasable configurations approaches 1.

In a tessellation structure which has erasable configurations, how large can the smallest inner array (corresponding to F in Fig. 4) be? All such tessellation structures which I have examined in detail have had erasable configurations where this inner array consisted of only one cell. Can this be proved to be always possible?

For a large machine which is either a tessellation model or a kinematic model of self-reproduction, how can the steps of reproduction be going on in parallel rather than serially? This would permit many parts of the machine to be produced at once, rather than the slower one-at-a-time procedure in the machines proposed by Penrose [20], von Neumann [16], and Jacobson [6].

Other references. References [1] through [31] of this paper are intended to give thorough coverage of what has been written on machine models of self-reproduction. References [32] through [48] give only a very small part of the literature on the somewhat related subjects of finite state machines, sequential circuits, and iterative circuits. Brief mention will be made in the next few paragraphs of those papers which have not already been cited in this paper.

Switching circuits whose circuit configuration extends repetitively throughout Euclidean n-dimensional space give an electrical realization of what have been called tessellation structures in this paper, and the circuits themselves are usually called *iterative circuits*. The work of Burks [2; 33] and of Church [5] deals with such circuits for arbitrary n. The work of Unger [47] and of Holland [38] deals with possible applications of computers built of such iterative circuits, chiefly for $n = 2$. Iterative one-dimensional circuits are treated in [37; 40; 41; 34].

The work of Burks [2; 3] is about tessellation structures, tessellation models of self-reproduction, and some of the problems about von Neumann's models. The work of Myhill [13] gives a partial description of a model intermediate in characteristics between von Neumann's kinematic model and his tessellation model. In [14], starting from some axioms assumed to describe the behavior of machines capable of producing other machines, a proof is given, using arguments from recursive function theory, that machines can build other machines which are improvements of themselves, although the improvement is in a sense which may be of more interest to a logician than to a biologist, since it involves the ability to print out larger classes of true theorems.

There are several other items [1; 10; 22; 23; 24] which give further explanations and developments of the models of Penrose.

Ulam [28] states a problem somewhat similar to the problem solved by Theorem 2 of this paper. Some practical and economic questions relating to construction of actual working models of self-reproducing machines are discussed in [9] and [31].

Kemeny [7] gives the most detailed discussion yet published of von Neumann's tessellation models of self-reproduction. Shannon [27] gives a general sketch of von Neumann's ideas, and the remaining papers [4; 8; 26; 30] make some reference to self-reproducing machines, although they mainly deal with other kinds of machines.

Ulam [29] considers combinatorial problems of describing the configurations which arise in various simple tessellation structures.

The abstract theory of machines which have only a finite number of states is treated in [35; 36; 43; 45; 46; 48], without emphasizing any electrical circuit realization of these machines. The electrical circuits which realize such machines are called *sequential circuits*, and are treated in [32; 34; 39; 40; 42; 44]. Many other references on sequential circuits, iterative circuits, and finite-state sequential machines are cited in [32; 35; 36; 41; 42; 45; 48], and others.

Acknowledgements. The author would like to acknowledge his indebtedness to C. E. Shannon, whose suggestions led to the statement of Theorem 2 and the basic idea for the method of proving it. He would also like to acknowledge suggestions given by H. O. Pollak, E. N. Gilbert, and J. B. Kruskal, which resulted in much simplification in the statement and proof of inequality (1), compared to earlier drafts of this paper.

Bibliography

1. Michael Abercrombie, *Introduction to replication*, New Biology, No. 31, pp. 10–24, Penguin Books, Harmondsworth, England, 1960.

2. Arthur W. Burks, *Computation, behavior and structure in fixed and growing automata*, Self-organizing systems, pp. 283–311, Pergamon, New York, 1960.

3. ———, *Summary of states and transition rules for von Neumann's self-reproducing automaton*, typescript (n.d., circa October 1960), 6 pp., to be published with [16].

4. Arthur W. Burks and Hao Wang, *The logic of automata*, J. Assoc. Comput. Mach. **4** (1957), no. 2, 193–218; no. 3, 279–297.

5. Alonzo Church, *Application of recursive arithmetic to the problem of circuit synthesis*, Summaries of talks presented at the Summer Institute of Symbolic Logic, Cornell University, 1957, 2nd ed., pp. 3–50, Communications Research Division, Institute for Defense Analyses, Princeton, N.J., 1960.

6. Homer Jacobson, *On models of reproduction*, American Scientist **46** (1958), no. 3, 255–284.

7. John G. Kemeny, *Man viewed as a machine*, Sci. Amer. **192** (1955), no. 4, 48–53; errata, ibid., no. 6, 6. Partially reprinted in *Automatic control*, pp. 132–146, Simon and Schuster, New York, 1955.

8. Lars Löfgren, *Automata of high complexity and methods of increasing their reliability by redundancy*, Actes du 1er Congrès International de Cybernétique, Namur, 26–29 Juin 1956, pp. 493–511, Gauthier-Villars, Paris, 1958. Also published in Information and Control **1** (1958), no. 2, 127–147.

9. Edward F. Moore, *Artificial living plants*, Sci. Amer. **195** (1956), no. 4, 118–126.

10. ———, review of four papers by Penrose [**19**; **20**; **21**; **22**], Trans. I.R.E. Professional Group on Electronic Computers **EC-8** (1959), no. 3, 407–408.

11. ———, *Machine models of self-reproduction*, Abstract 560–52, Notices Amer. Math. Soc. **6** (1959), 622–623.

12. Harold J. Morowitz, *A model of reproduction*, American Scientist **47** (1959), no. 2, 261–263.

13. John Myhill, *Self-reproducing automata*, dittoed typescript, Institute for Advanced Study (n.d., circa 1959), 15 pp.

14. ———, *Possibilities of favorable mutation in self-reproducing automata*, dittoed notes for a series of lectures at the University of Michigan, June 1960, 7 pp.

15. John von Neumann, *Theory and organization of complicated automata*, mimeographed typescript of a series of five lectures delivered in December 1949, 82 pp., to be published by the Univ. of Illinois Press.

16. ———, *The theory of automata: Construction, reproduction, and homogeneity*, uncompleted typescript of three chapters (n.d., circa 1952), 220 pp., to be published by the Univ. of Illinois Press.

17. ———, *The general and logical theory of automata*, *Cerebral mechanisms in behavior*, pp. 1–48, Wiley, New York, 1951. Partially reprinted in *The world of mathematics*, pp. 2070–2098, Simon and Schuster New York, 1956.

18. A. I. Oparin, *The origin of life on the earth*, 3rd ed., Academic Press, New York, 1957.

19. L. S. Penrose and R. Penrose, *A self-reproducing analogue*, Nature **179** (1957), no. 4571, 1183.

20. L. S. Penrose, *Mechanics of self-reproduction*, Ann. Human Genetics **23** (1958), part I, 59–72.

21. ———, *Automatic mechanical self-reproduction*, New Biology, No. 28, pp. 92–117, Penguin Books, Harmondsworth, England, 1959.

22. ———, *Self-reproducing machines*, Sci. Amer. **200** (1959), no. 6, 105–114, 202.

23. ———, *Developments in the theory of self-replication*, New Biology, No. 31, pp. 57–66, Penguin Books, Harmondsworth, England, 1960.

24. ———, *Automatic mechanical self-replication*, 16 mm. kodachrome movie film, filmed and distributed by H. A. Cresswell, Hemel Hempstead, Herts., England, 1958; 11 minutes.

25. Robert Rosen, *On a logical paradox implicit in the notion of a self-reproducing automaton*, Bull. Math. Biophys. **21** (1959), 387–395.

26. Claude E. Shannon, *Von Neumann's contributions to automata theory*, Bull. Amer. Math. Soc. **64** (1958), no. 3, part 2, 123–129.

27. ———, *Computers and automata*, Proc. I.R.E. **41** (1953), no. 10, 1234–1241. Reprinted as Bell Telephone System Monograph 2150 (1954), 8 pp. Also partially reprinted in Methodos **6** (1954), no. 21–22, 115–130.

28. S. M. Ulam, *A collection of mathematical problems*, Interscience, New York, 1960; cf. especially "A problem on matrices arising in the theory of automata," p. 30.

29. ———, *A study of combinatorial problems in patterns of growth*, this volume, pp. 215–224.

30. Hao Wang, *Universal Turing machines: an exercise in coding*, Z. Math. Logik Grundlagen Math. **3** (1957), no. 2, 69–80.

31. Dean A. Watkins and Edward F. Moore, *Letters*, Sci. Amer. **196** (1957), no. 2, 8–10.

References on sequential machines and iterative circuits

32. M. A. Aĭzerman, L. A. Gusev, L. I. Rozonoèr, I. M. Smirnova and A. A. Tal′, *On realization methods for a finite automaton of which the time pace is defined by the change of input state*, Avtomat. i Telemeh. **21** (1960), no. 12, 1576–1594 (Russian); translated in Automation and Remote Control **21** (1961), no. 12, 1126–1143.

33. Arthur W. Burks, *The logic of fixed and growing automata*, Proceedings of an International Symposium on the Theory of Switching, 2–5 April 1957, Part I, The Annals of the Computation Laboratory of Harvard University, Vol. 29, pp. 147–188, Harvard Univ. Press, Cambridge, Mass., 1959.

34. Samuel H. Caldwell, *Switching circuits and logical design*, Wiley, New York, 1958.

35. Calvin C. Elgot, *Decision problems of finite automata design and related arithmetics*, Trans. Amer. Math. Soc. **98** (1961), 21–51.

36. Seymour Ginsburg, *Connective properties preserved in minimal state machines*, J. Assoc. Comput. Mach. **7** (1960), no. 4, 311–325.

37. Frederick C. Hennie, *Analysis of bilateral iterative networks*, Trans. I.R.E. Professional Group on Circuit Theory **CT-6** (1959), no. 1, 35–45.

38. John Holland, *A universal computer capable of executing an arbitrary number of sub-programs simultaneously*, Proceedings of the Eastern Joint Computer Conference, December 1–3, 1959, no. 16, pp. 120–132.

39. D. A. Huffman, *The synthesis of sequential switching circuits*, J. Franklin Inst. **257** (1954), no. 3, 161–190; no. 4, 275–303.

40. ———, *Canonical forms for information-lossless finite-state logical machines*, Trans I.R.E. Professional Group on Circuit Theory **CT-6** (1959), supplement, 41–59.

41. E. J. McCluskey, Jr., *Assignment of carry variables in iterative networks*, Trans. Amer. Inst. Elec. Engrs. I. Communication and Electronics **79** (1961), no. 52, 772–778.

42. Gr. C. Moisil, *Algebraic theory of automatic machines*, Editura Tehnică, Bucharest, 1959. (Rumanian)

43. Edward F. Moore, *Gedanken-experiments on sequential machines*, *Automata Studies*, Annals of Mathematics Studies, No. 34, pp. 129–153, Princeton Univ. Press, Princeton, N.J., 1956.

44. David E. Muller and W. S. Bartky, *A theory of asynchronous circuits*, Proceedings of an International Symposium on the Theory of Switching, 2–5 April 1957, Part I, The Annals of the Computation Laboratory of Harvard University, Vol. 29, pp. 204–243, Harvard Univ. Press, Cambridge, Mass., 1959.

45. Sundaram Seshu and Myril B. Reed, *Linear graphs and electrical networks*, Addison–Wesley, Reading, Mass., 1961; cf. especially pp. 250–267, 301–310.

46. M. O. Rabin and Dana Scott, *Finite automata and their decision problems*, IBM J. Res. Develop. **3** (1959), no. 2, 114–125.

47. S. H. Unger, *A new computer oriented toward spatial problems*, Proc. I.R.E. **46** (1958), no. 10, 1744–1750.

48. Hao Wang, *Circuit synthesis by solving sequential Boolean equations*, Z. Math. Logik Grundlagen Math. **5** (1959), no. 3/4, 291–322.

Bell Telephone Laboratories, Incorporated,
Murray Hill, New Jersey