# John von Neumann, Arthur W. Burks - Theory of self-reproducing automata - 1966

Tom Rochette <tom.rochette@coreteks.org>

August 30, 2025 — 861fb9d0

### 0.1 Context

### 0.2 Learned in this study

### 0.3 Things to explore

- Higher class type implies it can describe a subset of the lower class type? For example, a program description (higher level) describing potential programs (lower level)
- For mutations to be probabilistically non-lethal, this means that most of the DNA code is likely to be a mutation chain that has expanded, while the basic mechanisms of self-reproduction are most likely minimal and left untouched

## 1 Overview

- The idea that communication between different systems is done through the translation of their respective language into the target's language (for example, from the sender's central nervous system language to the sender's language (English) to the receiver's language (English) and finally into the receiver's language central nervous system language)
  - Is the sender's CNS language different from his english language?
- The idea that digital computers are used to perform computations of exceedingly great logical depth (while it would be the opposite for natural automata) seems to imply that we should look for ANN architectures that are more parallel and less deep
- When faced with a very difficult problem, it is difficult to even define what the difficulties are

## 2 Notes

- (p14) A problem of any considerable complexity is composed of many subproblems, and flow diagrams and subroutines can be prepared for these in advance
- (p15) Thus the nervous system appears to be using a radically different system of notation from the ones we are familiar with in ordinary arithmetics and mathematics...
  ... whatever language the central nervous system is using, it is characterized by less logical and arithmetical depth than what we are normally used to.
  Thus logics and mathematics in the central nervous system, when viewed as languages, must be structurally essentially different from those languages to which our common experience refers.
  ... when we talk mathematics, we may be discussing a secondary language, built on the primary language truly used by the central nervous system
- (p19) Two basic questions of automata theory:
  - How can reliable systems be constructed from unreliable components?
  - What kind of logical organization is sufficient for an automaton to be able to reproduce itself?
- (p20) Three fundamental factors limiting the engineer's ability to build really powerful computers:

  - The size of the available components
  - The reliability of these components
  - A lack of a theory of the logical organization of complicated systems of computing elements
- (p22) Natural automata are more parallel in operation, digital computers more serial
- (p22) Von Neumann calculated the thermodynamical minimum of energy that must be dissipated by a computing element and concluded that in theory computing elements could be of the order of $10^10$ times more efficient in the use of energy than neurons
- (p23) Von Neumann thought that below a certain level, complexity is degenerative, and self-reproduction is impossible. He suggested that, generally speaking, in the case of simple automata a symbolic description of the behavior of an automaton is simpler than the automaton itself, but that in the case of exceedingly complex automata the automaton is simpler than a symbolic description of its behavior
- (p24) Pulses circulating in neuron cycles, the change of neural thresholds with use, the organization of the nervous system, and the coding of the genes together constitute such a (memory) hierarchy (similar to a digital computer memory hierarchy)
- (p25) Von Neumann thought that automata mathematics should be closer to the continuous (mathematics) and should draw heavily on analysis. He thought that the specific problems of automata theory require this, and he felt that there is a general advantage in an analytical as opposed to a combinatorial approach in mathematics
- (p26) Von Neumann wanted a probabilistic logic which would handle component malfunction as an essential and integral part of automata operation
- (p28) Von Neumann found an analog of thermodynamic degeneration in the theory of self-reproducing automata: below a certain minimum level, complexity and degree of organization are degenerative, but above that level they are not degenerative and may even increase.
- (p28) The efficiency of a computer depends on the proper balance of its different parts with respect to speed and size
- (p28) Von Neumann thought that the mathematics of automata theory should start with mathematical logic and move toward analysis, probability theory, and thermodynamics
- (p32) Of all automata of high complexity, computing machines are the ones which we have the best chance of understanding. In the case of computing machines the complications can be very high, and yet they pertain to an object which is primarily mathematical and which we understand better than we understand most natural objects
- (p38) The reason for using a fast computing machine is not that you want to produce a lot of information. After all, the mere fact that you want some information means that you somehow imagine that you can absorb it, and, therefore, wherever there may be bottlenecks in the automatic arrangement which produces and processes this information, there is a worse bottleneck at the human intellect into which the information ultimately seeps
- (p39) All these automata really consist of two important parts: the general switching part (an active part which affects the logical operations the automaton is supposed to perform), and the memory (which stores information, chiefly intermediate results which are needed for a while and are then discarded and replaced by others)
- (p41) There is one very important thing we do not know: whether nature has ever been subject to this handicap (the use of memory to store temporary computation), or whether natural organisms involve some much better memory device
- (p41) In any fast machine the memory you need is characterized by two data:
  - the capacity
  - the access time
- (p44) McCulloch and Pitts wanted to axiomatize an idealized neuron, which is much simpler than the real one. They believed that the extremely amputated, simplified, idealized object which they axiomatized possessed the essential traits of the neuron, and that all else are incidental complications, which in a first analysis are better forgotten
- (p44) A formal neuron is symbolically designated by a circle, which symbolizes the body of the neuron, and a line branching out from the circle, which symbolizes the axon of the neuron. A neuron has two states: it's excited or not
- (p45) No matter how you formulate your conditions, you can always put a neural network in the box

which will realize these conditions, which means that the generality of neural systems is exactly the same as the generality of logics

- (p47) Using the figures we have, which are not very good, but which are probably all right for an orientation, we conclude that apparently a network of about 2 billion relays does nothing but determine how to organize a visual picture
- (p47) There is a good deal in formal logics to indicate that the description of the functions of an automaton is simpler than the automaton itself, as long as the automaton is not very complicated, but that when you get to high complications, the actual object is simpler than the literary description
- (p48) Why has the digital notation never been used in nature, as far as we know, and why has this pulse notation been used instead? (in reference to neuron pulse)
  - Suggested answer: the frequency modulation scheme is more reliable than the digital scheme
- (p49) The memory requirements of the human nervous system are probably very large. Estimates have been made and they are of the order of 10^16 binary digits
- (p49) The main difficulty with the memory organ is that it appears to be nowhere in particular
- (p49) A lesser degree of complexity in an automaton can be compensated for by an appropriate increase of complexity of the instructions
  - Additional complexity is brought in by giving more elaborate instructions (separation of class types thus the avoidance of gödel incompleteness?)
- (p57) If you look at automata which have been built by men or which exist in nature you will very frequently notice that their structure is controlled only partly by rigorous requirements and is controlled to a much larger extent by the manner in which they might fail and by the (more or less effective) precautionary measures which have been taken against their failure. And to say that they are precautions against failure is to overstate the case, to use an optimistic terminology which is completely alien to the subject. Rather than precautions against failure, they are arrangements by which it is attempted to achieve a state where at least a majority of all failures will not be lethal. All we can try to do is arrange an automaton so that in the vast majority of failures it can continue to operate
- (p65) The human nervous system is roughly a million times more complicated than these large computing machines (ENIAC and IBM SSEC)
- (p65) The time in which a human nerve can respond is about 1/2 millisecond. However that time is not a fair measure of the speed of the neuron, because what matters is not the time in which the neuron responds, but the time in which it recovers, the time from one response to the next potential response. That time is, at best, 5 milliseconds
- (p65) Thus the nervous system has a million times as many components as these machines have, but each components of the machine is about 5 thousand times faster than a neuron
- (p66) The thermodynamical minimum of energy per elementary act of information
  - $3 \times 10^{-14}$ergs for the the thermodynamical minimum
  - $3 \times 10^{-3}$ergs for a neuron cell
  - $6 \times 10^{2}$ergs for a vacuum tube
- (p67) The disparity of $10^{11}$ between the thermodynamical minimum and the neuron might be due to something like a desire for reliability of operation
- (p68) There's no reason to believe that the memory of the central nervous system is in the switching organs (the neurons)
- (p69) The human organism is essentially a mixed system (digital and analog)
- (p69) It appears that digital mechanisms are necessary for complicated functions. Pure analog mechanisms are usually not suited for very complicated situations. The only way to handle a complicated situation with analog mechanisms is to break it up into parts and deal with the parts separately and alternately, and this is a digital trick
- (p70) The natural materials have some sort of mechanical stability and are well balanced with respect to mechanical properties, electrical properties, and reliability requirements
- (p72) In discussing a computing machine it is meaningless to ask how fast or how slow it is, unless you specify what type of problems will be given to it

## 2.1 Re-evaluation of the problems of complicated automata - Problems of hierarchy and evolution

- (p75) The medium which is fed to the automaton and which is produced by the automaton is completely different from the automaton. In fact, the automaton doesn't produce any medium at all; it merely modifies a medium which is completely different from it
- (p75) Draw up a list of unambiguously defined elementary parts. Imagine that there is a practically unlimited supply of these parts floating around in a large container. One can then imagine an automaton functioning in the following manner: It also is floating around in this medium; it essential activity is to pick up parts and put them together, or, if aggregates of parts are found, to them them apart
- (p77) I will introduce as elementary units neurons, a "muscle," entities which make and cut fixed contacts, and entities which supply energy
- (p77) What principles are involved in organizing these elementary parts into functioning organisms, what are the traits of such organisms, and what are the essential quantitative characteristics of such organisms?
- (p78) An object is of the highest degree of complexity if it can do very difficult and involved things
- (p78) Living organisms are very complicated aggregations of elementary parts, and by any reasonable theory of probability or thermodynamics highly improbably. That they should occur in the world at all is a miracle of the first magnitude; the only thing which removes, or mitigates, this miracle is that they reproduce themselves
- (p79) These organisms have the ability to produce something more complicated than themselves
- (p79) An automaton A, which can make an automaton B, must contain a complete description of B and also rules on how to behave while effecting the synthesis
- (p79) Complication, or productive potentiality in organization, is degenerative, that an organization which synthesizes something is necessarily more complicated, of a higher order, than the organization it synthesizes
- (p79) Complication is degenerative below a certain minimum level
- (p80) There is a minimum number of parts below which complication is degenerative, in the sense that if one automaton makes another the second is less complex than the first, but above which it is possible for an automaton to construct other automata of equal or higher complexity
- (p80) Von Neumann estimated this threshold to be in the millions of simple parts required to be over the minimum threshold
- (p81) Eight kinds of parts:
  1. A stimulus organ (p or q)
  2. A coincidence organ (p and q)
  3. An inhibitory organ (p and not-q)
  4. A stimuli producer
  5. A rigid member (can connect to other rigid members as well as to parts which are not rigid members)
  6. A fusing organ (when stimulated, welds or solders two parts together)
  7. A cutting organ (when stimulated, unsolders a connection)
  8. A muscle (used to produce motion)
- (p82) The construction automaton contains in its memory a description of the automaton to be constructed. Operating under the direction of this description, it picks up the parts it needs and assembles them into the desired automaton. To do this, it must contain a device which catches and identifies the parts that come in contact with it
- (p84) There is reason to suspect that our predilection for linear codes, which have a simple, almost temporal sequence, is chiefly a literary habit, corresponding to our not particularly high level of combinatorial cleverness, and that a very efficient language would probably depart from linearity
- (p85) The juxtaposition of two copies of the same thing is in no sense of higher order than the thing itself
- (p85) The general constructive automaton A has a certain creative ability, the ability to go from a description of an object to the object
- (p86) The general copying automaton B has the creative ability to go from an object to two copies of it

- (p86) Neither A nor B are self-reproductive
- (p86) The control automaton C is far from having any kind of creative or reproductive ability. All it can do is stimulate two other organs so that they act in certain ways, tie certain things together, and cut these things loose from the original system
- (p86) Yet the combination of the three automata A, B and C is auto-reproductive. Thus you may break a self-reproductive system into parts whose functioning is necessary for the whole system to be self-reproductive, but which are not themselves self-reproductive
- (p86) Let X be A + B + C + D, where D is any automaton. Then $(A + B + C) + \phi(A + B + C + D)$ produces $(A + B + C + D) + \phi(A + B + C + D)$. In other words, our constructing automaton is now of such a nature that in its normal operation it produces another object D as well as making a copy of itself
- (p86) The system (A + B + C + D) can undergo processes similar to the process of mutation
- (p86) If any element is changed at random in one of the automata A, B, or C, the system will usually not completely reproduce itself
- (p86) If there is a change in the description $\phi(A + B + C + D)$, the system will produce, not itself, but a modification of itself. Whether the next generation can produce anything or not depends on where the change is. If the change is in A, B, or C, the next generation will be sterile. If the change occurs in D, the system with the mutation is exactly like the original system, except that D has been replaced by D'. This system can reproduce itself, but its by-product will be D' rather than D. This is the normal pattern of an inheritable mutation

## 2.2 The Theory of Automata: Construction, Reproduction, Homogeneity

- (p92) Logical universality: When is a class of automata logically universal, i.e., able to perform all those logical operations that are at all performable with finite (but arbitrarily extensive) means?
- (p92) Constructability: Can an automaton be constructed, i.e., assembled and built from appropriately defined "raw materials," by another automaton?
- (p92) Construction-universality: Can any automaton be construction-universal, i.e., be able to construct every other automaton?
- (p92) Self-reproduction: Can any automaton construct other automata that are exactly like it? Can it be made, in addition, to perform further tasks, e.g., also construct certain other, prescribed automata?
- (p92) Evolution: Can the construction of automata by automata progress from simpler types to increasingly complicated types? Assuming some suitable definition of "efficiency," can this evolution go from less efficient to more efficient automata?
- (p93) Von Neumann considered five models of self-reproduction:
  - The kinematic model: Deals with the geometric-kinematic problems of movement, contact, positioning, fusing, and cutting, but ignores problems of force and energy.
    - Logical (switch) and memory (delay) elements: store and process information
    - Girders: provide structural rigidity
    - Sensing elements: sense objects in the environment
    - Kinematic (muscle-like) elements: move objects around
    - Joining (welding) and cutting elements: connect and disconnect elements
  - The cellular model: Self-reproduction takes place in an indefinitely large space which is divided into cells, each cell containing the same finite automaton. Each cell of the infinite structure of the cellular model contains a 29-state automaton
  - The excitation-threshold-fatigue model: The neuron has a designated threshold and a designated refractory period. The refractory period is divided into two parts, an absolute refractory period and a relative refractory period. If a neuron is not fatigued, it becomes excited whenever the number of active inputs equals or exceeds its threshold. The neuron cannot be excited during the relative refractory period; it can be excited during the relative refractory period, but only if the number of active inputs equals or exceeds a threshold which is higher than the normal threshold
  - The continuous model: The cellular model of self-reproduction is developed first, it is then reduced to the excitation-threshold-fatigue model, and finally, this model is described by non-linear partial differential equations

  - The probabilistic model: Automata in which the transitions between states are probabilistic rather than deterministic
- (p106) Von Neumann chose an infinite 2-dimensional array of square cells. Each cell is occupied by the same 29-state finite automaton, and each such automaton is connected to its four immediate neighbords in exactly the same way
- (p107) The 29 states each cell is capable of assuming falls into three categories:
  - unexcitable (1)
  - excitable (20)
    - 4 confluent states $C_{\epsilon\epsilon'}$, where $\epsilon$ and $\epsilon'$ range over 0 and 1, 0 symbolize quiescence, 1 symbolizes excitation
    - 8 ordinary transmission states $T_{0\alpha\epsilon}$, where $\alpha = 0, 1, 2, 3$ and $\epsilon = 0, 1$ as before
    - 8 special transmission states $T_{1\alpha\epsilon}$
  - sensitized (8)
- (p110) Ordinary stimuli are to be used for logical operations, taking the species of the neurons that are involved as fixed; that is, ordinary stimuli are to be used for the control and utilization of already satifactorily organized sub-assemblies
- (p110) Special stimuli are to be used for growth operations, involving the introduction of excitability, together with a new determination of the neuronic species, into previously unexcitable, or otherwise different areas
- (p116) Given a complete description of a secondary automaton, the first automaton can construct this second automaton
  - $\epsilon$: 0 if $x_1 \geq 0$, else 1
  - $\eta$: 0 if $y_1 \geq 0$, else 1
  - $|x_1|$: initial x coordinate of the bounding box containing the automaton
  - $|y_1|$: initial y coordinate of the bounding box containing the automaton
  - $\alpha$: length of the x side
  - $\beta$: length of the y side
  - $\lambda_{ij}$: cell state at coordinate i, j
- (p121) One cannot copy an existing automaton directly because by "testing" the state of each cell, the automaton runs into the trouble of potentially activating said cell, and thus possibly trigger a chain reaction to neighbor cells. This is why a description is used to create copies of automata instead of the automata themselves
- (p125) Two methods of obtaining a description of itself (or an automaton):
  - passive: the self-reproducing automaton contains within itself a passive description of itself and reads this description in such a way that the description cannot interfere with the automaton's operations
  - active: the self-reproducing automaton examines itself and thereby constructs a description of itself
- (p126) Von Neumann suggests that this second method would probably lead to paradoxes of the Richard type, and for this reason he adopts the first method
- (p128) While the secondary is being constructed, it must remain in its quasi-quiescent initial state until its construction is completed. At this point, the primary will deliver a single stimulus at the appropriate point of the secondary, which will transfer it i nto its normal mode of activity
- (p128) What does the primary do after it has completed the secondary and given it the starting stimulus?
  - Single-action: The simplest arrangement would be to let it return to a quasi-quiescent state which is identical with its original state
  - Sequential: An alternative possibility is to finish with the quasi-quiescent state, plus activity in a suitable terminal organ which imparts the starting stimulus again
- (p134) States are needed to express the properly logical or neuronal functions
- (p135) Since a line (of cells) must be able to pass a (neural) stimulus, each one of its cells must possess, for this purpose alone, a quiescent and an excited state. We call these the transmission states of the cell and designate them by the symbol T, with an index $\epsilon = 0$, 1 (quiescent, excited)
- (p135) The transmission (of a stimulus) must be a directed process, since the lines were directed to

connect definite points. Hence we must set up certain limitations. We may stipulate that a cell in a transmission state accepts a stimulus only from one, definite direction, its input direction. That is, an excited transmission cell brings an immediate neighbor into the excited transmission state only if the former lies in the latter's input direction. Alternatively, we may also stipulate that a cell in a transmission state emits a stimulus only in one, definite direction, its output direction. Finally, we may make both stipulation together
  - After trying various models along these lines, it appears most convenient to stipulate a definite output direction
  - In order to avoid certain uncontrolled, and hence undesirable, return-stimulation phenomena, it seems desirable, while not prescribing any particular input direction, to specify that the output direction is insensitive to inputs
- (p136) The + neuron (OR) can implemented by simply using a transmission cell. It will accept input on its four side, except the output side and output to its output side if any input side is excited
- (p136) The · neuron (AND) uses confluent states (C). Confluent states are are defined as cells for which 1, 2 or 3 of its immediate neighbors are transmission cells with their output directions pointing at it
- (p138) The - neuron (NOT) must be ordinarily excited but it must be made quiescent by an input stimulation
- (p139) The split allows an output to branch into multiple inputs. Using the confluent states with 1 input, you can have up to 3 possible ouputs
- (p139) The purpose of the special class of excited states is to induce growth, i.e., to transfer cells from unexcitable to excitable states, and within the latter category also to determine the state's species
- (p140) It is desirable not only to be able to effect transfers from the unexcitable state U into the excitable states (T and C), but also to have this process reversible, that is, to be able to effect transfers from the excitable states into U
- (p157) We have free timing when the emission of a certain stimuli (at a specified point and in a certain order) is provided for, but the time intervals or delays from one such emission to the next are not prescribed
- (p157) We have rigid timing when the delays of the above pattern are numerically specified
- (p159) Pulser: This organ has an input a and an output b. Upon a stimulus at a it will emit a prescribed sequence $i^1 \cdots i^n$ at b. This pulser has the symbol $\mathbf{P}(\overline{i^1 \cdots i^n})$. The sequence $i^1 \cdots i^n$ which can be described at will, is its characteristic; n is its order
- (p163) Periodic pulser: This organ has two inputs $a_+$, $a_-$, and an output b. Upon a stimulus at $a_+$, it will begin to emit a prescribed, periodically repeated sequence, say $\overline{\overline{i^1 \cdots i^n}}$ at b. upon a stimulus at $a_-$ this emission at b will be stopped. This period pulser has the symbol $\mathbf{PP}(\overline{\overline{i^1 \cdots i^n}})$. The sequence $\overline{\overline{i^1 \cdots i^n}}$ which can be described at will, is its characteristic; n is its order
- (p163) The simplest periodic repeat is a (closed) cycle of transmission states. The shortest cycle of this kind has a period of 4. Clearly, any (such) cycle must have an even period; hence the general cycle has the period 2l, l = 2, 3, ....
- (p164) An output b must be issued from this repeater. Hence, at least one of its cells must be able to stimulate in two directions; i.e. it must be a C (AND neuron with confluent states)
  - The C raises the length of the period to 2l + 1
  - If an even period is wanted, a second C must be inserted
- (p165) To stop the periodic repeater, we want to convert any cell part of the cycle into a U
  - However, if we want to restore the periodic repeater, we will have to restore this cell to its original state
  - Thus, it is simpler to convert a C than a T (which has a direction)
- (p166) To restore the U to a C, we need a 5 consecutive pulse $\mathbf{P}(\overline{11111})$
- (p175) A decoding organ will output a single stimulus at output b if a sequence of bits containing the characteristic of the organ is fed at input a ($\mathbf{D}(\overline{1001})$ will accept $\overline{1001}$, or $\overline{1111}$, but not $\overline{0100}$ or $\overline{0001}$)
- (p175) A decoding organ is to be distinguished from a "recognizing" device which produces an output for some particular sequence and no other ((X & Y) == Y vs X == Y, X being the input and Y being the characteristic/mask)
- (p176) Top row confluence states are used to detect coincidence, while bottom row confluence states

are used to introduce delay between paths of odd amounts

- (p176) Decoding: This organ has an input a and an output b. The ideal form would be one which, upon the arrival of a prescribed sequence, say $\overline{i^1 \cdots i^n}$, at a, and only then, will emit a single stimulus at b. However, for our specific applications of this organ a simpler requirement will do, and this is the one that we are going to use.
  - Let a sequence, e.g. $\overline{i^1 \cdots i^n}$, be prescribed. Upon arrival of any sequence $\overline{j^1 \cdots j^n}$ at a, which contains all the stimuli of $\overline{i^1 \cdots i^n}$, and only then, it will emit a single stimulus at b
- (p176) This decoding organ has the symbol $\mathbf{D}(\overline{i^1 \cdots i^n})$. The sequence $\overline{i^1 \cdots i^n}$, which can be prescribed at will, is its characteristic, and n is its order
- (p191) The coded channel performs a function which is necessitated by the peculiar narrowness of 2-dimensional space
- (p192) A multiple line-crossing organ that would be universally useful may be described as follows. It is a rectangular area A, on whose periphery there are various inputs and outputs. Every input is designated by a symbol $a_\nu$, $\nu = 1, \cdots, n$; every output is designated by a symbol $b_\nu$, $\nu = 1, \cdots, n$. There may be several $a_\nu$ with the same $\nu$, and there may be several $a_\nu$ with the same $\nu$. The order in which all of these $a_\nu$ and $b_\nu$ are arranged around the periphery of A may be prescribed in any manner whatsoever. It is desired that upon stimulation of an $a_\nu$ all the $b_\nu$'s (with the same $\nu$) should respond, with appropriate (and not necessarily equal) delays
- (p193) The obvious way to achieve this (signal crossing) is by coding and decoding

## 2.3 Design of a tape and its control

- (p201) The following units are involved in the tape and its control
  - A linear array **L** for storing information: "zero" is represented in cell $x_n$ by state **U** and "one" is represented by state $\downarrow 0$
  - A connecting loop $\mathbf{C_1}$ for reading an arbitrary cell $x_n$
  - A timing loop $\mathbf{C_2}$, used in modifying the length of the connecting loop $\mathbf{C_1}$
  - A memory control **MC**, used to control the operations of **L**, $\mathbf{C_1}$ and $\mathbf{C_2}$
  - The constructing unit **CU**, which controls **MC**
- (p201) The constructing unit **CU** sends a pulse to the memory control **MC** signifying that cell $x_n$ is to be read. This pulse causes the sequence $\overline{10101}$ to enter the connecting loop $\mathbf{C_1}$. The sequence $\overline{10101}$ then enters cell $x_n$ with the following effect:
  - if $x_n$ is in state **U** the sequence $\overline{1010}$ changes it into state $\downarrow 0$ and $\overline{1}$ returns to **MC**
  - if $x_n$ is in state $\downarrow 0$ the whole sequence $\overline{10101}$ returns to **MC**
- (p201) A $\overline{1}$ vs $\overline{10101}$ discriminator $\Psi$ detects the output and informs the constructing unit **CU** whether $x_n$ stored a "zero" or a "one." In either case cell $x_n$ is left in state $\downarrow 0$
- (p201) The constructing unit **CU** then tells the memory control **MC** whether the loop $\mathbf{C_1}$ is to be lengthened so as to pass through cell $x_{n+1}$ or shortened so as to pass through cell $x_{n-1}$, and whether cell $x_n$ is to be left in state **U** ("zero") or $\downarrow 0$ ("one"). Loop $\mathbf{C_1}$ is used to time the lengthening (or shortening) of loop $\mathbf{C_2}$. Then loop $\mathbf{C_2}$ is used to time the lengthening (or shortening) of loop $\mathbf{C_1}$. The new bit of information is written in cell $x_n$ while loop $\mathbf{C_1}$ is being lengthened (or shortened). At the end of the whole process the memory control **MC** sends a finish signal to the constructing unit **CU**
- (p202) The primary purpose of **CU** is to carry out the construction of a secondary automaton whose description is stored in **L**
- (p203) In order to make the external memory **L** useful, it is necessary to construct the means for "exploring" L. This "exploration" includes reading **L** at any prescribed place, and also altering it in any desired way at any prescribed place
- (p208) The linear array **L** is a linear sequence of cells $x_n, n = 0, 1, 2, \ldots$. We assume it to be horizontal, extending, and numbered, from left to right. We assume furthermore that **MC** lies to the left of **L**
- (p208) The obvious way for reading cell $x_n$ consists of having a line of ordinary transmission states leading to it from **MC**, and then back from it to **MC**
- (p208) Cell $x_n$ can be "read" by sending a stimulus into the $\mathbf{C_1}$ loop at its entry $v_1$, and observing what emerges at its exist $w_1$
- (p224) Altering $x_n$ in **L** (this procedure is done when lengthening or shortening **L**, but here we describe

the basic idea)
  - Modify the upper line from $\overset{0}{\rightarrow}$ into $\overset{1}{\rightarrow}$ until n-1 is reached
  - Modify the cell above n into $\downarrow 1$
  - Toggle the $x_n$ cell
  - Convert back the top row cells from $\downarrow 1$ to $\downarrow 0$ and $\overset{1}{\rightarrow}$ to $\overset{0}{\rightarrow}$
- (p226) The most important part of **MC** are the read-write-erase unit **RWE** and its control **RWEC**
- (p227) The **RWEC** contains 16 control units **CO**; each **CO** contains a $\mathbf{PP}(\overline{\overline{1}})$ which is active while that **CO** is in control. The **RWEC** also contains four $\mathbf{PP}(\overline{\overline{1}})$ which are used to store the bit which is to be written in cell $x_n$
- (p228) The basic operation of reading, writing, and lengthening (or shortening) the loops $\mathbf{C_1}$ and $\mathbf{C_2}$ is carried out in two stages. First, **CU** sends a signal to **MC** telling it to read; **MC** then reads cells $x_n$ and sends the result to **CU**. Second, **CU** sends signals to **MC** telling it what to write in cell $x_n$ and whether to lengthen or shorten the loops $\mathbf{C_1}$ and $\mathbf{C_2}$; **MC** then executes these instructions and sends a completion signal to **CU**
- (p266) A Turing machine has two main parts: a tape unit with an indefinite memory capacity, and a finite automaton which can interact with this tape unit
- (p266) A constructing automaton has two corresponding parts: a tape unit $\mathbf{MC}+\mathbf{L}$ and a constructing unit **CU** which directs the construction of a secondary automaton on the basis of the information stored in **L**. Thus **CU** is a finite automaton which interacts with the tape unit and also performs the function of construction
- (p270) Any abitrary Turing machine can be embedded as an initially quiescent automaton, and a fortiori a universal Turing machine can be embedded as an initially quiescent automaton
- (p271) Von Neumann's universal constructing automaton consists of a constructing unit **CU** combined with a tape unit $\mathbf{MC}+\mathbf{L}$. He calls this the "primary automaton," and he calls the initially quiescent automaton to be constructed the "secondary automaton"
- (p272) The primary automaton operates on the (possibly remote) secondary area by means of a "constructing arm" or information path which extends from the primary automaton to the secondary area.
  - The constructing unit **CU** first builds the constructing arm out to the secondary area.
  - Then **CU** sends signals down the arm which construct the secondary automaton and provide its starting stimulus
  - Finally, **CU** withdraws the constructing arm
- (p275) Five operations for the constructing arm:
  - Horizontal advance
  - Vertical advance
  - Horizontal retreat with $\gamma-\delta$
  - Vertical retreat with $\gamma-\delta$
  - Injection of the starting stimulus
- (p279) At this point I should like to speculate on von Neumann's thoughts concerning his design for a self-reproducing automaton at the time he stopped working on the manuscript. His design had turned out to be much more complex than he had anticipated. After developing his two-path construction procedure, he must have realized that it could be used on **L**, and that this would greatly simplify the design of **MC** and of the whole machine. Realizing this, he would have wanted to redesign is self-reproducing automaton along new lines
- (p280) The information concerning the location and size of the secondary automaton, as well as a complete description of the secondary automaton, is stored on the linear array **L**.
  - First comes a period.
  - Then come the location and size parameters $x_1$, $y_1$, $\alpha$, and $\beta$, each followed by a comma
  - Next comes the sequence of $\lambda_{i,j}$'s describing the secondary automaton cell by cell, for $i=0,\cdots,\alpha-1$ and $j=0,\cdots,\beta-1$
  - The sequence of $\lambda_{i,j}$'s is terminated by a period, which also marks the end of the information on the tape
- (p290) If part of the secondary automaton were active during construction, it could interfere with the construction process. Von Neumann solved this problem by stipulating that the initial state of the

secondary automaton is to be composed entirely of quiescent states. After the secondary automaton has been completed, it may be rendered active by a starting stimulus injected into its periphery
- (p291) The class of automata constructible by the universal constructing automaton is a proper subclass of the automata which can be specified as part of an initial cell assignment of von Neumann's cellular structure

# 3 See also

# 4 References


- Neumann, John von, and Arthur W. Burks. "Theory of self-reproducing automata." (1966).
- https://archive.org/details/theoryofselfrepr00vonn_0
- Richard's paradox