Alma Mater Studiorum · Università di Bologna

School of Sciences
Bachelor Degree in Computer Science

Von Neumann's Self-Reproducing Automaton

How a Theory is Born

Supervisor:
Chiar.mo Prof.
Simone Martini

Presented by:
Mattia Ferrarini

I Session
Academic Year 2024/2025

To my parents and my sister

Sommario

Tra il 1946 e il 1956, John von Neumann si dedicò allo studio degli automi con l'obiettivo di sviluppare una teoria generale che unificasse sistemi biologici e artificiali. Nel corso del lavoro, si interessò al problema dell'auto-riproduzione artificiale e ideò un automa cellulare auto-riproducente, il quale diede avvio alla teoria degli automi cellulari. Questa tesi esplora il lavoro di von Neumann, concentrandosi in particolare su motivazioni, obiettivi e metodologie.

Analizzando la vita e le opere dell'autore, individuiamo l'origine dell'interesse per gli automi nell'obiettivo di costruire calcolatori migliori e illustriamo come il riconoscimento della superiorità biologica portò von Neumann a tentare di raggiungere questo obiettivo tramite una teoria che comprendesse sia i calcolatori sia gli organismi viventi.

Successivamente, determiniamo le altre caratteristiche fondamentali del suo lavoro: il concetto di "complessità", ovvero la capacità degli automi di svolgere operazioni complesse, caratteristica che von Neumann riteneva fondamentale; l'ampio uso del metodo assiomatico, che prevede di assumere l'esistenza di alcuni componenti elementari con comportamenti ben definiti; il focus strutturale, che ignora i dettagli fisici per concentrarsi su quelli funzionali e organizzativi.

Trattiamo poi il problema dell'auto-riproduzione. Evidenziamo l'importanza della sua assiomatizzazione e colleghiamo il problema al concetto di complessità e all'unificazione di automi biologici e artificiali, ma mostriamo anche come esso rappresenti un allontanamento dagli obiettivi e dalla metodologia di von Neumann. Presentiamo quindi il progetto dell'automa auto-riproducente, mostrando come in esso confluirono le influenze di Turing, McCulloch e Pitts, oltre ai risultati del lavoro sull'EDVAC. Chiariamo infine le limitazioni di tale automa e cosa esso riuscì a dimostrare in relazione al problema dell'auto-riproduzione artificiale.

Nel complesso, la tesi mostra il carattere unico e innovativo del lavoro di von Neumann, il quale immaginò una teoria interdisciplinare, combinò obiettivi ingegneristici con un approccio formale, adottò metodologie matematiche prima ancora della nascita dell'informatica come disciplina e riuscì a utilizzare i pochi risultati sugli automi a sua disposizione per progettare un automa auto-riproducente.

Contents

Introduction **1**

1 Background **3**

1.1 Work on computing machines 3

1.1.1 A consulting career 3

1.1.2 Work during the war 4

1.1.3 The ENIAC and the EDVAC 4

1.1.4 Post-war work 7

1.2 Turing's work on computable numbers 9

1.2.1 The Turing machine 9

1.2.2 The universal machine 10

1.3 McCulloch and Pitts' neurons 10

1.3.1 A mathematical model of the nervous system 10

1.3.2 Von Neumann's interpretation 12

1.4 Relationship with the biomedical community 14

1.5 The interdisciplinary decade 15

2 A Theory of Automata **19**

2.1 A unifying theory 19

2.1.1 Artificial automata 20

2.1.2 Biological automata 21

2.1.3 The need for a theory 22

2.2 The concept of complication 24

2.2.1 The potentiality to do things 24

2.2.2 Complication as the unifying factor 25
2.3 Methodology 25
2.3.1 Axiomatization 26
2.3.2 Results of axiomatization 27
2.3.3 Interdisciplinarity 28
2.4 Logic and engineering 29
2.4.1 The organizational focus 29
2.4.2 The engineering goal 30
2.4.3 Computing for computing's sake 30
2.5 Summary of the theory's characteristics 31
2.6 Works of the theory 32

3 Self-Reproduction 33
3.1 The problem of self-reproduction 33
3.2 Axiomatization of the problem 36
3.2.1 The kinematic model 36
3.2.2 The cellular model 37
3.3 Why self-reproduction? 40
3.3.1 A more complete discussion of automata 40
3.3.2 Self-reproduction and complication 41
3.3.3 Connections to the theory of automata 42
3.4 Self-reproduction as a deviation 42
3.4.1 Deviation from the engineering goal 42
3.4.2 Methodological deviations 43

4 The Self-Reproducing Automaton 47
4.1 Synthetic description 47
4.2 Integral description 49
4.2.1 The 29 cell states 50
4.2.2 Ordinary and special stimuli 53
4.2.3 The automaton's description 54
4.2.4 The linear array L 54

4.2.5 The memory control MC 55
4.2.6 The constructing unit CU 56
4.2.7 The universal constructor 57
4.2.8 The self-reproducing automaton 59
4.3 Von Neumann's results 60
4.3.1 Logical universality 61
4.3.2 The class of constructible automata 62
4.3.3 Limitations of von Neumann's cellular model 63
4.3.4 Answers to the questions 65

Conclusion **67**

Acknowledgments **69**

Bibliography **71**

List of Figures

1.1 McCulloch and Pitts' nets 11
1.2 Attendees to the 10th Macy Conference 17
1.3 Participants to the Hixon Symposium 18

3.1 The cellular model 40

4.1 Ordinary transmission states 51
4.2 A confluent state 52
4.3 Sensitized states and the transformation tree 53
4.4 The linear array L and the memory control MC 56
4.5 The universal constructor M_c and the constructing arm 59
4.6 The self-reproducing automaton 61
4.7 A non-costructible automaton 63

Introduction

John von Neumann and his theory of automata

John von Neumann (1903–1957) was one of the pioneers of computing. Through his consulting work for the US government, he developed an interest in computing equipment, which ultimately led him to contribute to the design of the EDVAC—the world's first stored-program computer—which inspired the design of the first generation of all modern computers. He believed that computers could significantly advance various scientific fields, including engineering, physics, and mathematics, and looked for ways to design better ones. He recognized that biological systems, such as the human nervous system, exhibited capabilities beyond those of artificial systems and sought a theory that could unify both types of automata. Indeed, he believed that the lack of such a general theory was an obstacle to the creation of more complex computers.

Von Neumann worked on developing this theory, where a central role was played by the concept of complication, i.e., the complexity of an automaton's operations. Through the analysis of biological systems, he became interested in the problem of automata self-reproduction. He saw self-reproduction as a defining feature of living organisms and as a sign of complexity. Possibly aiming at unifying artificial and biological automata while also proving the former are capable of complex operations, he studied self-reproduction and designed a self-reproducing artificial automaton.

Purpose and structure of this work

Von Neumann carried out most of his research on automata between 1946 and 1956, at a time when computer science was yet to become a discipline and even the study of automata lacked organization and autonomy. He effectively tried to create a new theory from a very limited prior foundation. Therefore, it is historically significant to examine his motivations, objectives, and approach. Moreover, even though the unifying theory of biological and artificial automata he envisioned was never fully realized, his model of self-reproduction holds historical significance for establishing the foundations of cellular automata theory[1].

This thesis explores von Neumann's research on automata self-reproduction with a particular focus on his motivations, goals, and methodology. First, von Neumann's efforts towards a unifying general theory of automata are outlined. Then, the problem of self-reproduction is introduced, showing both its connections to the general theory and the ways in which it diverges from it. Finally, the self-reproducing automaton designed by von Neumann is presented. The remainder of this work is structured as follows:

- **Chapter 1** covers the main experiences and scientific results that influenced von Neumann's work on automata and self-reproduction;
- **Chapter 2** describes von Neumann's theory of automata, along with his motivations, goals, and methodology;
- **Chapter 3** introduces the problem of self-reproduction, shows how it ties to the general theory of automata, but also argues that it can be interpreted as a deviation from it;
- **Chapter 4** describes the self-reproducing automaton designed by von Neumann (and completed by Arthur W. Burks), which proves the feasibility of self-reproduction in artificial systems.

[1] See `https://plato.stanford.edu/entries/cellular-automata/`.

Chapter 1

Background

Von Neumann was a man of many interests, both theoretical and practical. During the course of his life, he made important contributions to various fields, including computing, mathematics, physics, and economics. In this chapter, we describe the main interests and influences that came together in his work on automata. We begin by briefly covering his experiences with computing machines. Then, we discuss two papers whose results became important building blocks of his studies: Turing's "On Computable Numbers, with an Application to the *Entscheidungsproblem*" (Turing 1937) and McCulloch and Pitts' "A Logical Calculus of the Ideas Immanent in Nervous Activity" (McCulloch and Pitts 1943). We conclude by describing von Neumann's relationship with the biomedical community and the interdisciplinary nature of the decade following World War II.

1.1 Work on computing machines

1.1.1 A consulting career

Von Neumann was born in Budapest, Hungary, in 1903. He earned a degree in chemical engineering from the Eidgenossische Technische Hochschule (ETH) in Zurich, Switzerland, and a Ph.D. in mathematics from the University of Budapest. In 1930, he emigrated to the USA, where he was later appointed professor at the newly founded Institute for Advanced Study (IAS) in Princeton. In 1937, he

became a naturalized US citizen and, soon after, he began working as a government consultant, mostly in defense research. His first occupation was at the Ballistics Research Laboratory (BRL) in Aberdeen, where he was introduced to military science, which in turn introduced him to applied sciences (Aspray 1990, 26).

1.1.2 Work during the war

During World War II, his consulting work intensified, eventually occupying all of his time. It was in this period that von Neumann became importantly involved with computing. Indeed, in 1943, he visited the Nautical Almanac Office (NAO) in Bath, England, where he had the opportunity to work with a computing machine to automate gas-dynamical calculations. This experience was certainly significant for him, as he wrote in a letter to Oswald Veblen about the visit that he had developed an "obscene interest in computational techniques" (Aspray 1990, 27).

The interest found realization when he joined the Manhattan Project in Los Almos in the autumn of 1943. There, he mostly worked on the hydrodynamics of implosion and explosions. The equations governing this phenomena could not be treated analytically, but needed to be treated either through experimentation or numerical simulations. Such simulations involved the integration of hyperbolic partial equations, which quickly grew the computing needs of the Los Almos Laboratory. This led von Neumann to seek high-speed computing equipment.

1.1.3 The ENIAC and the EDVAC

Through the research of high-speed equipment and the rather fortunate meeting with Herman Goldstine, a captain in Army Ordnance, von Neumann learned about the ENIAC and the work carried out at the University of Pennsylvania's Moore School of Electrical Engineering. The ENIAC was a complicated machine composed of thirty semiautonomous units which operated simultaneously. The computation to be carried out by the machine had to be manually defined before each execution by setting switches and connecting different units (Goldstine and Goldstine 1946). Von Neumann visited the Moore School of Electrical Engineering,

but most choices had already been made, so he did not contribute to the design of the ENIAC.

However, he collaborated to the design of its successor, the EDVAC. While there has been a lot of debate about his contributions to the project (Aspray 1990, 36–46), it is generally believed that he vastly contributed to the logical design of the machine (Aspray 1990, 38–40; von Neumann 1966, 9). Such design was described within the "First Draft of a Report on the EDVAC" (von Neumann 1945), which introduced the first stored-program machine and what later came to be known as "von Neumann's architecture". In the remainder of this section, we will focus on those elements and characteristics of the draft that will be most significant for treating von Neumann's theory of automata.

Structural approach The draft had a primarily *logical* focus, rather than an engineering one. It described the structure of the stored-program machine, the internal organization of its parts, and the relationships between these. Specifically, von Neumann designed the EDVAC to be composed of three main parts: a central arithmetical part (CA) for the execution of arithmetical operations; a central control (CC) for the logical control of operations; and a memory (M) for storing data. The machine was completed by three peripheral components: an external recording medium, input and output systems (von Neumann 1945, 3–7).

Analogy with the human nervous system The components of the EDVAC were described as "organs", which are in correspondence to those of the human nervous system:

> The three specific parts CA, CC [together C] and M correspond to the associative neurons in the human nervous system. It remains to discuss the equivalents of the sensory or afferent and the motor or efferent neurons. These are the input and the output organs of the device, and we shall now consider them briefly.
>
> von Neumann 1945, 6

Abstraction and neurons In attempting to design the functioning of the various organs, von Neumann realized that "the decisions regarding the arithmetical and the logical control procedures of the device, as well as its other functions, can only be made on the basis of some assumptions about the functioning of the elements" (von Neumann 1945, 21). Instead of introducing engineering considerations, he believed it was best to temporarily abstract the physical details of the components:

> The ideal procedure would be to treat the elements as what they are intended to be: as vacuum tubes. However, this would necessitate a detailed analysis of specific radio engineering questions at this early stage of the discussion, when too many alternatives are still open to be treated all exhaustively and in detail. Also, the numerous alternative possibilities for arranging arithmetical procedures, logical control, etc., would superpose on the equally numerous possibilities for the choice of types and sizes of vacuum tubes and other circuit elements from the point of view of practical performance, etc. All this would produce an involved and opaque situation in which the preliminary orientation which we are now attempting would be hardly possible.
>
> In order to avoid this we will base our considerations on a hypothetical element, which functions essentially like a vacuum tube—e.g. like a triode with an appropriate associated RLC-circuit—but which can be discussed as an isolated entity, without going into detailed radio frequency electro-magnetic considerations. We re-emphasize: this simplification is only temporary, only a transient standpoint, to make the present preliminary discussion possible. After the conclusions of the preliminary discussion the elements will have to be reconsidered in their true electromagnetic nature. But at that time the decisions of the preliminary discussion will be available, and the corresponding alternatives accordingly eliminated.
>
> von Neumann 1945, 21–22

The "hypothetical element" von Neumann decided to use was the formal neuron of McCulloch and Pitts (1943)[1]. This element has an "all-or-none" character and can be in two states: quiescence and excitation. It can receive stimuli from other neurons and emit a stimulus as a result. Due to its binary digital nature, the formal neuron is suitable to model circuit elements such as vacuum tubes.

1.1.4 Post-war work

Overall, as noted by (Aspray 1990, 48), "the war was an education in computing for von Neumann": through his work as a consultant, he understood the importance of automatic computation and, through his experiences with computing equipment, he learned the foundations of machine design. Following the end of the war, he kept working on these technologies, especially in the context of scientific research.

As a matter of fact, von Neumann believed computers to be instrumental to many areas of science, as he explained in the "Computing Machines in General" lecture (von Neumann 1949, 31–41) that he delivered at the University of Illinois in 1949. During the lecture, he claimed that numerical computing played a large role in engineering and that it could also have a large impact on physics:

> It's perfectly clear that numerical computing plays a large role in engineering. If more computing and faster computing could be done, one would have even more uses for computing in engineering. [...]
>
> However, effective computing plays a role in physics which is larger than the role one would expect it to have in mathematics proper. For instance, there are large areas of modern quantum theory in which effective iterative computing could play a large role. A considerable segment of chemistry could be moved from the laboratory field into the purely theoretical and mathematical field if one could integrate the applicable equations of quantum theory. Quantum mechanics and

[1]The work by McCulloch and Pitts will be covered in more detail in Section 1.3.

> chemistry offer a continuous spectrum of problems of increasing difficulty and increasing complexity, treating, for example, atoms with increasing numbers of electrons and molecules with increasing numbers of valence electrons. Almost any improvement in our standards of computing would open important new areas of application and would make new areas of chemistry accessible to strictly theoretical methods.
>
> von Neumann 1949, 32–33

He went on to explain how computing could also be beneficial for pure mathematics. According to him, applying methods of pure mathematics only becomes feasible once one has a "reasonably intuitive heuristic relation to the subject" (von Neumann 1949, 33). In von Neumann's view, computing could provide a more flexible way to develop a heuristic understanding than experimentation:

> If one could calculate solutions in certain critical situations like those we have mentioned, one would probably get much better heuristic ideas. I will try to give some indications of this later, but I wanted to point out that there are large areas in pure mathematics where we are blocked by a peculiar inter-relation of rigor and intuitive insight, each of which is needed for the other, and where the unmathematical process of experimentation with physical problems has produced almost the only progress which has been made. Computing, which is not too mathematical either in the traditional sense but is still closer to the central area of mathematics than this sort of experimentation is, might be a more flexible and more adequate tool in these areas than experimentation.
>
> von Neumann 1949, 34–35

During the lecture, von Neumann also mentioned non-linear problems as ones that could greatly benefit from advancements in high-speed computing. Indeed, following the war, he worked on numerical analysis in the context of several of these problems, including fluid dynamics, shocks, and weather prediction.

1.2 Turing's work on computable numbers

In the 1936 paper "On Computable Numbers, with an Application to the *Entscheidungsproblem*" (Turing 1937), Alan Turing addressed David Hilbert's *Entscheidungsproblem* (decision problem). The problem consists of determining if there exists a mechanical procedure that can decide if any given statement in first-order logic is universally valid. To answer the question, Turing introduced a computation formalism that became known as the Turing machine. Although Turing was interested in formal logic rather than in automata, as von Neumann himself noted (von Neumann 1949, 49), some aspects of his work were influential on von Neumann's theory of self-reproducing automata.

1.2.1 The Turing machine

In defining his theoretical machine, Turing took inspiration from the way a human computer would carry out the computation of a number. He especially noted that such a computer can observe only a finite number of symbols at any time and takes actions based on one of a finite set of "states of mind" (Turing 1937, 250). As a result, his machine has a finite set of states and uses symbols from a finite alphabet. It operates on an infinite tape, which is divided into "squares", each containing a symbol. At any time, the machine is scanning only one of the tape squares. Based on the symbol it reads and its state, it can write a new symbol in the scanned square, change its state, and shift the scanned square one position to the right or to the left (Turing 1937, 231–232).

The Turing machine is particularly significant because it "provided a mathematically precise characterization of the basic functions and components common to all computing automata" (Aspray 1985, 131). We will see in Chapter 4 which of these functions and components von Neumann equipped his self-reproducing automaton with.

1.2.2 The universal machine

Turing also devised a universal machine that, given a tape on which is written a description of a second machine M, will compute the same sequence as M (Turing 1937, 241–246). The idea of a machine's description will be fundamental in von Neumann's work on self-reproducing automata. Indeed, he considered Turing's results on the universal machine to be the most important of his work:

> [The universal machine] is able to imitate any automaton, even a much more complicated one. Thus a lesser degree of complexity in an automaton can be compensated for by an appropriate increase of complexity of the instructions. The importance of Turing's research is just this: that if you construct an automaton right, then any additional requirements about the automaton can be handled by sufficiently elaborate instructions.
>
> von Neumann 1949, 50

1.3 McCulloch and Pitts' neurons

1.3.1 A mathematical model of the nervous system

In 1943, the neurophysiologist Warren McCulloch and the logician Walter Pitts published the seminal paper "A Logical Calculus of the Ideas Immanent in Nervous Activity" (McCulloch and Pitts 1943), which suggested for the first time a relationship between neural networks and computation (Piccinini 2020, 108). By making some assumptions and simplifications about the behavior of human neurons, they proposed a mathematical model of the nervous system and showed it is equivalent to propositional logic.

The formal neurons and nets McCulloch and Pitts described human neurons as having a soma and an axon, with synapses connecting the axon of a neuron to the soma of another, allowing the propagation of an impulse from the former to the latter. Impulses can be excitatory or inhibitory. If the sum of the excitations

received by a neuron exceeds a certain threshold and no inhibitory impulse is received, an impulse is initiated.

The authors assumed that the neurons have an "all-or-none" behavior (they either fire or not) and that they fire at discrete time intervals (McCulloch and Pitts 1943, 101). Based on these assumptions, they developed a mathematical model of the human neurons. Their formal neurons fire at time t if and only if, at time $t-1$, they did not receive an inhibitory input, and the received excitatory input exceeded a certain threshold. Intuitively, formal neurons can be grouped in nets by connecting their inputs and outputs, as shown in Figure 1.1.

Relationship with propositional logic McCulloch and Pitts stated that "the 'all-or-none' law of nervous activity is sufficient to insure that the activity of any neuron may be represented as a proposition" (McCulloch and Pitts 1943, 100). They also showed how to write an expression describing the behavior of a network and how to construct a network behaving according to a certain logical expression.

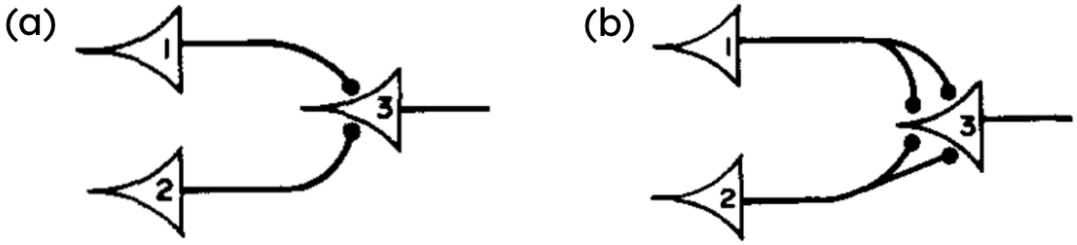

Figure 1.1: McCulloch and Pitts' nets from McCulloch and Pitts (1943). The numbered triangles are the neurons, each of which is assumed to have a threshold of two. The lines ending with a dot represent excitatory connections. (a) is the net for conjunction; whereas (b) is the net for disjunction.

Connecting artificial and biological computation Towards the end of their work, McCulloch and Pitts claimed that their nets, if augmented with an infinite tape and appropriate elements to act on it, can compute the same numbers a Turing machine can. Moreover, they set forth the "first published link between mathematical theory of computation and brain theory" (Piccinini 2020, 116–117):

> It is easily shown: first, that every net, if furnished with a tape, scanners connected to afferents, and suitable efferents to perform the necessary motor-operations, can compute only such numbers as can a Turing machine; second, that each of the latter numbers can be computed by such a net; and that nets with circles can be computed by such a net; and that nets with circles can compute, without scanners and a tape, some of the numbers the machine can, but no others, and not all of them. This is of interest as affording a psychological justification of the Turing definition of computability and its equivalents, Church's λ-definability and Kleene's primitive recursiveness: if any number can be computed by an organism, it is computable by these definitions, and conversely.
>
> McCulloch and Pitts 1943, 113

1.3.2 Von Neumann's interpretation

According to Aspray (1990, 180), von Neumann read McCulloch and Pitts' paper soon after its publication in 1943. He clearly found it valuable, as he used the formal neurons in the draft report of the EDVAC (see Section 1.1.3). Later, he also made the neurons a foundational element of his study of automata.

Among the authors' contributions, von Neumann was particularly interested in the axiomatic method and the result of "co-extensiveness" between logic and formal neural networks (von Neumann 1948). However, he also recognized several limitations in McCulloch and Pitts' model of the neuron.

Axiomatic method Von Neumann recognized the axiomatic method of mathematics in McCulloch and Pitts' approach, and, contrary to many members of the biomedical community, he considered the approach to be "justified" and of considerable help:

> They used what is known in mathematics as the axiomatic method, stating a few simple postulates and not being concerned with how

nature manages to achieve such a gadget.

They went one step further. This has been emphasized very strongly by those who criticize their work, although it seems to me that the extent to which they went further can be justified. They said that they did not want to axiomatize the neuron as it actually exists, but they 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. Now, I am quite sure that it will be a long time before this point is generally agreed to by everybody, if ever; namely, whether or not what one overlooks in this simplification had really better be forgotten or not. But it's certainly true that one gets a quick understanding of a part of the subject by making this idealization.

von Neumann 1949, 43–44

Co-extensiveness In "The General and Logical Theory of Automata", read at the Hixon Symposium in 1948 (von Neumann 1948), von Neumann stated that the main result of McCulloch and Pitts was showing that all logical propositions can be realized by a formal neural network:

McCulloch and Pitts' important result is that any functioning in this sense which can be defined at all logically, strictly, and unambiguously in a finite number of words can also be realized by such a formal neural network.

It is well to pause at this point and to consider what the implications are. It has often been claimed that the activities and functions of the human nervous system are so complicated that no ordinary mechanism could possibly perform them. It has also been attempted to name specific functions which by their nature exhibit this limitation.

> It has been attempted to show that such specific functions, logically, completely described, are per se unable of mechanical, neural realization. The McCulloch-Pitts result puts an end to this. It proves that anything that can be exhaustively and unambiguously described, anything that can be completely and unambiguously put into words, is ipso facto realizable by a suitable finite neural network. Since the converse statement is obvious, we can therefore say that there is no difference between the possibility of describing a real or imagined mode of behavior completely and unambiguously in words, and the possibility of realizing it by a finite formal neural network. The two concepts are co-extensive.
>
> von Neumann 1948, 309–310

Limitations Von Neumann was well aware that McCulloch and Pitts' formal neuron was "an oversimplification of the actual functioning of a neuron" (von Neumann 1948, 309). For instance, during the lecture "Rigorous Theories of Control and Information" (von Neumann 1949, 42–56), he noted that the model did not provide an explanation for the phenomena of fatigue and memory (von Neumann 1949, 48–49). Even earlier, in the "First Draft of a Report on the EDVAC", he had noted that following McCulloch and Pitts' approach would entail ignoring "the more complicated aspects of neuron functioning: thresholds, temporal summation, relative inhibition, changes of the threshold by after-effects of stimulation beyond the synaptic delay, etc" (von Neumann 1945, 12–13).

1.4 Relationship with the biomedical community

Section 1.1 briefly covered some of von Neumann's work: he was a mathematician by education, he worked on engineering problems related to computing equipment, and tackled several problems in physics. However, his interests were even broader. The biomedical one deserves particular attention in this context.

As reported by Aspray (1990, 181–183), von Neumann discussed biological top-

ics with numerous scientists, including, for instance, the biochemist Sol Spiegelman, the population biologist A. J. Lotka, the physical chemist K. F. Bonhoeffer, and the physiologist R. Lorente de Nó. He also talked about biological phenomena with close friends and colleagues, such as Julian Bigelow, Herman Goldstine, Stan Ulam, Norbert Wiener, and Eugene Wigner.

Von Neumann's interest in biomedical topics was not mere pleasure. He actively intended to learn about biological information processing. Indeed, in a 1946 letter to Norbert Wiener, he expressed his interest in studying the human nervous system. He realized it was very difficult to carry out such a study and also suggested a research program involving the study of the bacteriophage, which is a simpler organism, but still capable of some complex behaviors (Aspray 1990, 184–186).

Von Neumann even took part in some of the Macy Conferences on Feedback Mechanisms and Circular Causal Systems in Biology and the Social Sciences[2] (1946–1953), a series of conferences where scholars of diverse interests, including mathematicians, neurophysiologists, psychiatrists, psychologists, and sociologists, could discuss the "mechanisms underlying purposive behavior" (Aspray 1990, 186–187).

Aspray (1990) notes that von Neumann's biomedical interests and his interactions with the biomedical community have often been ignored when describing his work on computing and automata. However, the depth of the interests, the extent of the interactions, and the time overlapping with his research on automata suggest that the biomedical aspect influenced his ideas and efforts in other areas.

1.5 The interdisciplinary decade

Not only was von Neumann a clearly interdisciplinary scholar, but interdisciplinarity actually became a defining characteristic of the decade following World War II (Aspray 1985, 118-119). As a matter of fact, typical of this period was the organization of interdisciplinary conferences, working groups and symposia, such

[2] `https://www.asc-cybernetics.org/foundations/history/MacySummary.htm`

as the aforementioned Macy Conferences.

Von Neumann even created one such group with Howard Aiken and Norbert Wiener. In 1944, the three reunited the Teleological Society (Masani 1990, 239-240), a small group of scholars that discussed information processing. Its members came from various fields and included the astronomer Leland Cunningham, the mathematician H. E. Goldstine, the physiologist R. Lorente de Nó, the geophysicist and meteorologist E. H. Vestine, the neurophysiologist Warren McCulloch, and the logician Walter Pitts.

To understand the importance of the interdisciplinary meetings, it is worth noting that von Neumann even presented one of his works on automata at one of such events. Indeed, he read "The General and Logical Theory of Automata" (von Neumann 1948) at the Hixon Symposium[3], held at Caltech (Pasadena, California) in 1948. Once again, the symposium saw the participation of a diverse set of scholars.

Several of the frequent attendees to these events knew each other and often collaborated on problems at the intersection of their interests (Aspray 1985, 118-119). This was facilitated by the absence of rigid separations between fields. Indeed, in the decade following World War II, many areas of the study of information were just being born and, therefore, did not have established boundaries.

[3] https://www.lancaster.ac.uk/fas/psych/glossary/hixon_symposium/

[4] Image source: https://asc-cybernetics.org/foundations/history/Macy10Photo.htm.

[5] Image source: https://historyofinformation.com/detail.php?id=682. Caption from https://www.lancaster.ac.uk/fas/psych/glossary/hixon_symposium/.

Figure 1.2: Attendees to the 10th Macy Conference (1953)[4]. 1st row (left to right): T.C. Schneirla, Y. Bar-Hillel, Margaret Mead, Warren S. McCulloch, Jan Droogleever-Fortuyn, Yuen Ren Chao, W. Grey-Walter, Vahe E. Amassian. 2nd row (left to right): Leonard J. Savage, Janet Freed Lynch, Gerhardt von Bonin, Lawrence S. Kubie, Lawrence K. Frank, Henry Quastler, Donald G. Marquis, Heinrich Kluver, F.S.C. Northrop. 3rd row (left to right): Peggy Kubie, Henry Brosin, Gregory Bateson, Frank Fremont-Smith, John R. Bowman, G.E. Hutchinson, Hans Lukas Teuber, Julian H. Bigelow, Claude Shannon, Walter Pitts, Heinz von Foerster.

Figure 1.3: Participants to the Hixon Symposium[5]. Left to right: (seated) Halstead, Lashley, Klüver, Köhler, Lorente de Nó; (standing) Brosin, Jeffress, Weiss, Lindsley, von Neumann, Nielsen, Gerard, Liddell. Dr. McCulloch was unable to be present when this picture was taken.

Chapter 2

A Theory of Automata

Having introduced the main influences on von Neumann in Chapter 1, we will now see how they came together in his work on a theory of automata. We will begin by analyzing the origins of the theory and its intended purpose. This will lead us to discuss the central concept of complication[1]. Then, we will analyze von Neumann's methodology, focusing on the axiomatic procedure. Finally, we will remark on some aspects of the theory, provide a summary of its core characteristics, and list its works.

2.1 A unifying theory

Von Neumann's interest in automata theory arose from his war work on computing machines and his post-war efforts to apply computing to scientific research (see Sections 1.1.2, 1.1.3, and 1.1.4). Through these endeavors, he recognized the need for more advanced computers and a theory that could explain how to structure them. He observed that biological automata, such as the human central nervous system, exhibited greater capabilities than computing machines and he sought a theory that could unify both types of automata.

[1]By complication, von Neumann meant the complexity of an automaton's operations: a complicated automaton is one that can do very difficult things. See Section 2.2 on the concept.

2.1.1 Artificial automata

The starting point of the theory was the computing machines of the time. As von Neumann himself explained during the lecture "Computing Machines in General" (von Neumann 1949, 31–41), this initial focus was partly due to his expertise, but also to a balance of complexity and understandability in computing machines, which made them suitable objects of study:

> I am talking about computing machines partly because my interests in the subject of automata are mathematical and, from the mathematical point of view, computing machines are the most interesting and most critical automata. But quite apart from this ex parte argument from the mathematical side, there is the important question of automata of very, very high complexity. 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. Therefore, by considering computing machines, we can discuss what we know and what we do not know, what is right and what is wrong, and what the limitations are, much more clearly than if we discussed other types of automata.
>
> von Neumann 1949, 32

Von Neumann also believed that the length of the computations was a determining factor of the complexity of computing machines:

> While computing automata are not the most complicated artificial automata from the point of view of the end results they achieve, they do nevertheless represent the highest degree of complexity in the sense that they produce the longest chains of events determining and following each other.
>
> [...]

> I am not aware of any other field of human effort where the result really depends on a sequence of a billion (10^9) steps in any artifact, and where, furthermore, it has the characteristic that every step actually matters—or, at least, may matter with a considerable probability. Yet, precisely this is true for computing machines—this is their most specific and most difficult characteristic.
>
> von Neumann 1948, 291–292

2.1.2 Biological automata

Although von Neumann considered the machines of his time relatively complex, he was well aware that their complexity was considerably lower than that of natural organisms. One area where natural automata seemed superior was the number of components:

> With any reasonable definition of what constitutes an element, the natural organisms are very highly complex aggregations of these elements. The number of cells in the human body is somewhere of the general order of 10^{15} or 10^{18}. The number of neurons in the central nervous system is somewhere of the order of 10^{10}. We have absolutely no past experience with systems of this degree of complexity. All artificial automata made by man have numbers of parts which by any comparably schematic count are of the order 10^3 to 10^8. In addition, those artificial systems which function with that type of logical flexibility and autonomy that we find in the natural organisms do not lie at the peak of this scale. The prototypes for these systems are the modern computing machines, and here a reasonable definition of what constitutes an element will lead to counts of a few times 10^3 or 10^4 elements.
>
> von Neumann 1948, 290

During the lecture "The Role of High and Extremely High Complication" (von Neumann 1949, 64–73), when comparing the way computing machines handled

errors, i.e., by stopping and immediately correcting them, with the way living organisms handle them, i.e., by flexibly adapting to it, von Neumann also noted the superiority of the biological approach:

> It's very likely that on the basis of the philosophy that every error has to be caught, explained, and corrected, a system of the complexity of the living organism would not run for a millisecond. Such a system is so well integrated that it can operate across errors. An error in it does not in general indicate a degenerative tendency. The system is sufficiently flexible and well organised that as soon as an error shows up in any part of it, the system automatically senses whether this error matters or not.
>
> von Neumann 1949, 71

2.1.3 The need for a theory

The limits of artificial automata By comparing the capabilities of biological and artificial automata, von Neumann realized that "complication is limited in artificial automata, that is, the complication which can be handled without extreme difficulties and for which automata can still be expected to function reliably" (von Neumann 1948, 302). He identified the lower quality of the materials employed in computing equipment when compared to natural ones to be a cause of such limitations (von Neumann 1948, 300–302). However, he also believed that the lack of a formal theory of automata was making it impossible to create artificial automata of higher complexity:

> All of this re-emphasizes the conclusion that was indicated earlier, that a detailed, highly mathematical, and more specifically analytical, theory of automata and of information is needed. We possess only the first indications of such a theory at present. In assessing artificial automata, which are, as I discussed earlier, of only moderate size, it has been possible to get along in a rough, empirical manner without such

> a theory. There is every reason to believe that this will not be possible with more elaborate automata.
>
> [...]
>
> It is unlikely that we could construct automata of a much higher complexity than the ones we now have, without possessing a very advanced and subtle theory of automata and information. A fortiori, this is inconceivable for automata of such enormous complexity as is possessed by the human central nervous system.
>
> This intellectual inadequacy certainly prevents us from getting much farther than we are now.
>
> von Neumann 1948, 304–305

Automata as information processors It is clear from the above citation that von Neumann believed in the existence of a relationship between automata and information (see also Aspray 1985). He was well aware that artificial and biological automata were very different. He pointed out in several occasions that the two were characterized by different sizes, materials, and modalities of operation (von Neumann 1948, 288–302; von Neumann 1949, 64–73; von Neumann 1956). However, he believed they could both be treated under the common lens of information processing. Indeed, von Neumann most likely used the term "automata" to mean "information processor" (Aspray 1985, 133), an entity capable of producing an output from an input.

Von Neumann's results Although his early death prevented von Neumann from fully developing the theory he deemed so important, he outlined its core characteristics (von Neumann 1948, 302–306). He also conducted a comparative analysis of artificial and biological automata (von Neumann 1948, 288–302; von Neumann 1949, 64–73; von Neumann 1956), and addressed the problems of reliability (von Neumann 1949, 57–63; von Neumann 1952) and self-reproduction (von Neumann 1949). In what follows, we will focus on the general nature of his proposed theory and on his comparison of artificial and biological systems.

2.2 The concept of complication

As emerges from the descriptions of artificial and biological automata in Section 2.1.1 and Section 2.1.2, von Neumann was interested in a specific aspect of these entities: their complication (or complexity). Indeed, this can be seen as the defining characteristic of living organisms and the one trait von Neumann ultimately hoped to enhance in computing machines.

2.2.1 The potentiality to do things

Von Neumann referred to the concept of automaton's complication in several occasions, and, within the lecture "Re-Evaluation of the Problems of Complicated Automata – Problems of Hierarchy and Evolution" (von Neumann 1949, 74–87), he linked it to the "the potentiality to do things":

> There is a concept which will be quite useful here, of which we have a certain intuitive idea, but which is vague, unscientific, and imperfect. This concept clearly belongs to the subject of information, and quasi-thermodynamical considerations are relevant to it. I know no adequate name for it, but it is best described by calling it "complication". It is effectivity in complication, or the potentiality to do things. I am not thinking about how involved the object is, but how involved its purposive operations are. In this sense, an object is of the highest degree of complexity if it can do very difficult and involved things.
>
> von Neumann 1949, 78

As von Neumann himself noted, his idea of complication was "vague, unscientific, and imperfect". Although he attempted to find alternative means to quantitatively define the concept, for instance, in terms of the number of elementary parts (von Neumann 1949, 36), he never developed a mathematical definition. Nevertheless, this idea is present in all of his works on automata, whether explicitly or implicitly.

2.2.2 Complication as the unifying factor

While information represents the common lens through which artificial and biological automata can be analyzed, it is complication that makes the unification altogether reasonable and potentially profitable. In von Neumann's own words, it is the high complication that makes "a comparison between the computing machines and the operation of the natural organisms not entirely out of proportion":

> Thus a computing machine is one of the exceptional artifacts. They not only have to perform a billion or more steps in a short time, but in a considerable part of the procedure (and this is a part that is rigorously specified in advance) they are permitted not a single error. In fact, in order to be sure that the whole machine is operative, and that no potentially degenerative malfunctions have set in, the present practice usually requires that no error should occur anywhere in the entire procedure.
>
> This requirement puts the large, high-complexity computing machines in an altogether new light. It makes in particular a comparison between the computing machines and the operation of the natural organisms not entirely out of proportion.
>
> von Neumann 1948, 292

2.3 Methodology

In order to unify artificial and biological automata, von Neumann employed the same approach used in the design of the EDVAC (see Section 1.1.3): he abstracted away the physical details of the distinct entities to focus on their common characteristics. To achieve this, he used the axiomatic method and described automata in terms of the formal neurons introduced by McCulloch and Pitts (see Section 1.3).

2.3.1 Axiomatization

Two aspects of the problem Von Neumann believed that the study of complicated automata could be divided into two parts: the first concerned the elementary components of the automata; whereas the second focused on the combination and organization of such elements into complicated systems. He believed the second part of the problem to be the object of automata theory and considered axiomatization necessary to focus on it:

> The natural systems are of enormous complexity, and it is clearly necessary to subdivide the problem that they represent into several parts. One method of subdivision, which is particularly significant in the present context, is this: the organisms can be viewed as made up of parts which to a certain extent are independent, elementary units. We may, therefore, to this extent, view as the first part of the problem the structure and functioning of such elementary units individually. The second part of the problem consists of understanding how these elements are organized into a whole, and how the functioning of the whole is expressed in terms of these elements.
>
> The first part of the problem is at present the dominant one in physiology. It is closely connected with the most difficult chapters of organic chemistry and of physical chemistry, and may in due course be greatly helped by quantum mechanics. I have little qualification to talk about it, and it is not this part with which I shall concern myself here.
>
> The second part, on the other hand, is the one which is likely to attract those of us who have the background and the tastes of a mathematician or a logician. With this attitude, we will be inclined to remove the first part of the problem by the process of axiomatization, and concentrate on the second one.
>
> von Neumann 1948, 289

The axiomatic procedure By axiomatization, von Neumann meant assuming the existence of some elements to be treated as "black boxes" whose inner structures could be ignored. These elements are assumed to have a well-defined behavior, i.e., to produce outputs in response to inputs in a predictable and clearly-defined way (von Neumann 1948, 289).

2.3.2 Results of axiomatization

Equivalence of neurons and vacuum tubes An early and important result of the axiomatic method can be found in "The General and Logical Theory of Automata" (von Neumann 1948). In it, von Neumann, already aware of McCulloch and Pitts' work, considered living organisms as if they were "purely digital automata", and described their central nervous system as composed of formal neurons, organs characterized by a "black box" nature and an "all-or-none" response (von Neumann 1948, 296-298). Also considering the vacuum tubes of computing machines as purely digital, he claimed the equivalence of the two elements:

> The neuron, as well as the vacuum tube, viewed under the aspects discussed above, are then two instances of the same generic entity, which it is customary to call a "switching organ" or "relay organ" (the electromechanical relay is, of course, another instance). Such an organ is defined as a "black box", which responds to a specified stimulus or combination of stimuli by an energetically independent response. [...]
>
> The basic switching organs of the living organisms, at least to the extent to which we are considering them here, are the neurons. The basic switching organs of the recent types of computing machines are vacuum tubes; in older ones they were wholly or partially electromechanical relays.
>
> von Neumann 1948, 298–299

This result is an extension of the idea of modeling machine components with formal neurons, which he had already introduced in the "First Draft of a Report on the EDVAC" (von Neumann 1945) (see Section 1.1.3).

Automata of formal neurons By modeling switching organs with McCulloch and Pitts' formal neurons, von Neumann was also able to develop a formal definition of an automaton. During the lectures on "Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components" he gave at the California Institute of Technology in 1952 (von Neumann 1952), he described an automaton as "a 'black box' with a finite number of inputs and outputs" (von Neumann 1952, 44), and formally defined a single-output automaton:

> A single output automaton with time delay δ (δ is positive) is a finite set of inputs, exactly one output, and an enumeration of certain "preferred" subunits of the set of all inputs. The automaton stimulates its output at time $t + \delta$ if and only if at time t the stimulated inputs constitute a subset which appears in the list of "preferred" subsets, describing the automaton.
>
> von Neumann 1952, 45

Moreover, following the same approach of McCulloch and Pitts, he explained that single-output automata can be combined in networks by connecting their inputs and outputs in to create more complex automata. Since the same can be done with such automata, von Neumann informally described a general automaton as a network with several inputs and several outputs (von Neumann 1952, 46).

2.3.3 Interdisciplinarity

By unifying artificial and biological automata, von Neumann envisioned a study that would initially combine logic, communication theory, and physiology. However, he also believed that such study would eventually reveal itself as a separate discipline:

> The *formalistic* study of *automata* is a subject lying in the intermediate area between logics, communication theory, and physiology. It implies abstractions that make it an imperfect entity when viewed exclusively

> from the point of view of any one of the three above disciplines—the imperfection being probably worst in the last mentioned instance. Nevertheless an assimilation of certain viewpoints from each one of these three disciplines seems to be necessary for a proper approach to that theory. Hence it will have to be viewed synoptically, from the combined point of view of all three, and will probably, in the end, be best regarded as a separate discipline in its own right.
>
> von Neumann 1966, 91

Von Neumann's interdisciplinary interests and efforts, together with the overall interdisciplinary orientation of the years following World War II, probably contributed to shaping this vision of the theory of automata (see Section 1.4 and Section 1.5).

2.4 Logic and engineering

2.4.1 The organizational focus

Axiomatizing the elements allowed von Neumann to "throw half the problem out of the window" (von Neumann 1949, 77) and focus on the organizational aspects. According to him, these included the investigation of "the larger organisms that can be built up from these elements, their structure, their functioning, the connections between the elements, and the general theoretical regularities that may be detectable in the complex syntheses of the organisms in question" (von Neumann 1948, 289–290).

Therefore, the theory of automata was meant to be, at least initially, purely logical. It did not need to be concerned with engineering and physical details, but with structural problems. This is very similar to the approach he took in the design of the EDVAC, where he focused on the computer's composition, its functioning, and the organization of its components (see Section 1.1.3).

2.4.2 The engineering goal

In a period when computing was only treated as an engineering problem, von Neumann's logical approach was innovative and unique. However, this approach ultimately had an engineering goal: building better computers. It is important to highlight this point to fully understand why he dedicated so much effort and a considerable part of his work to the comparison of biological and artificial automata—including his last endeavor, "The Computer and the Brain" (von Neumann 1956).

Had von Neumann been solely interested in theoretical results concerning computability and automata, he may have focused on the study of formal neurons and Turing machines. Indeed, these formalisms were very powerful (see Sections 1.2 and 1.3), and they fitted the definition of information processors. Moreover, both could be employed to study computing machines, since neurons can model switching elements, and von Neumann's stored-program machine is a Turing machine.

However, von Neumann chose not to concentrate on these "axiomatic paper automata", which, as he noted, nobody was particularly concerned to build (von Neumann 1949, 43). Instead, he employed them to study and compare existing automata, such as the human brain and computing machines. In a sense, von Neumann looked for examples and inspirations in nature to resolve his engineering problems, viewing natural systems through the formal lens of information processing.

2.4.3 Computing for computing's sake

Since von Neumann frequently compared the computers and the human brain, one might assume he was hoping to build *intelligent* machines comparable to humans. However, there is no indication that this was his intention. This absence is especially significant considering that, around the same years he was working on automata's theory, Alan Turing proposed the idea of machines capable of *intelligent* behavior (Turing 1950). Given that von Neumann was familiar with Turing's work on computability and was very active in the computing community, it is reasonable to assume that he would have mentioned the possibility of creating

intelligent machines in his own works if he had considered it a possibility.

Instead, when discussing problems that more advanced computers could be applied to, he only mentioned purely computing ones. For instance, on top of those listed in Section 1.1.4, he mentioned quantum mechanical calculations on atomic and molecular wave functions (von Neumann 1949, 38), and the control of missiles and planes (von Neumann 1949, 69).

2.5 Summary of the theory's characteristics

Having already introduced several aspects of von Neumann's theory of automata, we now summarize the characteristics most relevant to the present work:

- **Logical character**: von Neumann intended to develop a logical theory of automata, focusing on their organization, structure, and functioning rather than the physical nature of their components;
- **Engineering goal**: the theory had the ultimate goal of facilitating the construction of more advanced computing machines;
- **Central role of complication**: to enable such developments, von Neumann focused on the complicated nature of automata;
- **Unification of biological and artificial**: in order to understand complication, he dedicated considerable effort to the comparison of biological and artificial automata;
- **Use of axiomatization**: axiomatization allowed the logical focus, as well as the unification of biological and artificial automata.

Von Neumann's intention to develop a theory that would eventually incorporate probabilistic elements closely related to thermodynamics was not addressed in this chapter, as it is not strongly tied to the idea of self-reproduction this work intends to explore.

2.6 Works of the theory

Von Neumann's work on the theory of automata consists of five pieces, produced beginning in the late 1940s:

1. **"The General and Logical Theory of Automata"**: read at the Hixon Symposium in September, 1948 (von Neumann 1948);

2. **"Theory and Organization of Complicated Automata"**: five lectures delivered at the University of Illinois in December, 1949 (von Neumann 1949);

3. **"Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components"**: lectures given at the California Institute of Technology in January, 1952 (von Neumann 1952);

4. **"The Theory of Automata: Construction, Reproduction, Homogeneity"**: manuscript written between 1952 and 1953, then completed and edited by Arthur Burks (von Neumann 1966);

5. **"The Computer and the Brain"**: a series of lectures left incomplete and prepared to be delivered at Yale University in 1956 (von Neumann 1956).

Chapter 3

Self-Reproduction

As introduced in Section 2.1.3, von Neumann's work on automata can be roughly divided into four parts: an outline of the theory's intended characteristics, a comparative analysis of artificial and biological automata, an investigation of reliability, and an examination of the problem of self-reproduction. The previous chapter addressed the former two topics, providing essential context for understanding the remaining issues. In what follows, we turn to the work on self-reproduction, as it is not only more developed than the one on reliability, but also a unique contribution.

We will begin by introducing the problem using von Neumann's own description, followed by its axiomatization. Then, we will attempt to explain how von Neumann came to the issue. First, we will provide an explanation that links it to the concept of complication. Finally, we will argue that the work on self-reproduction can actually be seen as a deviation from the rest of the theory, both in its goal and its methodology.

3.1 The problem of self-reproduction

Self-reproduction was a very recurring topic in von Neumann's work and public lectures on automata. He first introduced the problem in "The General and Logical Theory of Automata" (von Neumann 1948, 312–318). He later explored it in "Re-

Evaluation of the Problems of Complicated Automata – Problems of Hierarchy and Evolution" (von Neumann 1949, 74–87). Finally, he wrote the manuscript "The Theory of Automata: Construction, Reproduction, Homogeneity" (von Neumann 1966), where he designed an automaton capable of self-reproduction. He never finished this work, which, however, was completed and edited by Arthur Burks.

At the beginning of the manuscript, von Neumann presents the "main questions" he intended to answer about automata:

> **(A) Logical universality.**[1] 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? Also, with what additional—variable, but in the essential respects standard—attachments[2] is a single automaton logically universal?
>
> **(B) Constructibility.** Can an automaton be constructed, i.e., assembled and built from appropriately defined "raw materials", by another automaton? Or, starting from the other end and extending the question, what class of automata can be constructed by one, suitably given, automaton? The variable, but essentially standard, attachments to the latter, in the sense of the second question of (A), may here be permitted.
>
> **(C) Construction-universality.** Making the second question of (B) more specific, can any one, suitably given, automaton be construction-universal, i.e., be able to construct in the sense of question (B) (with suitable, but essentially standard, attachments) every other automaton?
>
> **(D) Self-reproduction.** Narrowing question (C), 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?

[1] List headings emphasized by the author of this work.

[2] An example of a "variable, but in the essential respects standard, attachment" is the arbitrarily extendible tape of a Turing Machine.

> **(E) Evolution.** Combining questions (C) and (D), can the construction of automata by automata progress from simpler types to increasingly complicated types? Also, assuming some suitable definition of "efficiency", can this evolution go from less efficient to more efficient automata?
>
> von Neumann 1966, 92

As von Neumann goes on to explain, "the answer to question (A) is known" (von Neumann 1966, 92), and it is provided by Turing's work (see Section 1.2). Indeed, the class of Turing machines is "logically universal", as these automata can perform all computations realizable with finite means[3]. Moreover, Turing's universal machine is itself a "logically universal" automaton since it can perform the computations of any other given Turing machine. All the remaining questions, however, had not been previously answered.

The core concept behind questions (B)–(E) is constructibility, i.e., the capacity of an automaton to build another one. This is a considerably different problem from the one of logical universality, as it expects automata to do more than just output pieces of information: it expects them to build something. Mirroring Turing's work on the universal machine (see Section 1.2.2), von Neumann also sought to determine whether there exists an automaton capable of constructing all others. Once constructibility is proven possible, questions (D) and (E) naturally arise. Indeed, if an automaton can create others, one could question whether it can also recreate itself, that is, self-reproduce. Assuming it can do so, can it also undergo changes across generations, that is, evolve?

Overall, questions (B)–(E) extend the treatment of automata beyond computing-only entities by introducing construction requirements. These, in turn, introduce several complications, which von Neumann addressed through axiomatization.

[3]See also Church-Turing's thesis: `https://plato.stanford.edu/entries/church-turing/`.

3.2 Axiomatization of the problem

When discussing self-reproduction, von Neumann was not thinking of "producing matter out of nothing" (von Neumann 1949, 75). Instead, he was imagining something similar to the assembly of a new entity from elementary parts, as suggested by question (B) above. While the logical problems of question (A) could be tackled with McCulloch and Pitts' neurons, question (B) forced von Neumann to take into consideration more elaborate elementary components:

> Question (A) involved merely logical determinations; therefore it required only (at least directly *only* [...]) organs with two states, true and false. These two states are adequately covered by the neural states of excitation and quiescence. Question (B), on the other hand, calls for the construction of automata by automata, and it necessitates therefore the introduction of organs with other than logical functions, namely with the kinematical or mechanical attributes that are necessary for the acquisition and combination of the organs that are to make up the automata under construction. To use a physiological simile, to the purely neural functions must be added at least the muscular functions.
>
> von Neumann 1949, 101

Evidently, a critical aspect of the problem's axiomatization is a proper abstraction of the aspects related to the "acquisition and combination of the organs". Over the years, various approaches were developed.

3.2.1 The kinematic model

Von Neumann had already developed a first and preliminary axiomatized version of the problem in the lecture "Re-Evaluation of the Problems of Complicated Automata – Problems of Hierarchy and Evolution" (von Neumann 1949, 74–87), where he had taken into consideration the kinematic aspects, and imagined a constructing automaton floating in a container with parts that could be used to assemble another automaton:

> In order to discuss these things, one has to imagine a formal set-up like this. 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; its essential activity is to pick up parts and put them together, or, if aggregates of parts are found, to take them apart.
>
> von Neumann 1949, 75

In the lecture, he had also noted that the validity of the axiomatization strongly depended on the choice of the elementary parts and recognized that there was not a "rigorously justifiable" way to make such choice (von Neumann 1949, 76–77). However, he also emphasized the importance of the axiomatic procedure to focus on the organization of the elementary parts into a functioning entity:

> The question that one can then hope to answer, or at least investigate, is: 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?
>
> von Neumann 1949, 77

What von Neumann introduced was referred to by Burks as the "kinematic model of self-reproduction" (von Neumann 1949, 82), as it dealt with geometrical and kinematic problems, such as movement, contact, and position. Clearly, all these problems made the treatment of the purely organizational aspects above difficult, ultimately leading von Neumann to shift his focus to the cellular model.

3.2.2 The cellular model

It was S. M. Ulam who suggested to von Neumann that a cellular model would be better suited to the logical and mathematical treatment of self-reproduction (von Neumann 1966, 94). Von Neumann recognized this, and indeed worked on a cellular automaton capable of self-reproduction in the manuscript "The Theory

of Automata: Construction, Reproduction, Homogeneity" (von Neumann 1966). In the first chapter of the work, titled "General Considerations", the author reasoned through several possible approaches to the treatment of the problem. When considering the appropriate level of abstraction, he noted that kinematic considerations should be initially avoided:

> Different degrees of abstraction are still possible; for example, one may or may not pay attention to the truly mechanical aspects of the matter (the forces involved, the energy absorbed or dissipated, etc.). But even the simplest approach, which disregards the above-mentioned properly mechanical aspects entirely, requires quite complicated geometrical-kinematical considerations. Yet, one cannot help feeling that these should be avoided in a first attempt like the present one: in this situation one ought to be able to concentrate all attention on the intrinsic, logical-combinatorial aspects of the study of automata. The use of the adjective *formalistic* at the beginning of Section 1.1.1.1 was intended to indicate such an approach—with, as far as feasible, an avoidance of the truly geometrical, kinematical, or mechanical complications. The propriety of this desideratum becomes even clearer if one continues the above list of avoidances, which progressed from geometry, to kinematics, to mechanics. Indeed, it can be continued (in the same spirit) to physics, to chemistry, and finally to the analysis of the specific physiological, physico-chemical structures. All these should come in later, successively, and about in the above order; but a first investigation might best avoid them all, even geometry and kinematics.
>
> von Neumann 1966, 102

Proceeding with the reasoning, he introduced the abstractions necessary to avoid such considerations (von Neumann 1966, 103–105, 148–149):

- **Stationarity and quiescence**: all objects should be stationary and normally in a quiescent state;

- **Discrete framework**: the objects are discrete elements of an infinite space;
- **"Crystalline" structure**: the medium has a regular structure;
- **"Functional homogeneity"**: all objects behave according to the same rules.

Von Neumann also assumed the time to be discrete, i.e., he assumed all events to happen at times t that are integers: $t = 0, \pm 1, \pm 2, \pm 3, ...$ (von Neumann 1966, 100).

Combining all these ideas, he developed his cellular model, which is an infinite grid of square cells, i.e., an infinite two-dimensional array. Each cell is identified by two coordinates and can be in one of 29 states, which determine its behavior. Each cell is also connected to its four neighbors (two along the vertical axis and two along the horizontal one). Stimuli from one or more neighbors can change a cell's state and may also be propagated in one or more directions.

The default state of a cell is unexcitability. In this state, the cell does not respond to the received stimuli by emitting itself. Instead, a series of stimuli can turn an unexcitable cell into an excitable one. Once it is excitable, it can either be quiescent, i.e., not stimulated and not emitting stimuli, or excited, i.e., stimulated and emitting stimuli. Special stimuli can cause the reverse process and turn an excitable cell into an unexcitable one.

In this framework, an automaton is a grouping of cells, and the problem of self-reproduction involves devising a configuration of cells that, when provided with an initial stimulus, will be able to recreate the same configuration in another area of the infinite space. This process of recreation happens through the transformation of existing unexcitable cells into excitable ones (von Neumann 1966, 109). Further details are postponed to Section 4.2, where the cellular self-reproducing automaton is described.

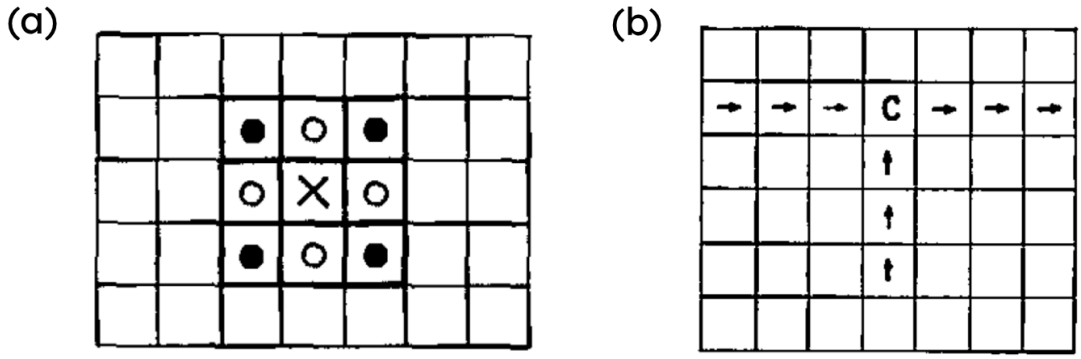

Figure 3.1: The cellular model from von Neumann (1966). All the squares are cells. (a) shows the neighbors of the cell X, which are marked with $\circ$. (b) shows a possible propagation of stimuli, represented by $\rightarrow$ and $\uparrow$, specifically through a cell with state C.

3.3 Why self-reproduction?

Usually, one does not expect information processors to be capable of self-reproduction. Similarly, when considering the questions in Section 3.1, one often only believes the first one to relate to computing automata. However, von Neumann considered the problem to be significant for a general theory of automata.

3.3.1 A more complete discussion of automata

At the beginning of the lecture "Re-Evaluation of the Problems of Complicated Automata – Problems of Hierarchy and Evolution" (von Neumann 1949, 74–87), von Neumann noted how the most significant artificial automata are "automata whose operations are not directed at themselves, so that they produce results which are of a completely different character than themselves" (von Neumann 1949, 74). This is the case for Turing machines, which modify a tape; for networks of formal neurons, which produce pulses; and for computing machines, which are fed and modify tape or magnetic memories (von Neumann 1949, 74–75).

Von Neumann believed that such behavior was not a requirement for the automata and that these could actually produce something like themselves. Indeed,

he believed that "a complete discussion of automata can be obtained only by taking a broader view of these things and considering automata which can have outputs something like themselves" (von Neumann 1949, 75).

While he did not mention biological automata in introducing the problem during the lecture, it is reasonable to assume he derived the idea of artificial automata capable of producing something like themselves from living organisms. Indeed, he believed self-reproduction to be a defining ability of living organisms:

> Anybody who looks at living organisms knows perfectly well that they can produce other organisms like themselves. This is their normal function, they wouldn't exist if they didn't do this, and it's plausible that this is the reason why they abound in the world. In other words, living organisms are very complicated aggregations of elementary parts, and by any reasonable theory of probability or thermodynamics highly improbable. 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.
>
> von Neumann 1949, 78

3.3.2 Self-reproduction and complication

Von Neumann believed self-reproduction and complication to be intimately connected to each other. Indeed, in the aforementioned lecture, it was the discussion of self-reproduction that led him to introduce the more general concept of complication (see Section 2.2), of which self-reproduction is an expression. He had already made the same connection in "The General and Logical Theory of Automata", where the idea of self-reproduction in nature led him to suspect the existence of a "concept of complication" (von Neumann 1948, 312).

In both cases, von Neumann noted how self-reproduction highlighted a gap between the complication of biological and artificial automata. According to him, self-reproduction in biological automata is clearly progressive, as living organisms have evolved and improved over time. On the other hand, self-reproduction seems

to be degenerative in artificial automata, and "everyone knows that a machine tool is more complicated than the elements which can be made with it" (von Neumann 1949, 79). Combining these two observations, von Neumann concluded that "complication is degenerative below a certain minimum level" (von Neumann 1949, 79).

3.3.3 Connections to the theory of automata

Drawing on the analysis from Sections 3.3.1 and 3.3.2, we conclude that, in tackling the problem of self-reproduction, von Neumann hoped to achieve two goals:

1. Unifying artificial and biological automata by demonstrating artificial ones possess the essentially biological ability of self-reproduction;

2. Proving artificial automata can exceed the minimum level below which complication is degenerative.

These goals are coherent with von Neumann's desire to develop a unifying theory of automata (see Section 2.1) and the importance of complication in such a theory (see Section 2.2).

3.4 Self-reproduction as a deviation

Having linked the problem of self-reproduction to essential aspects of von Neumann's theory, we now show how it can also be seen as a deviation from other core aspects. Specifically, we argue that it constitutes a deviation from the theory's intended engineering goal, as well as from von Neumann's own methodology.

3.4.1 Deviation from the engineering goal

In Sections 2.1.3 and 2.4, we discussed von Neumann's engineering goals, showing that his work on a general theory of automata was animated by a desire to build better computing equipment. This was more than a stated intention: it

was a goal he actively pursued, as demonstrated by his work on improving the reliability of artificial automata (von Neumann 1952).

However, reliability was not the only area where von Neumann believed computers needed to improve. As already introduced in Section 2.1.2, he also argued that machines were well behind biological automata in terms of the number of their parts. Simultaneously, they appeared to be disproportionately big for the number of their components (von Neumann 1948, 299–301). Later, he also pointed out that the human brain seemed to operate largely in parallel, whereas computing equipment of the time processed data sequentially (von Neumann 1956, 50–52).

Overall, von Neumann had a clear idea of areas where artificial automata needed improvement. Because of his engineering goals, it is reasonable to expect him to have addressed these areas. However, he decided to focus on the problem of self-reproduction, which did not resolve any of the limitations he had identified. This is especially clear when looking at the questions in Section 3.1: none of them appear to have any potential to improve the practical capabilities of machines. In this light, the work on self-reproduction represents a deviation from the stated engineering goals.

Of course, it would be unreasonable to expect all of von Neumann's work to have practical implications for computers. After all, he intended to develop a unifying theory of automata and, as explained in Section 3.3.3, tackling the problem of self-reproduction served this goal. We could also suppose that his work would have eventually led to engineering applications if he had been able to complete it. Nevertheless, the focus on self-reproduction resembles more an exploration of the theoretical capabilities of artificial automata than a pursuit of engineering developments.

3.4.2 Methodological deviations

Not only does self-reproduction represent a deviation from some of von Neumann's stated goals, but it also constitutes a deviation from his methodology, particularly in the use of axiomatization and in the relationship he had established between artificial and biological aspects.

The loss of axiomatization We showed in Section 2.3 that axiomatization played an important role in von Neumann's theory, since it allowed him to tackle the logical aspects of automata and unify biological and artificial automata. All of this was possible through the sole formalization of neurons. However, self-reproduction forced von Neumann to expand his framework in order to accommodate the kinematic nature of the problem, as evident in Section 3.2. Specifically, it introduced an axiomatization of space through the cellular model. Interestingly, this extension brought various complications not present in nature:

> Comparing these processes of construction and reproduction of automata, and those of actual growth and reproduction in nature, this difference is conspicuous; in our case the site plays a more critical role than it does in reality. The reason is that by passing from continuous, Euclidean space to a discrete crystal, we have purposely bypassed as much as possible of kinematics. Hence the moving around of a structure which remains congruent to itself, but changes its position with respect to the crystal lattice, is no longer the simple and elementary operation it is in nature. In our case, it would be about as complex as genuine reproduction.
>
> von Neumann 1966, 129–130

These complications could be interpreted as a sign that the axiomatization was preliminary and imperfect. However, they might also indicate that the theory was not developed enough to accommodate these developments. Indeed, the work on self-reproduction seems a narrowing of the theory on a specific aspect rather than an organic expansion of it. This is particularly significant given that the limitations of machines in Section 3.4.1 seem to be addressable, at least partly, with the sole axiomatization of computation via neurons.

Bringing the biological into the artificial The deviation from the engineering goal and the loss of axiomatization are deviations from goals and methods that von Neumann explicitly stated. However, the work on self-reproduction also represents a change in his general approach to the study of automata.

As frequently mentioned in this chapter, von Neumann aimed at unifying artificial and biological automata under a single general theory. In this process, he almost always considered problems from the artificial point of view. For instance, when unifying computing machines and the nervous system (see Section 2.3.2), he did not treat both of them as mixed systems (von Neumann 1948, 296–298). Instead, he assumed both to be digital, favoring the primarily digital nature of computers over the largely analogical one of living organisms. Moreover, when comparing artificial and biological automata, he exclusively considered aspects central to machines, such as memory, size, and number of components, altogether ignoring defining aspects of biological automata, such as collaboration, creativity, and intelligence (see also Section 2.4.3).

Evidently, treating self-reproduction involves the opposite approach. Indeed, it is an attempt to bring something exclusively biological into the artificial domain. The discrete and digital framework is maintained, but the problem is purely biological. In a sense, it was ignoring the engineering goal that made this altogether possible.

Chapter 4

The Self-Reproducing Automaton

When analyzing the work of Turing and that of McCulloch and Pitts, von Neumann recognized that there are two ways to describe automata: the synthetic and the integral approaches (von Neumann 1949, 43). Turing adopted the synthetic approach by axiomatically describing the function of his machine without specifying its components. On the other hand, McCulloch and Pitts followed the integral approach by axiomatically defining some very simple elements (the neurons) and showing how they can be combined in more complex systems (the nets).

Although not explicitly, von Neumann followed both approaches in designing his cellular self-reproducing automata, with which he addressed questions of constructibility and self-reproduction. In this chapter, we will explore these descriptions, beginning with the synthetic one. Afterwards, we will examine what his design achieved with respect to the questions of Section 3.1.

4.1 Synthetic description

Von Neumann developed a synthetic description for both the kinematic and the cellular models (von Neumann 1949, 82–87; von Neumann 1966, 118–119). These descriptions served as proof of the theoretical feasibility of self-reproduction. Since the descriptions share the same structure, in what follows, we will abstract the few model-dependent details and focus on the common features.

Automaton's description Von Neumann applied "Turing's trick" (von Neumann 1949, 83), and associated a logical description to each automaton (see Section 1.2.2). This description could define each element of the automaton or a plan for its assembly, and it is the artificial analog of a gene (von Neumann 1966, 130). Given an automaton X, we designate with $\phi(X)$ its description.

Constructing and the copying automata Von Neumann axiomatically defined a constructing automaton A, that, given a description $\phi(X)$, consumes it to create the corresponding automaton X. He also axiomatically defined the copying automaton B, which, given a description $\phi(X)$, creates a new copy of it.

Control automaton and universal constructor The functioning of A and B can be controlled by a third automaton C to construct an automaton X from its description $\phi(X)$. First, C causes B to duplicate $\phi(X)$; then, C activates A, which consumes one of the two copies of $\phi(X)$, and constructs X; finally, C ties X and $\phi(X)$ together. In the end, the complex $(A + B + C) + \phi(X)$ will have produced the entity $X + \phi(X)$. Therefore, it has effectively constructed an automaton from its description, while also preserving the description.

Since X can be any automaton, the complex $A + B + C$ is a universal constructor, and it provides an affirmative answer to question (C) in Section 3.1.

Self-reproducing automaton If we choose $X = (A + B + C)$ in the setting above, we obtain that $(A + B + C) + \phi(A + B + C)$ can produce $(A + B + C) + \phi(A + B + C)$. Therefore, $A + B + C$ is capable of self-reproduction. Moreover, by choosing $X = A + B + C + D$, where D is another automaton, we obtain that $(A + B + C) + \phi(A + B + C + D)$ produces $(A + B + C + D) + \phi(A + B + C + D)$. This automaton can produce a specific object D in addition to replicating itself.

Therefore, this procedure provides a positive answer to the question of self-reproduction (question (D) in Section 3.1). Furthermore, self-reproduction was achieved by combining parts (A, B, C) which are not "themselves self-reproductive" (von Neumann 1949, 86).

Evolving automaton Clearly, once the automaton has replicated itself, the newly generated one can do the same, creating an endless chain of self-reproduction. Von Neumann assumed a random change in $\phi(A+B+C+D)$ could happen during this process and identified two possibilities. The first one is a change in $A+B+C$, which would be lethal as it would hinder the reproduction process. The second one is a change in D, which would lead to the creation of a different entity D'. This is a case of "inheritable mutation" (von Neumann 1949, 87), which provides a (very) partial answer to the question of evolution (question (E) in Section 3.1).

Reduction of self-reproduction to constructibility An important result of the synthetic description is the reduction of the problem of self-reproduction to the one of universal constructibility. Indeed, the description showed how a universal constructor can be easily transformed into a self-reproducing automaton. This result will be used by von Neumann in his integral design of the self-reproducing automaton.

4.2 Integral description

Having defined the self-reproducing automaton synthetically, we now show how it can be realized within a cellular model. Von Neumann developed this integral description within the manuscript "The Theory of Automata: Construction, Reproduction, Homogeneity" (von Neumann 1966), which was completed and edited by Arthur Burks.

Von Neumann's design is fairly complex, as he was more interested in feasibility than in "optimality" and "minimality" (von Neumann 1966, 91). Additionally, some parts were only sketched, while others were developed by Burks. Therefore, to simplify the discussion, we will present the automaton in a more schematic fashion. We will describe the 29 states that the cells could be in. We will also describe the components of the self-reproducing automaton and their functioning, though we will not detail how they can be built from individual cells. These

construction aspects are not essential and could be modified without affecting the general structure.

4.2.1 The 29 cell states

In Section 3.2.2, we introduced von Neumann's cellular model, which consists of an infinite grid of square cells. Each cell has four neighbors (to the right, left, above, and below) with which it can exchange stimuli. Moreover, it can be in one of 29 states, which determine its behavior[1]. Such states fall into five categories: ordinary transmission, special transmission, confluent, unexcitable, and sensitized.

The default state of a cell is unexcitability. In this state, the cell does not respond to the received stimuli by emitting itself. Instead, a series of stimuli can turn an unexcited cell into an excitable one. Once it is excitable, it can either be quiescent, i.e., not stimulated and not emitting stimuli, or excited, i.e., stimulated and emitting stimuli. Special stimuli can cause the reverse process and turn an excitable cell into an unexcitable one.

Each cell is essentially a complex formal neuron that can communicate two types of stimuli: ordinary and special. Ordinary stimuli are used for logical functions (question (A) from Section 3.1), while the combination of ordinary and special stimuli can be used for constructing (and destructive) purposes (question (B)–(E)) (von Neumann 1966, 109–110, 140–143).

The remainder of the section describes the five types of states, while the purpose of the two types of stimuli is detailed in the next section. To lighten exposition, we will often say "state Y" or "cell Y" to mean "cell in state Y".

Ordinary transmission states Ordinary transmission states are used to communicate ordinary stimuli from one cell to another, analogous to the connections between some neurons' outputs and other neurons' inputs. In these states, a cell receives *disjunctively* ordinary stimuli from its neighbors and emits an ordinary

[1] In his manuscript, von Neumann actually referred to a 29-state automaton occupying each cell. Here, we associate the state directly to the cell to simplify the discussion and avoid ambiguity between the 29-state *automaton* and the self-reproducing *automaton*.

stimulus in one specific direction with a unit delay. Since there are four possible output directions and a cell can be either quiescent or excited, there are a total of eight ordinary transmission states.

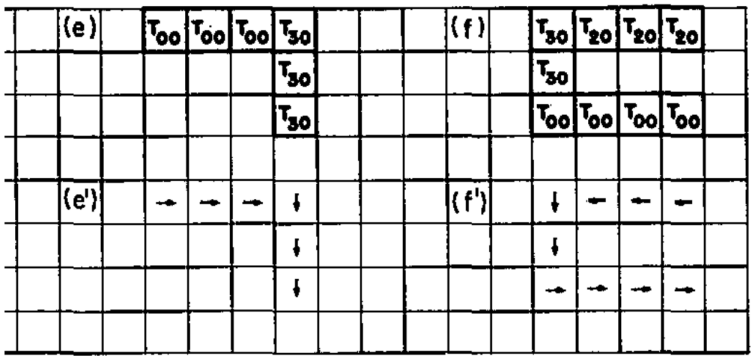

Figure 4.1: Two representations of ordinary transmission states from von Neumann (1966). The four output directions are encoded as the numbers $0, 1, 2, 3$, and quiescence/excitation are encoded as $0/1$. T_{ij} is an ordinary transmission state with output direction i and excitation j. (e') and (f') show the propagation directions of stimuli.

Confluent states When in confluent states, cells receive *conjunctively* from ordinary transmission states directed towards them and emit with double delay[2] to all transmission states—both ordinary and special (see below)—directed away from them. The double delay requires accounting for both the current excitation and the future one; therefore, there are four confluent states: quiescent and next quiescent; quiescent and next excited; excited and next quiescent; excited and next excited.

[2]Because the cellular model is an infinite grid, two distinct paths between cell A and cell B will differ in length by an even number. If each cell along these paths is in a transmission state, then the times to traverse the paths will also differ by an even number of time steps, as transmission states all have the same single delay. Injecting a stimulus at A therefore causes the arrival of two stimuli at B separated by an even delay. To allow odd delays, von Neumann introduced a double delay in confluent states.

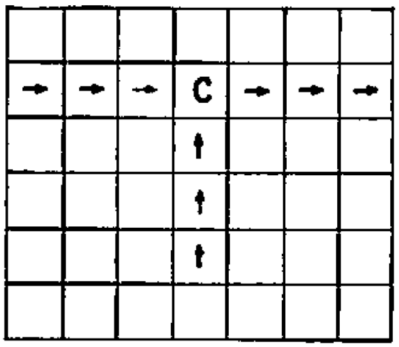

Figure 4.2: A confluent state C from von Neumann (1966). C has two ordinary transmission states with output directions directed towards it, one from the left ($\rightarrow$) and one from below ($\uparrow$). C also has an ordinary transmission state directed away from it ($\rightarrow$). If both the incoming $\rightarrow$ and $\uparrow$ are excited at time t, C will emit a stimulus along the outgoing $\rightarrow$ at time $t + 2$.

Special transmission states These are used to communicate special stimuli. They behave very similarly to the ordinary transmission states: they receive stimuli from their neighbors and emit in one direction. However, they can only receive special stimuli from other special transmission states or ordinary stimuli from confluent states. Moreover, they only emit special stimuli. Just like in the ordinary case, there are eight special transmission states.

Unexcitable state Unexcitability is the default state of each cell. In this state, the cell does not react to received stimuli by emitting. Instead, a received stimulus from a transmission state (ordinary or special alike) will turn an unexcitable cell into a sensitized one.

Sensitized states These are intermediate states between the unexcitable state and the quiescent ones. The transformation of the unexcitable state into an excitable one is realized in various steps through the reception of ordinary or special stimuli from transmission states. These stimuli can turn a sensitized state into another sensitized state or into the quiescent form of an ordinary transmission, special transmission, or confluent state. There are eight sensitized states.

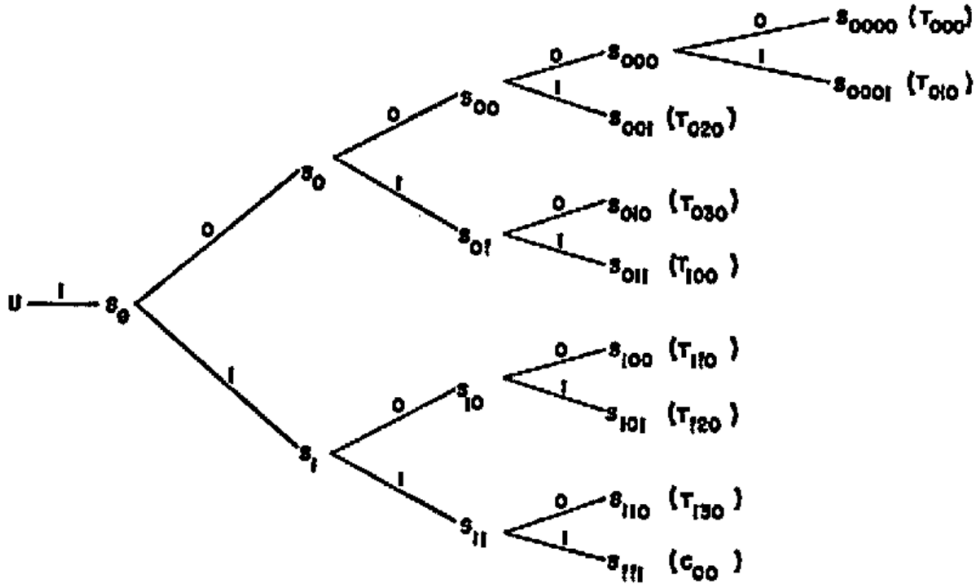

Figure 4.3: Sensitized states and the transformation tree from von Neumann (1966). The tree depicts how the unexcitable state U can be transformed into any quiescent state. The inner nodes S_ϵ are the sensitized states, whereas the leaves are quiescent transmission or confluent states. Each edge represents a single transformation step, and the number on it indicates if a stimulus is needed or not for the transformation to happen.

4.2.2 Ordinary and special stimuli

In the cellular framework, the cell is the elementary component, and an automaton is a configuration of cells. Constructing a new automaton involves turning an area of unexcitable cells into a prescribed configuration of excitable cells. This requires cells to possess both constructive and destructive abilities over each other (von Neumann 1966, 272). Constructive abilities are needed to create a path of transmission or confluent cells to the area where the new automaton should be created and to turn the unexcitable cells there into the prescribed states. Once this process is over, the destructive abilities are needed to remove the path by killing the cells along it, that is, reverting them to the unexcitable state.

Since the same stimulus cannot hold both constructive and destructive powers on the same state, von Neumann introduced two types of stimuli. The result is a dual setting (von Neumann 1966, 142), where ordinary stimuli, carried by ordinary transmission states, kill special transmission states; whereas special stimuli, carried by special transmission states, kill ordinary transmission states. Special stimuli are also destructive on confluent states. Importantly, confluent states serve as conversion points between ordinary and special stimuli: special transmission states can receive ordinary stimuli from confluent states without being killed and emit special stimuli.

Finally, both ordinary and special stimuli hold constructive abilities. Indeed, they can be both used to turn unexcitable states into sensitized ones, and these into excitable ones.

4.2.3 The automaton's description

In the cellular framework, an automaton X is a configuration of cells. A description $\sigma(X)$ can be easily obtained by considering the minimal rectangle that contains X and replacing each cell with a unique natural number representing its state. The description can also include the coordinates of the bottom-left corner of the rectangle, together with its width and height. This way, $\sigma(X)$ can be represented as a sequence of integers. This is the approach proposed by von Neumann (von Neumann 1966, 116–117).

Constructing the automaton X from its description $\sigma(X)$ is easier than copying the original X. Indeed, this second option requires determining the structure of X and the states of its cells. This could only be achieved by exploring the automaton through a series of stimuli, which could activate X in unpredictable ways (von Neumann 1966, 121–122).

4.2.4 The linear array L

Having already reduced self-reproduction to construction-universality (see Section 4.1), von Neumann focused on developing a universal constructor. Since the

constructor should be fixed, but the automata's descriptions could be arbitrarily large, these cannot be part of the constructor itself. To solve the problem, von Neumann applied again one of "Turing's tricks", and equipped his automaton with an arbitrarily extendible linear memory, which we will refer to as L.

L is a horizontal array of cells that extends to the right of the universal constructor. In it, the description $\sigma(X)$ is stored in binary format. Since each component of $\sigma(X)$ is an integer, a binary encoding is straightforward. Moreover, special encodings are used to separate subsequent elements within L and to mark its end (von Neumann 1966, 112–116).

Each cell of L holds a single digit (0 or 1). To represent the two values, different states are used: the unexcitable state represents 0, while the quiescent ordinary transmission state directed downwards represents 1 (von Neumann 1966, 202–203).

Since these two states transmit ordinary stimuli differently, reading the content of a generic cell x_n can be achieved by simply sending an appropriate sequence of ordinary stimuli through x_n from the cell above it, and analyzing the series of stimuli received in the cell below it (von Neumann 1966, 208–209). Indeed, von Neumann had designed an organ capable of discriminating series of stimuli (von Neumann 1966, 187–190). However, its treatment is beyond the scope of the present work.

4.2.5 The memory control MC

In von Neumann's design, the reading of L is carried out by a memory control organ, which we will refer to as MC. MC sends a series of stimuli through a cell x_n of L by using a connecting loop C_1 of ordinary transmission states, through which it also receives the sequence produced by x_n. When the constructor is started, we can assume C_1 to be passing through x_0, the very first cell of L.

Once MC has read cell x_n, it may need to read one of its immediate neighbors. To do so, C_1 needs to be lengthened or shortened, which requires changing the state of the last cell before x_n: to make C_1 longer, the cell should become an ordinary transmission state directed to the right; to make it shorter, it should become unexcitable. This is achieved by changing the whole loop between ordinary

and special transmission states (von Neumann 1966, 214–220). Doing so requires sending a series of stimuli through C_1. In this case, the number of impulses should be proportional to the length of the loop because each ordinary state needs a specific number of impulses to turn into a special state and vice versa. This process is managed by MC with the use of a timing loop C_2 that matches C_1's length. Together, loops C_1 and C_2 coordinate the timing for lengthening and shortening each other. The details involve a complex use of stimuli.

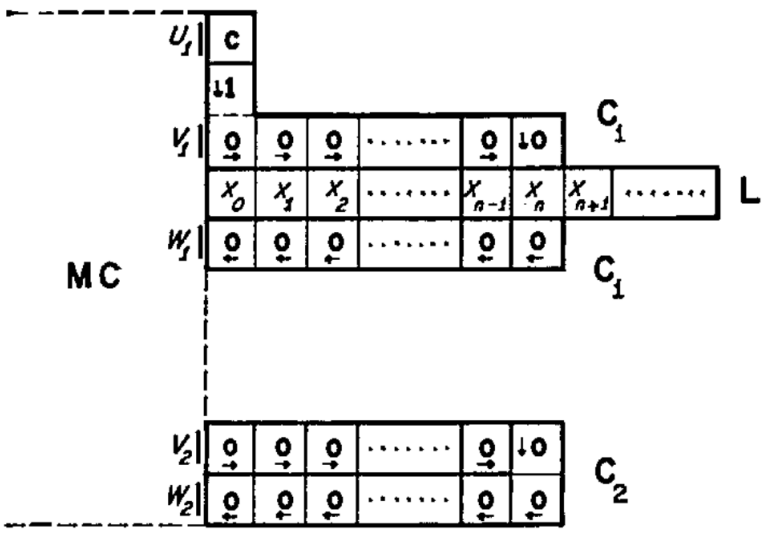

Figure 4.4: The linear array L and the memory control MC from von Neumann (1966). Cell x_n can be read by sending a series of impulses through C_1 from v_1 to w_1. The timing loop C_2 has the same length as C_1, and is used in the lengthening/shortening process.

4.2.6 The constructing unit CU

The final component of von Neumann's universal constructor is the constructing unit CU, which serves two purposes. The first one is controlling the operations of MC. It is CU that first starts MC, which reads one cell and communicates its value to CU. At this point, the constructing unit decides whether MC should move the connecting loop to another cell, and possibly restarts the reading process.

As the name suggests, the second purpose of CU is managing the construction of a new automaton X. We can assume this is built in an area to the top-right of the universal constructor (von Neumann 1966, 271). The area is reached via a constructing arm, which is essentially a path of transmission states. The arm needs to be extended and contracted to reach all the cells within the rectangle encompassing X. This is achievable with the same approach used for C_1 and C_2 in the memory control or with a more economical design, which, however, was only outlined by von Neumann (von Neumann 1966, 272–275).

CU operates the constructing arm based on the information present in L and received from MC (von Neumann 1966, 280–285). The information about the size of X is used to pass from one row or column to the following; whereas the information about the state of each cell is used to properly transform the unexcitable cells. X cannot be built in an excited state, as this could interfere with the construction process. Instead, X is built in a quiescent state and is activated, after full construction, by a specific initial stimulus.

Von Neumann did not develop a full design of the constructing unit. A functional outline was given by Burks and is reported in the next section. However, both von Neumann and Burks recognized that CU is a finite automaton, which can thus be built by translating its operational algorithm into machine design (von Neumann 1966, 285). A complete design was developed by Thatcher (1964) in "Universality in the von Neumann Cellular Model", and a working implementation was created by Pesavento (1995) in "An Implementation of von Neumann's Self-Reproducing Machine".

4.2.7 The universal constructor

Overall, von Neumann's universal constructor M_c is composed of two parts: a memory component (formed by MC, L, C_1, and C_2) and a constructing component (formed by CU and the constructing arm). This is very similar to the logical design of the EDVAC, with the constructing component being analogous to the combination of the central arithmetical and the central control parts (see Section 1.1.3).

The complete algorithm of M_c was developed by Burks in the final chapter of the manuscript "The Theory of Automata: Construction, Reproduction, Homogeneity" (von Neumann 1966, 283–285). It works as follows:

1. MC reads the coordinates of the bottom-left corner of X from L, together with its dimensions, and communicates them to CU.

2. Using the information from MC, CU extends the constructing arm to the top-left corner of the area where X will be built.

3. CU constructs X two rows at a time via the constructing arm. It uses the dimensional information from step 1 to determine when a row has been completed and when the whole automaton has been completed. In this process, CU communicates with MC to read the prescribed state of each cell from L[3].

4. CU injects an initial stimulus into X from a specific cell via the constructing arm. At this point, X becomes operational.

5. CU withdraws the constructing arm.

[3]To simplify this process, Burks assumed the description $\sigma(X)$ to be sorted according to the way it will be used by CU (von Neumann 1966, 280).

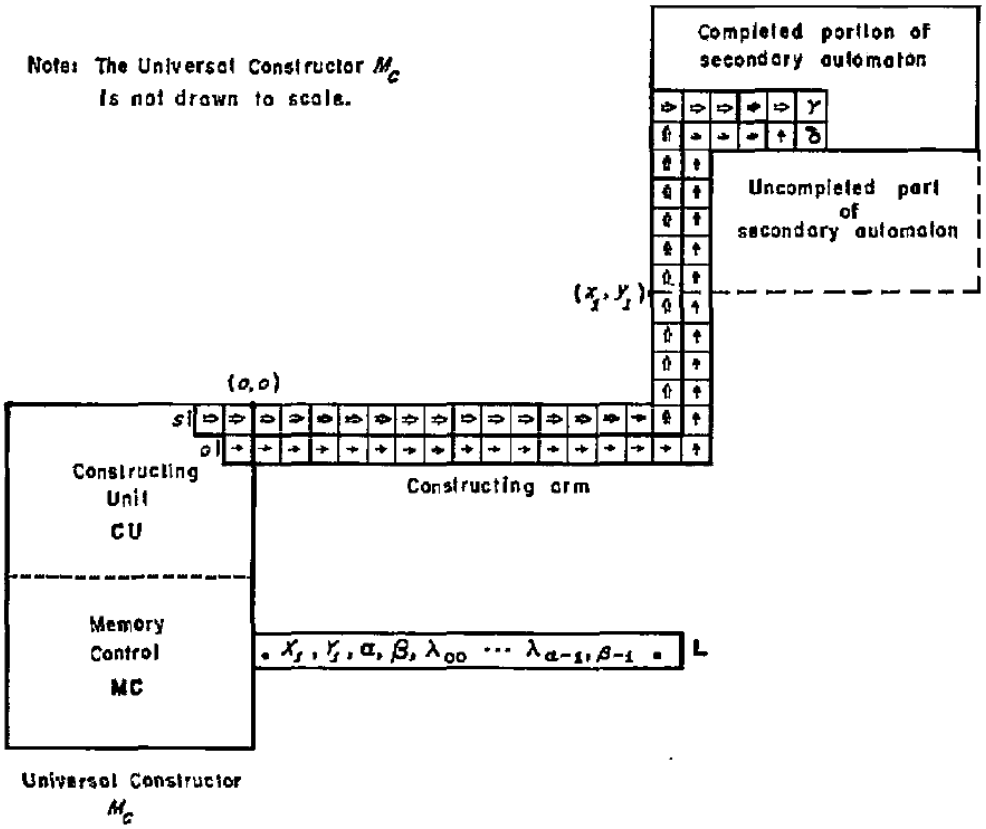

Figure 4.5: The universal constructor M_c and the constructing arm from von Neumann (1966). The universal constructor is composed of the memory control MC, which reads the linear array L, and the constructing unit CU. Through the constructing arm, CU builds a new automaton ("secondary automaton" in the picture) from top to bottom.

4.2.8 The self-reproducing automaton

Compared to the synthetic description of the universal constructor given in Section 4.1, the above design lacks a component B for the copy of $\sigma(X)$ (L in this case). This was noted by Burks, who pointed out that the constructing unit could be easily extended to copy the linear array L using the constructing arm (von Neumann 1966, 295).

A possible procedure involves scanning and copying L one cell at a time. Using the information on the position of the bottom-left corner of X, CU extends the

constructing arm to a cell at a fixed distance below it. After that, it communicates with MC to read the content of L one cell at a time from left to right. For each cell, CU uses the constructing arm to replicate the state it has read; then, it moves the constructing arm one position to the right. When the end of the array is reached, a copy of L will have been created.

By introducing this copying operation, we obtain a modified universal constructor M_c^*. With this change, the same functioning defined in the synthetic description can be achieved:

1. M_c^* constructs X following steps 1–3 described in the previous section;
2. M_c^* copies L next to X;
3. M_c^* carries out steps 4–5 of the previous section, starting X and withdrawing the constructing arm.

Since M_c^* can be completely defined, its description $\sigma(M_c^*)$ can be stored in L. In this case, M_c^* will replicate itself. Moreover, if L contains $\sigma(M_c^* + M_u)$, where M_u is another automaton, M_c^* will reproduce itself while also constructing M_u. This is the behavior of the self-reproducing automaton synthetically described in Section 4.1.

4.3 Von Neumann's results

Following the same functional approach he had used in the design of the EDVAC (see Section 1.1.3), von Neumann devised a self-reproducing automaton composed of two functional organs: a memory (formed by MC, L, C_1 and C_2) and a constructing component (formed by CU and the constructing arm). In his work, he integrated McCulloch and Pitts' idea of formal neurons (see Section 1.3) into the definition of his cellular model. Moreover, mirroring Turing's work on the universal machine (see Section 1.2.2), he equipped his universal constructor with an arbitrary extendible linear memory for storing automata's descriptions.

By developing the axiomatization of the cellular model, he was able to avoid most of the kinematic considerations and design a self-reproducing automaton

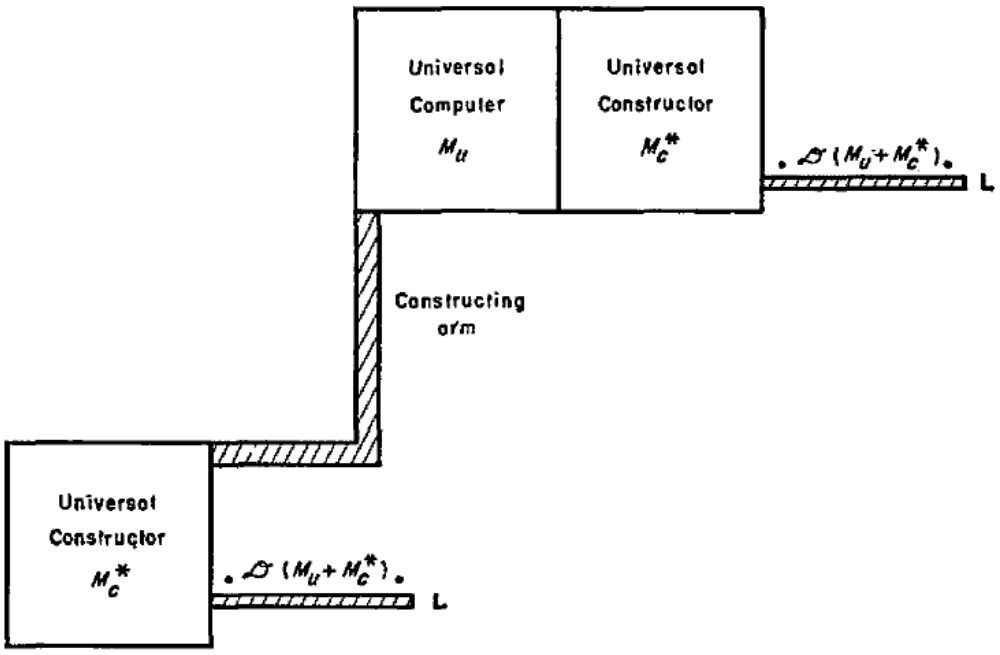

Figure 4.6: The self-reproducing automaton M_c^*, adapted from von Neumann (1966). If L contains the description $\sigma(M_c^* + M_u)$, M_c^* will replicate itself, but also construct a new automaton M_u. Since L is copied and attached to the newly constructed M_c^*, the process can be repeated.

that addresses the questions of constructibility, construction-universality, and self-reproduction. In this section, we will examine in greater detail what von Neumann was able to achieve, as well as the limitations of his results.

4.3.1 Logical universality

Although this work mostly focused on self-reproduction, von Neumann's cellular model is actually logically universal[4], as Turing machines can be embedded into it. This can be achieved by embedding the tape (together with the system to operate on it) and a finite automaton.

[4]In the sense of Section 3.1.

Embedding of the tape A linear array of cells L can be used as the tape, while the connecting loop C_1 can be used to simulate the head of the machine. Reading and moving operations are performed by the memory control MC as described in Section 4.2.5. Writing can be achieved by simply transmitting special stimuli through C_1 to the cell of L it is connected to.

Embedding of the finite automaton This was shown by Burks (von Neumann 1966, 266–270). The procedure involves simulating each state of the automaton with a copy of a state organ SO. All these organs are interconnected, via the usual transmission and confluent states, based on the transition function of the automaton. Through MC, each SO receives stimuli representing the content of the cell of L being read. However, only one organ is active at any time. The active organ directs MC to write a new symbol and change the length of C_1.

4.3.2 The class of constructible automata

Von Neumann's design constructs automata in an initially quiescent state and then activates them (see Section 4.2.6). Therefore, the class of constructible automata consists of those configurations that can eventually arise from an initially quiescent one after activation. As pointed out by Burks, this is a proper subclass of all finite automata that can exist within the cellular model (von Neumann 1966, 291). Indeed, there are automata that cannot be built from the universal constructor, such as the one shown in Figure 4.7 below.

Overall, limiting the universal constructor to only build initially quiescent automata appears to be a reasonable choice. From an engineering and control perspective, automata that can be activated as needed are easier to manage and integrate into larger systems. Indeed, the possibility of activating them from a known stationary state makes them more predictable. Moreover, a quiescent automaton avoids interferences with the constructor during the building process, as already discussed in Section 4.2.6. Finally, if non-constructible automata are as unstable as the one in Figure 4.7, they are of little practical utility.

While we are not aware of any works that comprehensively define the class of automata that cannot be constructed by the universal constructor, several studies do address those configurations that cannot be built within cellular models due to inherent limitations of the models themselves. This situation applies to von Neumann's framework and is discussed in the next section.

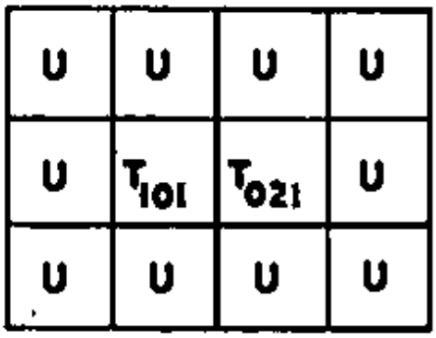

Figure 4.7: A non-costructible automaton from von Neumann (1966). T_{101} and T_{021} are a special and an ordinary transmission state, respectively, with output directions towards each other. Both are excited. The automaton is composed of these two cells and a surrounding layer of unexcitable cells U. Since special and ordinary transmission states hold reciprocal destructive powers, the configuration can only exist for a single time step, before T_{101} and T_{021} turn each other into U. A constructing arm can easily create the quiescent forms of T_{101} and T_{021}. However, it will not have time to withdraw from the surrounding area in the single timestep window between their activations and the mutual killing. Therefore, the complete configuration cannot be constructed.

4.3.3 Limitations of von Neumann's cellular model

The automaton of Figure 4.7 is non-constructible because it does not admit predecessors, that is, there is no configuration that will turn into it in a single timestep. Therefore, it can only exist as part of the initial global configuration at time $t = 0$. Adopting John W. Tukey's terminology, Edward F. Moore referred to this type of configurations as "Garden of Eden" (GOE) (Moore 1962, 23). We

will now present a theorem showing that the existence of such non-constructible configurations depends on the nature of the cellular model itself.

Amoroso and Cooper's theorem Amoroso and Cooper (1970) expanded on previous work by Moore (1962) and Myhill (1963) to develop a *necessary and sufficient* condition for the existence of GOE configurations in the finite case.

Their condition is based on the transformation function of the cellular model. Indeed, a cellular model is characterized by a (deterministic) local transformation[5] σ, which determines the state of each cell at time t as a function of its own state and those of its neighbors at time $t-1$. When applied simultaneously to all cells, σ induces a global transformation τ.

Amoroso and Cooper proved the following theorem: when considering only finite configurations, GOE configurations exist if and only if τ is not bijective.

Reversible cellular models Following von Neumann and Burks' work, several other cellular models were developed, including reversible cellular models (Toffoli 1977). These are characterized by bijective transformation functions and, consequently, by the absence of Garden-of-Eden configurations. However, Kari (1994) demonstrated that the problem of determining whether a given two-dimensional cellular model is reversible is undecidable. There exist ways to devise reversible cellular models, but these involve more complex approaches than the one used by von Neumann (Kari 2018, 152–157).

Significance for von Neumann's work Amoroso and Cooper's theorem applies to von Neumann's case, as his automata are all finite. The theorem proves that the existence of non-constructible configurations is at least partially due to the definition of the cellular model.

Unfortunately, we are not aware of any studies that explicitly define the relationship between the class of GOE configurations and the class of configurations that are non-constructible by the universal constructor.

[5] We verbally defined this function in Section 4.2.1.

Given Kari's undecidability result and the absence of prior studies on cellular automata, it would have been very difficult—if not almost impossible—for von Neumann to have devised a reversible cellular model. In fact, his work essentially laid the foundations for the study of cellular automata, making the aforementioned developments possible.

4.3.4 Answers to the questions

Having covered logical universality and the existence of non-constructible automata, we can now summarize the results von Neumann achieved with respect to the questions in Section 3.1:

(A) Logical universality: von Neumann devised a cellular model where Turing machines can be embedded, as shown in Section 4.3.1.

(B) Constructibility: making use of an arbitrarily extendible linear array L with a description $\sigma(X)$ and of a construction arm, the universal constructor of Section 4.2.7 can build an *initially quiescent* automaton X.

(C) Construction-universality: the universal constructor also provides a *partial* affirmative answer to the question of whether a single automaton can construct all other constructible automata. Indeed, it can construct all other *initially quiescent* automata.

(D) Self-reproduction: following the procedure in Sections 4.1 and 4.2.8, the universal constructor can be modified to build itself, as well as other automata, thus proving self-reproduction is possible.

(E) Evolution: von Neumann made a few considerations on evolution (see the last paragraph of Section 4.1), but did not address the problem in his design of the self-reproducing automaton.

Conclusion

Von Neumann's approach and methodology in the study of automata were extraordinary in many respects.

His efforts were deeply interdisciplinary. Beginning with the technological goal of building better computers, he attempted to create a logical theory that could bridge biology and engineering by unifying living organisms and machines. He incorporated results from logic and computability, such as those of Turing, but also from cognitive science, such as those of McCulloch and Pitts. Moreover, he founded his methodology on the mathematical process of axiomatization.

Many of his choices were ahead of their times. He saw the necessity of a structural study of machines based on abstraction at a time when computer science had yet to be born. He recognized the connections between the work of Turing and that of McCulloch and Pitts, but also attempted to expand their results by addressing constructibility. Not only did he see analogies between living and artificial systems, but he actually envisioned achieving equal complexity in both.

Thanks to his ability to connect different fields, as well as his own personal experience, von Neumann was able to make great use of very few prior results on automata and to design a universal constructor and a self-reproducing machine. These expanded the knowledge and study of automata, laying the foundations of cellular automata theory.

Overall, von Neumann's work on automata and self-reproduction shows that the combination of an engineering goal, an innovative approach, and a long-term vision can give rise to a new theory and meaningful results, even with very limited prior foundations.

Acknowledgments

Use of GenAI tools

GenAI tools, mainly ChatGPT (versions GPT-4o, GPT-4.1-mini, and o4-mini) and Claude (version Sonnet 4), were used as aids in the writing of this thesis to find synonyms, reword sentences, enhance the clarity of paragraphs, and proofread. The tools were not used to generate entire paragraphs from scratch, and their suggestions were carefully analyzed before being integrated into the work.

Personal acknowledgments

I am very grateful to my supervisor, Professor Simone Martini, for his personal and academic support in the process of writing this thesis.
I am deeply thankful to my family for their immense support in all forms.

Bibliography

Al-Hashimi, H. M. (2023). Turing, von Neumann, and the Computational Architecture of Biological Machines. *Proceedings of the National Academy of Sciences*, 120(25):e2220022120.

Amoroso, S. and Cooper, G. (1970). The Garden-of-Eden Theorem for Finite Configurations. *Proceedings of the American Mathematical Society*, 26(1):158–164.

Aspray, W. (1985). The Scientific Conceptualization of Information: A Survey. *IEEE Annals of the History of Computing*, 7(02):117–140.

Aspray, W. (1990). *John von Neumann and the Origins of Modern Computing*. History of Computing. MIT Press.

Goldstine, H. H. and Goldstine, A. (1946). The Electronic Numerical Integrator and Computer (ENIAC). *Mathematical Tables and Other Aids to Computation*, 2(15):97–110.

Kari, J. (1994). Reversibility and Surjectivity Problems of Cellular Automata. *Journal of Computer and System Sciences*, 48(1):149–182.

Kari, J. (2018). Reversible Cellular Automata: From Fundamental Classical Results to Recent Developments. *New Generation Computing*, 36:145–172.

Kemeny, J. G. (1955). Man Viewed as a Machine. *Scientific American*, 192(4):58–67.

Masani, P. R. (1990). *The Cybernetical Movement and von Neumann's Letter, 1946*, pages 239–250. Birkhäuser Basel.

McCulloch, W. S. and Pitts, W. (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, 5:115–133.

McMullin, B. (2000). John von Neumann and the Evolutionary Ggrowth of Complexity: Looking Backward, Looking Forward. *Artificial Life*, 6(4):347–361.

Moore, E. F. (1962). Machine Models of Self-Reproduction. In *Proceedings of Symposia in Applied Mathematics*, volume 14, pages 17–33. American Mathematical Society.

Myhill, J. (1963). The converse to Moore's Garden-of-Eden theorem. *Proceedings of the American Mathematical Society*, 14:685–686.

Pesavento, U. (1995). An Implementation of von Neumann's Self-Reproducing Machine. *Artificial Life*, 2(4):337–354.

Piccinini, G. (2020). The First Computational Theory of Cognition: McCulloch and Pitts's "A Logical Calculus of the Ideas Immanent in Nervous Activity". In *Neurocognitive Mechanisms: Explaining Biological Cognition*. Oxford University Press.

Thatcher, J. W. (1964). Universality in the von Neumann Cellular Model. Technical report, University of Michigan.

Toffoli, T. (1977). Computation and Construction Universality of Reversible Cellular Automata. *Journal of Computer and System Sciences*, 15(2):213–231.

Turing, A. M. (1937). On Computable Numbers, with an Application to the *Entscheidungsproblem*. *Proceedings of the London Mathematical Society*, s2-42(1):230–265.

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 59(236):433–460.

Ulam, S. (1958). John von Neumann 1903–1957. *Bulletin of the American Mathematical Society*, 64(3.P2):1 – 49.

von Neumann, J. (1945). First Draft of a Report on the EDVAC. Technical report, Moore School of Electrical Engineering, University of Pennsylvania.

von Neumann, J. (1948). The General and Logical Theory of Automata. In Taub, A. H., editor, *Design of Computers, Theory of Automata and Numerical Analysis*, volume V of *Collected Works*, pages 288–328. Pergamon Press. *Read at the Hixon Symposium in 1948. Published in 1963.*

von Neumann, J. (1949). Theory and Organization of Complicated Automata. In Burks, A. W., editor, *Theory of Self-Reproducing Automata*, chapter I, pages 29–87. University of Illinois Press. *Lectures delivered by von Neumann at the University of Illinois in 1949. Edited by Arthur Burks in 1966.*

von Neumann, J. (1952). Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components. In Shannon, C. E. and McCarthy, J., editors, *Automata Studies*, pages 43–98. Princeton University Press. *Lectures delivered at the California Institute of Technology in 1952. Published in 1956.*

von Neumann, J. (1956). *The Computer and the Brain.* Yale University Press, 1 edition. *Manuscript written between 1955 and 1956. First edition published in 1958.*

von Neumann, J. (1966). The Theory of Automata: Construction, Reproduction, Homogeneity. In Burks, A. W., editor, *Theory of Self-Reproducing Automata*, chapter II, pages 89–380. University of Illinois Press. *Manuscript written by von Neumann in 1952 and 1953. Completed and edited by Arthur Burks in 1966.*