VON NEUMANN'S CONTRIBUTIONS TO AUTOMATA THEORY

CLAUDE E. SHANNON

The theory of automata is a relatively recent and by no means sharply defined area of research. It is an interdisciplinary science bordered mathematically by symbolic logic and Turing machine theory, bordered engineering-wise by the theory and the use, particularly for general non-numerical work, of large scale computing machines, and bordered biologically by neurophysiology, the theory of nerve-nets and the like. Problems range from Godel-type questions (relating to Turing machines and decision procedures), to questions of duplication, of various biological phenomena in a machine (e.g., adaptation, self-reproduction and self-repair).

Von Neumann spent a considerable part of the last few years of his life working in this area. It represented for him a synthesis of his early interest in logic and proof theory and his later work, during World War II and after, on large scale electronic computers. Involving a mixture of pure and applied mathematics as well as other sciences, automata theory was an ideal field for von Neumann's wide-ranging intellect. He brought to it many new insights and opened up at least two new directions of research. It is unfortunate that he was unable to complete the work he had in progress, some of which is in the form of rough notes or unedited lectures, and for some of which no record exists apart from his colleagues' memories of casual conversations.

We shall not here discuss his tremendously important contributions to computing machines and their use—his ideas on their logical organization, [1; 3] the use of flow diagrams for programming, [3; 4; 5] methods of programming various problems such as the inversion of matrices, [2] the Monte Carlo method, and so on,—but restrict ourselves to the automata area proper.

Reliable machines and unreliable components. One important part of von Neumann's work on automata relates to the problem of designing reliable machines using unreliable components [10]. Given a set of building blocks with some positive probability of malfunctioning, can one by suitable design construct arbitrarily large and complex automata for which the overall probability of incorrect output is kept under control? Is it possible to obtain a probability of

Received by the editors February 10, 1958.

error as small as desired, or at least a probability of error not exceeding some fixed value (independent of the particular automaton)?

We have, in human and animal brains, examples of very large and relatively reliable systems constructed from individual components, the neurons, which would appear to be anything but reliable, not only in individual operation but in fine details of interconnection. Furthermore, it is well known that under conditions of lesion, accident, disease and so on, the brain continues to function remarkably well even when large fractions of it are damaged.

These facts are in sharp contrast with the behavior and organization of present day computing machines. The individual components of these must be engineered to extreme reliability, each wire must be properly connected, and each order in a program must be correct. A single error in components, wiring or programming will typically lead to complete gibberish in the output. If we are to view the brain as a machine, it is evidently organized for protection against errors in a way quite different from computing machines.

The problem is analogous to that in communication theory where one wishes to construct codes for transmission of information for which the reliability of the entire code is high even though the reliability for the transmission of individual symbols is poor. In communication theory this can be done by properly introduced redundancy, and some similar device must be used in the case at hand. Merely performing the same calculation many times and then taking a majority vote will not suffice. The majority vote would itself be taken by unreliable components and thus would have to be taken many times and majority votes taken of the majority votes. And so on. We are face to face with a "Who will watch the watchman" type of situation.

To attack these problems, von Neumann first set up a formal structure for automata. The particular system he chooses is somewhat like the McCullough-Pitts model; networks made up of a number of interconnected components, each component of a relatively simple type. The individual components receive binary inputs over a set of different input lines and produce a binary output on an output line. The output occurs a certain integer number of time units later. If the output were a function of the inputs, we would have a reliable component that might perform, for example, operations of "and," "not," "Sheffer stroke," etc. However, if the output is related only statistically to the input, if, for example, with probability $1-\epsilon$ it gives the Sheffer stroke function and with probability ϵ the negative of this, we have an unreliable component. Given an unlimited number of such unreliable elements, say of the Sheffer stroke type, can one construct a reliable version of any given automaton?

Von Neumann shows that this can be done, and in fact does this by two quite different schemes. The first of these is perhaps the more elegant mathematically, as it stays closely within the prescribed problem and comes face to face with the "watchman" problem. This solution involves the construction from three unreliable sub-networks, together with certain comparing devices, of a large and more reliable sub-network to perform the same function. By carrying this out systematically throughout some network for realizing an automaton with reliable elements, one obtains a network for the same behavior with unreliable elements.

The first solution, as he points out, suffers from two shortcomings. In the first place, the final reliability cannot be made arbitrarily good but only held at a certain level ϵ (the ϵ depending on the reliability of the individual components). If the individual components are quite poor the solution, then, can hardly be considered satisfactory. Secondly, and even more serious from the point of view of application, the redundancy requirements for this solution are fantastically high in typical cases. The number of components required increases exponentially with the number n of components in the automaton being copied. Since n is very large in cases of practical interest, this solution can be considered to be only of logical importance.

The second approach involves what von Neumann called the multiplex trick. This means representing a binary output in the machine not by one line but by a bundle of N lines, the binary variable being determined by whether nearly all or very few of the lines carry the binary value 1. An automaton design based on reliable components is, in this scheme, replaced by one where each line becomes a bundle of lines, and each component is replaced by a sub-network which operates in the corresponding fashion between bundles of input and output lines. Von Neumann shows how such sub-networks can be constructed. He also makes some estimates of the redundancy requirements for certain gains in reliability. For example, starting with an unreliable "majority" organ whose probability of error is 1/200, by a redundancy of 60,000 to 1 a sub-network representing a majority organ for bundles can be constructed whose probability of error is 10^{-20}. Using reasonable figures this would lead to an automaton of the complexity and speed of the brain operating for a hundred years with expectation about one error. In other words, something akin to this scheme is at least possible as the basis of the brain's reliability.

Self-reproducing machines. Another branch of automata theory developed by von Neumann is the study of self-reproducing machines—is it possible to formulate a simple and abstract system of "machines" which are capable of constructing other identical machines, or even more strongly, capable of a kind of evolutionary process in which successive generations construct machines of increasing "complexity." A real difficulty here is that of striking the proper balance between formal simplicity and ease of manipulation, on the one hand, and approximation of the model to real physical machines on the other hand. If reality is copied too closely in the model we have to deal with all of the complexity of nature, much of which is not particularly relevant to the self-reproducing question. However, by simplifying too much, the structure becomes so abstract and simplified that the problem is almost trivial and the solution is unimpressive with regard to solving the philosophical point that is involved. In one place, after a lengthy discussion of the difficulties of formulating the problem satisfactorily, von Neumann remarks: "I do not want to be seriously bothered with the objection that (a) everybody knows that automata can reproduce themselves (b) everybody knows that they cannot."

Von Neumann spent a good deal of time on the self-reproduction problem, discussing it briefly in the Hixon Symposium paper [**8**] and later in more detail in uncompleted manuscripts [**12**].

He actually considered two different formulations of the problem. In the Hixon Symposium paper and in earlier lectures on this subject, a model is discussed in which there are a small number of basic components from which machines are made. These might be, for example, girders, a sensing organ (for sensing the presence of other parts), a joining organ (for fastening other parts together), etc. Machines are made by combinations of these parts and exist in a geometrical environment with other similar parts freely available.

Certain machines, made from these parts, are capable of gathering and working with components from the environment. It is possible also to construct "programmed" machines which follow a long sequence of instructions much as a computer does. Here, however, the instructions relate to manipulating parts rather than carrying out long calculations. The situation is somewhat analogous to that of Turing machines and indeed there is a notion of a *universal constructing machine* which can, by proper programming, imitate any machine for construction purposes. Von Neumann indicates how such a universal machine, together with a program-duplicating part can be made into a self-reproducing machine.

This model is a very interesting one but, involving as it does complex considerations of motion of parts in a real Euclidean space, it would be tremendously difficult to carry out in detail, even if one ignored problems of energy, noise in the environment, and the like. At any rate, von Neumann abandoned this model in his later work in favor of a simpler construction.

The second type of self-reproducing system is described in an unfinished book for the University of Illinois Press. This second model is perhaps a little more suggestive of biological reproduction in the small (say at the cellular or even molecular level) although it is not closely patterned after any real physical system. Consider an infinite array of squares in the Euclidean plane, each square or "cell" capable of being in any of a number of states. The model that von Neumann developed had cells with twenty-nine possible states. Time moves in discrete steps. The state of a cell at a given time is a function of its state at the preceding time and that of its four nearest neighbors at the preceding time. As time progresses, then, the states of all cells evolve and change according to these functional relations. A certain state of the cells is called "quiescent" and corresponds to an inactive part of the plane. By proper construction of the functional equations it is possible to have groups of neighboring "active" cells which act somewhat like a living entity, an entity capable of retaining its identity, moving about and even of reproduction in the sense of causing another group of cells to take on a similar active state.

In addition to the self-reproducing question, he considers to some extent the problem of "evolution" in automata—is it possible to design automata which will construct in successive generations automata in some sense more efficient in adapting to their environment. He points out the existence of a *critical size* of automaton built from a given type of component such that smaller automata can only construct automata smaller than themselves, while some automata of the critical size or larger are capable of self-reproduction or even evolution (given a suitable definition of efficiency).

Comparison of computing machines with the brain. A field of great interest to von Neumann was that of the relation between the central nervous system and modern large-scale computers. His Hixon Symposium paper relates to this theme as well as to the problem of self-reproducing machines. More particularly, the Silliman Memorial Lectures [**11**] (which he prepared but was unable to deliver) are largely concerned with this comparison.

While realizing the similarities between computers and nerve-nets, von Neumann was also clearly aware of and often emphasized the many important differences. At the surface level there are obvious differences in order of magnitude of the number and size of components and of their speed of operation. The neurons of a brain are much slower than artificial counterparts—transistors or vacuum tubes, but on the other hand they are much smaller, dissipate less power and there are many orders of magnitude more of them than in the largest computers. At a deeper level of comparison von Neumann stresses the differences in logical organization that must exist in the two cases. In part, these differences are implied by the difference in the kind of problem involved, "the logical depth," or the number of elementary operations that must be done in sequence to arrive at a solution. With computers, this logical depth may reach numbers like 10^7 or more because of the somewhat artificial and serial method of solving certain problems. The brain presumably, with more and slower components, operates on a more parallel basis with less logical depth and further, the problems it confronts are much less of the sequential calculation variety.

In the Silliman lectures, von Neumann touches briefly on a curious and provocative idea with some relevance to the foundations of mathematics. Turing, in his well known paper on computability, pointed out how one computing machine could be made to imitate another. Orders for the second machine are translated by a "short code" into sequences of orders for the first machine which cause it to accomplish, in a generally roundabout way, what the first machine would do. With such a translating code the first machine can be made to look, for computing purposes, like the second machine, although it is actually working inside in a different language. This procedure has become a commonplace and very useful tool in the everyday use of computers.

If we think of the brain as some kind of computing machine it is perfectly possible that the external language we use in communicating with each other may be quite different from the internal language used for computation (which includes, of course, all the logical and information-processing phenomena as well as arithmetic computation). In fact von Neumann gives various persuasive arguments that we are still totally unaware of the nature of the primary language for mental calculation. He states "Thus logics and mathematics in the central nervous system, when viewed as languages, must be structurally essentially different from those languages to which our common experience refers.

"It also ought to be noted that the language here involved may well correspond to a short code in the sense described earlier, rather than to a complete code: when we talk mathematics, we may be discussing a *secondary* language, built on the *primary* language truly used by the central nervous system. Thus the outward forms of our mathematics are not absolutely relevant from the point of view of evaluating what the mathematical or logical language *truly* used by the central nervous system is. However, the above remarks about reliability and logical and arithmetic depth prove that whatever the system is, it cannot fail to differ considerably from what we consciously and explicitly consider as mathematics."

In summary, von Neumann's contributions to automata theory have been characterized, like his contributions to other branches of mathematics and science, by the discovery of entirely new fields of study and the penetrating application of modern mathematical techniques. The areas which he opened for exploration will not be mapped in detail for many years. It is unfortunate that several of his projects in the automata area were left unfinished.

Selected Bibliography

1. *Preliminary discussion of the logical design of an electronic computing instrument.* Part I, Vol. I. With A. W. Burks and H. H. Goldstine. Report prepared for U. S. Army Ord. Dept. under Contract W-36-034-ORD-7481 (June 28, 1946, 2d ed. Sept. 2, 1947), 42 pp.

2. *Numerical inverting of matrices of high order.* With H. H. Goldstine. Amer. Math. Soc. Bull. vol. 53 (1947) pp. 1021–1099.

3. *Planning and coding of problems for an electronic computing instrument.* Part II, Vol. I. With H. H. Goldstine. Report prepared for U. S. Army Ord. Dept. under Contract W-36-034-ORD-7481, 1947, 69 pp.

4. *Planning and coding of problems for an electronic computing instrument.* Part II, Vol. II. With H. H. Goldstine. Report prepared for U. S. Army Ord. Dept. under Contract W-36-034-ORD-7481, 1948, 68 pp.

5. *Planning and coding of problems for an electronic computing instrument.* Part II, Vol. III. With H. H. Goldstine. Report prepared for U. S. Army Ord. Dept. under Contract W-36-034-ORD-7481, 1948, 23 pp.

6. *The future of high-speed computing.* Proc., Comp. Sem., Dec. 1949, published and copyrighted by IBM, 1951, p. 13.

7. *Numerical inverting of matrices of high order,* II. With H. H. Goldstine. Proc. Amer. Math Soc. vol. 2 (1951) pp. 188–202.

8. *The general and logical theory of automata.* "Cerebral mechanisms in behavior—The Hixon symposium," September 1948, Pasadena, edited by L. A. Jeffress, John Wiley and Sons, Inc., New York, 1951, pp. 1–31.

9. *Entwicklung und Ausnutzung neuerer mathematischer Maschinen.* Arbeitsgemeinschaft für Forschung des Landes Nordrhein-Westfalen, Heft 45. Düsseldorf, September 15, 1954.

10. *Probabilistic logics and the synthesis of reliable organisms from unreliable components,* "Automata studies," edited by C. E. Shannon and J. McCarthy, Princeton University Press, 1956, pp. 43–98.

11. The Silliman Memorial Lectures. (In press). Yale University Press. To be published, 1958.

12. Unfinished manuscript for the University of Illinois Press relating to the theory of automata and, particularly, self-reproducing machines.

Massachusetts Institute of Technology