SPECULATIONS

ON

PERCEPTRONS AND OTHER AUTOMATA

by

I. J. Good

Royal Naval Scientific Service

IBM
Research Lecture
RC-115
June 2, 1959

International Business Machines Corporation
Research Center
Yorktown Heights, New York

Printed in U.S.A.

"Mental experiences are the events in the universe of which we have the most direct knowledge."

W. Russell Brain
"Mind, Perception and Science"

"I think, therefore I am Brain."

SPECULATIONS ON PERCEPTRONS AND OTHER AUTOMATA†

Harold Jeffreys once said that the brain may be an imperfect thinking machine, but is the only one available. For over 2,000 years men have devised tools that aid or to some extent replace thought. Most of these men were philosophers. There were Aristotle's logic, Boole's "Laws of Thought"; the theory of probability founded by Cardan, Leibnitz, Pascal, Fermat and others; simple calculating machines designed by Leibnitz and Pascal; and a complicated calculating machine designed by Charles Babbage but never completed, owing to the short-sightedness of the British Treasury. But it was not until the advent of electronic computers that many scientists began to suspect that robots may be feasible. The justification for this view is not so much that existing electronic computers have yet been programmed to do much that is very close to thinking, as that electronics provides a very rapid and reliable method of handling information.

There are logically two different approaches to the problem of designing intelligent machines. The first is to try to imitate the physiology of the brain to some extent; the other is to analyse various types of activity usually described as intelligent, or requiring the use of intelligence, and then to simulate this behavior by any type of machine (and in particular by means of a program run on a general-purpose computer). Most of the work has been in the latter direction, as for example in the checkers and chess-playing programs, and the programs for proving theorems in logic and geometry. On the face of it one can hardly hope to imitate intelligence artificially without accurately understanding what is meant by intelligence, but this impression is not logically watertight since it may be possible to copy the structure of the brain without understanding it.

Frank Rosenblatt is one of a large number of people who believes that those who are interested in the possibility of artificial intelligence may learn something useful from the physiology of the brain. One of his contributions to the field was the invention of the name "Perceptron".

Imagine a newly born baby who is shown a large number of squares and circles. Two questions arise. The first is whether the baby will in some sense form the concepts of squares and circles, and the second is whether it will do so if it is rewarded for giving

† This paper is based on a lecture which was sponsored by the Machine Organization Department, IBM Mohansic Research Laboratory in Yorktown Heights, New York, December 17, 1958.

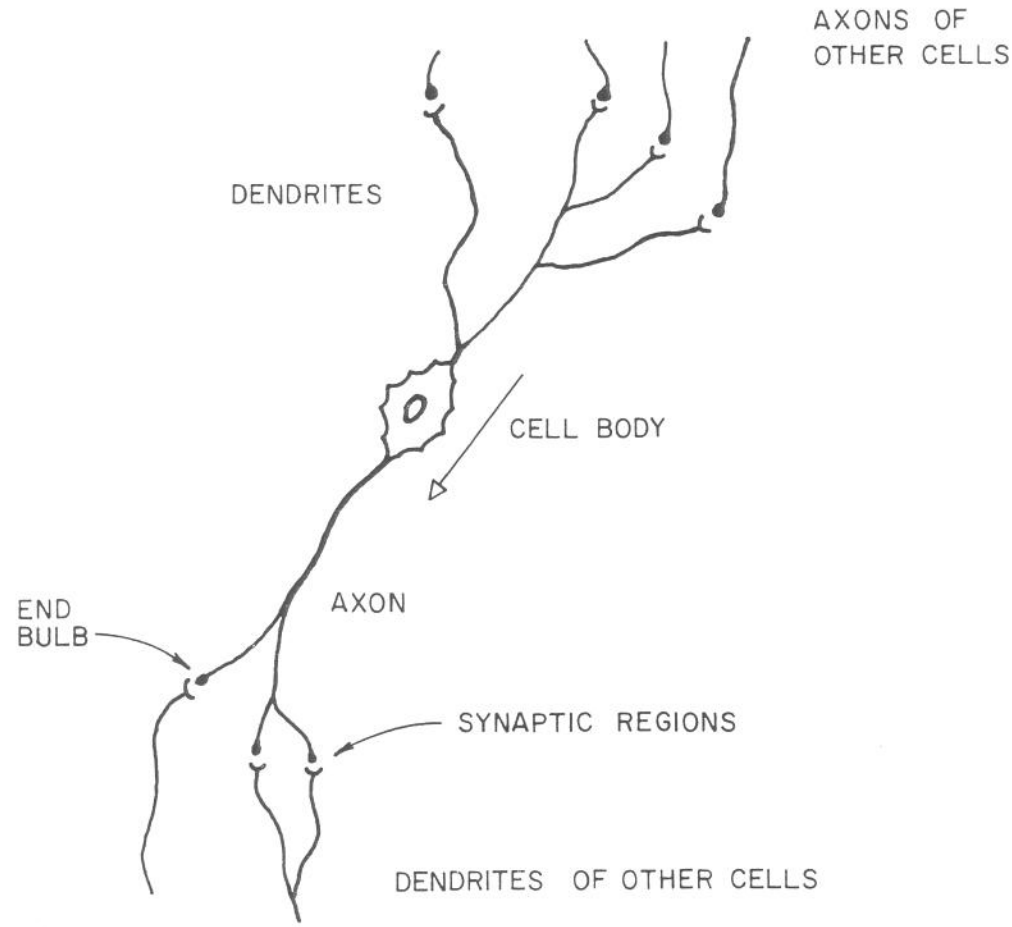
correct responses and/or punished for giving wrong ones. This question can be formulated for a machine as well as for a baby, and perhaps even more precisely.

First we may replace the retina by a mosaic of sensory points, in the form of a two-dimensional lattice. A precise definition of what is meant by squares and circles on this lattice can be formulated. Since this formulation is precise, the recognition could be carried out by means of a suitably programmed general purpose computer. Rosenblatt's intention is to get away from this approach and design a machine that would discriminate between various classes of stimuli without precise programming, and even without previous understanding on our part of the nature of the various classes. In other words he wants a machine that will form new concepts. For this reason perhaps the name "Conceptor," previously suggested by Rochester, Holland, Haibt, and Duda,[1] may be more appropriate than "Perceptron."

If the construction of a machine is to be inspired by the structure of the brain, we need to know something about the brain, both about its organization and about the units of which it is composed. The unit is the nerve cell or neuron, consisting of a cell body with an axon and dendrites branching out from the cell body. Nerve impulses can travel in both directions along both axons and dendrites. Axons are capable of giving output to other neurons. The junctions between an axon of one neuron and dendrites of another one are called synapses or synaptic regions. An axon branches out and ends at "end-bulbs." Culbertson,[2] who talks about hypothetical neurons, calls the end of a dendrite a "synapse" and admits that this usage is incorrect for real neurons. McCulloch and Pitts[3] also found it convenient to talk about hypothetical neurons. Norbert Wiener told me that this work was based on his previous unpublished work.

A neuron and its contacts is illustrated in the diagrammatic sketch on the following page.

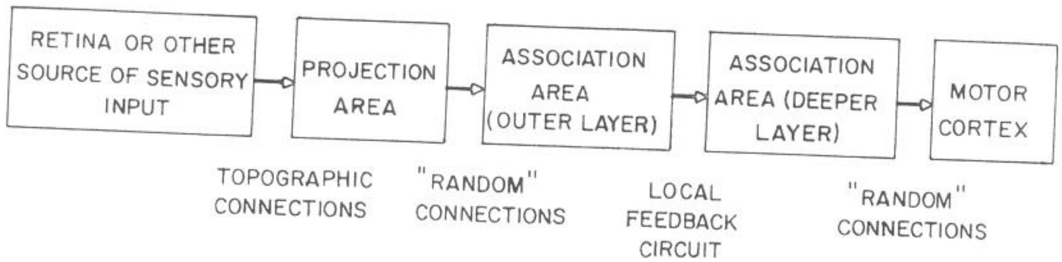
Some of the synapses are stimulatory, and some are inhibitory. If, during a suitable period of time, the number of stimulations of a neuron exceeds the number of inhibitions by an amount exceeding some threshold, the neuron fires. (This is a slight oversimplification for real neurons.) Firing is an all-or-none process and is followed by a refractory period of the order of a millisecond during which it cannot fire.[4] In the McCulloch-Pitts or



Wiener model, time is made discrete.

For an up-to-date account of real neurons, see Eccles, "Physiology of the Nerve Cell."[5] A good source of numerical information about neurons is Sholl, "The Organization of the Cerebral Cortex."[6] An estimate for the number of neurons in the cerebral cortex was made as long ago as 1899, and has not been modified much since. The present estimate is about $1/2 \times 10^{10}$. For the whole brain the estimate is 10^{10}, and for the whole nervous system about 10^{11}. In the cerebral cortex the number of dendrites per neuron is of the order 20 to 80. In the peripheral nervous system there is usually only one dendrite per neuron.

The organization of the nervous system is made especially complicated by the variety of possible stimuli and responses, but in very general terms there is the following oversimplified structure.



The projection area more or less preserves the topology of the sensory input.

Another aspect of the organization is that it seems to be a mixture of fairly specific areas and generalized activity. Thus there is a visual area, but an animal with the visual area removed need not be blind. In the human the removal of a part of the cerebral cortex may have little noticeable effect, though it tends to decrease the general intelligence. This fact supports the idea that the action of the brain is partly statistical.

An interesting analogy is with the method of superimposed coding, of which ZATOCODING is an example. This is a method of coding of information for information-retrieval purposes. Suppose we wish to identify a document by means of r index terms. Each term is represented by means of n punched holes in a card containing N locations each of which can be punched or not punched. If the total number of distinct index terms is R there would be no difficulty as long as N exceeded R, since we could allocate one location to each term. In many practical problems, however, R is much larger than N. But if

$$\binom{R}{r} < 2^N, \quad \left(R < \frac{r\,2^{N/r}}{e} \text{ approximately} \right)$$

then an unambiguous coding is certainly possible. The question is how to design it. One method is to list all the $\binom{R}{r}$ possible selections of r index terms and assign distinct binary numbers to each selection. If $\binom{R}{r}$ is large, this procedure is laborious, and if it is done by means of a formula then it may be arithmetically tedious to do.

Instead, for each of the R terms we may select n locations out of the N at random. The representation of the joint occurrence of r index terms is then simply the Boolean sum of the r individual punchings of n locations each. This is known as superimposed coding or ZATOCODING. In the application to information retrieval if we extract all the cards punched in the n locations corresponding to any given term, we may get some cards that are irrelevant, by chance. If N is large, and n is suitably selected, mistakes need seldom occur. In fact it is natural to arrange that:

$$\left(1-\frac{n}{N}\right)^{r} \;\underline{\Omega}\; \frac{1}{2}$$

$$\text{i.e.,}\quad n \;\underline{\Omega}\; \left(1-2^{-1/r}\right) N$$

This must be the best value of n since to have half the holes punched gives the largest variety of possible punchings.

By analogy, Nature's most economical usage of the brain would be for a reasonable proportion of it to be "in operation" at any one time, rather than having "one concept, one neuron." In other words there must be a very great deal of "parallel working." This is one advantage that the brain probably has over existing general-purpose electronic computers.

If you take two identical twins (i.e. two twins from the same ovum, and therefore having identical genetic composition) even their fingerprints are not identical. The differences may be attributed to chance. You might say that a man is a product of his heredity, his environment, and chance development after the genes have been determined (or you may classify chance development as part of the environment). The structure of the cerebral cortex is a lot more complicated than the patterns of the fingerprints, hence it is reasonable to suppose that the local structure at birth is very largely random. The cortex does have a great many non-random features; certain areas are associated with specific sensory and motor activities. (See E. G. Fulton, p. 399 for a mapping by A. T. Rasmussen and W. G. Penfield.) Furthermore there are at least seven different types of neurons (see Sholl,[6] p. 45), and the cortex has, to some extent, an onion-like structure. The randomness is partial and local.

The idea of randomness in the local structure of the cerebral cortex at birth, and in the initial structure of the automata, has been around for about the last twelve years. See, for example, Ashby, "Design for a Brain,"[7] and McCulloch and Pitts, "The Statistical Organization of Nervous Activity."[8] Ashby does not much care what the initial structure is provided that there are a large number of intercommunicating units.

An argument in favor of building a machine with initial randomness is that, if it is large enough, it will contain every network that will ever be required. The same aim can be achieved by means of a network consisting of a large number of elements, each connected to all the others, but for a given cost in interconnections a random network would contain more variety. For economical deterministic designs there is a danger that the determinateness will exclude some possible networks. Planning removes freedom. Pseudo-random networks would have the advantage of giving assurance that all automata produced had identical initial states, just as the use of pseudo-random numbers facilitates checking in Monte Carlo methods.[9]

A further argument in favor of having randomness is that it may solve some problems of design by simply evading them, just as in ZATOCODING.

Two of the writers who have emphasized the idea that concepts and behavior correspond to the activity of large assemblies of neurons are D. E. Hebb, "The Organization of Behavior,"[10] and Lashley, "In Search of the Engram."[11] The extreme alternative theory is that individual neurons are of the essence, as in electronic computers, so that the cutting of a single neuron in the cerebral cortex would often have a large effect on behavior. This extreme alternative theory seems most unlikely. Intermediate theories are concerned with obtaining reliability with unreliable components.[12, 13, 14] The notion of cell assemblies is highly analogous to superimposed coding in that it assumes that each neuron can be used in a great many distinct assemblies.

A student of Hebb, namely P. M. Milner, decided that allowance must be made for inhibitory synapses, as well as excitatory ones, and that within a cell assembly most synapses are excitatory, while between assemblies most are inhibitory. (See Rochester, Holland, Haibt, and Duda.[1])

Rosenblatt decided to try to work out mathematically the consequences of various assumptions in simplified structures called "perceptrons." He stated at the recent N. P. L. Symposium that his new mathematical work[15] supersedes his earlier mathematical work,[16, 17] and he stated further that the new work is also incorrect. (Information obtained from J. W. Backus and W. G. Bouricius.) In order to judge the merit of the designs we are therefore concerned largely with the experimental results, but unfortunately these have not yet been properly reported as far as I know.

In order to understand the sort of thing that Rosenblatt has in mind, the following quotation is illuminating:

> "The perceptron does not recognize forms by matching them against a stored inventory of similar images, or performing a mathematical analysis of characteristics. The recognition is direct, and essentially instantaneous, since the "Memory" is in the form of new pathways through the system, rather than a coded representation of the original stimuli." (Rosenblatt,[17] p. 16)

In spite of this remark, cross-classification is a part of the general planning. (See Rosenblatt,[17] pp. 14-16 and Chapter IX.) Presumably cross-classification can also be performed "instantaneously" by means of "pathways through the system." It seems to me that conscious cross-classification is another matter. Logic is a conscious process, whereas immediate recognition is performed without any conscious reasoning. You can recognize a man's face without being able to draw it. The main difficulty in acquiring a new skill seems to be the transfer of conscious understanding partially to the unconscious. On the other hand the difficulty of designing automata is in expressing in conscious terms what is unconscious.

It seems reasonable to conjecture that the hierarchy of larger and larger abstractions, and of more and more detailed cross-classification, corresponds to lower and lower levels of the onion-like structure of the cortex.

The 64-million dollar question is to what extent one should imitate living processes in the design of automata. How much importance should be attached to each of the following properties of the cerebral cortex:

a. The vast number of neurons.

b. The number of connections between neurons. (I think it may be significant that 32, which is the logarithm to base 2 of the total number of neurons in the cortex, is also the right order of magnitude for the number of dendrites. Perhaps, as the brain grows by successive subdivision of a single cell, about one new dendrite is formed at each splitting.)

c. The all-or-none property of neuron firing.

d. The existence of inhibitory connections as well as stimulatory ones.

e. The apparently very large amount of parallel working.

f. The probable re-use of neurons in a great many different sub-assemblies.

g. The fact that we can feel pain and pleasure.

h. The fact that we usually learn by successive abstraction, building up from simple beginnings.

i. The (perhaps related) fact that the cortex has an onion-like structure.

j. The apparently large amount of local randomness in the connection between neurons in the cerebral cortex.

k. The fact that the neurons of the cortex have been classified into at least seven distinct types.

l. The fact that the cortex is the result (though not necessarily the final result) of an evolutionary process that has been going on for many millions of years.

m. The belief that there are "local reverberatory circuits."

n. Drugs have a wide variety of effects.

o. People can be mentally ill in a variety of ways and can be psychoanalysed.

I should like to comment first on headings (j), (h), and (i). J. Z. Young[18] has emphasized how the young have more randomness in their brains, and therefore have more flexibility. As they grow older this randomness is partly used up, in fact learning uses up randomness and leads to some loss of flexibility. If we combine this idea with (h) and (i) it suggests that the using up of randomness, and the learning, occurs first in the outer layers of the cortex and gradually works inwards as successively higher abstractions are formulated.

Perhaps this is a convenient point to give a definition of learning. Given an organization with an aim, and a procedure for trying to achieve this aim, the organization is said to learn if it modifies its procedure in the light of experience in such a manner as to tend to improve the procedure.

In this definition the method for modifying the procedure may also be called a procedure, but it is better to call it a procedure of type II. Then learning of type II will involve the modification of the type II procedure in the light of experience, with some tendency to improve it, and so on.

It is reasonable to conjecture that the successive layers of the cortex correspond to some extent to higher and higher types of procedures and of learning.

It is necessary here to distinguish between two kinds of experience, namely external and internal. By internal experience I mainly have in mind conscious thought, including reasoning, though there may also be some sort of unconscious thinking. Understanding, or the formation of concepts that are often intended to represent nature, is dependent on both thought and experiment, but modifications in behavior may depend on other things apart from understanding, such as pain and pleasure. One of the parameters in the design of an automaton may be the importance that is attributed to "understanding" by the automaton, as distinct from the nerve reinforcements of procedures that give rise to correct responses. The same problem arises when teaching humans. There is the extreme school of swimming instructors who throw novices into the water, and there may be another extreme school who first give lessons in hydrodynamics. It is better to steer a middle course, but where it should be will depend on the student.

The distinction between understanding and merely behaving right, for a human being, is I think mainly a question of the ability to verbalize. One of the best breast-stroke swimmers at my old school was quite incapable of verbalizing his swimming activity, even though he was an exceptionally intelligent person. For automata we can define understanding as adequate symbolizing, or coding, or classification, of a class of stimuli; what is sometimes called "representational activity", though I prefer the word "symbolic" to "representational".

If an automaton had some method of detecting when it had found an improved classification of its experiences it could "reward" itself, i.e. reinforce the successful procedures. In this case it could sometimes improve itself without giving any external responses to stimuli. Similarly a child has pleasure merely in understanding things, and not just in satisfying its more obvious instinctive needs. (I can remember that this was so for me when I was three years old. Also the word "meaning" had great fascination for me.)

An important question therefore is how a machine can recognize that it has improved its methods of classification. A program for this type of recognition can be written when the class of stimuli are points in n dimensions and fall into two or more clusters. The methods of separation into clusters involve matrix manipulation and are given, for example, by Rao.[19] An analytic criterion can be provided for measuring how good the separation is. For more general classes of stimuli it is not easy to provide such a criterion. One of the things that is required is a measure of simplicity of a description. A report of a symposium on measures of simplicity is shortly to be published.[20] The simplicity and cogency of a hypothesis should determine its initial probability, and then it may be possible to make estimates of its final probability by means of an application of Bayes's theorem. (Also non-Bayesian statistical methods may be applied.)

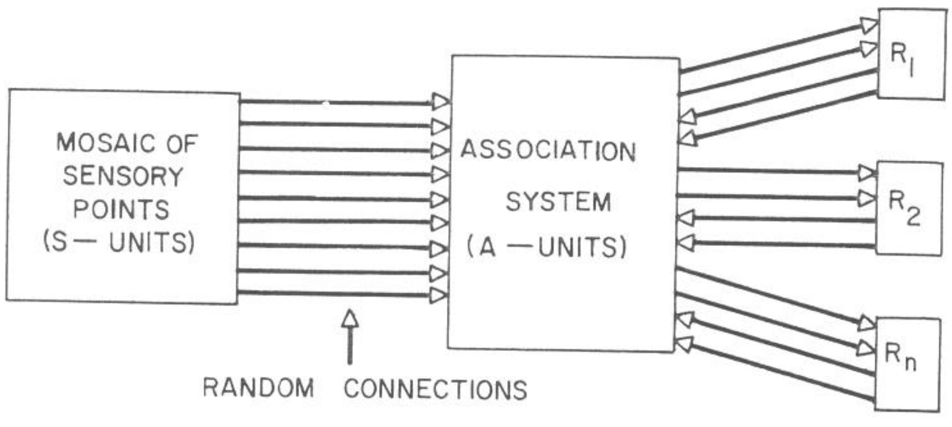
In default of an adequate explicit measure of initial probability in terms of simplicity and cogency, we have to fall back on a feeling of aesthetic pleasure in the simplicity and cogency. Therefore an automaton, too, must have a tendency to reinforce circuits corresponding to simple hypotheses. This remark may be relevant for all the work now being carried out on artificial intelligence. It is important to have some measure of simplicity, even if it is nowhere near optimal.

It has been estimated that a unit concept survives in the mind for about a quarter of a second (and so probably corresponds to about 6 to 10 bits of new information at most). If the time of operation of a neuron is of the order of a millisecond, it is reasonable to suppose that there are closed feedback circuits enabling the nervous impulse to circulate say 100 or 200 times. These are the "local reverberatory circuits" mentioned, for example, by J. Z. Young[21] who suggested that they are required in order that some permanent change may be effected (such as a lowering threshold for the neurons involved in the circuit, or an increase in the size of the end-bulbs).

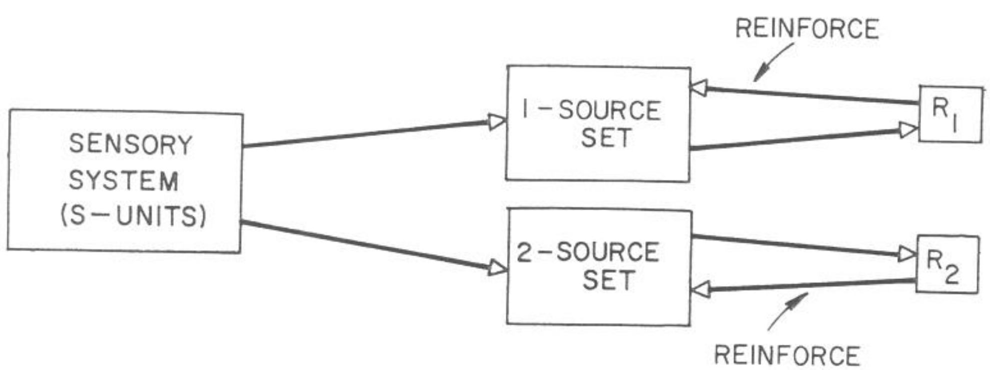
Perhaps the physiological event corresponding to a unit concept is that there is not much change in the cell assemblies that are reverberating. In this way unit concepts would get to some extent consolidated and would more readily recur in the future.

Any newly formed hypothesis that corresponds to a unit concept, i.e., that is simple enough to be all present simultaneously in a set of reverberating circuits, would tend to become consolidated. But new thought that was more complicated would be less well consolidated. This may explain why we naturally accept the simpler hypotheses, other things being equal. We think of them more readily in the first place, and they more easily become consolidated.

I should like to say more about some of the aspects (a) to (n) of real nervous systems, but first I had better say something about perceptrons. This name has been used by Rosenblatt in a number of different senses. They are all based on the following diagram.



Here R_1, R_2 . . . , R_n are n "response-units", each of which is capable of two states, 0 and 1. The sensory points are also each 0 or 1 (off or on). The simplest perceptron is



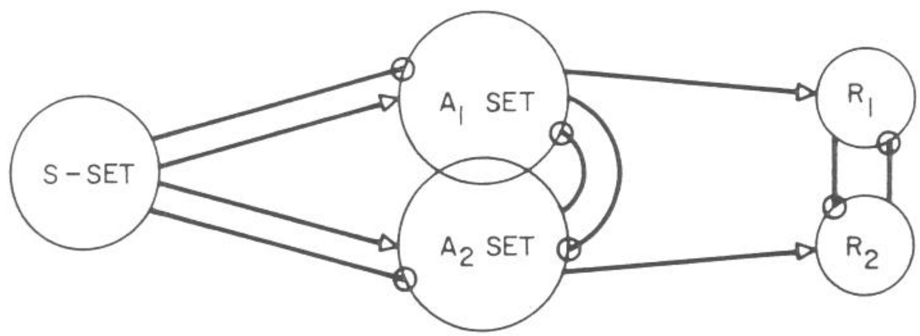
Here each A-unit is connected either to R_1 or R_2. The experimenter can choose which of R_1 or R_2 is to send back reinforcement so as to try to train the machine to give what the experimenter regards as a satisfactory response. The whole of the large January 1958 report[22] was concerned with this training problem, and not with the problem of a self-organizing perceptron. The connections from the S units may be excitatory (positive) or inhibitory (negative), and each A-unit may have connections from several S-units. If an A-unit is connected positively from x active S-units and negatively from y active S-units then its input is

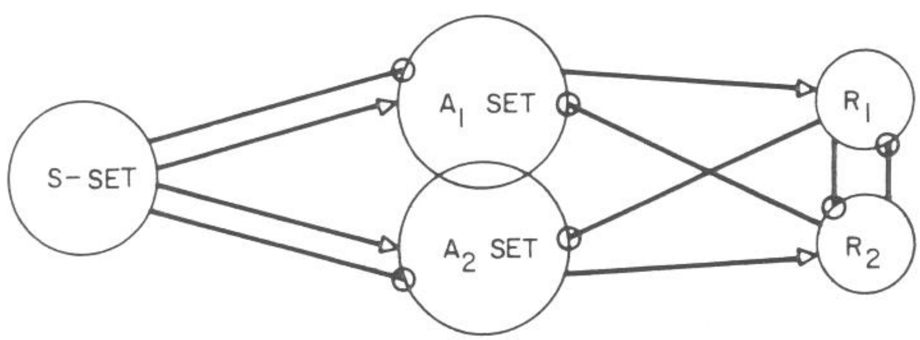
$$\alpha = x - y,$$

and its output is (x - y)v where v is the current so-called "value" of the A-unit. For real neurons the formula is more like

$$\begin{cases} 1 & \text{if } x - y \geq \theta \quad (\theta = \text{threshold}) \\ 0 & \text{if } x - y < \theta \end{cases}$$

"Reinforcement" means increasing the value of v. In the January 1958 paper,[22] time was taken as discrete and the following two simple perceptrons were considered.





In both these perceptrons, when a stimulus is applied there will be a tendency for either A_1 or A_2 to gain dominance over the other, because of the mutual inhibitions. The sets A_1 and A_2 are, so to speak, in competition with each other for each stimulus. This mutual inhibition idea is Milner's modification of Hebb's hypothesis of cell assemblies. It seems very reasonable to me.

At the recent N. P. L. symposium on the mechanization of thought processes Rosenblatt introduced a third perceptron, this time with continuous time because he hoped that this would simplify the mathematical analysis. Now the input to an A-unit is

$$\alpha = \Sigma x - \Sigma y$$

where the stimuli $x_1, x_2, \ldots, y_1, y_2, \ldots$ are real numbers, and the output of the A unit is again αv, where v is its value. The response units send back reinforcement signals, with sum $\Sigma\rho$, to the A-unit, and three types of reinforcements are considered:

$$\frac{dv}{dt} = \alpha\Sigma\rho, \qquad \frac{dv}{dt} = \alpha\Sigma\rho - \delta v, \qquad \text{and } \frac{dv}{dt} = \Sigma\rho\ (\alpha - \delta v).$$

Note, that for the third type of reinforcement, when there is no reinforcement the value of the A-unit decays exponentially. This device had been previously used in a little machine built by A. M. Uttley, consisting of nine photocells inter-connected by means of leaky condensers. He succeeded in training it to recognize T-shaped shadows.[23]

The sorts of theorems that Rosenblatt is interested in are theorems of "statistical separability," namely that certain classes of perceptron tend to states in which they tend usefully to separate out classes of stimuli. Since he seems to have withdrawn all his theorems there is no point in quoting any of them precisely. A similar theorem was given by McCulloch and Pitts.

It seems to me, however, that this is a reasonable thing to try to do. If everything is left to simulation experiments, the amount of experimentation required could be very large, and any correct theoretical results would, as usual, cut the simulation work down considerably.

On the experimental side it is not easy to find out what has been done from the publications. A perceptron seems to be capable of distinguishing between shapes in the left hand and right hand halves of the sensory mosaic, and there was a rumour that it could distinguish squares from circles. The reason why this may be possible, with a trained perceptron, is that if it is shown a large number of squares and circles, each square shown to it later may resemble in the area it covers at least one of those shown to it before. If this resemblance is great enough, and if there are a great many A-units, then the perceptron may work, by analogy with a Zatocoding (superimposed coding) system in which there is a very great redundancy (because there are so many A-units).

It seems to me that the main contributions made so far by Rosenblatt are that

i) He selected a good name, the "perceptron," and obtained a lot of publicity for the idea, thus stimulating interest, including a grant of $100,000 for future work;

ii) He has done a lot of thinking about artifical neurons with inhibition (but in his most recent perceptron, the one with continuous time, the function of inhibition seems to have lost most of its relevance);

iii) He has tried to work out mathematical theorems about aggregates of these artifical neurons, and has thus suggested the desirability of work being done in this field by mathematicians. (He is himself a psychologist rather than a mathematician.)

If I support this last type of work I am also in favor of the other type of work on artifical intelligence, namely the attempts to mechanise logic, geometry, chess, and the reading of handwriting. There is also the further type of work (mentioned to me by John Griffith) of simulating the nervous system of some animals having very few neurons. I am rather shaken by the idea of a large steel lobster wandering about Westchester County. If an artifical lobster could be built with only 20 neurons, perhaps many other less important neurons could be added in order to make an improved lobster, i.e., one capable of learning.

I should like to refer back to the conjecture made under heading (b). Suppose that the entire cortex were developed from a single cell by 32 binary fissions. The original cell could be represented by a decimal (or binary) point. It splits into two cells denoted by .0 and .1, ultimately to become the left and right hemispheres of the cortex. This goes on until we have 2^{32} cells, each represented by a 32-bit binary fraction. This notation suggests a number of possible conjectures concerning the connections in the cortex. In the first place two fractions that differ by a small amount are close together in the fully developed cortex and are also closely related on the "genealogical tree." We may conjecture that at each binary fission each daughter cell has about one more axon and one more dendrite than the parent cell. We may have, say,

.011101
.011100

mutually inhibiting. A plausible model leads to the further conjecture that $a_1a_2 \ldots a_ra_{r+1} \ldots a_{32}$ has probability 2^{r-32} of having an inhibitory connection to $.a_1a_2 \ldots a_rb_{r+1} \ldots b_{32}$, where $a_{r+1} \neq b_{r+1}$. A model along these lines may perhaps have the property that a cell assembly will tend to inhibit the rest of the cortex, especially the surrounding cells, and may explain why one hemisphere usually gains dominance.

This conjecture would need modification in order to say something about how the cortex is connected to the peripheral nervous system, and why there are several types of cells.

If, say, six splittings contribute to the thickness of the cortex, it would be 64 cells thick. This seems to be about right for the frontal polar region, whereas 128 looks a better estimate for the extra-pyramidal region, judging by the diagrams on page 384 of Krieg.[24] We may also estimate the average thickness, measured in neurons, by the following argument. The volume of the cerebral cortex is about 570 c.c. $\pm$ 40 c.c. (Sholl,[6] p. 32), so the average amount of space used up per neuron is $(570 \div 1/2 \cdot 10^{10})$ c.c. $= (1/20 \text{ mm})^3$ The average thickness of the cortex is about 2.7 mm (Sholl,[6] p. 34). So if we think of the space around each neuron as a little cube (which admittedly is far from true) we get an average thickness of 54.

When considering how much should be spent on developing artifical brains, it is necessary to try to guess the probability of various degrees of success. My own guess for the time it will take to develop a really useful artifical brain is 20 years multiplied or divided by 1 1/2, if it is done during the next 100 years (which I think is odds on). An argument against is the large number of neurons in the cortex of the human brain (and the chimpanzee has almost as many). On the other hand electronic units may work more reliably and perhaps a million times as fast in ten years time, and very small units will exist and will be inexpensive. Moreover natural evolution is extremely wasteful and blind, whereas artifical selection can be controlled intelligently. It is true that evolution has gone on for a long time and over a wide area, and it could be argued that to make life by chemical means may be easier than to construct an artifical brain. On the other hand the simulation of intelligence on general-purpose computers has already met with some limited success. We can analyse conscious thought by introspection, a method that does not apply to biochemistry, and also does not apply very well to unconscious thought. I understand that some schizophrenics are good at analyzing unconscious thoughts; perhaps, as occupational therapy, they should be given work on the construction of artifical brains.

My own guess is that, ultimately, efficient machines having artifical intelligence will consist of a symbiosis of a general-purpose computer together with locally random or partially random networks. The parts of thinking that we have analyzed completely could be done on the computer. The division would correspond roughly to the division between the conscious and unconscious minds.

Once a machine is designed that is good enough, say at a cost of $100,000,000, it can be put to work designing an even better machine. At this point an "explosion" will clearly occur; all the problems of science and technology will be handed over to machines and it will no longer be necessary for people to work. Whether this will lead to a Utopia or to the extermination of the human race will depend on how the problem is handled by the machines. The important thing will be to give them the aim of serving human beings.

It seems probable that no mechanical brain will be really useful until it is somewhere near to the critical size. If so, there will be only a very short transition period between having no very good machine and having a great many exceedingly good ones. Therefore the work on simulation of artifical intelligence on general-purpose computers is especially important, because it will lengthen the transition period, and give human beings a chance to adapt to the future situation.

REFERENCES

1. N. Rochester, J. M. Holland, L. H. Haibt, and W. L. Duda, "Tests on a Cell Assembly Theory of the Action of the Brain, Using a Large Digital Computer," I.R.E. Trans. on Information Theory, Vol. IT-2, No. 3, pp. 80-93, Sept. 1956.

2. J. T. Culbertson, Consciousness and Behaviour, Wm. C. Brown, Dubuque, Iowa, 1950.

3. W. S. McCulloch, and W. Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," Bulletin of Mathematical Biophysics, 5, pp. 115-133, 1943.

4. J. F. Fulton, Physiology of the Nervous System, New York, Oxford, 1938, 1949.

5. J. C. Eccles, Physiology of Nerve Cells, Baltimore, John Hopkins Press, 1957.

6. D. A. Scholl, The Organization of the Cerebral Cortex, Methuen, Wiley, 1956.

7. W. R. Ashby, Design for a Brain, Wiley, 1954.

8. W. S. McCulloch, and W. Pitts, "The Statistical Organization of Nervous Activity," Biometrics 4, pp. 91-99, 1948.

9. Cf my review of "Symposium on Monte Carlo Methods" in Mathematical Tables and Other Aids to Computation, pp. 44-46, January, 1957.

10. D. O. Hebb, The Organization of Behavior, Wiley, Chapman and Hall, 1949.

11. K. S. Lashley, "In Search of the Engram," Symp. Soc. Experimental Biology, 4, pp. 454-482, 1950. "Lashley argues convincingly that millions of neurons are involved in any memory recall, that any memory trace or engram has multiple representation; that each neuron or even each synaptic joint is built into many engrams."(Quoted from Ref. 5, p. 266.)

12. J. von Neumann in Automata Studies

13. E. W. Moore and C. E. Shannon, J. Franklin Institute 262, p. 151, 1956.

14. W. S. McCulloch, Recent NPL Symposium on the Mechanism of Thought Processes.

15. Frank Rosenblatt, "Two Theorems of Statistical Separability in the Perceptron," NPL Symposium on the Mechanisation of Thought Processes, paper 1-3, November 26,27, 1958.

16. Frank Rosenblatt, "Theory of Statistical Separability," Cornell Aeronautical Laboratory, January 1958.

17. Frank Rosenblatt, "The Design of an Intelligent Automaton. The Perceptron Program," About October, 1958.

18. J. Z. Young, Doubt and Certainty in Science, Oxford, 1951.

19. C. R. Rao, Advanced Statistical Methods, Wiley, 1952.

20. Report of a Symposium on Measures of Simplicity. Philosophy and Phenomenological Research, forthcoming.

21. J. Z. Young, "Growth and Plasticity in the Nervous System," Ferrier Lecture, Proc. Roy. Soc. B, 139, pp. 18-37, 1951.

22. I. J. Good, "Can a Machine Make Probability Judgments?" Computers and Automation, forthcoming, (A lecture given in Jan. 1958 to the Philosophy of Science group of the British Society for the History of Science).

23. A. M. Uttley, "Contribution to a Study Programme in Biophysical Science," to be published in Review of Modern Physics, Jan. 1959 (Reference obtained from Ed. Quade).

24. W. J. S. Krieg, Functional Neuroanatomy, Blakiston Co., New York and Toronto, 2nd. Edit., 1953.

25. J. C. Eccles, The Neurophysiological Basis of Mind, Oxford, 1953.