types of modifications that might prove useful for a given type of expression.  
The problem of sample-size requires mention: How large a sample of experience is necessary to obtain learning? Or better: How much information about the effects of behavior is necessary to successfully modify the behavior? Chess affords a good example of this problem. It is extremely doubtful whether there is enough information in "win, lose, or draw" when referred to the whole play of the game to permit any learning at all over available time scales. There is too much behavior. For learning to take place, each play of the game must yield much more information. This is exactly what is achieved by breaking the problem into components. The unit of success is the goal. If a goal is achieved, its subgoals are reinforced; if not, they are inhibited. (Actually, what is reinforced is the transformation rule that provided the subgoal.) This is so whether the game is ultimately won or lost. Each play gives learning information about each goal that is generated. This also is true of the other kinds of structure: every tactic that is created provides information about the success or failure of tactic search rules; every opponent's action provides information about success or failure of likelihood inferences; and so on. The amount of information relevant to learning increases directly with the number of mechanisms in the chess playing machine.

**CONCLUSION**

"As every design engineer knows, the only difference between a good design and the actual machine is time and effort." This adage is a byword in the field of robots and thinking machines. The scheme presented here is not far from a "good design." One can estimate the man-hours necessary to draw up the detailed flow diagrams from which machine coding follows as routine chore. But this is not sufficient. These mechanisms are so complicated that it is impossible to predict whether they will work. The justification for the present article is the intent to see if in fact an organized collection of rules of thumb can "pull itself up by its bootstraps" and learn to play good chess.

**Comments on Session on Learning Machines**

Walter Pitts (Research Lab. for Electronics, M.I.T.) The speakers this morning are all imitators in the sense that the poet in Aristotle "imitates" life. But, whereas Messrs. Farley, Clark, Selfridge, and Dineen are imitating the nervous system, Mr. Newell prefers to imitate the hierarchy of final causes traditionally called the mind. It will come to the same thing in the end, no doubt, but for the moment I can only leave him to Mr. Miller, and confine my detailed remarks to the others. It is of great value to the neurophysiologist to have people invent machines that perform some of the same tasks as parts of the nervous system, especially when the machines operate along the same general lines. This would be true at any time, but is especially so now, which I can make evident by a thumbnail sketch of the present state of the science.

Anatomy has divided the brain into internally homogeneous parts, called nuclei, has named and classified them, and has traced the more conspicuous pathways, or fiber-tracts, connecting them. After adding some substantial contributions from electrophysiology, the result is a block-diagram of the nervous system. Although incomplete in detail in most places, particularly with respect to connections employing few, small, or slow fibers, it is still a generally reliable guide to research. Thus, for part of the visual system, anatomy draws a diagram running somewhat like Fig. 1.

The next chief problem is to find out what transformation the arriving information suffers within the nuclei. There are several ways of doing this. The oldest was to remove a particular nucleus in an experimental animal, and watch the animal after recovery to see what capacities it had lost. The limitations of the method are very great. At most, one can infer that a nucleus is necessary for a capacity, never that it is sufficient. The more complicated disabilities often disappear gradually, by a process, such that one cannot tell whether it was the removal of the nucleus or the incidental disturbance of something else during operation that caused the disability called the "vicarious assumption of function." And the method cannot be applied to men, in whom alone the more complicated forms of thinking can easily be detected. (Accidental human brain damage is rarely well localized, and one never has tested the man before the accident.) This line of attack is therefore becoming obsolete, except among experimental psychiatrists wishing to discover the seat of the soul, the unconscious, the emotions, or other metaphysical entities. A more useful method is to apply various
stimuli to tracts of fibers leading into a nucleus and record from those leading out. This procedure is less generally successful than one might expect. Most fiber tracts are not compact bundles, easily reached for local recording; moreover, even if they are, one cannot tell what “information” he has sent in, or what has come out, unless he knows the coding system. A tract will consist of thousands or millions of fibers, carrying a sequence of impulses in time which is arbitrary except in not having too many impulses too close to one another. In principle, we ought to record from the impulses from each fiber separately, since it is possible for the relevant information to lie (1) in the intervals of time between successive impulses in a train proceeding along each single fiber; (2) in which fibers are carrying the impulses; (3) in the intervals in time between impulses, or trains of impulses, in different fibers. To do this is technically impossible, nor could one interpret the results if he had them. In favorable cases, we can find the rate of things: record for all fibers of a tract in parallel (one then cannot tell in which fibers a discharge originates), and record from one fiber in a tract. Both of these give useful results, as we shall see. But unfortunately, in very large cases, the rate of impulses is so large, and the field so large, that they occur in nuclei close to the primary sensory input to the system and the ultimate motor outflow from it. Even for these nuclei, in many cases the signal along a particular fiber depends upon a few of the input fibers, and it is quite impossible to find these and isolate them for stimulation. This brings us to the remaining line of investigation, the direct study of the nuclei themselves. Here we have two kinds of knowledge to start with. The historical knowledge of Cajal and his successor has told us in some detail the structure and mutual arrangements of the neurons in a nucleus and the principal lines of connection among them. The kind of information provided may be seen in Fig. 2, from Ramon y Cajal. The second method is a semi-microscopic one. It endeavors to follow the transmission of a group of impulses from place to place within a nucleus, or along a group of fibers connecting one part to another, without recording from individual units. This procedure also requires microelectrodes, but is not so fine as the other. The object is to secure a record of changes in the general potential field at each of a large number of points, without having the records unduly perturbed by a cell or fiber immediately adjacent to the electrode tip. This potential field cannot be interpreted directly in terms of nervous activity (although many of the older electro-physiologists thought it could), because the neurons are immersed in a good conductor, and it usually happens that activity of any part of a nucleus generates currents and potentials observable throughout the entire nucleus. When several parts, or all of them, as is usually the case, are simultaneously active, the potential at all points depends upon the activity at all points, and the distribution of activity cannot be read off from the potentials by inspection. In these cases one must compute the divergence of the potential field, which does represent the distribution of activity. (An impulse in a neuron consists of a region on its bounding surface where current is being absorbed from the external medium into the cell, flanked by regions where it emits current to the outside: for the potential field measured in the external medium, therefore, an impulse acts like a sink of current flanked by sources. The divergence of the field is therefore a measure of the number of impulses at a given place and time.) 

It is obvious that this method likewise can provide a direct experimental check of computations made with models. It too has considerable limitations in that the simultaneous potential field required cannot be obtained: one must record from different points of the nucleus seriatur, repeating the stimuli along the afferent channels as exactly as possible, and taking all possible
precautions to detect any change in the condition of the nucleus during measurement.

These new experimental methods put a new complexion on the problem of understanding the nervous system. Formerly we began by throwing away, as soon as we could, the gross behavior with the actual behavior, and then making more assumptions to complete what was known about their interconnections. We noted that none of these assumptions were precise and detailed an experimental check than an ultimate comparison of the calculated behavior with the actual behavior, the piling-up of dubious hypotheses on one another, until it became clear that, even if the result should then agree, the hypotheses would gain so little verisimilitude that rational men hardly cared to venture so far from the truth. The results of this procedure were the successful model neurons, still randomly connected, but according to more interesting probability distributions. I have reasons for thinking that this possible generosity of theirs might not be ill-advised for their own secret purpose. (I suppose everyone else has guessed it as I have; they mean to invent a cheaper and more efficient substitute for Generals of the Air Force and me; one must not forget that these latter are organisms with nervous systems.)

More seriously, though, it will be exciting to be the neurophysiologist if they repeat very many times the experiment of letting their random net acquire a definite structure by learning and can then discover what is common in the "effectively structural" changes in the network so produced; we might imagine something new to look for in the erudite brain to distinguish it from the unordered.

Messrs. Selfridge and Dinneen are less humble imitators of the animal nervous system, particularly in their learning mechanism; but in their elementary operations they are closely indeed. They have two kinds of experimental results which seem interesting to mention here. The first concerns the response of single fibers in the optic nerve, as recorded by microelectrodes, to a flash of light suddenly appearing to the whole retina. It turns out that the fibers fall into three principal groups: the "on-off" fibers, which respond by a short train of impulses when the light is turned on, and another when it is turned off; the "on-and-off" the discharge of which continues throughout the presence of the light; and the "off" fibers, which respond only after the light is turned off, and then continue for some while in cats, and probably in primates, the vast majority of fibers are of the first, or "on-off" variety, and therefore respond principally to changes in the degree of illumination, rather than to illumination itself. The next result is the well-known fact that we see boundaries between light and dark regions far more prominently than what is inside them; indeed, we can even recognize a line-drawing of a face, landscape, a feat nobody would ever believe possible on the basis of simple considerations of information-theory. The suggestion arises from the so-called "physiological nystagmus" —the nervous fine tremor of the eyes that goes on all the time. This causes the boundaries of regions of different brightness to pass back and forth continually over the receptors situated near it, so that the effect of continuous changes of illumination, and will respond strongly. The receptors situated in the middle of regions of nearly uniform brightness will experience little change, and therefore show only a small response except for the relatively few "on-and-off" fibers.

This provides in itself a good corollate of Selfridge and Dinneen's "edging" operation, and their "averaging" is represented as an extension of the operation in the opposite direction. The optic nerve responds to illumination of a fairly sized "patch" of receptors. Anatomical knowledge of the successive convergence of fibers in higher nuclei provides ample opportunity to discover the simplest assumptions, to provide an iteration of these operations, but to what extent this actually occurs we cannot yet say.

Here we might leave the matter, but the recent experiments of Kuffler provide so much more direct a correlate to these operations, in the retina itself, that I must say a little about them. Kuffler succeeded in recording from a single ganglion cell in the retina—these cells are the origin of the fibers of the optic nerve; the receptors (rods and cones) are connected to them indirectly and explored, with a very small spot of light, the effect of illuminating receptors at various distances from it on its response. He found that all the receptors in a certain "receptor-field" surrounding the ganglion cell could excite it; but the ones in the center of the field close to the cell affected it quite oppositely from the ones farther away. The "on-center" ganglion cells, as he called them, gave an "on"-response when he illuminated only the center of its "receptor-field" response when he illuminated only the periphery. The "off-center" ganglion cells behaved in exactly the reverse fashion. Illuminating at intermediate distances generally caused an "on-off" behavior. He discovered that the receptors in the center of the receptor-field and those at the periphery generally inhibited one another; the effect on the ganglion cell, that is, if both were illuminated, was that its characteristic response to each group was reduced or abolished.

When we recall that these "receptor-fields" for different ganglion cells form a set of overlapping patches covering the whole retina, and that the ganglion cells are of two opposite types, the general result becomes rather complicated, particularly when one includes the "physiological nystagmus." Nevertheless, it is apparent that there is a premium on inequality of illumination, that is, that such ganglion cells as respond most have the greatest inequality of illumination between the receptors immediately adjacent to them and those at a little distance, so that the net result must constitute a large degree of "edging" as well as "averaging."

It is worth reminding you, in conclusion, that these resemblances between work on the neurophysiology of vision and Messers. Selfridge and Dinneen's fundamental operations on the alphabet are not at all wonderful or surprising. They know these physiological facts as well as I do, and have deliberately designed their proposed machine to imitate them. Nor is this at all unreasonable, when one reflects that they mean to imitate a human activity, so that it is merely economical to begin by throwing away, as soon as possible, that vast bulk of "information" present in every visual image, which we begin by throwing away.

I had meant here to reflect on certain fundamental presuppositions determining the thinking of man, and apparently all other mammals as well, which a machine capable of inductive reasoning offers the first hope of really escaping, and therefore understanding. These are the epistemological limits of first discursive operations. As I once heard Robert Oppenheimer, who has a rare gift for the appropriate expression, say that good science is just that work that your colleagues are grateful for you to do. On that basis, at least, I am sure that we have been listening to good science here this morning—it must be good, because I feel deeply grateful to these men for doing it. Yet I want to do here today is to look at what these men are doing from the point of view of a psychologist. Now as a psychologist, concerned with understanding the behavior and the minds of living organisms, I have a lot of problems. In fact I often think I have nothing but problems. So when I see other people trying to solve some of these difficult problems for me, I am inclined to become almost pathetically grateful.

Broadly speaking, psychologists have two somewhat different classes of problems which, for simplicity, I will call descriptive problems and functional problems. A descriptive problem is one in which we try to
state the properties that characterize an animal’s behavior. Thus we speak of a man as being honest, as having self-esteem, as being intelligent, well-adjusted, introverted, dominant, etc. These same properties can be, and often are, attributed to machines. We say that our automobile is stubborn for the same reasons we say a mule or a man is stubborn—it won’t do what we want it to. Such properties are often attributed to computing machines.

I have heard a certain amount of discussion as to whether or not it is proper to ascribe such human properties to computing machines. To my mind, these arguments are completely irrelevant. No matter which side wins the argument, the practical business of designing new machines is not changed in any way. So I will not waste my time arguing over distinctions that have no consequences.

When we come to the functional problems, however, the situation is far more interesting. A functional problem is one in which we try to state how an organism performs a certain function. Thus we speak of a man detecting a faint sound, recognizing an object, learning a skill, remembering a name, solving a problem, performing a job, or communicating an idea. These are all psychological functions and we would like to know how it is that living organisms carry them out.

With regard to such functional problems there is a long history in psychology of interest in machines, in robots, that can perform the same functions. Descartes may have been the first to think in this way. The value of such analogies for the psychologist is that they often help to clarify his thinking. If he can express the problem in terms of a machine, then he feels that he has defined his terms scientifically and has eliminated any possible mysticism or magic from his discussion.

A discussion phrased in the ordinary language of everyday conversation is often very ambiguous. In order to be more precise we often have recourse to the language of mathematics where we must define our terms clearly and perform our deductions according to established patterns of logic. Psychologists, however, are often unable to make a precise mathematical statement, and so they have often resorted to this dodge, this halfway house on the road to precision, the concept of a model. A model is certainly superior to a purely verbal argument, and is inferior to a complete mathematical theory. Thus many psychologists have tried to design machines that would perform psychological functions.

This morning we have heard three attempts to design machines for carrying out psychological functions. Selfridge and Dinneen have a machine that recognizes objects. Farley and Clark have a machine that learns by being rewarded. Newell has a machine that solves problems in chess. In all these cases, the psychologist expects to learn whether or not such machines, designed according to certain stated principles, actually perform the functions intended. In that way we get an indirect verification of the adequacy of our psychological postulates.

For a concrete example, consider the question of recognizing objects. In trying to conceive how a machine might do this job, many psychologists have imagined that some memory trace of every object ever perceived in the past is stored somewhere in the nervous system. Then when a new object is presented, its image is correlated with all the stored images, and that stored image which gives the highest correlation with the perception of the present object determines what we call it. However, when we in fact get around to really building such a machine, the inadequacy of such a model quickly becomes apparent. The machine cannot remember so much and it cannot compute so many correlations in the time allowed. Therefore, we are at once forced to perform certain operations on the image that extract from it the significant features, and these features are then compared with a much smaller memory. Such a machine is practical to build, it is in better conformity with the known facts of human perception, and Selfridge and Dinneen have shown that it can work to give the same results.

Similar remarks are also true of the chess machine. It is not good machine design to explore all possible consequences of all possible responses to a problem. And it is not good psychological theory either. So Newell considers a machine that forms inductive memories and remembers rules and does not try to store away every item of information it has received.

To me, a psychologist, this is the principal lesson to be learned from the papers presented here today. Our problem, our joint problem, is to discover what transformations must be made on the available data in order to preserve intact the significant features and to discard the irrelevant details. When we have discovered such transformations or operations, we shall be a great deal closer to building better machines, and also a great deal closer to understanding human behavior.

A New Nondestructive Read for Magnetic Cores

R. THORENSEN† AND W. R. ARSENAULT†

Summary—A simple but effective method has been developed for reading information from ferrite storage cores without destroying the information. A single ferrite core containing a small radial hole is used per stored bit. A wire is threaded through the hole and a current pulse is passed down the wire to interrogate the core. The interrogation current temporarily disturbs the residual flux in the core and induces a signal in an output winding. The output signal is unique for each core, and it is based on the hysteresis loop and upon removal of the core will come to rest at the point 1. The input signal is applied.

The read operation can be extremely fast. Complete read signals may be as short as 0.3 microsecond, making interrogation cycles of 0.5 microsecond feasible. Laboratory tests have proven these cores to have all the properties associated with normal coincident current memories plus nondestructive readability.

Fabrication of these cores should prove practical and economical as compared to nondestructive read arrangements known to the authors. Normal square loop ferrite cores can be altered to the new design or it should be possible to fabricate the radial hole at the time of manufacture.

† Res. Labs., The Magnavox Co., Los Angeles, Calif.

Introduction

The application of ferrites to the electronic computer field for storing information is well known and has been well covered in the literature. For those not completely familiar with present methods of inserting and removing information from such storage systems, a brief review is included.

Fig. 1 shows a toroidal core and its hysteresis properties. Fig. 1(b) is a somewhat idealized curve of a square loop ferrite. The core is shown with three windings, two carrying switching current and the third, a sensing signal. A current of (-I) will bring the core to a point marked (-\(\phi_m\)) on the hysteresis loop and upon removal of (-I) the core will come to rest at the point 1. The core is then said to store a binary 1. In order to sense the information stored in the core, a current of (I) is applied. The flux in the core then traverses the loop to (0). The