The Contribution of Information Theory to Pathological Mechanisms in Psychiatry

W. Ross Ashby

BCL 164, fiche #9, 16-31

Information theory is sometimes presented as a new philosophy; here it will be presented as an essentially practical branch of science. Its essence occurs when an engineer says: "You can't get ten maneuvers out of that satellite when you've only five signals." He is thinking along the usual and well understood lines of cause and effect, but using a rather unusual approach: instead of trying to relate each cause to its particular effect (e.g., "what is *the* cause of tuberculosis?"), he is bringing a *set* of (five) causes into some relation with a *set* of (ten) effects. Throughout this article, information theory will be used in accordance with what I believe to be its true nature — that it is the body of knowledge developed to help when we have problems in which large numbers of causes are related in some way to large numbers of effects.

Problems that involve causes and effects in large numbers fall into two natural classes: 1. where the various causes are distributed over space (or equivalent dimensions), as, for instance, *patterns* of light spots distributed over the retina, and 2. when the various causes are joined in long chains in time, each effect becoming the cause of the next. This second case corresponds to the activities of the modern computer, which might be described simply as a device for performing accurately throughout an extremely long chain of causes and effects. For this reason, computer-theory and information-theory are hardly separable, and I shall refer to both freely. Reference to both is specially necessary in this article, for the brain shows the double complexity of accepting, through the senses, complex *patterns* of stimulation, and then carrying them through long *chains* of processes. Modern theories of information and computers attempt to say something useful about such broad and lengthy processes.

It should perhaps be noticed at this point that these theories, of information and computers, are entirely objective in their methods. Though words such as *information*, *memory*, *control*, and *recognize* have an introspective aspect (practically the only aspect considered in the psychologies of the previous century), they are used in these theories of today (and in this article) solely to refer to objectively demonstrable facts of behavior. Thus, the geneticists and molecular biologists today speak freely of the "information" on the DNA molecule: this "information" has no reference, of course, to any "knowing" by the DNA: it is simply a reference to the various causes, exertable by the DNA as a physical system, over the various effects that can be shown on, say, the proteins synthesized.

The theories of information and computers are thus essentially concerned with general principles that should hold, or that should guide one, when one has to deal with a system in which the processes at work are extremely complex and lengthy. For this reason it seems that psychiatry must inevitably be related in some degree to these theories. Is anyone more likely to say "this complexity is becoming unmanageable" than the psychiatrist? I need therefore offer no apology for attempting to trace some relation, to explore the borderline, between these two sciences.

When the complexity is lacking — when one relates, say, ten maneuvers to five signals — the theories of information and of computers may give little help, for the worker needs no help. They tend to become increasingly useful as the complexities grow. When does this happen? — the work of Shannon and Weaver (1949) has shown that the main factor leading to the complexities referred to here is *combination*: where many parts have relations to one another. Thus, these theories are likely to be useful whenever we encounter such concepts as:

- an organization (of units);
- a net of neurons;
- a society of persons;
- a system of parts;
- interactions between parts;
- co-ordination of parts to a goal;
- integration of parts to form a whole;
- a patter, gestalt (of units).

It is just when dealing with the higher functions of the brain that these theories may be specially useful.

Memory as "transmissions"

If these theories are to be used, however, one must learn to see some old phenomena from a new angle. As illustration, consider the subject of "memory". A century it was defined as the power of evoking past images, or by some similar phrase that appealed essentially to introspection. The man in the street today (and the beginner-student) still tends to think of "memory" in that way. Psychologists, however, long ago found that to study the subject one must treat it as a phenomenon of correction between past events and present behaviors. Now "cause and effect" may well have the cause in one place and the effect in another (e.g., closing a switch here lights a lamp there). If a change of dimension is allowed we may consider a cause now as having an effect later in time (e.g., setting the alarm clock now will make it ring tomorrow). With this approach, the phenomena of memory (in its objective aspects) can be treated by the theories of information and computers when we regard "memory" as corresponding simply to the existence of demonstrable "transmission" (here understood as "having a correlation") between events that were appreciably separated in time. To illustrate the theme, I will give two examples to show two sentences, one that has a "memory" for exactly one word back (so that adjacent pairs of words are related) and the other one for exactly three words back (related in fours). The particular span was ensured by the method of generation, which was a follows (in the three word case:

Three words were written to get started. These were then shown to a person who was asked to add one word that would be related naturally to the visible three. The first word was then covered, and the remaining three words were shown to another person, who was asked similarly to add a natural next word. Then the second word was covered, a third person asked, and so on. In this way each word could be related directly only to just three previous words. (The one-word case differed by having only one word visible as each was added.)

Here is the sentence with each word related to only *one* word back: —

> *paper bag of the time which was coming to be friendly atmosphere was never forget when suddenly he stretched himself and all swollen neck line drawing water hole below deck playing games played with effect on him with only you are well within the nature boy and believing that was already for another pair of pay as before long...*

The triple "... line drawing water ..." shows clearly how "drawing" follows as "line drawing", and "water" as "drawing water"; but "water", after "line drawing" shows that, at the moment when "water" was selected, "line" was playing no effective part in the selection.

And here is the sentence with each word related to just three words back:

> *the costume had holes in my socks I forgot to remind the writer to tell here who would not like the enemy who ran wildly when who should come in handy regardless of the crowd and cameraman seemed pleased with me when suddenly it exploded with great force of gravity was getting lower and lower until at last it gave...*

Here the phrase "...exploded with great force of gravity..." shows that when "gravity" was selected, "force of" was obviously operative, but "exploded" was clearly not operative.

Such examples, as a "synthetic psychosis", show how "memory", as an objective fact showing in behavior, can be treated without any reference to it introspectional aspects. They suggest also that the application of these methods may give insight into some pathological mechanisms.

Determinate or probabilistic?

Another question that has been much clarified by the modern studies of computer theory is whether the brain may be regarded as a machine in particular, whether it should be regarded as determinate or probabilistic. Shepherdson's recent survey (1967) shows how thorough is today's understanding of just what is implied by the idea of "machine". The experience of the last twenty years has shown that, apart from mathematical subtleties, all the various attempted definitions of "machine" prove to be practically identical, even though the various workers have started from very different branches of science. All have tended to the various forms of "semi-group": a set of states (the machine) and an operator (its laws) such that unlimitedly repeated action by the operator on the states cannot generate a state outside the set. The usual ideas about machines either lead to this definition or are derivable from it. What has emerged is that the definition is largely indifferent to whether the operator is determinate (the ordinary "law") or probabilistic (with Markovian transitions). And it has been shown (e.g., Shepherdson, 1967) that the distinction between them is, in essence, small. What either type of machine can do, the

other can do. This fact makes the writing of this article rather simpler: by referring only to the determinate forms, I shall in fact be including most of what is worth saying of the probabilistic forms.

Recent Advances

The theories of information and computers are, as I have said, essentially those of causes and effects when they occur in great numbers. One way of approaching the application of these theories to pathological mechanisms in psychiatry is to look first at what these theories have led to, for here we can see the theories actually at work. I will therefore review their main achievements over the time since my last review in this journal (Ashby, 1958), omitting those that are hardly applicable to psychiatry.

Perhaps the outstanding event, in its philosophical and theoretical importance, occurred on 12 July, 1962, when the Connecticut champion of draughts ("checkers" in the U.S.), R. W. Nealey, was defeated by the computer programmed by A. L. Samuel (1963). The point is that Samuel has no major skill at draughts: he programmed the machine to develop its own strategies, on the basis of its own experience. The process at work was in no way mysterious; the machine was told: use a random generator to suggest random strategies, test them (either in actual play or against published games by masters), keep those that lead to success, and reject or modify (again at random) those that lead to failure. The process is abstractly identical with that of natural evolution: mutation, with preservation of the better and rejection of the worse, leading to ever better skill against Opponent Nature. Samuel demonstrated that this process is intrinsically capable of generating the skills involved in draughts-playing: his work lay chiefly in ensuring that the process, as it actually occurred in the IBM 7090, did not excessively waste the resources available.

Studies such as Samuel's, aiming at what may be called "artificial intelligence", have shown (e.g., Feigenbaum and Feldman, 1963) that the fundamental principles of such processes are often basically quite simple. Why they have not been developed faster, in industry say, is because a great deal of selection within the details of the process is necessary if it is not to be of abysmally low efficiency. (One may remember here that the *principle* of the steam engine was known to the Greeks, but it took many centuries for its efficiency to be raised to a point of usefulness.) Thus the studies of today are turning from the question of the principle (which seems to be often simple) to the question, of much greater difficulty and practical importance, of what factors and methods will raise its efficiency. As Minsky (1963) puts it: "The real problem is to find methods which significantly delay the apparently inevitable exponential growth of search trees."

As was said above, it is precisely when the process is richly combinatorial that the demands for information-processing are most apt to increase excessively. Prominent among the methods for reducing the demands (i.e., for increasing the efficiency) is that of breaking the process into stages. The importance of this method has been strikingly confirmed by the work of Simon and Simon (1962). They took, in the game of chess, the problem of devising a computer program that should work through the final moves to mate. They looked for a process that could be specified by a relatively small number of rules and that would then show relatively great power in the terminal game. They showed that a goal (e.g., mate in eight moves) that might be utterly unachievable by its complexity if the process attempted the whole analysis might be quite readily achievable if the final goal could be reached via a few intermediate ("sub") goals. They developed a number of simple rules relating to such sub-goals (e.g.: give priority to a check that adds a new attacker to the list of active pieces), and showed that such a process was capable of paralleling many of the "brilliancies" in the recorded literature. In fact, the improvement in efficiency was so great that the huge modern computer was hardly required: mere hand-simulation with pencil and paper was sufficient in many cases. They also discovered the interesting fact that were their program showed weakness, apt to overlook the best move, some of the historic mistakes of master play had consisted in overlooking exactly the same move. It seems possible that the cerebrations of master play may be carrying out a process similar to that specified in their paper. But in any case, their work showed clearly the importance of the *method* of thinking (i.e., via sub-goals) and the relative sterility of mere speed and quantity.

Theorem-proving

Another activity commonly regarded as of "higher intellectual" form is that of theorem-proving. Here again the goal can hardly be simpler: one seeks a process, going by steps of accepted validity, to join the available axioms to the final deduction. In the last ten years some major discoveries have been made about the nature of such processes. At first almost all researchers took for granted that the process was deductive, and they devised computer programs to perform the operations. They achieved some success: Wang's program (1960) with the aid of a big computer, successfully proved 220 theorems in 3 minutes. But these theorems were all of simple type, and it became clear that such methods would demand utterly unacceptable times if the theorems were to become mod-

erately complex. (Again the process that seemed obvious proved to be extremely inefficient.) It was then discovered, by Newell, Shaw and Simon (1957), that the process could be enormously speeded if it were changed from an imitation of the deductive method given in the text books to one of: Guess the theorem (i.e., "Is it true that…"), and then search among the axioms (and previously established theorems) for *any* set that justifies the result. This method, implausible though it seems, proves in practice to be enormously superior to the other. After a discovery, of course, it is easier to see reasons, and they wrote: "… the efficiency of working backward may be analogous to the ease with which a needle can find its way out of a haystack, compared with the difficulty of someone finding the lone needle in the haystack." By that as it may, the brain in evolution has clearly encountered many different *ways* of thinking: it seems likely that today our brain has developed, not merely good biochemical and electrical methods, but also some expertness in the construction of methods of information-processing, constructions that are at an entirely higher level than those of the events in the transistor or neuron.

Is the machine original?

At this point the reader may raise the question whether these chess brilliancies or theorem proofs are to attributed to the machine or to the programmer-designer. The experience of Minsky (1967) may help to clarify the matter. He developed a computer-program to prove theorems in geometry. One of its first productions was a proof for the *pons asinorum* that was unknown to Euclid, new to Minsky, and of high mathematical quality. It is so brilliantly simple that it can be given in a few lines. (Triangle ABC has AB = AC; prove that $A\hat{B}C = A\hat{C}B$). The machine proceeded: Compare triangle BAC and CAB!

△	B	A	C		△	C	A	B
	B	A		=		C	A	
		A	C	=			A	B
	B	Â	C	=		C	Â	B

By Euclid's immediately preceding theorem ("two sides and the included angle") the two triangles are equal, so angle $A\hat{B}C$ is equal to the angle that corresponds to it: $A\hat{C}B$ (Q.E.D.). Euclid's proof, with its extensions of sides and construction of extra triangles, looks absurdly clumsy beside this one, which sets aside as irrelevant (as it is) the fact that the two triangles have been derived from a common source. (The proof cannot be claimed as wholly original, as it was known apparently to Pappus (A.D. 300), but it was certainly unknown to Minsky and those working with him.)

The question may now be asked whether the proof was "really" produced by the computer or by Minsky. In fact, however, there was no computer! — Minsky was trying out his program by pencil-and-paper simulation when the simulation process led to his writing down the proof! To attribute it to Minsky is true in some obvious sense, but the allocation would be very misleading: his activities were directed at producing a process, not a proof. If the proof is to be attributed to anything it must be attributed essentially to the process, for wherever that *process* occurs, whether in a computer, or in Minsky's brain, or perhaps in Pappus's brain, that proof is capable of emerging. Thus our original question, of "man-versus-machine" type, has been found to be misdirected. The interesting question of today, valid for both man and machine, is of the type: What processes tend to generate what results?

Pattern recognition

Another branch of "higher" information processing that has been much studied recently is that of "pattern-recognition". Some of the work has been explicitly so directed: the ten digits to be read from a cheque, the 26 letters on an envelope, the ten digits when spoken into a telephone, and so on. Other work has involved it implicitly — e.g., is this position at chess suitable for an advance in the center?, will this geometrical proof be helped by drawing a circle?

The opinion seems to be emerging that *every* pattern-recognizer must ultimately be a special purpose one, designed (whether by man, machine, or natural selection) to perform a certain grouping because that grouping useful. With this opinion the writer agrees: there can no more be a general-purpose recognizer than there can be a general purpose map (for *all* countries!); but perhaps not all workers in the subject would agree with me. Be this as it may, pattern recognition by machine is today being used industrially in the simpler operations. (Its development to a major degree will depend on the effects of the difficulty referred to later.)

Error correction

Another technique that has grown greatly in the past decade is that of "error-correcting codes". One of Shannon's first discoveries, with his new theory of information, was that no matter how much messages might be disturbed by noise (i.e., by effects due to irrelevant and undesired causes), there always exists a code, i.e., a way of sending the messages, that shall reduce the disturbance to insignificance. The catch is that the code must be matched in its general characteristics to those of the

noise, and the code may be very difficult to find. Now the brain is obviously subject, in its work at any moment, to many effects from "irrelevant and undesired causes" — think of the car driven at night in town, surrounded by flickering lights of all colors, only a few of which are traffic signals. Even in more ordinary situations, a large part of what comes to our retina is simply irrelevant to the work in progress, and must be nullified. It is likely therefore that the brain, during evolution, has developed many special methods for combating noise. Unfortunately, most of the studies of error-correcting codes have been made either for telephone/radio channels or for processes in the computer. What has been done to help understand these codes in the brain has been reviewed by Arbib (1964), but it is clear that much remains to be discovered. The psychiatrist might well find that his clinical material, viewed from this angle and suitably interpreted, gives invaluable evidence about the brain's processes.

Adaptive machines

Another aspect of "intelligence" that has quite lost its mystery is the power of the brain to change its organization. From the earliest surgical experiments (e.g., Marina, 1915), to the wearing of reversing spectacle (e.g., Taylor, 1962), it has been known that the brain has a remarkable power, when faced by a new external situation (e.g. a reversal of the attachments of the ocular muscles), of itself altering its ways so that it *compensates* for the external reversal. Analysis of the theory of machines, however, (e.g. Ashby, 1940, 1947, 1952) showed that the "mystery" was due only to our thinking of machines in too simple a form. As soon as one considers the case in which the whole is formed of two sub-machines, of which one performs the obviously visible part while the other performs tasks showing only in the structure of the first machine, then one has a whole that may, if one wishes, be regarded as "a machine that changes its own organization". Today, "adaptive controls", as they are called, have a developed theory and a growing technology. Their theory is today simply a part of the modern theory of machines. It is shown most strikingly perhaps in the latest types of computer. In the early forms, the computer simply performed a computation — solving a set of equations, say, — and it was the human programmer who managed the machine's progress from problem to problem. Today, the computer that solves the equations is only a part, becoming even a minor part, of the total machinery. Behind it is another computer that acts only to manage the primary computer. The "second level" computer (the "manager") accepts a variety of problems, arranges them in order of priority, brings forward necessary subroutines, find suitable storage locations, tells the primary computer what to compute, and may well order it in the middle of its job to lay that job temporarily on one side so that another job of higher priority can be put through. In fact, the enormously increased power of modem computers is not so much due to faster speed in electronics as to vastly better organization of its work. (We could say, less politely, that modem machines are not so appallingly inefficient as the early machines, which would perform the computation in, perhaps, a second, and would then wait for ten minutes while the human programmer supplied the next problem.) It is not impossible (or perhaps is likely) that the human brain is characterized not only by what it can do in the immediately manifest way, but by its exceptional power of planning what it will do, at what time, and under what conditions. If so, the modern computer, this plans its computations ahead, may be developing along the same lines as the living brain. (The subject is referred to again below.)

THE THEORY OF MACHINES

Some achievements described above have been possible only because the past ten years have seen the emergence of a general "theory of machines". Books have been written in the past with this title, but they have referred only to the purely mechanical. The new theory of machines is based on a property that, though suspected or accepted for two centuries, is only beginning to show its power: the idea that a machine is any system such that its state at one instant determines its subsequent behavior. This property, taken for granted by Laplace, was explicitly denied by the ancients, who held that an event now might be determined by what happened many years earlier (the laying of a curse, for instance) regardless of the events in between. Two centuries of science, however, have shown that, in every system adequately studied, its future behavior has been found to depend on just its present state and its present surroundings. The consequences of this "law" (unnamed but universal) are beginning to be traced.

It is not yet easy to say of this new "theory of machines" to what degree it may be useful in psychiatry: what is sterile to one worker may be rich in possibilities to another. Here I will attempt to sketch some possibilities.

It is now known that all behaviors that are clearly and objectively describable can be produced by a machine, in the sense given (McCulloch and Pitts, 1943; von Neumann, 1951); so the question: "Can a machine do it?" is dead, for the answer is always "Yes". (I exclude here certain purely logico-mathematical complications.) Related to this result is the recent proof by Steiglitz (1965) that the analogue and digital modes of processing informa-

tion are essentially isomorphic: whatever can be done by one can, perhaps clumsily, be done by the other. In some sense, therefore, any argument about whether the brain is "really" analogue or "really" digital is of minor interest, for all the higher processes of intellectual activity could be achieved by either mode. Further, all the main theorems provable in one mode must have a corresponding form true in the other. Those who are interested only in the higher processes may thus justifiably ignore the distinction: it becomes significant only when one considers the actual working details.

Some of the results to be anticipated from this theory of machines can be seen from the work of Gill (1962). Among the problems he considered were two that obviously may have application to psychiatry. The first he called the Problem of Diagnosis (he was thinking largely of the computer, but the medical parallel was obvious): given that a system, whose laws of behavior are known, is in some one of a *set* of states, devise a sequence of actions on it, and observations from it, that shall enable the observer to identify the state it is (or was) in. (The theory accepts that the making of the observations will usually change the system's state.) To prove his theorems, Gill showed that we must distinguish between the *simple* experiment, where the machine is unique and non-expendable (like a human patient), and the *multiple* (like a laboratory rat), where the system may be returned repeatedly to the same state (by just starting again with a new rat). Also to be distinguished were the *pre-set* experiment, in which the experimenter would declare beforehand all that he would do, and the *adaptive*, in which later stages of the experiment would be dependent on what had been observed in the earlier stages. He proved a number of theorems some of which showed that certain diagnosis problems were essentially unsolvable; others, essentially unsolvable by a pre-set experiment, would become solvable if the experiment were adaptive.

He also considered the Problem of Homing: to so act on the system as to bring it to a desired state (e.g., to one corresponding to "health"). Here again theorems have been proven about when a pre-set "treatment" may succeed and when the treatment *must* be adaptive, i.e., based on information gathered during progress.

These results are, at the moment, somewhat remote from immediate application. Nevertheless, by making clear the *principles* that must guide the therapist in his interactions with his patient, they may well lay the foundations for a science of the therapy of complex systems, replacing methods based somewhat on intuition and rules of thumb. All the results are, in a sense, dominated by information theory, for they treat the situation of the compound "therapist + patient" as one system subject to basic laws of cause and effect: the patient obviously so, and the therapist also restricted in that he cannot become "knowing" except as the actions or behaviors of the patient make him so.

Consciousness

In this discussion of persons treated as machines, the reader may feel that some essential element of "consciousness" is being ignored. It is true that, in the past, the distinction between man and machine was so obvious that even the slightest resemblance was astonishing; but the point of view has changed much in the last twenty years. By "a machine" is today meant "that which behaves as a machine"; so far as a man behaves like a machine, so far *is* he a machine — to other observers. If a woman dislikes her husband coming home late, but is always put into a good mood by being given flowers, then if her husband *is* late and puts her into a good mood by bringing her flowers, it is merely a verbal matter whether we say he is treating her as a machine or as a woman. From the operational, and from the entirely objective point of view of information and computer theories, all scientific knowledge of dynamic systems is knowledge of the aspect that is machine-like. Nevertheless, the questions are still being asked: Can a machine know it is a machine? Has a machine an internal self-awareness? Can it feel pain? These questions are of the greatest difficulty. One should notice that "consciousness" is sometimes used in a sense that is not intended here; after a motor accident, say, a victim may be "unconscious" in the sense of being simply non-reactive: pricked with a pin he makes no movement. There is no difficulty about *this* use of the word: any dynamic system may be demonstrably reactive or non-reactive. The difficulty enters when "consciousness" is used to refer to personal introspective awareness, to direct "self-knowledge".

It is sometimes held (e.g., Culbertson, 1950) that we have only to extend our scientific knowledge a little further and all will be explained. I can only say here that my opinion is quite otherwise. The work of the last twenty years seems to me only to have repeatedly emphasized the profound difference between those aspects of a system that an observer can discover from its outside, by interacting with it (giving it stimuli and receiving stimuli in return from it) and those aspects accessible to the system about itself. The difficulty seems to be that science deals only with what is communicable (to other scientists and thus to the body of collective knowledge). A system can thus yield to science only such aspects of itself as are communicable. Some aspects, e.g., its weight, are readily communicable, but what Eddington described as "my taste of mutton" is not so: he can transmit to another

only his reaction to mutton. As soon as one attempts to probe this matter thoroughly one comes, it seems to me, directly at the fact of solipsism. If I have no absolute certainty whether a starfish feels pain when it is pricked, or a mimosa, or a balloon, (though all three react), I have to admit that I am exactly as devoid of certainty if what is pricked is a twin-brother: of only one object in the universe have I the direct certainty. Self and not-self are, from this point of view, entirely, and not just quantitatively different. It seems to me, therefore, that the last twenty years' work in cybernetics, far from bridging the gap between *knowledge of self* and *knowledge of other*, has only strengthened our appreciation of its profundity.

In one aspect, however, the theory of machines helps to support the psychotherapist in his conviction that empathy can be useful. The *isomorphism* of systems has not yet been studied much beyond the elementary cases in engineering and physics, but if patience and therapist have a similar background of childhood and experience, the "structure" of knowledge and normal adaptation in the therapist is clearly available as reference for the differing structure in the patient. The subject is too large to develop here; but it may well provide the possibility of a fully scientific basis for the very high-level interactions between patient and psychotherapist.

Information theory of many variables

This digression to the subjective may give clarity to a brief discussion of whether the methods of Shannon are adequate to represent the many ways in which "information in general" enters into such subjects as psychology, psychiatry, sociology, and everyday life. Here there is space only for me to record my opinion that it *is* sufficient, and that most of the dissatisfaction with it comes either from the wish to introduce introspectional aspects (consistently excluded from scientific work) or from a failure to appreciate the great range of ideas and methods that Shannon's basic work has opened up. Here it must be admitted at once that we in the biological sciences have been little helped of recent years by the mathematicians and engineers, most of whose developments of the theory have been directed at the telephone and the computer. The developments in directions meaningful in the biological sciences have largely yet to be made. As example, take the fact that most developments are from the case of sender-receiver: two variables. Now two variables is an absurdly small number for most biological systems. McGill (1954) showed that the extension of information theory to any number of variables is straightforward, and his methods have proved capable of further extension to matters of real interest to the biologist (e.g., Garner, 1962; Ashby, 1965). It is, for instance, now possible to measure informational transmission not merely in-and-out, as a passive telephone wire or an optic nerve treats it, but as the amount that is processed *internally*, as a computer works or a man thinks. (It should be remembered here that "internally" may be interpreted not only as "internal to the organism but also, if one wishes, as "internal to the system of organism-and-environment", the interaction between the two being the real focus of interest.)

One consequence of this development is that it provides a direct and objective measurement for the amount of co-ordination or integration in a system (Ashby, 1968a). To make the idea clear, let us consider the pianist who has the skill to play scales in any key. "Co-ordination", muscular and nervous, is clearly involved, and is objectively demonstrable, for any disturbance by drugs or disease would show objectively as a failure to keep to the appropriate eight notes of the possible twelve. Now while he is playing (correctly) in, say, G major, the twelve notes are obviously not being produced at random, i.e., with statistical independence. (An example in detail is given below.) This lack of independence corresponds to a calculable quantity of "transmission" that *must* exist in some form between the various finger movements if the coordination is to be achieved. The transmission may be effected by a great variety of possible mechanisms (and one must not jump to the assumption that it must all be mediated by nerve fibres, for, e.g., mechanical forces may in fact be used) but the total quantity of transmission *must* be demonstrable if the process is not to be achieved by non-material magic. Thus, *every well-defined set of actions showing coordination specifies a definite quantity of internal transmission* that must be performed if the coordination is to be successful. The quantity is as basic as, say, the quantity of work that a man weighing 150 pounds must do if he is to climb a 20-foot ladder. Because of its fundamental nature, this quantity of information associated with coordination and integration may well prove a useful index when coordination and integration fail.

Dynamic nets

Computer science, too, is today intensively but narrowly specialized, since it is still largely concerned only with prodigiously long chains of simple addings and multiplyings. In this particular form of "complex cause-and-effect relations" it tends to be of small direct interest to the psychiatrist. Designers, however, are aware that more advanced forms of information processing will require something more complex than simple chaining: many more parts must be active simultaneously. *Illiac IV* (at the University of Illinois) is now being designed so as

to be able to carry on 256 operations simultaneously. A very different style of programming will have to be developed, and the programmers themselves will have to think along somewhat new lines, but the extension will doubtless continue. Little, however, is being done in the direction of exploring the "computer" that is brain-like in the sense of using nearly all its parts nearly all the time. To understand such a system, we need to know much more about what might be called "generalized dynamics — the dynamics of systems that are supplied freely with energy (and so are not restricted by its conservation) but which have laws to rule them because they are state-determined and either isolated or subject to a determinate input. It has been proved (Ashby, 1959) that *habituation* will tend to be shown by a very wide variety of such systems, and some further properties have been indicated (Ashby, 1960), but progress is slow.

The study of such systems by the methods of classical mathematics leads quickly to quite un-manageable complexities. Modern studies are turning increasingly to the method of modeling such processes on a computer and simply seeing what happens. The behaviors of nets of randomly connected units were studied in this way by Walker and Ashby (1966). Certain general trends were found, useful perhaps for further studies in the same direction.

Studies of such dynamic systems have repeatedly encountered a phenomenon that may well be of psychiatric interest. It was encountered by Friedberg (1958), and by the writer at about the same time, and was called the "mesa" phenomenon by Minsky and Selfridge (1961). It is apt to occur whenever some change of conditions acts on a large and complex dynamic system: as the system is made larger and larger, so do the consequences of a change in conditions tend to pass from the more or less smooth to having either no effect at all or to having a sudden and large consequence at one critical value. In other words, the response curve changes, as the system is made larger, from a steady slope to a step-function.

This tendency seems to be inherent in a very wide class of systems; and although each particular system has its own particular physical mechanism at work, yet the tendency is so widespread that it may be a general system property. Gardner (1968), for instance, has found it to occur in linear dynamic systems as the richness of internal connection is increased. He asked: what is the probability that the system will be stable?, and then found how this probability changes with increasing richness of internal connection. He found that when the system was small (five or fewer components), increasing connections caused a *steady* fall (in the probability of stability). As the system was made larger (to ten variables or more) the probability tended to stay high as the connections were increased until suddenly it fell to almost zero. What this means is that the large system's behavior will depend critically on the richness of its internal connections, with the dependency very sensitive near the critical value. Thus, in his examples with ten elements, a change of 2 per cent, in the richness of connection caused a change of nearly 100 per cent, in the probability of stability.

It is obvious that a system as complex and dynamic as the brain may provide many aspects at which this "mesa" phenomenon may appear, both in aetiology and in therapy. There is scope for further investigation into this matter, both in its theory and its applications.

The nature of memory

Such studies in the theory of machines have forced into prominence, and have helped to clarify, what is meant by "memory". As was said earlier, this word must be interpreted, under its operational and objective aspects, as equivalent to "transmission between variables significantly separated in time". If the separation is small one may prefer to call it "delay"; if long, "recording". Computer science today ranges over the gamut, lumping them all as "memory", (though, of course, it uses the different types of memory with discrimination).

Saying that memory is a form of, or is homologous with, transmission is more than just using a phrase. It implies that all the discipline of information theory, and all its theorems, are applicable. Thus the theorems of error-correcting transmission become theorems about how to store with immunity to specified types of disturbance; the theorems of channel-capacity now hold over storage-capacity, and theorems of transmission by code become applicable to methods of storage in code.

This new point of view is likely to open up entirely new approaches to the old problems of memory. Von Forster, for instance, has given a most suggestive illustration in his paper "Memory without Record" (1965). The title might suggest the purely mental memory of the Middle Ages, but such is not his intent. He points out that when we ask a child: what is 3×7?, and the child answers 21, nothing is more obvious than that the child somewhere has a "record" (? engram) of the fact that $3 \times 7 = 21$. Suppose now that we want to be able to obtain on demand the product of all 10-digit numbers by all 10-digit numbers. If we assume that the record is to be in the form of an ordinary book, a little figuring soon shows that it will have to be about a billion miles thick! Yet in fact all such products are obtainable on demand from an object about a foot across — called a desk-computer! The point is, of course, that products can be generated

actively: passive storage is not the only way of keeping them.

At the time of writing, the topic of "memory and its storage" is attracting many workers, and research in the future will undoubtedly explore the subject extensively. Yet almost all the work at the moment is envisaging an essentially static trace rather than an active regeneration. Yet everyone knows that few systems have quite such rich facilities for active regeneration as the brain. No neurophysiology's can say that the suggestion of an *active* process is absurd.

The new possibility has major consequences. Suppose, for instance, a child of six or so sees a chess-board, and then shows later that he remembers (can produce on demand) its characteristic pattern. As 64 squares that may be either black or white, the chess-board has, of course, a very high redundancy, in the sense that after we have seen a few, 10 say, we can predict the colors of the remaining 54. Thus the actual information to be used by the boy is not the full 64 bits but 10 or less, and the memory only *need* use the 10 or less. Here is an obvious occasion for the memory to be regenerated: a few bits' storage can hold the initial conditions, and then the remainder can be regenerated (by the rule that the change to each next square calls for a reversal of the color). Thus the answer to: How does this child store the pattern of the chess-board? would be: He doesn't — he regenerates it. And, of course, an experimenter who looks for anything static that resembles a chessboard would find nothing. Only when the brain acts will the *process* develop the pattern.

Another example can be given showing how this approach to memory, as transmission, can be illuminating (Ashby, 1968b). As was said above, every act of coordination implies a certain quantity of transmission. If the coordination extends over times as well as space (later events correlated with earlier) then a calculable quantity of memory is required. This minimal quantity is demanded by the coordination as such, and is quite independent of whatever mechanisms may be used to achieve the coordination. In the article mentioned (1968b), an example is given of coordination in piano-playing, in which is selected the specially simple case in which the player must play some two of three notes A, B, C, and must play on one of two beats and he silent on the other. In this selected example it is easy to show that the total achievement demands a total transmission of 2.92 bits (per bar), of which some must be transmission between the fingers and the remainder between the times. Two different mechanisms are considered in the article and it is shown that they partition this total of 2.92 bits in different ways. One obvious method is to use a "memory store" to record whether the chord was or was not played at the first beat, requiring 1.00 bit, and to provide the other 1.92 bits as transmission between fingers. Another, less obvious, method is to give each finger (or some corresponding nervous center) a memory store of its own, and then coordinate the fingers' trajectories. This last method demands 0.25 bit for memory at each finger, and 2.17 bits between trajectories. What is noteworthy is that the second method (or form of mechanism), though it demands three stores, is less demanding on storage (0.75 bit) than the first method (1.00 bit) which uses only one store. If, therefore, storage were very expensive (in some sense), the method with three stores should be chosen, not that with one. The example shows clearly how the ideas of information theory, appropriately used; can give an unusual insight into the fundamentals of "memory theory".

BREMERMANN'S LIMIT

It remains for me to mention one other fact, a cloud on the horizon no larger than a man's hand, that I think will in time become of dominating importance in the science of information and computers.

Computers today have various limitations. Some are easily removable, with a little more time and money; others are much more profound. Perhaps the most fundamental is that identified by Bremermann (1965). He showed that two of the most basic relations in physics — the mass-energy relation and Heisenbergian uncertainty — together put an absolute limit to the quantity of information that can be transmitted by matter. It certainly covers the industrial computer; and if the scientist's thinking is carried on by some material process in his brain (and no physiologist doubts this), his thinking is also absolutely so bounded.

The actual value of the limit is 10^{41} bits per gramme per second. This quantity may seen1 too large to be of any importance, but in fact examples are easily given (e.g., Ashby, 1963, 1964, 1966a, 1966b) showing how readily quantities exceeding this limit may be demanded. They tend to occur when the information comes from systems having actions in combination (such as were listed earlier). Thus, as soon as one tries to devise processes that will carry out actions of some real complexity — playing a game of chess, driving a car from London to York, writing an article of 10,000 words — one is apt to find, not merely that the demand goes far beyond the limit but that the demand goes beyond it by vast orders of magnitude. The limit in fact is found to be extremely restrictive, so grossly restrictive as to make clear that either our brains are not using ordinary matter (hardly a serious suggestion today) or they are using methods that are of far higher efficiency than those used in today's comput-

ers.

Lest the difficulty be left looking like a hopeless paradox, a few words may be useful to indicate a possible solution. Many processes are known today that compute answers at first sight far beyond even the biggest machine's capacity: they all work by breaking the whole process down into sub-processes (and they can be used only in such problems as *can* be analyzed into sub-processes). Thus, while the game of chess in the strict sense, i.e., with faultless play towards mate, demands quantities of information-processing far beyond Bremermann's limit, yet quite a good game can be played by machine (and man) as a sequence of sub-processes: 1, mobilize your pieces; 2, control the center; 3, get the rooks working; and so on. Such a division into sub-processes, each of which can be carried out without reference to the other sub-processes, has the effect, when it can be done, of lowering by great orders of magnitude the demands for information-processing.

It is not unlikely that the human brain uses this method extensively, of carrying out the total process *as a sequence of sub-processes*. But here the worker who is devoted to the idea that the brain acts as a whole (no one more so than the writer!) must beware of going too far. The Gestalt psychologists who insisted that the brain acts as a whole were perfectly right to oppose the previous generation's attempt to see the brain as a bundle of independent atomistic reflexes. But the introduction of wholeness can go too far. All the studies of the last twenty years, including those studies of nets mentioned above, show that systems should be only moderately connected internally, for in all cases too rich internal connection leads to excessive complexity and instability. The psychiatrist knows well enough that no one can produce associations so quickly or so wide-ranging as the acute maniac; yet his behavior is inferior, for knowing what associations to avoid, how to stick to the point, is an essential feature for effective behavior.

Some evidence in this direction has come to light as the result of an investigation of the amount of information that is processed by an average person in the course of his everyday activities (in contrast to his maximal capacity when stressed on a special task) (Ashby *et al*., 1968a). This investigation attempted to assess the quantity of information processed during the following action (with emphasis on the information required for the coordination and integration involved): (The human subject is given as being engaged in reading when he encounters an unfamiliar French word.)

> **Action**: *He walks across the room to his book shelf (avoiding a chair that is in his path), finds his French dictionary (among 100 other books), finds the word, reads the English translation, and writes down the corresponding English word.*

Details are given in the paper. What is of interest here is that the final estimate for the rate came out at about 2 bits per second (not likely to be in error by more than a factor of 2). Now this quantity seems at first to be astonishingly low. Each optic nerve alone, with half a million fibers, can transmit at least 500,000 bits per second — where is all this information going to, or why is it collected?

The interpretation of these facts is not yet certain, but there is one interpretation that may be related to the limit (Bremermann's) just mentioned. The estimate of 3 bits per second refers to what is necessary for the defined action. Were a robot made to carry out just this action and nothing more, then fully efficient design should not demand more than the 3 bits per second. But a human being, of course, while carrying out this action, is treating this action as only one of a great number of other possible actions; so there must also be information processing to decide the answer to: should *this* action continue? Thus, should the telephone ring during the course of the action, the normal person will no longer elect to persist in this action but will switch to another. And, what information theory has made abundantly clear, the refraining from going to the telephone when it is *not* ringing demands information-processing capacity just as absolutely as the going when it is ringing.

This difference, between the millions of bits entering by all the senses and the 3 bits per second used directly in the action, suggests that what goes on in the brain may be responsible for the obviously visible ("tactical") actions to a minor degree, and to a far greater degree may be concerned with the less visible ("strategic") question of the choices between the various possible actions. Such a method, separating the details of the tactical action from the processes controlling the strategic organization of many actions would achieve just the reduction of combinations that would be appropriate to Bremermann's limit.

There is a striking parallel here with the developments in computers since they were introduced. As was said above, today's machines are enormously more effective than the earlier, though their operations of computation are little altered. Their vast superiority is due to the fact that they organize their work so much better. The triumph of the last decade is not a faster addition but the development of (say) time-sharing. Today one computer can keep a dozen departments happy, dovetailing all their demands together with a skill like that of a juggler who keeps a dozen balls in the air. The further theory of "higher information-processing" will, I

suspect, be at this organizational level rather than at the unit-operational. Psychiatry may perhaps be able to learn something from the computer-scientists; but it is just as likely that the computer-scientists will be able to learn from the psychiatrists. What is chiefly necessary at the present time is that they should learn to speak something of each other's language.

Looking back over the last twenty years one is tempted to think that information theory promised too much, and failed to deliver the goods. Yet the fact remains that information theory is essentially the science of complex dynamic systems, with complex weavings of causes and effects in great numbers. Those who would study such systems need information theory (in some form) just as surveyors need some form of geometry. What has happened, I think, is that so much effort has gone into its development for the telephone and computer that it has developed along lines little suited to the real needs of workers in the biological sciences. Take for instance the fact that almost all information theory developed so far deals with the very tidy case in which the system is going to use an exactly defined set of symbols — the 26 letters of the alphabet, the 10 digits, the distinct voltages between -3 and $+3$, for instance — a very useful case in much engineering. In the biological cases, however, the "alphabet" is not sharply limited, but tails off almost indefinitely. Again, Shannon's basic method is to consider one set (e.g., all possible ten-word phrases) with the messages as a population to be sampled. But in "content analysis" one has the essentially opposite situation: a unique message has been received and one wants to discuss the various sets from which it might have come. Thus a patient might utter just "Doctor, I hate you", leaving the real question whether this message is from the set of those expressing opposition, or whether it is from the various ways of saying "At last I can be frank and reveal what is troubling me." Content analysis thus provides a direction which the basic ideas of information theory can be developed in a way of real interest to those in the biological science. A start has been made (e.g., Krippendorff, 1967) but the field is almost entirely unexplored.

His work, and the other advances described above, suggest that information theory is at last beginning to be forced in the directions appropriate to the needs of the biological sciences. Twenty years' experience has helped to make the topic more realistic. The time is now ready for the researcher who can appreciate sympathetically the work done by the engineers, and who then, with the needs of the biological worker firmly in mind, can force its development in the directions appropriate to psychiatry. Perhaps this article may help to suggest what these new directions may be.

ACKNOWLEDGMENT

The work on which this article is based was jointly supported by the U.S. Air Force Office of Scientific Research under 7-67, the U.S. Air Force Systems Engineering Group under contract 33(615)-3890, and the National Aeronautics and Space Administration.

REFERENCES

Arbib, M. A. (1964). *Brains, Machines and mathematics*. New York.

Ashby, W. Ross (1940). "Adaptiveness and equilibrium." *J. ment. Sci.*, 86, 478–483.

— (1947). "Principles of the self-organizing dynamic system." *J. gen. Psychol.*, **37**, 125–128.

— (1952). *Design for a Brain*. London.

— (1958). "Cybernetics." In: *Recent Progress in Psychiatry* (ed. Fleming). **3**, 94–117. London.

— (1959). "The mechanism of habituation." In: *N.P.L. Symposium on the Mechanization of Thought Processes* (ed. Cherry). London. **4–4**, 1–21.

— (1960). 2nd edition of 1952.

— (1963). "Systems and information." *Trans. IEEE*, MIL-7, 94–97.

— (1964). "Modelling the brain." In: *IBM Symposium on Simulation Models*. Pp. 195–208. New York.

— (1965). "Measuring the internal informational exchange in a system" *Cybernetica*, **8**, 5–22.

— (1966a). "Mathematical models and computer analysis of the function of the central nervous system." *Ann. Rev. Physiol.*, **28**, 89–106.

— (1966b). "Some consequences of Bremermann's limit for information-processing systems." *Bionics Symposium*, Dayton, May 3–5. (In the press.)

— (1968a). "Information-processing in everyday human activity." *BioScience*. (In the press.)

— (1968b). "Measuring memory." In: *Festschrift for Prof. P. K. Anokhin*. (In the press.)

Bremermann, H. J. (1965). "Quantal noise and information." In: *5th Berkeley Symposium on Mathematical Statistics and Probability* (ed. Neyman). vol. 4.

Culbertson, J. T. (1950). *Consciousness and Behavior*. Dubuque.

Feigenbaum, E. A., and Feldman, J. (eds). (1963). *Computers and Thought*. New York.

Friedberg, R. M. (1958). "A learning machine." *IBM J. Res. & Devel.*, **2**, 2–13, and **3**, 282–287.

Gardner, M. R. (1968). Thesis for M.S. University of Illinois.

Garner. W. R. (1962). *Uncertainty and Structure as Psychological Concepts*. New York.

Gill, A. (1962). *Introduction to the Theory of Finite-state Machines*. New York.

Krippendorff, K. (1967). Thesis for Ph.D., University of Illinois.

McCulloch, W. S., and Pitts, W. (1943). "A logical calculus of the ideas immanent in nervous activity." *Bull. Math. Biophys.*, **5**, 115–133.

McGill, W. J. (1954). "Multivariate information transmission." *Psych—et* [illegible], **19**, 97–116.

Marina, A. (1915). "Die Relationen des Palaeoencephalons sind nicht fix." *Neurol. Centrabl.*, **34**, 338.

Minsky, M. (1963). "Steps toward artificial intelligence." Reprinted in Feigenbaum and Feldman (q.v.). 406–450.

— (1967). (Personal communication.)

— and Selfridge, O. G. (1961). "Learning in random nets." in *Proc. 4th London Symp. on Inf. Theory* (ed. Cherry). London.

Newell, A., Shaw, J. D., and Simon, H. A. (1957). "Empirical explorations with the logic theory machine." Reprinted in Feigenbaum and Feldman (q.v.), 109–133.

Samuel, A. L. (1963). "Some studies in machine learning using the game of checkers." In: Feigenbaum and Feldman (q.v.), 71–105.

Shannon, C. E., and Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana.

Shepherdson, J. C. (1967). "Algorithms, Turing machines and finite automata." In: *Automaton Theory and Learning Systems* (ed. Stewart), 1–22. London.

Simon, H. A., and Simon, P. A. (1962). "Trial and error search in solving difficult problems." *Behavioral Sci.*, **7**, 425–429.

Steiglitz, K. (1965). "The equivalence of digital and analog signal processing." *Inf. & Control*. **8**, 455–467.

Taylor, J. G. (1962). *The Behavioral Basis of Perception*. New Haven.

Von Foerster, H. (1965). "Memory without record." In: *The Anatomy of Memory* (ed. Kimble), 388–440. Palo Alto.

Von Neumann, J. (1951). "The general and logical theory of automata." In: *Cerebral Mechanisms of Behavior* (ed. Jeffress), 1–32. New York.

Walker, C. C., and Ashby, W. Ross (1966). "On temporal characteristics of behavior in certain complex systems." *Kybernetic*, **3**, 100–108.

Wang, H. (1960). "Toward mechanical mathematics." *IBM J. Res. & Devel.*, **4**, 2–22.

W. Ross Ashby, M.D., D.P.M., *Professor Of Electrical Engineering and Biophysics, University of Illinois, Urbana, Illinois, U.S.A.*

(*Received 15 January, 1968*)