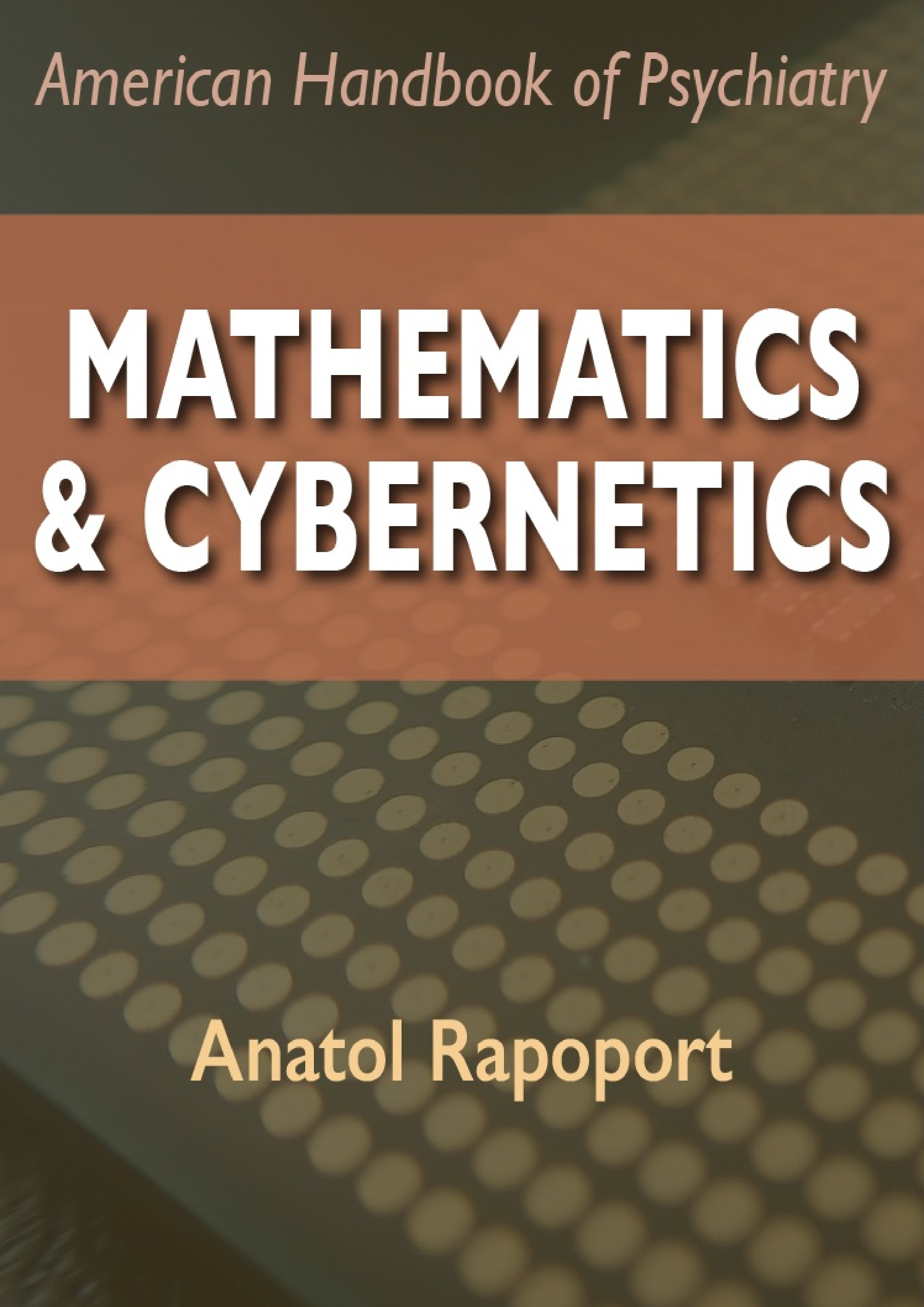

American Handbook of Psychiatry

MATHEMATICS & CYBERNETICS

Anatol Rapoport

MATHEMATICS AND CYBERNETICS

Anatol Rapoport

e-Book 2015 International Psychotherapy Institute

From *American Handbook of Psychiatry: Volume 1* edited by Silvano Arietti

Created in the United States of America

Table of Contents

MATHEMATICS AND CYBERNETICS

Anatol Rapoport

The fundamental contribution of mathematics to science has been to provide a precise and contentless language in which to describe events, to formulate generalizations, and to deduce consequences of assumptions. Precision and independence from content are interdependent. The vocabulary of everyday language depends on the way perceptions and concepts are organized; for instance, on the particular way objects are classified or relations among them are interpreted. In attaching names to objects, properties, or actions, we fix the categories in which we think. These categories are of necessity too crude to capture the infinite variety of events that constitute “objective reality.” Thus a con- tent-bound language may impose a structure on our perceptions of the world and on the abstract concepts we form, and this structure may or may not correspond to the structure of reality.

Because mathematical language is content- less, that is, totally abstracted from perceptions, its structure is entirely transparent. In the exact (mathematicized) sciences the structure of a mathematical theory is constantly compared with the structure of a portion of the world under study. Mathematics itself, however, is concerned with the structure of relations independent of empirical content.

As an example consider the equation (a mathematical statement) relating the area of a circle to its radius, $A = \pi R^2$. It says that whatever be the radius of a circle, the ratio of the area to the square of the radius is always constant, equal approximately to 3.1415926. The statement is actually a composite of a potentially infinite number of statements, since it specifies the magnitude of the area of *all* possible circles. Assuming that the radius can be specified with infinite precision, the area can also be specified with infinite precision, because the number *v* can be calculated with infinite precision. However, the statement cannot refer to anything in the empirically observable world, because there are no perfect circles and because physical measurements cannot be made with infinite precision. The statement refers only to idealized objects in an idealized mathematical world.

The scientific revolution of the seventeenth century was a consequence of a discovery that certain real events could be *almost* precisely described by idealized mathematical models, in the first instance, the motions of heavenly bodies and the behavior on physical bodies subjected to specified forces under controlled conditions. Thus the first mathematicized science was born —mechanics.

The scientific revolution of the seventeenth century is generally recognized as the impetus that stimulated the Industrial Revolution of the eighteenth century and consequently the immense social changes that came in its wake. To appreciate the significance of this impetus fully, it is necessary to recognize the *conceptual* impact of mathematicized science. First, the world of matter appeared to be governed by physical laws. These laws, however, could no longer be stated as metaphysical principles like "Nature abhors a vacuum," "There is no effect without a cause," or "All things consist of substance and form." Physical laws are invariably stated as mathematical equations— relations among quantities—-and the quantities themselves represent results of specified measurements, that is, concrete operations with meter sticks, balances, clocks, thermometers, barometers, potentiometers, and the like. Implied in each physical law are predictions of what will be observed under *specified* conditions. Both the conditions and the observations having been specified as quantities, that is, readings on instruments, the truth or the falsehood of the assumptions can in principle be determined by independent observers. Thus *philosophical* arguments about the validity of generalizations become irrelevant. In the final analysis the truth of an assertion becomes a matter of objective verification of observations. Therefore, the first result of the scientific revolution was that of fixating the specific meaning of "truth" in the context of scientific discourse, making it independent of pronouncements of authority, of speculations couched in verbal arguments, or metaphysical concepts.

Second, mathematical language has greatly expanded the scope of *deduction.* Deduction is a process by means of which, assuming the truth of some assertions, we can assert the truth of other assertions. Syllogistic reasoning is an example of deduction applied to assertions involving class inclusion. For instance, assuming that no A is B and some C are A, we can conclude that some C are not B. Rules of mathematical operation vastly expand the range of deductive reasoning. Thus, from mathematical equations expressing physical laws, a vast number of other quantitative relations can be deduced by chains of mathematical reasoning. Empirical verification of the deduced relations corroborates the validity of the laws. Empirical falsification of the deductions necessitates a search for the roots of the discrepancy. At times it is discovered that certain conditions had not been taken into account. At times the formulation of the laws is modified to bring them into closer correspondence to reality.

In this way science has changed fundamentally the old conception of knowledge as a collection of insights of wise men to be absorbed by studying texts. Scientific knowledge revealed itself as constantly growing and constantly being revised in the light of new observations and new interpretations of what is observed. The most important single factor effecting this change has been the adoption of mathematics as the language of the exact sciences. Assertions in that language leave no doubt about what is asserted (and consequently what must be done to test the assertions) and, moreover, bind the assertions into logical interdependence, the organic structure of scientific theories.

The role of the exact sciences in technology is obvious. Their role in the development of scientific medicine is no less apparent. Diagnostic procedures have come to depend more and more on refined observations made possible by instruments and laboratory procedures. In fact, scientific diagnosis is largely formulated in quantitative terms: temperatures, blood pressures, concentrations of substances in body fluids, shapes of electrocardiograms and electroencephalograms. Chemotherapy and physiotherapy are extensions of chemical and physical technology to medicine. Genetic etiology of diseases is discovered by statistical techniques. Effectiveness of drugs and other forms of therapy is evaluated by statistical inference. Indeed, the bulk of contemporary scientific medicine stems from the conception of the living body as a material system and of its living process as a complex network of physical and chemical processes that preserve a certain dynamic balance. The balance can in principle be described by certain limits within which the parameters of the process may vary. Disease can be defined in terms of deviations from these limits. If the deviations are reversible the “normal state” can be restored. Otherwise, death eventually occurs, which means that the dynamic processes that characterize the living organism can no longer be re-established.

To the extent that psychiatry is rooted in knowledge of organic structure and function, the same methods and conceptualizations apply to its findings. Neural anatomy and histology, neurophysiology, biochemistry, and genetics have all contributed to scientific psychiatry and so have demonstrated the relevance of the contributions of mathematics. No less important are the contributions of statistics (a branch of applied mathematics), an indispensable tool in studying gross trends and in evaluating results of therapeutic procedures on populations of patients.

In short, wherever psychiatry is concerned with physical events or with assessment of causes and effects on a gross scale, mathematics (including statistics) contributes to it as it does to any other science.

The Mind-Body Problem

The cleavage that still persists between psychiatry and other branches of science, including scientific medicine, is rooted in the Cartesian mind-body dualism as reflected, for instance, in the distinction between “organic” and “functional” mental disorders. From the standpoint of the essentially materialistic world outlook embodied in at least classical natural science, the mind-body dichotomy is not essential. It is seen not as a reflection of a dualism of reality but simply as an idea induced by our direct introspective knowledge of our state of consciousness, which seems different from the sort of knowledge we obtain through our senses about the external world. For the materialist, “mind,” “consciousness,” and so forth are only aspects of material events; for instance, nervous activity perceived “from the inside” as it were, rather than from the outside. From this point of view “thoughts,” “concepts,” “memories,” “emotions,” and the like are assumed to be the subjective aspects of objectively observable events, in principle describable in physiological terms.

The qualification “in principle” frees the adherent of this view from the necessity of demonstrating its validity in each specific instance. He is content to search for physiological correlates of mental activity, and whenever he finds apparent correlates, he is satisfied that the discovery corroborates the basic reductionist assumption.

The question remains of what constitutes a correlate of mental activity. Some phenomena clearly deserve the name; for instance, reports by individuals of thoughts, feelings, and the like reproducibly evoked by stimulating specific areas of the brain (as in experiments performed on patients undergoing brain surgery). Other evidence is obtained from ablation experiments on animals, where reproducible behavioral changes are effected. Here, since we have no access to the animals’ mental state via reports, the corroboration of the reductionist hypothesis must depend on a tacitly assumed linkage between mental activity (not directly observable) and behavior (directly observable). Since, however, the materialist takes this linkage for granted, he is satisfied that reproducible correlations between anatomical structures and physiological events, on the one hand, and behavior patterns, on the other, corroborates the identification of “mind” with material events. The task of reduction, accordingly, becomes that of disclosing the “mapping” of neural events upon behavioral events.

The simplest “mappings” of this sort go back to the discovery of the reflex are. A large advance is associated with Pavlov’s discovery of the conditioned reflex. Thereby the extreme flexibility of behavior patterns of higher animals appeared explainable in principle. Objections to what appeared to be a mechanistic conception of behavior (and, by implication, of mental activity) revolved around the so- called purposefulness or “goal-directedness” of animal behavior, which, it was said, eluded all explanations based on mechanical models. The argument is similar to that of vitalists, who would subsume all living processes (not only behavior) under “goal-directed” ones, to be clearly differentiated from mechanical (not goal-directed) processes characteristic of the nonliving world.

It is true that classical physical science expelled goal-directedness from its conceptual repertoire. However, the conception of instantaneous local “causality,” devoid of teleological components, is not confined to classical mechanics. It pervades all mathematicized physical science. Processes governing chemical reactions and the propagation of electromagnetic waves are typically formulated in differential equations, which relate magnitudes of variable quantities to their rates of change. The solutions of these equations are time courses of the variables. Thus, if the totality of these magnitudes and the relations among them are taken to be the description of a system, then each instantaneous state of the system is, in a way, the "cause" of the immediately succeeding state. "Causality," then, when analyzed completely, turns out to be acting "here and now" without reference to future states or "goals."

A detailed examination of *some* aspects of the living process showed that they could be explained in terms of obeying known physical and chemical laws. In particular, the early contentions of the vitalists that the energetics of living processes cannot be derived from the law of conservation of energy proved to be groundless. Also main- of the regulating physiological processes, which keep temperatures, concentrations of substances, and so forth within certain limits, turned out to be the effects of homeostasis, the preservation of nonequilibrium steady states. It has been shown that nonliving systems can also be regulated by homeostasis as long as they are permeable to exchanges of matter and energy with the environment (open systems).

A much more serious difficulty in the way of extending the mechanistic paradigm to apply to living systems is the conspicuously goal-directed nature of gross behavior. Only if such apparently purposeful behavior could be exhibited in a system where nothing but established physical laws were known to operate, could the mechanistic conception be extended to the behavior of living organisms in relation to the outside world.

The problem of vindicating the wider applicability of the mechanistic view of nature (to include at least some aspects of living behavior) became linked with the problem of constructing machines that would exhibit purposeful behavior. The actual construction of such machines was spurred on by other than philosophical motives. The need was for machines that could transcend the limitations of the human brain so as to guide the performance of other machines that transcended the limitations of the human muscle. This need is being met by modern automation technology. The brilliance of this technological achievement, however, should not obscure the importance of its philosophical implications, namely, a corroboration (not a proof, of course!) of the conjecture that the behavior of organisms can be explained in terms of known physical laws.

Servomechanisms

Machines capable of what appears to be goal- directed behavior are called *servomechanisms.* The branch of technology dealing with their construction developed especially rapidly during and since World War II and has been christened *cybernetics.* Figuratively speaking, cybernetics deals with the "intelligence" of machines. The engines of the precybernetic era had no "intelligence" to speak of. Vehicles had to be steered, guns had to be aimed; power had to be turned on or off by human operators as conditions demanded. Even in the early days of the industrial era, however, certain simple cybernetic devices were known. Steam engines had governors that controlled the speed of the flywheel by automatic action triggered by a critical speed. In the thermostat, another familiar device, the source of heat is turned off when the column of mercury in the thermometer reaches a critical height and turned on when it sinks below it. When the rudder of a ship is set in a certain position, the ship will eventually assume a prescribed course, since there will be a torque on the hull as long as the ship is *not* on the prescribed course.

These examples illustrate the fundamental principle of cybernetics, namely, the utilization of *error* in correcting the error. Every machine is designed to respond in prescribed ways (emit certain outputs) to given conditions in the environment (inputs). In servomechanisms the performance of the machine itself, or rather the comparison between its performance and some prescribed end state, serves as an input. In a way a servomechanism can be viewed as a machine that keeps asking "How am I doing?" Through a system of closed loops, called feedback loops, a servomechanism responds not only to the environment but also to its response to the environment, to the response to the previous response, and so on. This circularity of responses creates the impression that a servomechanism is guided by a preset “goal” and so simulates the purposeful behavior of a living organism.

Similarity is a symmetric relation. If servomechanisms can be said to behave in some ways like living organisms, then living organisms can be said to behave in some ways like servomechanisms. Once this analogy is noticed, new methods of investigation suggest themselves in psychology and in the behavioral sciences in general. For the theory of cybernetics, linked with rich engineering experience, gives rise to concepts, hypotheses, and conjectures that often can be translated in behavioral terms.

To take an example, consider the concept of the transfer function, central in system engineering. An engineer’s system is designed to give a prescribed output to each of the inputs to which it is sensitive. For instance, the input may be the image made by the path of an airplane and the output an appropriate aiming of the antiaircraft gun. The motion of the plane is described in terms of its instantaneous position, the instantaneous rate of change of position (velocity), the rate of change of the rate of change (acceleration), and so on ad infinitum. Clearly there is a limit on how rapidly the output can change appropriately. It would be difficult for a gun weighing several tons to follow the motions of a swallow. The inertia of the gun is one limitation; another is the speed with which information inputs can be processed. The capacity of a servomechanism to respond to inputs is determined by its transfer function, which depends, in turn, on a system of interconnections of its parts, an analogue of a “nervous system” processing the inputs and translating them into outputs. The structure of this “nervous system” is, of course, completely known to the designer of the machine. Indeed, his task is to design servomechanisms with prescribed transfer functions or else to calculate the characteristics of a transfer function capable of achieving the purpose for which the servomechanism is designed.

Consider now the inverse problem: given the performance of a servomechanism, to infer the structure of its “nervous system.” This problem, called the “black box” problem, is central to the task of a physiological psychologist, who seeks to infer at least the general features of a nervous system that could account for some observed behavior pattern.

As a rule inverse problems are harder than direct ones, and their solutions often are not unique. That is, a great many arrangements can give the same transfer function, so that even its precise determination does not give much information about the underlying structure. Thus it would be hopeless to try to infer the vast collection of servomechanistic arrangements in a human brain by noting correspondences between stimuli and responses arbitrarily chosen, or chosen for their supposed importance in human behavior. In some situations, however, the transfer function itself is an object of interest, the determination of which depends on our ability to describe the inputs and the outputs in precise mathematical terms. When this *can* be done such situations are singled out for study, not because they are necessarily behaviorally important, but because they are analyzable by the methods at our disposal and so can serve as stepping stones in the development of the theory. In the mathematicized sciences the choice of a problem is, of necessity, often guided by tractability.

So-called tracking problems are of this sort. Their investigation was motivated partly by the need to understand the performance of the human component in man-machine systems, but their theoretical tractability was an additional impetus. The usual tracking experiment involves the task of following a target by moving a lever. The input (the motion of the target) is fully describable in terms of superimposed simple motions. Thus the complexity of the input is a controllable quantity. The output (the tracking motions of the subject) are likewise analyzable. From the mathematical relations between the input and the output, the subject’s transfer function can be determined. This knowledge is useful to the engineer designing a man-machine system. It also provides theoretical leverage for the black box problem. On the basis of the inferred transfer function, the neurophysiologist can at least make guesses about structural features in the nervous system that can account for the transfer function.

There have been suggestions for using cybernetic methods in diagnostic procedures. Already in the earliest formal treatment of the subject, N. Wiener called attention to the similarity between certain kinds of nervous pathology and servomechanism malfunctioning, particularly the oscillations accompanying the loss of motor control. The corresponding causes of the malfunctioning in servomechanisms being known or inferable, it appeared to Wiener that such knowledge might be transferable to the neurological situation. The work of L. Stark and T. N. Cornsweet on the servomechanistic analysis of the pupil reflex is an example. If a sinusoidally varying light intensity impinges on an eye, the pupil will respond by periodic contractions and dilations. This is essentially a "tracking" task. From these oscillations the corresponding transfer function has been computed. As the gain of the system (the decrease of intensity due to the contraction divided by the increase of applied intensity) is increased past a certain threshold, the system becomes unstable, and the pupil oscillates at its "natural" frequency. This frequency is calculated from the transfer function and turns out to be about 72 cycles per minute. The actually observed "natural" frequencies in human subjects ranged from 62 to 80 cycles per minute (in some 80 subjects). But in 70 pupils of patients with multiple sclerosis, these oscillations averaged only 41 cycles per minute.

Since oscillations are clinical manifestations of a wide variety of neurological diseases (tremor, ataxia, clonus, nystagmus) and since servomechanistic analysis leads to specific neurological hypotheses, Wiener's early conjecture concerning the diagnostic value of the cybernetic approach may be a valuable guide to research on the working of the nervous system.

Information Theory

The central concept underlying the technology of the First Industrial Revolution had been that of energy. The primary function of an engine is to utilize a source of energy, such as fuel, to do work, to move masses of matter, for instance. The central concept underlying the (cybernetic) technology of the Second Industrial Revolution is that of information. The primary function of a servomechanism is to process inputs to convert them into appropriate outputs. This is also the central problem in the technology of telecommunication.

Information theory (or, more properly, the mathematical theory of communication) deals with that which is carried by signals abstracted from what the signals are made of or signify. A signal can be sent by producing an air disturbance, or an electrical disturbance, or a light. The method of sending the signal is not important in information theory. What is important is what "knowledge" the signal conveys, or rather the quantitative aspect of that knowledge. The most fundamental idea in information theory is that the "amount of information" depends not on *what* is said in a message but on what *could* have been said. In much the same way the probability of an event is associated not with the event itself but rather with the whole context in which the event could occur. Thus the probability of drawing a particular ball from an urn depends on the number of balls in the urn. Hence it is not the ball in question that determines the probability but the number of balls that *could* be drawn. Similarly the amount of information in a message is not defined unless one can specify all possible messages from which the message in question is selected. This is the meaning of the “amount of information” in the mathematical theory of communication. The meaning is made precise by abstracting totally from the content of the messages. For example, the telegraph operator is not concerned (indeed, his professional ethics do not allow him to be concerned) with the content or meaning of the messages he sends. To him the messages are only sequences of signals. The amount of information in a message is calculated in terms of the *a priori* probability of that message being selected from all the possible messages that could have been sent

This concept of the “amount of information” allows the telecommunication engineer to design efficient and economical equipment for transmitting expected information loads over channels. In this context information becomes something that flows over channels, much like power flows over power lines or oil flows through pipes. It makes sense to speak of capacities and volumes of flow and efficient “packing methods” (which in telecommunication become “coding”) quite in the same way that one speaks of the flow of traffic, people, or material goods where efficiency depends on scheduling or packaging. These considerations apply to all forms of telecommunication, telephone, radio, and television.

Here, then, is still another view of the nervous system—that of a telecommunicative device. It is that, of course, in the literal sense and has long been recognized as such. What information theory has achieved is to have created a powerful and precise language for describing the performance of telecommunication devices. This language is now used as a tool in constructing theories of nervous function, which, it is hoped, can be extended to theories of behavior.

The formulation of the essentials of information theory has inspired numerous psychological experiments based on the central concepts of the theory. In these experiments the individual is treated as a “channel” whose overall characteristics—for example, channel capacity—are to be estimated. One way of doing this is by pumping information through the individual, that is, by making him a link in a telecommunication channel, a transducer. For example, if the individual is required to respond differentially to each of a collection of signals presented in random sequence, his channel capacity is expected to put an upper limit on the rate and on the accuracy of his performance. (Information theory also deduces the mathematical relation between rate and accuracy.) Now the amount of information per signal can be varied at will by varying the number of signals from which selections are made, by varying the relative frequencies with which the different signals are sent, and by varying the sequential probabilities of the signals. Thus it is possible to have the same average amount of information per signal in several different situations, involving different numbers of signals, different relative frequencies, or different sequential probabilities. The conception of the individual as a link in a communication channel suggested that his performance should be determined by the rate of information flow rather than by the particular way this rate is achieved. Early experiments on choice reaction times provided some corroboration for this hypothesis. Of equal or even greater importance, however, were the discrepancies that could not be accounted for by the channel model. These led to the design of more refined experiments and to a more detailed analysis of reaction times, which revealed some distinct inadequacies of the information theory approach and advanced alternative interpretations of the experimental results. Thus in its very failure information theory served in a constructive role *as a point of departure* for a theory of information processing in the nervous system.

Automata

In addition to its central role in telecommunication, information theory is also important in the theory of automata, a class of machines to which the high-speed computers belong. For what is called the “memory” of a digital computer is simply a storage facility for information, a reservoir, to which information is shunted to be recovered when needed. In cybernetics, too, information theory ideas are important. The “conditionality of response” of a piece of automatic equipment, that is, the complexity of instructions it can “understand,” is also measurable in information units. The “intelligence” of machines thus becomes a measurable quantity, just as in the early days of technological evolution, mechanical advantage and, later, horsepower were standard evaluative units. The ability of computing machines to perform not only arithmetical calculations but also complex logical operations has induced their classification (partially in jest, one supposes) as “thinking machines.” There is no question, of course, that in some respects automata simulate the thought process. Again, turning the simile around, we might ask whether in some respects our thinking organs may not function on the principle of computing machines.

W. S. McCulloch and W. Pitts showed that a model of the functional logical processes, as denoted by the operations of symbolic logic, is entirely consistent with certain simplified assumptions concerning the interaction of neurons.[1] Suppose we picture a neuron as a unit that can exist in only one of two possible states—“firing” and “nonfiring.” (This is not factually correct, of course, but is an idealized version of the “all-or-none” law.) Suppose further that the firing of a neuron is occasioned by the impingement on its dendrites or cell body of the summed activities of other neurons, transmitted via axones to the terminal buttons. Let the threshold of firing of a neuron be defined as the minimal number of active terminal buttons sufficient to fire it. Finally suppose that some of the terminal buttons are inhibitory; that is, their activity subtracts from rather than adds to the firing potential impinging on the neuron. These characteristics are sufficient to represent any conditionality of response of any neuron or set of neurons by an appropriate arrangement of excitatory and inhibitory connections. It follows not only that one-to-one stimulus- response relations can be represented in an idealized nervous system (this could be done already with the old “telephone switchboard” models), but also the dependencies of responses on “inner states,” be they interpreted as memories, accidental associations, or random fluctuations, can be included.

Following the completely abstract “logical” model of the nervous system, several “engineering models” of neurons were proposed and built in connection with experiments simulating the activity of neurons or neural nets.-' The engineering models, in turn, stimulated theoretical analysis of information processing in elements assumed to have the characteristics of “real” neurons, for example, membrane potential, absolute and relative refractoriness, and so forth.

The behavior of automata has been shown to be capable of far greater variability and flexibility than had been imagined. Modern computing machines do not just perform specified listed operations in order; they are capable of “making decisions,” as is evident from the programs that guide their operations; for example, “Add column 6 to column 13, compare the result with the last entry in column 2; if the sum is greater, extract square root of column 10, otherwise proceed to step 7, etc.” Computing machines can solve logical problems such as this one: If bandits don’t drink beer only if the sun shines and the moon is in the first quarter, and if, whenever the sun shines, shrimps cannot whistle unless the moon is either in the second or third quarters, and if bandits drink beer at the same time when shrimps do not whistle only when gophers go skating in the moon’s last quarter, it being understood that when the last-mentioned does not occur it does not mean that bandits cannot drink beer if shrimps whistle or that shrimps must whistle if bandits do not drink beer; what may or must be the phase of the moon when gophers go skating on a cloudy day while the shrimps remain silent?

Certainly some of the thinking we do is of this type (though not as complicated). It had been almost taken for granted, until the theory of automata showed otherwise, that “thinking” is necessarily a different sort of activity from what machines are able to do. Indeed, machines had been habitually looked upon as strong but stupid. In some circles an argument rages about whether the technology of automata has refuted this view, whether computers “really” think. The theory of automata has shown that once we have described the thinking process with sufficient precision, we can build an automaton to simulate it. Nor does the simulated thought process need to be rigid. For the *rules* of inference can be made to change in consequence of the automaton’s “experience.” The “lifelike” character of automata equipped with simple servomechanistic regulatory units and just one or two “motivation” mechanisms has been dramatically demonstrated. A “turtle” that persists in following white lines randomly drawn on the floor seems to have an “aim in life.” It seems even more lifelike when it is observed to run to electrical outlets to get recharged as its batteries threaten to run down, and even more so when it changes its behavior patterns after being “spanked.”

None of these demonstrations is sufficient to change the minds of those who insist that “machines can’t think.” It is always possible to keep revising the definition of “thinking” so as to keep it in the residual area of what has *not* yet been successfully simulated. But this sort of procedure may be a rationalization of an aversion to equating men with machines rather than the discovery of the basic difference between men and machines.

Turning to the possibility of applying the theory of automata (as it pertains to logical operations) to a theory of specific nervous activity involved in thinking, we find that the practical difficulties are enormous. Whereas, in the case of cybernetic and information theory approaches, it was possible to ferret out gross concepts reasonably applicable to nervous regulatory and signal-transmitting activity (for example, transfer function, channel capacity, and so forth), we find that, in viewing the nervous system as an automaton, we must postulate the existence of specific units and specific relations among them. The basis of the theory is the correspondence between the fundamental logical operations and certain arrangements of relays. These are arrangements corresponding to the logical operations of “and,” “or,” “implies,” “not,” and the like. In this way every logical function consisting of binary variables (propositions and their negations) and logical operations can be mapped (not uniquely, though!) upon certain networks of relays, which one may interpret as “neurons.” Even if the real neurons obligingly acted in every way like those automaton units, it would be all but inconceivable with our present techniques to identify the particular arrangement responsible for even a modest range of behavior patterns of a living organism. At best this is possible in the simplest instances. For example, B. Hassenstein and W. Reichhardt have studied the responses of a beetle to stimuli impinging upon the separate contiguous facets of its complex eye with the aid of a neural model essentially of the McCulloch-Pitts type. The responses were sufficiently simple so that they can be fully analyzed; yet they contain sufficient conditionality to necessitate an apparatus more complex than a simple aggregate of reflex pathways.

Specifically the beetle responds with rotations of its body to various patterns of stimulus incidence, depending on (1) the order of stimulation of contiguous facets, (2) the relative intensity of the successive stimuli, and (3) the time interval between the presentations. To account for all these aspects of behavior (remarkably consistent), Hassenstein and Reichhardt have postulated the simplest conceivable arrangement of automaton units, of which only a few are required to serve each pair of facets. As a consequence of this arrangement the prediction is made and verified that stimuli impinging consecutively upon facets separated spatially by more than one facet do not interact with each other. By and large the model is mainly an explanatory one; that is, its theoretical significance is confined to a schematization of neural elements to account in the simplest way for observed behavior. The model thus serves as a *possible* solution to a black box problem.

Naturally one expects rather more from a model. If, for example, the postulated arrangements were identified anatomically or at least indirectly by further consequences not observed in the preliminary investigations, the theoretical force of the model would have been greatly enhanced. On the other hand, the conceptual value of automaton models is not to be underestimated. It is instructive to note how “much” can be done with only a few “neurons.” “Much” is put in quotes advisedly. Richness of conditionality of response is to be distinguished from ordinary complexity of response. A response may be marvelously complex in the sense of having many components and yet not necessitate any complicated neurological mechanism. This would be so if each step in the sequence were rigidly determined by the preceding step. To be sure, neural connections would be required to link the steps sequentially. But there would be no need of information-processing and decision-making units. It is the *conditionality* of behavior that necessitates complex automation, behavior described in terms of "if so, then so, unless so, in which case so, provided this or that but not both . . ." and so on. The few hundred neurons of the ant must be sufficient to provide it with all the conditionality at its disposal. This relatively small number reflects the circumstances that, although the behavior of the ant may seem quite complex, the conditionality of its behavior patterns must be rather small compared to that of animals with enormously larger numbers of neurons.

Mathematical Theories of Neural Nets

It appears, therefore, that the weakest link in the application of automaton theory to the anatomy and physiology of the brain is the specificity of automaton models. True, for any pattern of behavior of any prescribed conditionality, an automaton to simulate it can be in principle constructed. But if the model is simply a translation from logical propositions to networks of relays (as it is in the McCulloch-Pitts model), the loss of a single unit may radically change the entire behavior pattern of the automaton.

It is inconceivable that such sensitivity to single units (neurons in this case if the analogy applies) should characterize the living brain. We are constantly impressed by the plasticity and adaptability of living behavior. Specific failures traceable to specific excisions are still exceptions rather than the rule. On the other hand, building in sufficient alternative connections to forestall every conceivable specific failure would probably necessitate more neurons than are available in the largest brains.

Another feature that distinguishes living behavior from that of precisely constructed automata is the “approximate” character of the former. Actions of living organisms are not mathematically precise; nor are they necessarily the most direct and efficient. They are “adequate,” with wide error margins. Moreover they are often recognized as responses to “fuzzy” stimuli. An object is recognized by a higher animal as “itself” from different visual angles, in different orientations, and at different distances, a circumstance emphasized in Gestalt psychology. These synthesizing and abstracting functions of the nervous system cannot be accounted for by assuming simple one-to-one correspondences between elementary stimuli impinging on specific elements and determinate responses of the latter.

Attempts to simulate Gestalt phenomena are reflected in the construction of networks of elements designed to recognize *patterns,* that is, gross features of events regardless of perturbations. Examples of this approach are found in the work of Rosenblatt, D. Rutovitz, R. Narasimhan, and many others. Work along these lines is clearly inspired by the ambitions of automation technology. One can well imagine a typist-automaton that takes oral dictation; that is, is able to recognize words regardless of accent, vocal characteristics, or speech peculiarities of the person dictating. Also the theoretical spinoffs of these investigations may be considerable. Constructing pattern-recognizing automata may suggest ideas about how living organisms synthesize information carried by impinging stimuli.

The immense plasticity of living behavior has led several workers concerned with the theory of the nervous system to attempt to construct "probabilistic" (statistical, stochastic) models. Here connections or stimulus-response relations are not specified, but only their probabilities. Experiences of the organism (learning, metabolic changes, and so forth) are supposed to operate on these probabilities. From these probabilities one infers only gross aspects of behavior, not its details, and the variability of behavior—its continuous rather than discrete character—can be attributed to statistical fluctuations in the "functional structure" of the system. To cite an analogy, the general outline of a fountain persists, but it is not rigid, nor does it depend on the path of each individual water drop.

A convenient starting point of a probabilistic model of a nervous system is a "random net," formally defined as a collection of nodes (neurons) among which the synaptic connections are indicated only as probabilities. Statistical computations then give the gross connectivity characteristics of such a net; for example, the expected number of paths between an arbitrary pair of neurons, the expected number of neurons so many synaptic connections removed from each neuron, and so forth. Given such gross statistical features and certain assumed laws of synaptic transmission, the activity of such a net, resulting from some initial input, can also be calculated. For example, given the probability distribution of the number of axons emanating from each neuron and the probability distribution of their targets, the “critical input” can be established, one that if exceeded results in the spread of excitation through the net and if not exceeded results in a dying away of the initial excitation.-

A physical demonstration of a systematic, even systematically modifiable, behavior of a servomechanism with a randomly connected “nervous system” is provided by the “homeostat,” which illustrates the so-called principle of *ultrastability*. The stability of a servomechanism depends, of course, on its connections. If the connections of a thermostat, for example, were reversed, it would become unstable: rising temperature would result in even more heat from the furnace until something would “give.” But if a servomechanism switched its connections whenever some variable exceeded a certain limit, it would have ultrastability. In the case of a thermostat we could initially connect the leads randomly. If we happened to make it unstable, the rise of temperature would switch the connections and make it stable, after which the connections would no longer be switched, because a critical temperature would not be exceeded. In the homeostat, whenever certain voltages are exceeded, the connections are randomly shuffled until stability is achieved. This property, besides insuring stability, even enables the homeostat to exhibit simple learning behavior. If a certain response pattern is "punished" (by increasing voltages beyond the tolerated limits), connections will be switched until the right ones for the situation are found. The principle of learning thus exhibited is that of random search and fixation on the correct response.

The theory of probabilistic automata also underlies much of the work of pattern recognition.

In short, concepts derived from cybernetic technology have been a rich source of ideas in theories of neural structure and function, which, it is hoped, will strengthen the still tenuous links between physiology and theories of mental phenomena.

Mathematical Linguistics and Psycholinguistics

Let us now see how a computer would solve the above-mentioned logical problem, involving bandits, shrimps, and gophers. The "givens" of the problem must first be stripped of all semantic meaning (which only interferes with the reasoning). Then it can be presented in a language the computer can “understand,” the language of two-valued symbolic logic—essentially a branch of mathematics where the variables can assume only either of two values, 0 (representing “false”) and 1 representing “true”). The variables, symbolized by letters, stand for propositions. For instance, b stand for “bandits drink beer”; s for “the sun is shining”; m_i for the “moon is in the i-th quarter”; g for “gophers go skating.” The denials of the propositions are symbolized by corresponding letters with bars over them. For instance, $\bar{b}$ stands for “bandits don’t drink beer.” Besides the symbols representing propositions, the language of symbolic logic contains symbols representing relations among propositions. These are “$\wedge$,” meaning “and”; “$\vee$,” meaning “and/or” $\rightarrow$ meaning “implies”; and parentheses for punctuating sentences. We can now represent the entire information given in the problem by the following “sentences” written in the language of symbolic logic:

$$\bar{b} \rightarrow s \wedge m_1$$

$$s \rightarrow (w \rightarrow m_2 \vee m_3)$$

$$b \wedge w \rightarrow g\ m_4$$

The computer is programmed to perform certain operations on the symbols in accordance with specified rules. The result of these operations leads to the following sentence:

$$g \wedge \bar{s} \wedge \bar{w} \rightarrow m_4$$

which, retranslated into English, says "When gophers go skating while the sun is not shining, and the shrimps do not whistle, the moon must be in the fourth quarter," the required solution.

We can say, therefore, that to solve the problem the computer must be presented with it in a language it "understands." The "grammar" of that language (in this case the rules of operation of symbolic logic) is built into the computer, and this is what we mean by saying that the computer "understands" it.

One of the problems attacked by computer technology was that of automatic translation. Automatic translation would be simple if sentences could be translated from one language to another word by word. For then the only "rules" that would have to be programmed would be those that link each word in one language with its equivalent in another. As is well known, however, the problem is vastly complicated, not only by the fact that most words have more than one meaning but also by the fact that grammars of even closely related languages are different. On the other hand, substituting whole sentences will not do, since the number of possible sentences is potentially infinite. (It is safe to assume that the sentence you are now reading has never been spoken or written before.) Thus, the problem of automatic translation is that of giving a complete description of a grammar of a natural language in a language accessible to a computer. The immense difficulty of this task has now been realized.

Although automatic translation still seems to be a thing of the distant future, the “theoretical spin-offs” of the associated problems have been considerable. The attention of linguists (who are only incidentally or not at all interested in automatic translation) has been turned to one of the most challenging and, possibly, one of the most important problems of human psychology: to describe rigorously (not intuitively) the internalized grammatical rules that enable a human being, only a few years after birth, to produce and comprehend a practically unlimited number of sentences in his native language.” “Comprehension” in this context means much more than associating words with their referents; for language is much more than assigning labels to objects or situations.

Mathematical linguistics is, in part, concerned with the construction of rigorous theories of grammar. “Mathematical” in this context does not mean “quantitative,” as it does in classical physical science. Here, the relevant branches of mathematics (for example, set theory, symbolic logic, the theory of semigroups, etc.) deal not with measurable quantities but only with rigorous rules of symbolic transformations. The term “mathematical” is justified in view of the definition of mathematics as a contentless language of rigorous description and deduction. To put it another way, mathematical linguistics is concerned with the abstract relational “framework” of language, which determines meaning by “shaping” the content that is poured into it.

Another mathematical approach to language behavior is via statistical linguistics. The verbal output of an individual, or of a population of individuals such as a speech community, can be viewed as a vast number of minute sequentially produced units, for example, words selected from the lexicon of a language. Because of their large size, these collections exhibit certain statistical regularities. The smaller the units, the less are the statistical characteristics of these large samples dependent on content. For example, in large samples of printed English the relative frequencies of the letters of the alphabet are very nearly the same, regardless of source. The frequencies of larger units (for example, words and phrases) will, of course, be more dependent on the source or content. Nevertheless, certain statistical features common to all large corpuses can be abstracted also on these levels. G. K. Zipf particularly stressed the repeated observation of the following relation. Let the different words in a large corpus (a sample of verbal output) be ordered in the order of the frequency of their occurrence, so that rank 1 ($r = 1$) is assigned to the most frequently occurring word, rank 2 ($r = 2$) to the next most frequently occurring word, etc. To each rank corresponds the actual frequency of occurrence in that corpus, denoted by f. Then in all large corpuses the product $f \times r^{\gamma}$ is approximately constant, where γ is a number close to 1 and, in almost all cases, somewhat larger than 1. This rank-frequency relation (in other contexts the rank-size relation) was observed in a great many widely disparate situations and was attributed by Zipf to an underlying universal law, which he called the principle of least effort. Zipf's justification of the law and its consequences was often extremely vague and cannot be considered as a significant theoretical contribution. Nevertheless, the basic idea—that of examining the "statistical profile" of verbal outputs—has remained fruitful. The point is that these statistical profiles are determined by certain parameters (indices) and can serve as a basis of objective comparisons. Thus Zipf noted that the verbal output of schizophrenics is characterized by unusually large values of γ, the characteristic parameter of the rank-frequency relation.

Comparison of statistical profiles involving more than just rank-frequency relations has been used in determining the authorship of texts. In fact, there are cases on record where disputes concerning the authorship of texts have been decided by such comparisons. The validity of these methods depends on the circumstance that, although an individual may well exercise voluntary control over detailed actions, *in the large* his patterns of behavior, including his verbal outputs, are much more determined by habits, predispositions, and the like. Thus the statistical profile of an individual's verbal output "reveals" his identity in the same way as his handwriting or the spectral characteristics of his voice. In a way the verbal output is a “secretion”; therefore, it seems reasonable to develop methods of analyzing this secretion parallel to those developed in scientific medicine for analyzing physical secretions. The implications for psychiatry are obvious.

Content analysis is essentially an extension of statistical methods to include the “semantic” features of a verbal output. Part of its task is the development of coding techniques, which map “meanings” on objectively identifiable units. These techniques require considerable competence in the subject matter of the verbal output undergoing analysis. However, after the coding procedure has been designed and the “content” translated into a statistical profile, analysis becomes entirely objective. For instance, the statistical profiles of two or more outputs or their trends over time can be compared in the same way as spectra of different light sources. In this way “hard” content analysis can be used to supplement the conclusions of “soft” content analysis, which depends on intuitive conjectures of the analyst (for example, interpretation of dreams, literary or musical criticism), and to put the theories of the latter to scientifically objective tests.

Examples of the application of content analysis, both hard and soft, ranging from analysis of international crises to shifts of emphasis in grade school readers, can be found in Gerbner, *et al.*

In the “semantic differential,” an instrument based on factor analysis techniques, the object is to construct the “semantic space” of a subject or of a population of subjects. The theory is based on the observation that a great many adjectives can be characterized by the connotations they evoke on three principal axes: a value axis, along the good-bad scale; a potency axis, along the big-little or strong-weak scale; an activity axis, along the active-passive scale. Moreover, a great many other words, especially those with strong emotional overtones, can also be so characterized. Thus, from a subject’s associative responses to a set of “concepts,” a (connotative) “semantic space” can be constructed, in which each concept appears in a definite position, determined by its three coordinates on the three axes. The semantic differential has been used in comparative studies of such semantic spaces characterizing different individuals, populations of individuals of different cultural backgrounds, and the same individuals at different times. For instance, of particular interest to psychiatrists may be a study, undertaken by Osgood, Suci, and Tannenbaum, involving a comparison of the semantic spaces associated with the three components of a “split personality” (“The Three Faces of Eve”).

Exploration of Ideas: Opportunities and Dangers

Progress in science depends essentially on successful generalizations that unite apparently disparate phenomena into unified theoretical schemes. The evolution of physical science illustrates this process most clearly. The law of conservation of energy, first established in classical mechanics, was later extended to unite mechanics with thermodynamics. Electrical and magnetic phenomena were united in electrodynamics and extended to include all forms of radiant energy. Statistical mechanics revealed the deep connection between information and entropy via the mathematical expressions of the “amount of order” (or disorder). There are also dangers lurking behind attempted generalizations guided by metaphorical instead of rigorous mathematical analogizing. Every model is, of course, an analogy. What makes a model heuristically useful is its conception as a point of departure rather than of arrival. Unfortunately the richly suggestive ideas of mathematicized theories are often used as explanatory props rather than as raw material for constructing testable hypotheses. This is probably inevitable as long as in many lines of inquiry “theory” continues to be understood as a collection of mental images or figures of speech, which, it is somehow felt, harmonize with intuitive feelings of what constitutes an “explanation.” To take an example at random, Freud’s “hydraulic” model of psychodynamics is a “theory” in that sense. In essence many sociological theories are collections of definitions, that is, invitations to organize experience in a particular way. Although this is not the way the term “theory” is used in natural science, it would be rash to consider the construction of such theories altogether useless. After all, the organization of thought along certain lines is often a prerequisite of any progress toward insight. Metaphorical theorizing is not of itself necessarily misleading. It can become misleading when hazy notions are coupled with precise-sounding terminology. The use of the latter may give an impression (to the theorizer himself, as well as to his audience) that precision has been achieved when, in fact, concepts that are precise in proper contexts have been muddled by metaphorical transformations of meanings.

The Homeostasis Metaphor

The concept of homeostasis was formulated in the context of physiological regulation by W. B. Cannon. In such regulations homeostatic mechanisms operate so as to keep certain variables (concentrations, pressures, temperatures) of the organism’s “internal environment” within certain limits of tolerance. Homeostasis is also a central principle of cybernetics, since the regulation activity of servomechanisms can be described in the same terms as the regulation of the physiological processes. In general, homeostasis operates on the principle of feedback. In negative feedback the “restoring force” is always opposite to the error, so that the variable in question tends to some equilibrium value. In positive feedback the error is self-enhancing, so that either a variable increases without bound or oscillations of ever increasing amplitude result. Homeostasis in system engineering is attained by a proper arrangement of feedback loops. (Positive feedback loops also have their place, where it is required that the system pass quickly from one steady state to another.)

As long as the variables represent real measurable quantities and the network of influences among them is actually observed or specifically assumed, one may speak of homeostasis in a great variety of situations; for instance, in engineering, where the variables are voltages or tensions, or water levels; in physiology, as described above; in ecology, where the variables may be populations, gene frequencies, etc., and where the "forces" are statistical trends, which, of course, do not have the physical characteristics of forces but have similar mathematical properties. One can, in the same spirit, speak of homeostasis in economics, where the variables are prices, interest rates, trade volumes, etc. In these instances the concept of homeostasis is, indeed, a unifying principle of several widely disparate areas. It is a "general systems" principle par excellence.

When the operational meaning of the variables is lost sight of and replaced by intuitive notions, the terminology associated with homeostasis becomes at best metaphorical. The models become paraphrases of impressions and cease to have theoretical significance, as this significance is understood in "hard" science. The various models of behavior, personality, and society, couched in terms borrowed from theories of homeostasis, give the illusory impression that powerful and rigorous methods are being applied to the study of man.[2]

To speak of the defense mechanisms of the individual, or of the mutual impact of political systems or cultures, in the language of homeostasis may be subjectively enlightening, but there is no way of knowing whether such enlightenment is any different from the sort experienced by philosophers who, in the days before the advent of physical science (and often afterward), "explained" the physical, the biological, and the social universes by picturing them as manifestations of metaphysical laws that reflect no more than grandiose verbiage.

The Information-Entropy Metaphor

There is a link between information theory and thermodynamics that carries a tantalizing suggestion of being of prime importance for theoretical biology, along with all the dangers of speculative promiscuity. The formal resemblance of the mathematical expression for the average amount of information per signal to the expression for the entropy of a physical system, as calculated in statistical mechanics, was noted by N. Wiener and C. E. Shannon, who laid the foundations of cybernetics and information theory, respectively. The definition of entropy is highly technical: very roughly speaking, entropy can be taken as a measure of disorder present in a system. It is this disorder in the motions of molecules that makes it impossible to convert heat energy *fully* into mechanical energy without other changes accompanying the process. Stated in another way (as the famous Second Law of Thermodynamics), in a system completely isolated from its environment, the total entropy can never decrease; it keeps increasing until the system is in thermodynamic equilibrium. When this state is attained the heat energy of the system can no longer be converted into "useful work."

When thermodynamic considerations first began to be applied to biological systems, some biologists forgot the important qualification "isolated from its environment" and argued that living systems violated the Second Law (an argument for vitalism!) since such systems tended, at least in their development, toward "greater organization," rather than toward chaos as the Second Law demands. Since no living system is isolated from its environment, the argument rested on a non sequitur. At any rate E. Schroedinger pointed out that life must "feed on negative entropy," by which is meant simply that organisms must ingest substances rich in "free energy," in other terms, low in entropy.[3] In metabolism this free energy becomes "degraded"; that is, entropy increases, and this surplus of entropy, dumped in excretion upon the outside world, "pays" for the decreases in entropy (increased organization) that the organism effects within itself. In this way entropy-lowering life processes can go on without the Second Law being violated in the end result.

So far the argument has been presented in thermodynamic terms, and its relevance for information theory is far from evident. A clear connection can be found, however, in an early paper by L. Szilard. Szilard analyzed the operation performed by Maxwell’s demon, a hypothetical creature posited by James Clerk Maxwell in 1869. The demon is supposed to be able to “see” the molecules of an enclosed volume of gas, mechanically and thermally isolated from the environment (an isolated system). By sorting them he can “increase the order” in the system and so lower its entropy in apparent violation of the Second Law. By considering the simplest possible system of this sort, consisting of a single molecule, Szilard was able to analyze completely the nature of the demon’s intervention. He showed that if the Second Law does hold, the demon himself (being part of the isolated system) must suffer an increase in entropy that at least compensates for the decrease he effects in the rest of the system. This conclusion is simply a logical consequence of the assumption that the Second Law does hold. The remarkable feature of Szilard’s analysis is the exact quantitative relation between the “amount of information” that the demon must utilize in his operation and the resulting decrease of entropy. He showed that in utilizing one bit[4] of information the demon lowers the entropy of the system (excluding himself) by $k \log_e 2$ ergs per degree, where k, the so-called Boltzmann’s constant, is 1.37×10^{-16} ergs per degree. Thus a transformation factor connecting entropy and information was established, analogous to the transformation factor connecting a unit of work and a unit of heat, discovered almost a century earlier. In 1951 L. Brillouin was able to show that the demon must indeed suffer at least the prescribed increase of entropy, regardless of the method he uses in determining the position or the velocity of a molecule.

The implications of these theoretical results for events in biological systems *on the molecular level* are now being actively investigated. There the connection between information and entropy is quite clear: what appears in the language of gross thermodynamics as entropy (units: energy over temperature) appears in the statistical formulation (the mechanical basis of thermodynamics) as information (units: pure numbers, logarithms of probabilities). Trouble arises when results are extrapolated in attempts to apply the concepts to information in its vernacular meaning. It is taken for granted by many writers that the quantitative information measure can be applied to the *content* of communications, so that the number of bits in this chapter, for example can be stated with as much precision as the weight of the paper it is printed on. In a way this is true but irrelevant to the informative content of the chapter. Quantity of information is defined with reference to the *statistical* properties of the source from which the signals are chosen. To be sure, the “information” of any verbal output can be measured in this way by reference to the statistical distributions of its units—say, letters, or phonemes, or words—but only by disregarding the meaningful content of the corpus. Thus one bit of information is gained by someone who is told the outcome resulting from a toss of a fair coin; one bit was gained by Paul Revere when he saw two lights appear in the tower of the Old North Church. What the last-mentioned “bit of information” meant for the American Revolution is irrelevant from the point of view of information theory. Indeed, the amount of information conveyed by a meaningless scramble of randomly selected letters is actually greater than that conveyed by a meaningful text of the same length, because the random selection is subjected to fewer statistical constraints. From the point of view of telecommunication this makes sense, because it would take more channel capacity to transmit random combinations of signals at a given rate than statistically constrained combinations. From the point of view of the recipient, however, who considers that he gets information when he is “informed,” the statistical definition makes no sense.[5] Therefore, no operational meaning can be assigned to a statement such as “A has received so many bits of information and has thereby lowered his entropy (increased his internal order) by so many units.” The statement can acquire meaning only if it is shown just how A has utilized this information and how the decrease of entropy was compensated by an increase elsewhere. However, the seductive power of metaphors is great, as evidenced by an abundance of loose talk about the relation of “entropy” and “information” in human affairs. So far extrapolations of the information-entropy identity to regions where receiving information means being informed have dissolved into vague and, one suspects, sterile speculations.

Do Machines “Think”?

Simulation of “thought” by machines has raised some questions of philosophical and ethical import. Crudely put, the fundamental question is whether it is proper to ascribe “thought” to machines or, conversely, to picture man as a complex machine. Inevitably the posing of these questions is charged with affect. Answers in the affirmative seem to some to imply a denial of man’s humanity, while others see in the erasure of distinction between living and nonliving systems another step toward the unification of science and toward the abandonment of anthropomorphism—a continuation of a maturing process instigated first by the heliocentric theory and later by the theory of evolution.

It is possible to by-pass the emotional overtones of these questions by a careful distinction between different meanings of “thinking.” There are some things that information-processing machines can demonstrably do; for instance, solve logical and mathematical problems and exercise control over physical processes. At one time it was thought inconceivable that inanimate systems might be capable of performing these apparently “intelligent” tasks. There are also some things that presently existing information-processing machines cannot do. However, the limitations are more difficult to *spell out* than the achievements. The difficulty is that, once the limitations are specifically spelled out, ideas are suggested on how to overcome them. It has been said that a computer cannot compose a poem or a quartet. Promptly computers were programmed to compose “poems” and “quartets.” The objection that these products are not really works of art, *because* they have been programmed, can be met by a powerful challenge, namely, to distinguish the machine- made products from some contemporary examples of man-made ones.

Arguments to the effect that what goes on in computers is not “thought” because the processes are “preprogrammed” are not conclusive. The analogy between information processing by man and by machine rests on the assumption that the processes in man’s nervous system are *also* preprogrammed, namely, by the structure of the nervous system and its physiological state at a given moment. The admittedly vast difference in complexity between the two kinds of processes is not sufficient reason for dismissing the analogy. Nor are arguments about “free will,” supposedly possessed by man but not by the machine, relevant to the issue, if what is wanted is evidence to resolve it one way or another. Our conviction of having “free will” stems from a metaphysical (or religious) position or is induced by introspection. Metaphysical positions are impermeable to evidence. Introspection is accessible only to the introspecting subject; hence there is no way of knowing whether the machine does or does not “introspect.”

There remains only the *ethical* basis for distinguishing between man and machine. The real meaning of the questions “Do machines think?” or “Are men machines?” is embodied in another question: “Shall our attitudes toward men and machines be similar or different?” This question is obviously valueoriented and should be frankly posed and recognized as such. It has substantial ethical import in a civilization where the lives of human beings are to a large extent organized by work in the services of machines. Comparing men to machines does deny man’s humanity in the sense of turning attention to man as an instrument: “Machines can in principle do everything men can do.” This sort of comparison turns attention away from man’s *intrinsic* worth, which, unlike his instrumental worth, resides not in what he can do but in what he is, namely, man. The fact that we communicate with other human beings *without* knowing analytically how this is done; the fact that we ascribe consciousness to other human beings, not on the basis of “evidence” (we have no access to another being’s consciousness), but intuitively, by identifying with them, puts relations between human beings into a unique category. Insistence on the uniqueness of these relations is a manifestation of certain values, and the adherence to these values is the only significant meaning underlying the refusal to identify human thought with automated information processing.

Conclusions

Physical science, with its formidable methodological machinery in which controlled experiment, induction, and mathematical deduction are meshed, has nourished the life sciences almost from their inception. As the methods of physical science are becoming extended to areas where not matter and energy but organization and information processing are of central interest, the basis for integration becomes even firmer and a hope emerges of extending the integration to include those aspects of the life process that have been considered absolutely *sui generis,* aspects involving “psychical” rather than physical events. Such an integration would lead to the final dissolution of the mind-body duality in the context of scientific investigations.

Those who attempt to realize such integration borrow from these latest developments of physical science their methods, their ideas, and their language. The conditions for a fruitful extension of method are explicit. Mathematicized science deals with exactly specifiable structural relations. Whenever such structural relations can be unambiguously defined (in terms of observations, operations, or mathematical manipulations), the method of mathematical deduction can serve as a powerful tool of theory construction. Discrepancies between theory and observation serve to initiate the cyclical process of hypothesis-deduction-verification-new hypothesis. Therefore, initial accuracy of assumed relations is not essential; only the unambiguous specification of variables and relations is a prerequisite for extending the mathematical method to new areas.

When such specification cannot be made, the heuristic value of the ideas immanent in mathematicized science may still remain. Therefore, rather than attempt explicit mathematical modeling, some behavioral scientists seek to adapt the general ideas emerging from mathematical analysis to theories of behavior. The value of such adaptations is an open question. They may be enlightening or they may be misleading. To illustrate take the so- called uncertainty principle of atomic physics. The principle sets limits to the precision with which the position and the momentum of a particle can be simultaneously measured. As such it is a principle of theoretical physics and nothing else. However, the principle has philosophical implications. One implication has to do with the failure of strict causality on certain submicroscopic levels of events. This had led to contentions that the uncertainty principle “proves” the existence of “free will,” largely a play on words, “free will” being the verbal antithesis of “strict causality.” The irrelevance of such conclusions to science need hardly be pointed out. However, there is another implication suggested by the uncertainty principle, namely, that events may be affected by being observed. These effects have been long felt to operate in psychology. As stated in quantum mechanics, the uncertainty principle is exact and explicit. It singles out pairs of so- called complementary quantities, position and momentum being one such pair, energy and time another. A specified amount of precision in determining one member of the pair introduces a specified minimum amount of uncertainty in the other. The principle could be of genuine heuristic value in psychology if analogous complementary pairs were sought and discovered. For an example of a rigorous treatment of the uncertainty principle in the context of signal detection, see C. W. Helstrom.

In short, an idea is scientifically fruitful if it serves to stimulate thinking that leads to discoveries. Such thinking may well start with consideration of analogies, provided they are not merely suggested by metaphorical use of language but are rooted in some aspect of reality.

The line between fruitful and sterile ideas is hard to draw. Some wild speculations of today may contain the germs of fundamental theoretical formulations of tomorrow. A typical sample of rather free-wheeling theorizing about the wider implications of cybernetics, ranging from a comparison of human and automated chess playing to an analysis of conscience and liberty, is contained in a volume published to commemorate Norbert Wiener's seventieth birthday in November 1964.- Wiener died in March of that year, and the book came out as a memorial volume.

On occasion some theorizers have simply borrowed the *language* of modern developments in the exact sciences. The most serious dangers of speculative promiscuity are rooted in this practice. Neologisms being, for some reason, more distasteful to physical scientists than to others, physical scientists tend to adapt common words to highly technical usage. The everyday connotations of these terms remain. For instance, "information," "feedback," "stability," "redundancy," "noise," etc., terms common in system engineering and cybernetics, are not entirely unrelated to the meanings of the corresponding common usage words. But common usage words are also heavy usage words; that is to say, they are rich in marginal, metaphorical meanings. It is here that the tendency to "theorize" by juggling words in their various contexts is greatest among those who are impressed but not disciplined by the spirit of the exact sciences. Whereas in information theory "noise" is defined as precisely as "heat" is in physics, in psychological speculations spiked with cybernetic terminology, "noise" often assumes a range of meanings stretching from the noise of traffic to the disturbances in the mental processes of a psychiatric patient.

The high prestige of science in a society dominated by technology can and has been utilized by quacks and charlatans to exploit the gullible. A notorious result of this practice was the "dianetics" fad that swept the United States in the 1950's. Dianetics was an amalgam of vulgarized notions lifted from psychoanalysis and a mumbo jumbo potpourri of terms common in cybernetics. It was offered as a sure-fire, cheap method of psychotherapy guaranteed not only to cure mental and emotional disorders but also to raise the intelligence of customers to genius levels. No better example is needed to demonstrate the seductiveness of scientific terminology in the role of word magic.

"Scientism" (simulation of scientific rigor by the misuse of technical terms) is especially harmful in psychiatry, which is concerned with events and conditions that elude precise objective analysis. By creating the impression that a new, powerful arsenal of concepts is being applied, scientism detracts from the important aspect of psychiatry as an art, where intuitive insights and empathetic understanding continue to be indispensable. The most important contributions of mathematics and cybernetics are not so much to the practice of the healing art as to the scientific infrastructure that underlies our understanding of the life processes, including mental activity and behavior.

Behavioral scientists and psychiatrists who feel that mathematics, cybernetics, and allied subjects have something of value to contribute to their area of concern will do well to draw from those fields of knowledge something of their discipline as well as inspiration. To the extent that this is done, the transplant of ideas may bear fruit. After all, the ideas of cybernetics are themselves transplants. Their origin was in biological science. They are the "organismic" ideas that had been banished from classical mechanics and whose absence in physical science has been so eloquently deplored by A. N. Whitehead." The early exclusion of "organismic" concepts from physical science was justified. The soil of early physics could not have nourished these concepts. Only when this soil was sufficiently enriched could the seeds sprout. Cybernetics, rooted in the physical sciences, is the result. Now the young shoots may be ready for transplantation back to the biological and behavioral sciences where they belong. The only question is whether the present soil in those areas can nourish them properly.

Bibliography

Afanasyeva, L. G., and Petunin, Yu. I., "Teoreticheskiy Analog Obrabotki Impulsnoy Informatsii Neyronom" ("Theoretical Analysis of the Processing of Impulse Information by a Neuron"), *Kihernetika,* 3: 74-81, 1971.

Allanson, J. T., "Some Properties of a Randomly Connected Neural Network," in Cherry, C. (Ed.), *Information Theory,* Academic Press, New York, 1956.

Ashby, W. R., *Design for a Brain,* John Wiley, New York, 1952.

_____, *Introduction to Cybernetics,* John Wiley, New York, 1956.

Bar-Hilliel, Y., "Semantic Information and Its Measures," *Tr. 10th Conference on Cybernetics,* Macy Foundation, New York, 1955.

Brazier, M. A. B., "How Can Models from Information Theory Be Used in Neurophysiology," in Fields, W. S., and Abbott, W. (Eds.), *Information Storage and Neural Control,* Charles C Thomas, Springfield, Ill., 1963.

Brillouin, L., "Maxwell's Demon Cannot Operate: Information and Entropy, I," *J. Appl. Phys.,* 22:334, 1951.

Cannon, W. B., *The Wisdom of the Body,* Norton, New York, 1939.

Chomsky, N., *Syntactic Structures,* Mouton, The Hague, 1957.

Echols, H., "Genetic Control of Protein Synthesis," in Fields, W. S., and Abbott, W. (Eds.), *Information Storage and Neural Control,* Charles C Thomas, Springfield, Ill. 1963.

Gerard, R. W., "Instruments and Men," *Instruments, 18:10,* 1945.

Gerbner, G., Holsti, O. R., Krippendorff, K., Paisley, W. J., and Stone, P. J. (Eds.), *The Analysis of Communication Content,* John Wiley, New York, 1969.

Hassensteix, B., and Reichhardt, W., "Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers Chlorophanus," Z*tschr. Naturforsch., 11b:* 513, 1956.

Helstrom, C. W., "Detection Theory and Quantum Mechanics (I)," *Information and Control, 10:2*54; (II), *Information and Control,* 13.156, 1968.

Hiltz, F. F., "Analog Computer Simulation of a Neural Element," *Tr. Inst. Radio Engrs.,* BME-9, 1:12-20, 1962.

Hyman, R., "Stimulus Information as a Determinant of Reaction Time," *J. Exper. Psychol.,* 45:188, 1953.

John, E. R., "Neural Mechanisms of Decision Making," in Fields, W. S., and Abbott, W. (Eds.), *Information Storage and Neural Control,* Charles C Thomas, Springfield, 1963.

Kornblum, S., "Sequential Determinants of Information Processing in Serial and Discrete Choice Reaction Time," *Psychol. Rev.,* 76:113, 1969.

McCulloch, W. S., and Pitts, W., "A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bull. Math. Biophys.,* 5:114, 1943.

Mazzetti, P., Montalenti, G., and Soardo, P., "Experimental Construction of an Element of a Thinking Machine," *Kybernetik,* 1:170, 1962.

Medvedev, Z. A., "Oshibki Reproduktsii Nukleinovvkh Kislot, Belkov i Ikh Biologicheskoye Znaehenie" ("Errors of Reproduction of Nucleic Acids and Nucleoproteins and Their Biological Significance"), *Prohlemy Kibernetiki,* 9:241, 1963.

Meyerhof, O., "Thermodynamik des Lebensprozesses," *Handbuck der Physik,* 9: 238, 1926.

Narasimhan, R., "Labeling Schemata and Syntactic Description of Pictures," *Information and Control,* 7:151, 1964.

Nasu, M., and Honda, H., “Fuzzy Events Realized by Finite Probabilistic Automata,” *Information and Control,* 12:284, 1968.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H., *The Measurement of Meaning,* University of Illinois Press, Urbana, 1957.

Rapoport, A., “Ignition Phenomena in Random Nets,” *Bull. Math. Biophys., 14:*35, 1952.

Rosenblatt, F., “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain,” *Psychol. Rev.,* 65.386, 1958.

Rutovitz, D., “Pattern Recognition,” *J. Roy. Statistical Soc.,* Ser. A., 129:504, 1966.

Schroedinger, E., *What is Life?* Macmillan, New York, 1945.

Shannon, C. E., “A Symbolic Analysis of Relay and Switching Circuits,” *Tr. Amer. Inst. Elec. Engrs.,* 57:713, 1938.

_____, and Weaver, W., *The Mathematical Theory of Communication,* University of Illinois Press, Urbana, 1949.

Sperry, R. W., “Orderly Function with Disordered Structure,” in Von Foerster, H., and Zopf, G. W. (Eds.), *Principles of Self- Organization,* Pergamon, New York, 1962.

Stark, L., and Cornsweet, T. N., “Testing a Servoanalytic Hypothesis for Pupil Oscillations,” *Science,* 127:588, 1958.

Steinbuch, K., *Automat und Mensch,* Springer Verlag, Berlin, 1961.

Szilard, L., “Ueber die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen,” *Ztschr. Physik.,* 53:840, 1929.

Toch, H., and Hastorf, A. H., “Homeostasis in Psychology,” *Psychiatry, 18:81,* 1955.

Von Neumann, J., “The General and Logical Theory of Automata,” in Jeffress, L. A. (Ed.), *Cerebral Mechanisms in Behavior: The Hixon Symposium,* John Wiley, New York, 1955.

Walter, W. G., *The Living Brain,* Norton, New York, 1953.

Whitehead, A. N., *Science and the Modern World,* Pelican Books, New York, 1948.

Wiener, N., *Cybernetics,* John Wiley, New York, 1948.

_____, *God and Golem Inc.,* M.I.T. Press, Cambridge, 1964.

_____, and Schade, J. P. (Eds.), *Progress in Biocybernetics,* Vol. 2, Elsevier, Amsterdam, 1965.

Zipf, G. K., *Human Behavior and the Principles of Least Effort,* Addison-Wesley, Cambridge, 1949.

Notes

[1] This work was anticipated in the context of electrical switching circuits by C. E. Shannon. Subsequently J. Von Neumann gave an extensive and lucid exposition of the theory in a general context.

[2] For an extensive critique of the use of the homeostasis concept in psychology, see, for example, H. Toch and A. H. Hastorf

[3] Schroedinger's fortunate phase has become a byword and has made facile speculation in biological thermodynamics fashionable. The basic idea, however, was already formulated 20 years earlier.

[4] The "hit" is a unit of information, the amount conveyed in a decision between two equally probable alternatives.

[5] In strictly limited contexts some progress in constructing a theory of semantic information has been made.