CHAPTER

5

A Simple Cybernetic Model

Mechanic, organismic, and historical models were based, substantially, on experiences and operations known before 1850, even though many of their implications were worked out more fully only later. A major change in this situation began in the 1940's. Its basis was in the new development in communications engineering, with its extensive use of self-monitoring, self-controlling, and self-steering automatic processes. By making equipment that fulfills the functions of communication, organization, and control, significant opportunities were gained for a clearer understanding of the functions themselves.

These new developments in science and engineering were the beneficiaries of long-standing developments in social organization. Communication was social before it became elaborately technological. There were established routes for messages before the first telegraph lines. In the nineteenth century, factories and railroads required accurate coordination of complex sequences of human actions—a requirement that became central in the assembly-line methods and flow charts of modern mass production. The same age saw the rise of general staffs, and of intelligence organizations for diplomatic as well as for military purposes. These staffs and organizations, just as the modern large-scale industrial research laboratory itself, represent in a very real sense assembly lines of information, assembly lines of thoughts. Just as the division of manual

labor between human hands preceded the division of labor between human hands and power-driven mechanisms, so the increasing division of intellectual labor between different human minds preceded today's divisions of labor between human minds and an ever-growing array of electronic or other communications, calculating, and control equipment.

What have the new machines of communication and control to offer for a further understanding of historical and social processes? For thousands of years, the operations of communication and control were largely carried on inside the nerve systems of human bodies. They were inaccessible to direct observation or analysis. They could be neither taken apart nor reassembled. In the new electronic machines of communication and control, messages or control operations can be taken apart, studied step by step, and recombined into more efficient patterns.

THE VIEWPOINT OF CYBERNETICS

The science of communication and control, which has been derived from this technology and which Norbert Wiener has called "cybernetics," is therefore a new science about an old subject. In investigating the old subject of communication and control, it uses the facilities of modern technology to map out step by step the sequence of actual events involved.[1]

Cybernetics, the systematic study of communication and control in organizations of all kinds, is a conceptual scheme on the "grand scale," in J. B. Conant's sense of the term.[2] Essentially, it represents a shift in the center of interest from drives to steering, and from instincts to systems of decisions, regulation, and control, including the noncyclical aspects of such systems. In its scope, it is comparable to Lavoisier's stress on quantitative chemistry, or to Darwin's concept of evolution. As to its performance and success, the future will have to tell, but it is perhaps safe to say that social science is already being influenced by the interests implicit in cybernetics at this time.

The fundamental viewpoint of cybernetics and its relevance to social science have been well expressed by Norbert Wiener:

> The existence of Social Science is based on the ability to treat a social group as an organization and not as an agglomeration. Communication is the cement that makes *organisations*. Communication alone enables a group to think together, to see together, and to act together. All sociology requires the understanding of communication.
>
> What is true for the unity of a group of people, is equally true for the individual integrity of each person. The various elements which make up each personality are in continual communication with each other and affect each other through control mechanisms which themselves have the nature of communication.
>
> Certain aspects of the theory of communication have been considered by the engineer. While human and social communication are extremely complicated in comparison to the existing patterns of machine communication, they are subject to the same grammar; and this grammar has received its highest technical development when applied to the simpler content of the machine.[3]

In other words, the viewpoint of cybernetics suggests that all organizations are alike in certain fundamental characteristics and that every organization is held together by communication. Communication is a process different from transportation on the one hand and from power engineering on the other. Transportation transmits physical objects such as liquids in pipelines, or boxes or passengers in trains or on escalators. Power engineering transmits quantities of electric energy. Communication engineering, by contrast, transmits neither tons of freight nor kilowatts of power. It transmits messages that contain quantities of information, and I shall say more about this concept of information later in this chapter. It is communication, that is, the ability to transmit messages and to react to them, that makes organizations; and it seems that this is true of organizations of living cells in the human body as well as of organizations of pieces of machinery in an electronic calculator, as well as of organizations of thinking human beings in social groups.[4] Finally, cybernetics suggests that steering or governing is one of the most interesting and significant processes in the world, and that a study of steering in self-steering machines, in biological organisms,

in human minds, and in societies will increase our understanding of problems in all these fields.

ANALOGIES AND CONVERGENT DEVELOPMENTS

Why should anyone think that this viewpoint represents a conceptual scheme and not a mere analogy? Actually, the meaning of the term *analogy* is often poorly understood. Analogy means limited structural correspondence. All mathematics is based on analogies, and so is a large part of every science. Darwin himself tells us that it was his perception of the analogy between Malthus' theory of human population and certain processes in the animal kingdom that led him to his theory of evolution. When scientists speak disparagingly of "mere analogies," they mean, more accurately, "false analogies" or "poor analogies." The test by which we discriminate between a false analogy and a good analogy consists in the extent of actual structural correspondence between the two systems from which the analogy is drawn.[5] How many and how significant are the instances in which the analogy holds good, and how numerous and how important are the instances in which it fails to work? These are the questions by which we test analogies and which serve to unmask the many false analogies which look plausible at first glance but fail completely after the early stages of the application. The test of a good analogy, conversely, is that it continues to be confirmed after we have penetrated more deeply into the subjects it purports to connect and that it becomes more fruitful of new ideas and of new investigations as we continue to apply it. Darwin's analogy with the work of Malthus was a good analogy in this sense, and so was Torricelli's analogy between the atmosphere and a "sea of air." It is suggested that cybernetics is currently proving itself a good analogy or conceptual scheme in a similar manner.

The rise of the viewpoint of communications in the present period has not been fortuitous. Rather, it has been the result of convergent developments in a whole series of different sciences. Among these trends is the development of mathematical and statistical methods

for the study of randomness and order, and thus of probability, leading to the mathematical theory of communication as developed by Norbert Wiener, Claude Shannon, and others. During the same decades the concept of homeostasis was developed by Claude Bernard, and later by Walter B. Cannon and Arturo Rosenblueth in physiology. This medical work found its parallel in the mathematical and empirical studies of control mechanisms, from Clark Maxwell's early paper on the governor in steam engines to the highly developed automatic control engineering of today. Problems of flow in various organizations were studied in production engineering, traffic engineering, city planning, and the design of telephone systems. Advances in the design of automatic switchboards eventually merged with the long-standing efforts to design effective calculating machines, from the early days of Leibniz and later of Charles Babbage to the analogue computor constructed by Vannevar Bush and the big digital computors of today. These advances in mathematics and the study of physical systems were paralleled by Ivan Pavlov's emphasis on the material nature of psychological processes, and on the discrete structure of the conditioned reflexes on which many of them were based. This emphasis was balanced by the rise of the school of *Gestalt* psychology led by Kurt Koffka and Wolfgang Köhler, emphasizing the importance of pattern and order, and the rise of the depth psychology of Sigmund Freud and his followers.

It is the experience of this new group of sciences that finds its reflection in some of the major ideas of cybernetics, such as the notion of the physical reality of patterns and of information and of the statistical nature of the latter, as well as the related notions of the physical nature of control processes, memory, and learning.

Taken together, the new experiences and notions promise to replace the classic analogues or models of mechanism, organism, and process, which so long have dominated so much of scientific thinking. All three of these models have long been felt to be inadequate. Mechanism and the equilibrium concept cannot represent growth and evolution. Organisms are incapable of both accurate analysis and internal rearrangement; and models of historical processes lacked inner structure and quantitative predictability.

In the place of these obsolescent models, we now have an array of self-controlling machines that react to their environment, as well as to the results of their own behavior; that store, process, and apply information; and that have, in some cases, a limited capacity to learn.

None of this is *thought* in the human sense of the word, as we find it in the behavior of individuals or groups, but it has significant parallels to it. Above all, the storage and treatment of information in machines, and its application to the control of the machines themselves, are taking place under conditions where every step can be traced distinctly and where every system can be taken apart for study and reassembled again. This is a research advantage that it would be neither easy nor entirely desirable to parallel in the case of human beings.

The test of the usefulness of this new science, as that of any science, must be its results. In the field of scentific theory it must offer new concepts rather than mere explanations. The analogies cybernetics may suggest between communication channels or control processes in machines, nerve systems, and human societies must in turn suggest new observations, experiments, or predictions that can be confirmed or refuted by the facts. They must be meaningful, that is, capable of being tested by practicable operations, and they should be fruitful, that is, lead to new operations and new concepts.

THE GENERAL CONCEPT OF A SELF-CONTROLLING SYSTEM

To the extent that we can demonstrate that such analogies exist, and that they are fruitful in research, we may derive from them a generalized concept of a *self-modifying communications network* or "*learning net.*" Such a "learning net" would be any system characterized by a relevant degree of organization, communication, and control, regardless of the particular processes by which its messages are transmitted and its functions carried out—whether by words between individuals in a social organization, or by nerve cells

and hormones in a living body, or by electric signals in an electronic device.[6]

How does a modern communications mechanism look and what concepts can be derived from it?

Let me refer here to a brief sketch I gave elsewhere:

> A modern radar tracking and computing device can "sense" an object in the air, interacting with its beam; it can "interpret" it as an airplane (and may be subject to error in this "perception"); it can apply records of past experience, which are stored within its network, and with the aid of these data from "memory" it can predict the probable location of the plane several seconds ahead in the future (being again potentially subject to error in its "recollections" as well as in its "guess," and to "disappointment," if its calculation of probability was correct, but if the airplane should take a less probable course); it can turn a battery of antiaircraft guns on the calculated spot and shoot down the airplane; and it can then "perceive," predict, and shoot down the next. If it should spot more than one airplane at the same time, it must become "infirm of purpose," or else decide ("make up its mind") which one to shoot down first. . . .
>
> Man made machines actually operating or designable today have devices which function as "sense organs," furnish "interpretations" of stimuli, perform acts of recognition, have "memory," "learn" from experience, carry out motor actions, are subject to conflicts and jamming, make decisions between conflicting alternatives, and follow operating rules of preference or "value" in distributing their "attention," giving preferred treatment to some messages over others, and making other decisions, or even conceivably overriding previous operating rules in the light of newly "learned" and "remembered" information.
>
> None of these devices approach the overall complexity of the human mind. While some of them excel it in specific fields (such as the mechanical or electronic calculators), they are not likely to approach its general range for a long time to come. But, as simplified models, they can aid our understanding of more complex mental and social processes, much as sixteenth century pumps were still far simpler than the human heart, but had become elaborate enough to aid Harvey in his understanding of the circulation of the blood.[7]

What are some of the notions and concepts that can be derived from this technology? Perhaps the most important is the notion of information.

THE CONCEPTS OF INFORMATION, MESSAGE, AND COMPLEMENTARITY

Power engineering transfers amounts of electric energy; *communications engineering transfers information.* It does not transfer events; it transfers *a patterned relationship between events.* When a spoken message is transferred through a sequence of mechanical vibrations of the air and of a membrane; thence through electric impulses in a wire; thence through electric processes in a broadcasting station and through radio waves; thence through electric and mechanical processes in a receiver and recorder to a set of grooves on the surface of a disk; and finally played and made audible to a listener—what has been transferred through this chain of processes, or channel of communication, is not matter, nor any one of the particular processes, nor any significant amount of energy, since relays and electronic tubes make the qualities of the signal independent from a considerable range of energy inputs. Rather it is *something* that has remained unchanged, invariant, over this whole sequence of processes.

The same principle applies to the sequence of processes from the distribution of light reflected from a rock to the distribution of black or white dots on a printing surface, or the distribution of electric "yes" or "no" impulses in picture telegraphy or television. What is transmitted here are neither light rays nor shadows, but information, the pattern of relationships between them.

In the second group of examples, we could describe the state of the rock in terms of the distribution of light and dark points on its surface. This would be a *state description* of the rock at a particular time. If we then take a picture of the rock, we could describe the state of the film after exposure in terms of the distribution of the dark grains of silver deposited on it and of the remaining clear spaces, that is, we should get another state description. Each of the two state descriptions would have been taken from a quite different object—a rock and a film—but a large part of these two state descriptions would be identical whether we compared them point by point or in mathematical terms. There would again be a great deal of identity

between these two descriptions and several others; such as the description of the distribution of black and white dots on the printing surface, or of the electric "yes" or "no" impulses in the television circuits, or the light and dark points on the television screen. The extent of the physical possibility to transfer and reproduce these patterns corresponds to the extent that there is "*something*" unchanging in all the relevant state descriptions of the physical processes by which this transmission is carried on. That "something" is *information—those aspects of the state descriptions of each physical process that all these processes had in common.*[3]

To the extent that the last state description in such a sequence differs from the first, information has been lost or distorted during its passage through the channel. From the amount of information transmitted as against the information lost, we may derive a measure of the *efficiency* of a channel, as well as of the relative efficiency or *complementarity* of any parts or states of the channel in relation to the others.

These patterns of information can be measured in quantitative terms, described in mathematical language, analyzed by science, and transmitted or processed on a practical industrial scale.

This development is significant for wide fields of natural and social sciences. Information is indeed "such stuff as dreams are made on." Yet it can be transmitted, recorded, analyzed, and measured. Whatever we may call it, information, pattern, form, *Gestalt*, state description, distribution function, or negative entropy, it has become accessible to the treatment of science. It differs from the "matter" and "energy" of nineteenth-century mechanical materialism in that it cannot be described adequately by their conservation laws.

But it also differs, if not more so, from the "idea" of "idealistic" or metaphysical philosophies, in that it is based on physical processes during every single moment of its existence, and in that it can and must be dealt with by physical methods. It has material reality. It exists and interacts with other processes in the world, regardless of the whims of any particular human observer; so much so that its reception, transmission, reproduction and in certain cases its recognition, can be and sometimes has been mechanized.

These, then, were the main developments that came to a head after 1940. Cybernetics as the science of communication and control arose in response to a technological and social opportunity. It was made possible by advanced and parallel developments in neurophysiology and psychology, in mathematics, and in electrical engineering, and by the growing need for cooperation among these and other sciences.[9] The result of these developments was a new body of experience, going beyond classic organism in its rationality, that is, in its ability to be retraced step by step in its workings.

The concept of information grew out of this new body of experience, and particularly out of the separation of communications engineering from power engineering. Information is what is transferred in telephony or television: it is not events as such, but a patterned relationship between events. Information has physical, "material" reality; without exception, it is carried by matter-energy processes. Yet it is not subject to their conservation laws. Information can be created and wiped out—although it cannot be created from nothing or destroyed completely into nothingness.[10] Finally, it differs from the classic notion of "form" in that it can be analyzed into discrete units that can be measured and counted.

Information consists of a transmitted pattern that is received and evaluated against the background of a statistical ensemble of related patterns. The classic example for this is the standardized birthday telegram transmitted by telegraphing a single two-digit number indicating the message to be selected from the limited set of prefabricated messages held ready by the company. All information at bottom involves the indication of some pattern out of a larger statistical ensemble, that is, an ensemble that is already stored at the point of reception.

From this it follows that recognition can be treated as a physical process and can, in fact, be mechanized in many instances. Current mechanical devices embodying operations of matching and recognition range all the way from the lowly fruit-grading and candling machines to the Moving Target Indicator and the Friend and Foe Identification device of the Armed Forces.[11] Similar standardized recognition processes are embodied in processes that have been only

partly mechanized thus far and that still embody standardized human operations at some stages. Examples of such semimechanized recognition processes include qualitative analysis in chemistry and the Crocker-Henderson odor classification scheme and its successor, the flavor profile, according to which each of five hundred well-known smells can be identified by a four-digit number.[12] These recognition devices have grown up empirically, but the application of the theory of information forms part of the current development work on more complex devices of recognition—devices that are to be used to permit the deaf to understand spoken messages and the blind to read printed books, as well as for work on machines that will transcribe dictation or translate printed matter from one language into another.[13] Work on all these problems has been under way since the 1950's at several institutions of research.

MEMORY AND RECOGNITION

Since information has physical reality, its storage, that is to say, *memory*, is also a physical process. Most processes of thought can be represented in terms of a seven-stage process:

First, abstraction or coding of incoming information into appropriate symbols;

Second, storage of these symbols by means of quasi-permanent changes in the state of some appropriate physical facilities, such as the patterns of electric charges in certain electronic devices, the activity patterns of cells in nervous tissue, or the distribution of written marks on paper;

Third, dissociation of some of this information from the rest;

Fourth, recall of some of the dissociated items, as well as of some of the larger assemblies;

Fifth, recombinations of some of the recalled items into new patterns that had not been present among the input into the system;

Sixth, new abstraction from the recombined items preserving their new pattern, but obliterating its combinatorial origin. Steps five and six together make up the operation of creating *novelty;* and

Seventh, transmission of the new item to storage or to applications to action. This application of novelty to behavior we may call *initiative*.

A similar multistep sequence could consist of matching an incoming pattern of information against another pattern recalled from storage or memory. Both patterns are exposed to a *critical process*, that is to say, a physical process the outcome of which depends critically on the degree of correspondence between the two patterns to which it is applied. If the difference between the two patterns is smaller than a certain threshold given by the process, the critical process will have one result; if the difference is larger, the process will have another outcome. The process of recognition is completed by applying the outcome of the critical process to the behavior of the system.

AN OPERATIONAL APPROACH TO QUANTITY AND QUALITY

In a much more general aspect the notions of information and complementarity might be used to clarify the notion of quality that has sometimes baffled social scientists. From Plato and Aristotle to Oswald Spengler, Otto Strasser, and Ernst Jünger, authoritarian philosophers and political theorists have invoked qualitative judgments, the "all or nothing" reactions of taste or esthetic appreciation, as weapons against rationality or democracy. Only the coarse and simple things of social life can be counted and measured, so the argument has run, while all truly subtle and important things defy quantification and step-by-step analysis. Their imponderable and incommensurable peculiarities make them a law unto themselves, and a proper analogue to the superiority of mankind's privileged individuals, classes, or races. By contrast, if those social scientists who favored privilege have invoked quality in its defense, some of those who attacked privilege have tried to ignore problems of quality altogether.

Perhaps definitions of quantity and quality might be developed in

terms of the operations from which they are derived, that is, from the operations involved in the processes of recognition and of measurement. At Massachusetts Institute of Technology a preliminary survey has been made of six processes involving *recognition*. These six systems are in actual use. Two of them are mechanical: the Yale lock and the automatic sorting of punched cards. Two are electronic: the equipment for the Identification of Friend or Foe (I.F.F.) and the Moving Target Indicator (M.T.I.). The last two involve biological or chemical processes within a systematized sequence of steps of human labor: the Crocker-Henderson odor classification scheme, and the scheme of qualitative analysis in chemistry.

In all cases it was found that the critical step in the recognition of quality was the establishment of a structural correspondence between a part of the recognizing system and the system that was recognized, and the testing of that correspondence by a *critical process*, that is, a physical process the outcome of which depends critically on the extent of that correspondence.[14] Quality is recognized, therefore, by the matching of two structures. The decisive step is to establish whether or not such matching has occurred.

Quantity in this view would appear to be really a more complicated notion than quality. It can be measured only *after* some qualitative matching has occurred or has been established; and it consists, then, in the matching of these matchings, that is, in comparing these operations of matching with each other, so as to derive a result of "more" or "less" from this comparison, or in comparing them with some counting structure, so as to record the number of complete matchings.

Quality in this view is derived from simple matching; quantity is derived from second-order matching. Despite this fact, quality has appeared to some writers as the more complex of the two notions, since quantitative measurement occurred only in those relatively well-understood situations where qualitative recognition had already taken place and where the latter could, therefore, already be taken for granted. The situations where qualitative problems were conspicuous were precisely those more difficult cases where structural matching had not yet been accomplished well enough to permit

quantitative comparisons. There is reason to suspect that many of the qualitative problems in social and political science may turn out to be problems of matching and complementarity in social communication.[15]

FEEDBACK AND EQUILIBRIUM

Another significant concept elaborated since the 1940's is that of the "feedback." The feedback pattern is common to self-modifying communications networks, whether they are electronic control devices, nerve systems, or social organizations. "In a broad sense [feedback] may denote that some of the output energy of an apparatus or machine is returned as input. . . . [If] the behavior of an object is controlled by the margin of error at which the object stands at a given time with reference to a relatively specific goal . . . [the] feedback is . . . negative, that is, the signals from the goal are used to restrict outputs which would otherwise go beyond the goal. It is this . . . meaning of the term feedback that is used here."[16] "By output is meant any change produced in the surroundings by the object. By input, conversely, is meant any event external to the object that modifies this object in any manner."[17]

In other words, by feedback—or, as it is often called, a servomechanism—is meant a communications network that produces action in response to an input of information, and *includes the results of its own action in the new information by which it modifies its subsequent behavior.* A simple feedback network contains arrangements to react to an outside event (for example, a target) in a specified manner (such as by directing guns at it) until a specified state of affairs has been brought about (the guns cover the target perfectly, or the automatic push-button tuning adjustment on a radio has been accurately set on the wavelength approached). If the action of the network has fallen short of reaching fully the sought adjustment, it is continued; if it has overshot the mark, it is reversed. Both continuation and reversal may take place in proportion to the extent

to which the goal has not yet been reached. If the feedback is well designed, the result will be a series of diminishing mistakes—a dwindling series of under- and over-corrections converging on the goal. If the functioning of the feedback or servomechanism is not adequate to its task (if it is inadequately "dampened"), the mistakes may become greater. The network may be "hunting" over a cyclical or widening range of tentative and "incorrect" responses, ending in a breakdown of the mechanism. These failures of feedback networks have specific parallels in the pathology of the human nervous system ("purpose tremor") and perhaps even, in a looser sense, in the behavior of animals, men, and whole communities.[18]

This notion of feedback—and its application in practice—is at the heart of much of modern control engineering. It is a more sophisticated concept than the simple mechanical notion of equilibrium, and it promises to become a more powerful tool in the social sciences than the traditional equilibrium analysis.

If we say that a system is in *equilibrium,* we make a number of rather specific suggestions. We suggest that it will return to a particular state when "disturbed"; that we imagine the disturbance is coming from outside the system; that the system will return with greater force to its original state the greater has been the disturbance; that the high or low speed with which the system reacts or with which its parts act on each other is somehow irrelevant (and we term this quality "friction" to denote that it is a sort of imperfection or blemish that has no proper place in the "ideal" equilibrium); and finally we suggest that no catastrophes can happen within the limits of the system, but that, once an equilibrium breaks down, next to nothing can be said about the future of the system from then on.

Such equilibrium theories are based on a very restricted field of science, called "steady state dynamics." They are not well suited to deal with so-called *transients;* that is, they cannot predict the consequences of *sudden* changes within the system or in its environment, such as the sudden starting or stopping of a process. Altogether, in the world of equilibrium theory there is no growth, no evolution;

there are no sudden changes; and there is no efficient prediction of the consequences of "friction" over time.

On all these points the feedback concept promises improvements. Instead of pushing the effect of "friction" into the background, feedback theory is based on the measurement of *lag* and *gain*. *Lag* is the time that elapses between the moment a negative feedback system reaches a certain distance from its goal and the moment it completes corrective action corresponding to that distance. *Gain* means the extent of the corrective action taken. An inexperienced automobile driver tends to have slow reflexes: he responds tardily to the information of his eyes that his car is heading for the right-hand ditch. His lag, in feedback terms, is high. Yet when he acts, he may turn his steering wheel sharply—with a high gain—and head for the left-hand ditch until he notices the overcorrection and corrects his course again. If we know three quantities—the speed of his car and extent of his lags and of his gains—we can try to predict the wobbliness of his resulting course.

Lag and *gain*, in the feedback approach, are the most important variables to work on. Of the two, *lag* is the more important. It can be reduced by improving the system, as when our novice driver learns to react faster; or lag can be compensated for by a lead—a prediction of a future distance from the goal—as when an experienced driver compensates for an anticipated skid at the first sign of its onset. What *lag* still remains will permit control engineers to calculate just how much gain—how drastic a self-correction at each step—the system can afford under known conditions without endangering its stability.

To sum up, equilibrium analysis is based on a restricted part of dynamics; it is restricted to the description of steady states. Cybernetics is based on full dynamics including changes of state; and it combines these full dynamics with statistics. Cybernetics is the study of the full dynamics of a system under a statistically varying input. The potential usefulness of this approach to such economic problems as, for example, the so-called "cobweb theorem" has been stressed by some economists.[19]

From a historical point of view, the rise of equilibrium analysis meant the neglect of problems of purpose. Cybernetics offers not only a gain in technical competence but also a possibility of restoring to problems of purpose their full share of our attention.

LEARNING AND PURPOSE

Even the simple feedback network shows the basic characteristics of the "learning process" described by John Dollard in animals and men. According to Dollard, "there must be (1) drive, (2) cue, (3) response, and (4) reward." In a man-made feedback network, "drive" might be represented by "internal tension," or better, by mechanical, chemical, or electric "disequilibrium"; input and output would function as "cue" and "response"; and the "reward" could be defined analogously for both organisms and man-made nets as a "reduction in intensity" (or extent) of the initial "drive" or internal disequilibrium.[20]

A simple feedback mechanism implies a measure of "purpose" or "goal." In this view a goal not only exists within the mind of a human observer; it also has relative objective reality within the context of a particular feedback net, once that net has physically come into existence. Thus a *goal* may be defined as "a final condition in which the behaving object reaches a definite correlation in time or in space with respect to another object or event."[21]

This definition of a goal, or purpose, may need further development. There is usually at least one such external goal (that is, one relation of the net as a whole to some external object) that is associated with one state encompassing the relatively lowest amount of internal disequilibrium within the net. Very often, however, an almost equivalent reduction of internal disequilibrium can be reached through an internal rearrangement of the relations between some of the constituent parts of the net, which would then provide a more or less effective substitute for the actual attainment of the goal relation in the world external to the net. There are many cases of such surrogate goals or *ersatz* satisfactions, as a short circuit in an elec-

tronic calculator, intoxication in certain insects, drug addiction or suicide in a man, or outbursts against scapegoat members of a "tense" community. They suggest the need for a distinction between merely internal readjustments and those that are sought through pathways that include as an essential part the reaching of a goal relationship with some part of the outside world.

This brings us to a more complex kind of learning. Simple learning is goal-seeking feedback, as in a homing torpedo. It consists in adjusting responses, so as to reach a goal situation of a type that is given once for all by certain internal arrangements of the net; these arrangements remain fixed throughout its life. A more complex type of learning is the self-modifying or *goal-changing* feedback. It allows for feedback readjustments of those internal arrangements that implied its original goal, so that the net will change its goal, or set for itself new goals that it will now have to reach if its internal disequilibrium is to be lessened. Goal-changing feedback contrasts, therefore, with Aristotelian teleology, in which each thing was supposed to be characterized by its unchanging *telos*, but it has parallels in Darwinian evolution.[22]

We can now restate our earlier distinction as one between two kinds of goal-changing by internal rearrangement. Internal rearrangements that are still relevant to goal-seeking in the outside world we may call "learning." Internal rearrangements that reduce the net's goal-seeking effectiveness belong to the pathology of learning. Their eventual results are self-frustration and self-destruction. Pathological learning resembles what some moralists call "sin."

Perhaps the distinction could be carried further by thinking of several orders of purposes.

A first-order purpose in a feedback net would be the seeking of *immediate satisfaction*, that is, of an internal state in which internal disequilibrium would be less than in any alternative state, within the range of operations of the net. This first-order purpose would correspond to the concepts of "adjustment" and "reward" in studies of the learning process. Self-destructive purposes or rewards would be included in this class.

By a second-order purpose would be meant that internal and external state of the net that would seem to offer to the net the largest probability (or predictive value derived from past experience) for the net's continued ability to seek first-order purposes. This would imply *self-preservation* as a second-order purpose of the net, overriding the first-order purposes. It would require a far more complex net.[23]

A third-order purpose might then mean a state of high probability for the continuation of the process of search for first- and second-order purposes by a group of nets beyond the "lifetime" of an individual net. This would include such purposes as the *preservation of the group* or "preservation of the species." Third-order purposes require several complex nets in interaction. Such interaction between several nets, sufficiently similar to make their experiences relevant test cases for one another, sufficiently different to permit division of labor, and sufficiently complex and readjustable to permit reliable communication between them—in short, such a *society*—is in turn essential for the higher levels of the learning process that could lead beyond third-order purposes.

Among fourth-order purposes we might include states offering high probabilities of the *preservation of a process* of purpose-seeking, even beyond the preservation of any particular group or species of nets. Such purposes as the preservation or growth of "life," "mind," "order in the universe," and all the other purposes envisaged in science, philosophy, or religion, could be included here.

The four orders overlap; their boundaries blur; and there seems to be no limit to the number of orders or purposes we may set up as aids to our thinking. Yet it may be worthwhile to order purposes in some such fashion, and to retain, as far as possible, the model of the feedback net that permits us to compare these purposes to some degree with physical arrangements and operations. The purpose of this procedure would not be to reduce intellectual and spiritual purposes to the level of neurophysiology or mechanics. Rather it would be to show that consistent elaboration of the simpler processes can elevate their results to higher levels.

VALUES AND THE CAPACITY TO LEARN

The movements of messages through complex feedback networks may involve the problem of "value" or the "switchboard problem," that is, the problem of choice between different possibilities of routing different incoming messages through different channels or "associative trails"[24] within the network. If many alternative channels are available for few messages, the functioning of the network may be hampered by indecision; if many messages have to compete for few channels, it may be hampered by "jamming."

The efficient functioning of any complex switchboard requires, therefore, some relatively stable operating rules, explicit or implied in the arrangements of the channels. These rules must decide the relative preferences and priorities in the reception, screening, and routing of all signals entering the network from outside or originating within it.

There are many examples of such rules in practice: the priority given fire alarms in many telephone systems; or the rules determining the channels through which transcontinental telephone calls are routed at different loads of traffic; these last include even the "hunting" of an automatic switchboard for a free circuit when the routing channels are fully loaded. They illustrate the general need of any complex network to decide in some way on how to distribute its "attention" and its priorities in expediting competing messages, and how to choose between its large number of different possibilities for combination, association, and recombination for each message.

What operating rules accomplish in switchboards and calculating machines is accomplished to some extent by "emotional preference" in the nervous systems of animals and men, and by cultural or institutional preferences, obstacles, and "values" in groups or societies. Nowhere have investigators found any mind of that type that John Locke supposed "to be, as we say, white paper." Everywhere they have found structure and relative function.

In much of the communications machinery currently used, the operating rules are rigid in relation to the content of the information

dealt with by the network. However, *these operating rules themselves may be made subject to some feedback process.* Just as human directors of a telephone company may react to a traffic count by changing some of their network's operating rules, we might imagine an automatic telephone exchange carrying out its own traffic counts and analyses, and modifying its operating rules accordingly. It might even modify the physical structure of some of its channels, perhaps adding or dropping additional microwave beams (which fulfill the function of telephone cables) in the light of the traffic or financial data "experienced" by the network.[25]

What seems a possibility in the case of man-made machinery seems to be a fact in living nerve systems, minds, and societies. The establishment and abolition of "conditioned reflexes" have long been studied in animals and men, and so have the results of individual and group learning. Such processes often include changes in the "operating rules" that determine how the organism treats subsequent items of information reaching it.

Any network whose operating rules can be modified by feedback processes is subject to *internal conflict* between its established working preferences and the impact of new information. The simpler the network, the more readily internal conflicts can be resolved by automatically assigning a clear preponderance to one or another of two competing "channels" or "reflexes" at any particular moment, swinging from one trend of behavior to another with least delay. The more complex, relatively, the switchboards and networks involved, the richer the possibilities of choice, the more prolonged may be the periods of indecision or internal conflict. Since the net acquires its preferences through a process of history, its "values" need not all be consistent with each other. They may form circular configurations of preference, which later may trap some of the impulses of the net in circular pathways of frustration. Since the human nervous network is complex, it remains subject to the possibilities of conflicts, indecision, jamming, and circular frustration. Whatever pattern of preferences or operating rules govern its behavior at any particular time can only reduce this affliction, but cannot abolish it.[26]

Since the network of the human mind behaves with some degree

of plasticity, it can change many of its operating rules under the impact of experience. It can learn, not only superficially but fundamentally: with the aid of experience the human mind can change its own structure of preference, rejections, and associations. And what seems true of the general plasticity of the individual human mind applies even more to the plasticity of the channels that make up human cultures and social institutions and those particular individual habit patterns that go with them. Indeed, this cultural learning capacity seems to occur in some proportion to the ability of those cultures to survive and to spread.

Since all learning including changes in goals or values consists in physical internal rearrangements, it depends significantly on material resources. The *learning capacity* of any system or organization, that is, the range of its effective internal rearrangements, can thus be measured to some extent by the number and kinds of its *uncommitted resources*. Such resources need not be idle; but they must be reassignable from their current functions. There is a qualitative element in learning capacity, since it depends not only on the amount of uncommitted resources but also on their configurations. Yet, since learning capacity consists in an over-all performance, a particular configuration of internal elements can be replaced, in many cases, by some functionally equivalent configuration of others. This is the more probable, the richer the range of available rearrangements, and thus, again, the greater the amount of uncommitted resources, and of facilities for their quick and varied recommitment.

Learning capacity can be tested by two independent sets of operations: first, by outside tests of a system's over-all performance in a given situation, much as the learning capacity of rats is tested in a maze and that of armies is tested in battle; and second, by analysis of its inner structure. Thus the greater learning capacity of rats compared to frogs can be predicted from the greater size and complexity of the rat's central nervous system, and the greater learning capacity or adaptability of one army relative to another can be predicted if, other things being equal, it has greater facilities of communication and transport and a greater "operational reserve" of uncommitted man power and equipment. Since over-all performance tests are

cheap in rats, but expensive in armies, or in the defense of cities against atom bombs, the prediction of probable learning capacity from structural analysis and the suggestions for probable improvements by the same method may have considerable practical importance.

So far we have described two kinds of feedback: "goal-seeking," the feedback of new external data into a net whose operating channels remain unchanged; and "learning," the feedback of external data for the changing of these operating channels themselves. A third important type of possible feedback is the feedback and simultaneous scanning of highly selected internal data, analogous to the problem of what usually is called "consciousness."

CHAPTER

6

Consciousness and Will as Patterns of Communication Flow

Consciousness may be defined, as a first approximation and for the purposes of this discussion, as a collection of internal feedbacks of secondary messages. *Secondary messages* are messages about changes in the state of parts of the system, that is, about primary messages. *Primary messages* are those that move through the system in consequence of its interaction with the outside world. Any secondary message or combination of messages, however, may in turn serve as a primary message, in that a further secondary message may be attached to any combination of primary messages or to other secondary messages or their combinations, up to any level of regress.

In all these cases, secondary messages function as symbols or internal labels for changes of state within the net itself. They are fed back into it as additional information, and they influence, together with all other feedback data, the net's subsequent behavior. "Consciousness" does not consist merely in these labels, but in the processes by which they are derived from the net and fed back into it, and in the processes by which two or more such secondary messages are brought to interact with each other.

CONSCIOUSNESS IN SOCIAL ORGANIZATIONS

Feedback messages about some of the net's internal states occur in simple form in electronic calculators where they serve important functions in recall. They may occur, in extremely complex patterns, in the human nervous system, where they are not easy to isolate for study. But they also occur, and can be studied with relative ease, in the division of labor of large human teams that process information and collectively fulfill certain functions of thought. We find such teams in industrial research laboratories, and in political or military intelligence organizations.

We can observe how guide cards and index tabs are added to the information moving through, or stored within, the filing systems, libraries, card catalogues, or "document control centers" of intelligence organizations (such as the State Department or, during World War II, the Office of Strategic Services), and how these secondary symbols influence the further treatment of the information. The heads, policy boards, or project committees of such organizations cannot deal with all the vast information in the original documents. They deal mostly with titles, description sheets, summaries, project requests, routing slips, and other secondary symbols, while a great deal of the material continues to be processed "below the level of the consciousness" of the guiding and policy-making parts of the organization. Only those feedback circuits and decisions that are "picked up" through the attachment and feedback of secondary symbols become directly "conscious" for the organization.

To be sure, the selective function of any network is by no means limited to this "conscious" zone of secondary symbols. On the contrary, what reaches that zone for separately labeled and recorded processing depends in turn on what has been selected or rejected, associated or disassociated, routed or blocked, recorded or misfiled or erased within the rest of the system. There is some automatic screening carried out by the reporter on the beat, and by the desk analyst in the intelligence organization; and we may suspect similar screening processes in the "nonconscious" remembering and for-

getting, the "aversions" and "hunches" of the individual mind, as well as in many of the "unverbalized" conventions and assumptions, preferences or taboos of human societies and cultures.[1]

The powers of the "nonconscious," internally unlabeled processes within a network, can be positive as well. An experience may be built up into a perception and recorded in memory, two and two may be put together, new associations, discoveries and insights may be put together "nonconsciously" without the intervention of secondary symbols, until secondary symbols are attached to the new combination, and suddenly the image of the new synthesis breaks through into the realm of consciousness, seeming all ready and armored, like Pallas Athene springing forth from the head of Zeus in the Greek legend.[2]

By attaching secondary symbols to some of our steps in a calculation or sequence of behavior we may change its outcome. For these secondary symbols are fed back into the net, and the message of which the net has become "conscious" may then appear in the net with greater frequency than its unlabeled alternatives, and it may remain more readily available for preferred treatment. This treatment may be preferred association, recording, transmission, blocking, or suppression, according to the current operating rules of the system.

If secondary symbols become attached to parts or connections in the net that embody these operating rules, then these rules themselves become "conscious" for the net. By being fed back into it, they become statistically reinforced for more effective application; and they may be changed more easily if this possibility is included in the net. The effects of such internal labeling may be thought of as to some extent comparable to the effect of dramatic symbols or publicity devices in a society. Once attached to particular ideas, practices, or laws, they may lift them from their previous obscure existence into the crossfire of public attention.[3]

The ensemble of secondary symbols may easily misrepresent the net's actual content. Some primary symbols may be "overrepresented" by ample feedback, while others may not be made "conscious" at all. Consciousness, therefore, may be false consciousness, much as the actual personality of a man may be quite different from what he thinks it is. Similarly, by attaching suitable symbols and feedbacks to selected aspects of their behavior, groups or nations can be given highly misleading ideas about their own character.

CONFRONTATION FOR SIMULTANEOUS INSPECTION

Consciousness, however, involves not one operation but at least two. It requires first of all a high degree of selection and abstraction from the stream of primary or lower-order messages, and their highly condensed and abridged mapping into a much smaller number of higher-order messages. But it also implies, as a rule, the more or less simultaneous scanning or inspection of as much of this abridged second-order information as can be encompassed in the "focus of consciousness"—or the "span of attention" or "span of control"—of a person, or in the effective range of surveillance or control of an organization.

Physical examples of such condensed and concentrated arrangements of secondary symbols for their simultaneous—or nearly simultaneous—inspection are abundant. They include the nineteenth century military staff maps on which colored pins and other movable symbols represented troops; the underground plotting center in the Battle of Britain in 1940 where, on a large simplified map of southern England and the Channel, wooden counters were moved with rakes by army personnel, so as to represent the strength, position, direction, and speed of attacking and defending aircraft and to permit quick decisions about the best use of still disposable British fighter defenses; and the transparent plastic screen in the antiaircraft control center of post-World War II vessels in the premissile age, when the quickly changing reported numbers and movements of attacking enemy aircraft were chalked in color on one side of the plastic, so as to permit the officer on the other side to encompass at one glance the rapidly mounting attacks from many directions against his ship,

and to decide on the best allocation of his own antiaircraft batteries, and perhaps fighter planes, for his defense; and the "situation room" of the early 1950's, where the President of the United States was reportedly briefed almost daily by his subordinates on the changing conditions and crises around the world, was similarly designed to facilitate simultaneous inspection. A current offspring of all these simpler devices could perhaps be seen in the vast warning and computing systems of the late 1950's and early 1960's—such as the SAGE system—that are designed to collect and compare a large number of highly abbreviated data, radar readings, and the like, so as to make or keep their operators aware of all actual or apparent movements of aircraft or missiles toward the territory of the United States.

The last stage in this approximation to certain aspects of consciousness would be reached when these early-warning and computing systems were made fully automatic. In this stage, not only would the collection, abridgment, and collation of the primary messages be carried out by electronic equipment, but so would their confrontation with selected data automatically recalled from the memory banks of appropriate computers, and, further, the interpretation of data and memories, as confronted with each other, for purposes of defensive or retaliatory action. Here the automatic simultaneous inspection of abridged symbols may lead to highly fallible decisions about war and peace, and thus about the life and death of nations and perhaps of all mankind, much as the quick, simultaneous inspection of the few matters of which an individual is aware may lead him to some highly fallible decision about his own fate. Here our man-made machinery of quasi-consciousness shows once again its close relation to problems of command and government. The subject of a recent thoughtful paper in communication engineering and command problems in national defense, entitled "Emergency Simulation of the Duties of the President of the United States," indicates some of the more remote perspectives that might possibly follow someday from trends now under way.[4]

The dangers from a premature and poorly understood extension of these techniques to the making of high-speed and almost inevi-

tably shallow decisions about the life and death of millions are appalling. No one has warned more emphatically and cogently against this overextension of engineering in advance of broader and more fundamental knowledge than one of the intellectual fathers of much of this technology, Norbert Wiener.[5] The realism of his warnings was illustrated by the news that, on November 24, 1961, at a tense moment during the Berlin crisis, bombers of the United States Strategic Air Command had been moved to the runways in response to an erroneous signal.[6]

There is no reasonable way in which we can transfer tasks of individual and social thinking, decision-making and consciousness to some aggregation of electronic machinery, unless we first have taken good care to understand this particular task or aspect of thought and consciousness in its social and individual setting. Before we hand any task of consciousness to a machine, we should understand at least to some extent what consciousness is; what the processes are that it describes; what the differences are that it makes in a set of probable outcomes; and what the facilities are that it requires. The answers to these questions are not at all complete; but the notion of consciousness, derived here from the viewpoint of communications, may help us to get more and better answers in the course of time.

How does this feedback notion of consciousness compare with other approaches? In the behaviorist school of psychology, we are told that "consciousness" and conscious processes "are excluded as not subject to scientific investigation, or . . . reinterpreted as covert language responses."[7] In social science writings, consciousness is often stressed, and ascribed to groups, but usually this is done without definition or description in any but intuitive terms.[8] Two recent writers describe individual consciousness as follows:

> [The] integrative (regnant) process in the brain . . . according to the findings and speculations of neurophysiologists . . . are capable of self-awareness (as if they had a mirror in which to see themselves). During the passage of one event many, but not all, of the regnant processes have the property of consciousness, at the moment of their occurrence or soon afterwards if recalled by retrospection. Thus the

stream of consciousness is nothing more than the subjective (inner) awareness of some of the momentary forces operating at the regnant level of integration in the brain field.[9]

This is a suggestive description in the language of everyday life in which processes behave like small individuals who "reign," "see themselves as if they had a mirror," and "have the property of consciousness" that "is nothing more than . . . subjective (inner) awareness." But it is not very helpful as a concept from which we might derive new observations and experiments.

In contrast to this, what are the operational implications of our feedback model? First of all, if consciousness is a feedback process, then it requires material facilities and is carried on at some material cost in terms of facilities and time. Some of the facilities that are tied up, and some of the delay imposed on primary processes, should be capable of measurement.

Second, feedback processes have structures, circuits, channels, switching relationships, incompatibilities, and discontinuities that might be susceptible of mapping.

In the third place, if we cannot isolate the physical facilities involved, we might devise functional tests for possible patterns, limits, and discontinuities in the performance of the process of consciousness. If these tests should yield a map of discontinuities in performance, we might derive a basis for further inferences about the structure of the underlying facilities and processes themselves.[10]

Similar considerations might apply to the processes of "consciousness" in nations, classes, or other social groups. If there are such processes, how are they organized and patterned? What are the manpower, facilities, symbols, learning processes, and teamwork relations by which they are carried on? If consciousness resembles a feedback, does it also resemble the feedback's peculiar kinds of instability? A small change in a feedback circuit can bring about a large change in its over-all performance. Are there analogies for this in social life?

The feedback model of consciousness is more than a verbal explanation. It is a concept. For it suggests many questions that

sooner or later should be answered, one way or another, by observation and experiment.

WILL AND THE CLOSURE OF DECISION SYSTEMS

Consciousness seems related to "will"—or to that sense of conation, or making autonomous decisions, which we mean when asserting that "our will is free." This notion of will includes not only decisions with internal labels attached to the very moment of action, or to several steps within an action. It also includes mere decisions to start an action, now or on a later signal, with the actual parts of the action following automatically without any "conscious" labels attached to them.[11] *Will* in all these cases may be tentatively defined, in any sufficiently complex net, nervous system, or social group, as the set of internal labels attached to various stages of certain channels within the net, which are represented by these labels as relatively unchanging, so that "we merely trip the purpose and the reaction follows automatically."[12]

In other words, *will* may be called the set of *internally labeled decisions and anticipated results, proposed by the application of data from the system's past and by the blocking of incompatible impulses or data from the system's present or future.* Since the net cannot foretell with certainty either the outcome of the subsequent trains of its own internal messages and switching orders, or the outcome of its own efforts to inhibit information incompatible with the "willed" result, it knows only what it "will do," not what it "shall do." It may "know its mind," but it cannot know with certainty whether or when it will change it.[13]

A fundamental problem of "will" in any self-steering network seems to be that of carrying forward and translating into action various data from the net's past, up to the instant that the "will" is formed (the determination becomes "set" or the decision "hardens"), while blocking all subsequent information that might modify the "willed" decision. Will resembles the "deadline" in a newspaper: it could be called the *internally labeled preference for predecision*

messages over postdecision ones. The "moment of decision" might then be seen as that threshold where the cumulative outcome of a combination of past information begins to inhibit effectively the transmission of contradictory data.

This general problem of "will" seems to apply, at least to some extent, to man-made devices whose operations can be accurately specified. Automatic pilots or steering mechanisms exclude or compensate for subsequent "experiences," such as gusts of storm, which might deflect them from their course. Guided missiles, homing torpedoes, proximity fuses, and similar weapons involve in their design problems of this kind.

A primitive once-for-all process of a somewhat comparable type seems to be involved in the process of "learning by imprinting" that has been observed in goslings and certain other young birds:

> Lorenz first made known the curious fact that whatever a gosling . . . first sees in the hours after hatching, be it bird, beast or man, the gosling will follow as it normally would follow its mother. The fixation is demonstrated as . . . essentially . . . on any first moving object perceived. . . . How persistent is the fixation was shown in another experiment in which the first thing presented to the eyes of a budgereegah remained forever its only object of attachment and its days of courtship were spent in trying to make love to a ping-pong ball. . . .
>
> A working model of the mechanism is simple. A feedback trigger circuit provides channels for the reception of a number of possible stimuli, and is so constructed that, when any one of them is activated, it locks on, and a common reflex puts all the other channels out of action—that is, excepting the one already locked.[14]

This process of "learning by imprinting" resembles what we have called "will" only to the extent that a stage of openness of the acting system to different messages is followed by its closure after a threshold period, and/or a threshold message, so that all competing or conflicting messages are thereafter excluded. "Will" differs from imprint learning, however, in that it is usually internally monitored and labeled, and thus appears as a conscious decision to the acting person himself and possibly to other observers. Moreover, will often involves repeated loosening and hardening of decisions, that is, the setting and resetting of goals, purposes, or courses of action

that remain fixed only for limited lengths of time. Imprint learning, so far as we can tell, occurs without any awareness on the part of the animal; and it lasts for a lifetime in some of the experiments described. It is thus a kind of extreme caricature of will: a blind slipping into an irrevocable decision about one's own further preferences and desires—an extreme form of learning *not* to learn in the future.

Isolating the pattern of "will" in feedback machines may help us to recognize it in men and communities. Men may shut out the experiences of pain or fear or doubt or pity that might deflect them from their "fell purpose." Cultures or states, ever since the days of the Spartans, have often put taboos or legal prohibitions in the way of all messages that might change their previously determined patterns of behavior. Modern nations, governments, or political parties in war or peace may strive to perpetuate their policies by blocking all incompatible experiences from the life of their community through all means at their disposal—legislation, indoctrination, pressure, censorship, police, or propaganda. It is in that sense, perhaps, that the concept of "will" can be applied meaningfully to the behavior of political movements, peoples, and social organizations. In government and politics, will is a pattern of relatively *consolidated preferences and inhibitions, derived from the past* experiences of a social group, *consciously labeled* for a relevant portion of its members, *and applied* to guide the actions, to restrict the subsequent experiences of that group and its members.[15]

THE NATURE OF "FREE WILL"

In what sense is this "will" free?

First of all, this will is relatively free from the pressures of the outside world at any one moment, since it represents the stored outcome of the net's past now being fed back into the making of present decisions. Without effective feedback of its past, the net's behavior would be determined largely by outside pressures.

It would not steer, but drift, in both its external and internal arrangements.

As long as it has autonomy, the net wills what it is. It wills the behavior patterns (the "personality") that it has acquired in the past and that it is changing and remaking with each decision in the present. Thanks to what it has learned in the past, it is not wholly subject to the present. Thanks to what it still can learn, it is not wholly subject to the past. Its internal rearrangements in response to new challenges are made by the interplay between its present and its past. In this interplay we might see one kind of "inner freedom."

In its external actions, the net does what it can do. Its outward behavior will be the result of the interplay between the orders transmitted to its effectors and the feedback data about their results among the pressures of the outside world. In this type of interplay we may see a kind of external freedom for the net to continue to seek its goal.

Freedom in a feedback network could go further. A chess-playing machine could be constructed that would rapidly compute all admissible moves on both sides for two or three moves ahead, and choose the ones more profitable for its side, according to a schedule of values derived from the rules of the game. It would play mediocre chess. It could be improved, however, by giving it a suitable memory and additional circuits, so that it could learn to modify its play on the basis of experience. The quality of its playing would then depend largely on that of its experience. If all its past opponents were mediocre, the machine might never learn to play brilliantly. It would remain imprisoned by the limitations of its past. But it could be aided to play better by building into it a device to break or sometimes override the patterns learned from its past, giving the machine a chance for initiative and creativity.

This function of autonomous internal habit-breaking could be fulfilled by building into the machine a circuit breaker controlled by some "internal receptor," such as the flipping of a coin, that is, by some element of the network whose state would not altogether be determined by the previous states of other parts of the net.[16] Such

a device could be so connected as to break up established connections or patterns of response from time to time, and to permit new combinations within the net to be formed, recorded internally in memory and carried through into external action.

The results might resemble those of a "spontaneous impulse." Like all "spontaneity," they would be subject to limitations. All they could do would be to replace an old or highly probable configuration by a new or less probable one, *provided that the elements for the new configuration were already present in the net* at the critical moment—even though they might have got there only through some input in the immediately preceding instant.[17] The range of possible new combinations would therefore also depend, among other factors, on the range of possible new input information from the outside world and on the effectiveness of the inner "habit breaker" in breaking up blocks against the integration of such new inputs with other data in the net. Apart from facilitating this inflow of new information, "spontaneity" could only bring out a wider flow of the potentialities already contained within the net.

This type of feedback network might provide an analogue for the problem of "free will." Such an analogy might be found in a machine combining a determinate store of memories with a randomly varying inner receptor in the circuits governing recall and combination. The random effects of the inner receptor (or "sudden impulse") are then limited by the statistical weight of alternatives as well as of critical recognition processes or patterns available from the stored past of the machine (its "personality"). Such a machine might act "freely," with initiative, but "in character."[18]

The analogy suggests that moral responsibility is conferred by the determinate, cumulatively learned element in the combination. To treat a man as "responsible" is to treat him on the assumption that his learning process has not been disrupted. Each of us is responsible for what he is now, for the personality he himself has acquired by his past free decisions.[19]

Aristotle.

CHAPTER

7

Political Power and Social Transactions

Will is related to *power*. Hardening a decision—that is, closing the decision-making system against any further messages by which that decision might possibly be modified—is insignificant in practice if there are no facilities to put it into effect against possible external resistance, or in any case, to put it into effect in such a manner as to make some appreciable difference to the ensemble of outcomes in the environment that would have occurred anyway.

THE INTERDEPENDENCE OF POWER AND WILL

Will is thus ineffective without power; but power is only randomly effective without will. Power cannot accomplish more than a succession of random impacts on the environment, unless there is some relatively fixed goal or purpose, some decision or strategic class or sequence of decisions, by which the application of power can be guided and directed. This guidance is indispensable for the sustained effectiveness of any system that applies power to its environments; and any such system must receive this guidance from its memories, its past decisions, its will, or somewhat more generally, from its character.

It is perhaps in connection with will and character—with the more or less stable inner program of a system, a person, an organization, or a government—that the concept of the power of any of these actors must be understood and that our present understanding of the problems can be most promisingly developed.[1]

The point may bear elaboration. In their internal relations, an individual or an organization may give preference to the value or behavior patterns that correspond most closely to the structure of the habits and memories they have acquired in the past. In their dealings with the outside world, individuals and organizations may try to act in "character" but they may not succeed in doing so. By the *power* of an individual or organization, we then mean the extent to which they can continue successfully to act out their character. Differently put, by power we mean the ability of an individual or an organization to impose extrapolations or projections of their inner structure upon their environment. In simple language, to have power means not to have to give in, and to force the environment or the other person to do so. Power in this narrow sense is the priority of output over intake, the ability to talk instead of listen. In a sense, it is the ability to afford not to learn.

Power in this narrow sense is conceived on the analogy of the hardness scale of minerals, of the scratching of glass by a diamond, or of the "pecking order" in a chicken yard.[2] It should not be confused with strength or growth. When carried to extremes, such narrow power becomes blind, and the person or organization becomes insensitive to the present, and is driven, like a bullet or torpedo, wholly by its past. The extolling of power by certain conservative writers, often in preference to its analysis, may not be unrelated to this pattern.

The simple view of power can be restated in probabilistic terms. Gross power can be thought of as the probability of a system acting out its internal program by imposing a given amount of changes upon the environment; and net power can be derived from this as the difference between the probability of these changes imposed on the outside world, and the probability of another critical or relevant amount of changes occurring in the inner structure of the system.

CONFLICT AND ITS MEASUREMENT

From this view of power, a concept of *conflict* can be developed that would, in principle, be susceptible to measurement. Conflict between two acting systems, A and B, could be measured in terms of:

1. the probable extent of *incompatibility* between their respective programs for the future; and

2. the probable *costs of avoiding collision* between them, wholly or in part. Each of these could be measured in terms of expected changes in the structure of the two acting systems; and the sum of these changes—perhaps expressed in measures of information—would measure the scope and intensity of conflict.

Incompatibility between two acting systems can be measured in terms of the sum of the probable changes—that is, the probable changes in inner structure—that would occur in System A, and of the changes in System B, if the inner programs of each of these two systems were carried out. The incompatibility of the proposed courses of two ships, or of two states, would thus be measured by the probable collision damage if both ships or governments stayed on their respective courses.

The costs of avoiding collision could then be measured in terms of the probable changes in the inner structure of System A, if its program, of course, were to be altered sufficiently to avoid collision, even though B's course remained unchanged; or in terms of the corresponding changes in the inner structure, and hence the program, of System B so as to avoid collision, even without any change in A's program; or in terms of the aggregate changes that would have to be distributed between the inner structures and future programs of both A and B if collision were to be avoided. How much change in terms of information patterns, as well as perhaps in terms of energy, strains, or internal organizational arrangements, would be required in either or both of two ships, or of two governments and bodies politic, in order to get them off collision course? Clearly, there may be in each case not one answer to this question, but a series of answers that should be susceptible to ordering.

It should be possible in many cases to state what would be the probable minimum costs of avoiding conflict if the most efficient of the possible patterns were selected.

An impending conflict would then be reckoned to be the more serious the greater the amount of expectable changes that this conflict, if it were joined, would impose upon the structures of one or both of the acting systems; and, also, the greater the changes that would be required in one or both of these systems if this conflict were to be avoided. Ships' captains and automobile drivers usually find it less trouble to change course than to let their vehicles run full tilt into collisions, but there are occasions when the physical or psychic changes required—the costs of effective steering—are prohibitive, and catastrophe results. In the case of governments, the costs of physical, social, or psychological change may appear even higher; and it may seem less "unrealistic" to political decision-makers to let their countries run into war—as they did in 1914, and as the Axis rulers did in 1939 and 1941—rather than to take the risks of the changes in policy, and internal programs and structure, that might still avoid the collision.[3]

None of the measurements or probability estimates proposed here, it should be clear, would exhaust the nature of conflict. Yet they might be useful in illuminating some of its aspects, and perhaps in bringing the entire problem of conflict somewhat nearer to more adequate rational and quantitative analysis in the future.

In fact, the current analysis of conflict and of power has been carried much further. There is a large and growing literature on power in the political and social sciences, to which the approach of cybernetics and the study of the problems of control systems may in the future offer some contributions.[4]

THE ANALYSIS OF POWER BY DAHL AND LASSWELL

Some of this literature offers interesting contrasts, as well as bridges, to the approach suggested here. Robert A. Dahl considers political power as measurable in two respects. The first is the ability to

produce a change in the probability distribution of a class of repetitive outcomes; the second is frequency of association of an actor with outcomes that appear "successful" from his presumable viewpoint. Both tests tend to merge: if bills endorsed by Senator X have a markedly greater probability—that is, frequency—of passing, then Senator X may be thought to have caused this higher probability of their passing, or else he may be a political chameleon who seems influential to us only by the prudent timing of his rushes to the assistance of the victors.[5] Dahl's measurement of power would tend to eliminate the manifestly powerless: those unwilling to conform to the expectable majority, yet unwilling to change it. But among the seemingly powerful it becomes unambiguous only when other information is added, such as the timing of the senator's first endorsement of the bill; or each particular bill's autonomous probability of passing on grounds other than his endorsement (which is different from the frequency of passing of the possibly less popular bills he did not choose to endorse); or the senator's reputation among his colleagues and among seasoned observers for being either a political chameleon or else an independent leader.[6]

Lasswell and Kaplan define power as "participation in the making of decisions," with a "decision" defined as "a policy involving severe sanctions."[7] Such effective participation is founded upon an actor's control over any one or several of a range of "base values." These are things, relationships, or situations that are strongly desired by other persons who, in order to avoid sanctions or obtain rewards in terms of some base value, are willing to subordinate their behavior in regard to some less strongly desired "scope value" to the commands, hints, or even the anticipated wishes, of the power-holder.[8] Any value can serve as a base value, or as a scope value, or as both: control over wealth may be used to command respect, or control over respect may be used to win riches, or riches may be employed so as to make their owner still richer.[9]

None of the views given thus far, however, undertakes explicitly to measure the cost of power to the power-holder. They measure at best the narrowed choices of the ruled, but not the narrowed choices of their rulers. What, if anything, must Dahl's senator

give up to improve the chances of passage of the bill he favors? What sacrifice, if any, did Lasswell's controller of base values have to make—in terms of matters valuable to himself—in order to obtain a certain amount of control over the scope values depending upon the behavior of another person? How many pecks or scratches did the chicken high in the pecking order have to take, and what was the cumulative wear on the diamond that scratched so many windowpanes?

In the draft of a new work, Dahl discusses the "opportunity cost" of influence to those who wield it. He makes this the basis of his concept of the "influence gap" that prevails ordinarily between the political influence an individual actually exercises and the influence he could exercise if he devoted all his disposable time and resources to this sole purpose. The same reasoning as for influence holds, of course, for power. These gaps between the actual and the potential influence and power of most people in ordinary times account, in Dahl's view, for the great potential instability of politics whenever crises or changed circumstances induce many persons to shift more of their resources to the pursuit of political goals.[10]

Questions of this sort may lead us to a concept of net rather than gross power, and to an appreciation of the power of the powerless. "The only way the white man in the South can keep the Negro in the ditch," Booker T. Washington is reported to have said, "is to stay in the ditch with him." A classic example of a similar process is given by the historian Arnold Toynbee, who reports that the Spartan conquerors of Messene found the continued subjugation of their Messenian Helots a full-time job that turned them into prisoners of their own victory and distorted permanently and fatally their politics and culture.[11] In our own time, the "price of greatness" is frequently mentioned by the governments of all great powers around budget time.

A concept of net power might define it as a difference—the difference between the amounts of changes imposed and changes accepted by the actor. The changes imposed are those imposed by the actor upon his environment, including changes imposed on relevant

antagonists. The changes accepted are those accepted in the values and in the communication and action systems of the actor. If the first group of changes is measurable, at least in principle, then so should be the second group. The greater the difference between gross and net power, the greater might be the proportion of national income, and perhaps also the proportion of national attention in the mass media, and the proportion of time of political leaders, that must be allocated in order to maintain some given gross power position—at the expense, of course, of learning capacity and of whatever other adaptations or goals the society might have pursued. In this view there are, for any particular task or conflict, not two states of power for an actor—adequate or inadequate—but three: (1) comfortable, (2) tightly strained and partly overcommitted, and (3) bankrupt. Many nations and governments find themselves in the middle of these three categories, with corresponding sacrifices of their ability to steer their own actions and to remain masters of their fate—the very things the pursuit of power originally seemed to promise.

Dahl has opened the way to the quantitative measurement of power in certain types of situations. Lasswell has put power into the context of a network of other possible base and scope values, and has thus clarified its meaning over a wide range of conditions. Talcott Parsons, in his current work, is taking the next step. He is putting power into the dynamic context of a flow of interchanges between the main functional subsystems of society; and, in so doing, he has perhaps opened a path to a more fundamental reinterpretation of power than has been possible since the days of Hobbes and Locke.

THE GENERAL INTERCHANGE MODEL OF TALCOTT PARSONS

Parsons distinguishes four functional prerequisites for any social system: (1) the maintenance or reproduction of its own basic patterns; (2) adaptation to the environment and its changes; (3)

the attainment of whatever goals the system has accepted or set for itself; and (4) the integration of all the different functions and subsystems within it into a cohesive and coordinated whole.

Each of these four main functions—pattern maintenance, adaptation, goal attainment, and integration—must be served to some extent by all subsystems of the society, but to each of the four basic functions there corresponds a major subsystem of the society, devoted to a markedly greater extent to activities serving that particular function rather than any of the rest.

The function of pattern maintenance is thus primarily served, in Parsons' view, by the subsystem of families and *households*, which are the child-rearing, labor-force-restoring and kin-group-preserving elements of the society. To the task of adaptation there corresponds in the main the *economy*, including its scientific and technological aspects. Goal attainment is mainly served by the *polity*, the political subsystem of the society, and particularly by the government. Finally, the integrative function is carried on to the largest extent by the subsystems of *culture*, including the social institutions of public and quasi-public education, religion, and mass communications.

Among these four main functional subsystems, which may be conveniently pictured as four corners of a square, there are six possible major flows of interchange, corresponding to the four sides and the two diagonals of the square, and connecting each of the four main subsystems with the three others. Thus, in the most simple case the households may be viewed as delivering labor to the economy, and eventually receiving consumer goods from it, in a flow of barter-type transactions of services for goods.

In a more advanced system, however, transactions are made more flexible and general by a social mechanism that is "narrowly specialized in generality." Such a mechanism we may call a *currency*. In this case, the currency is money, in the form of gold, paper, or checks; and it enhances greatly the flexibility of interchange and the ranges of choice open to the participants. Household members now exchange their labor for money wages; these wages are turned into consumer spending, either at once or after some delay; and for

these consumer expenditures, goods are then obtained. There are thus two transaction flows in each direction, one in terms of the physical things and acts concerned, and the other in its generalized and flexible form in terms of the currency employed. The households put first labor, and later consumer spending, into the economy, while the economy furnishes to the households first money wages and then goods.[12]

The physical flows, and even more the flows of currency, lend themselves to quantitative study in terms of equilibrium analysis, familiar to economists, and their dynamic changes might perhaps be traced fruitfully with the aid of feedback models, which could be used to trace the cumulative shift of the system from certain levels and states to others.

Similar flows of transactions can be envisaged among all the other pairs of major subsystems. Those interchanges may also occur in two or more forms, once, directly in terms of the matters, facilities, or information conveyed or allocated, and second, indirectly through the reciprocal flows of symbols in terms of some "currency," analogous to the economy-and-households case.

SOME INTERCHANGES OF THE POLITICAL SYSTEM

As regards the political system, households may be considered in the most simple case as making specific demands upon the political system. They offer specific support to rulers who in turn use this support to make and enforce binding decisions of the kind desired by their supporters. Thus, in effect, specific support appears exchanged for dependable specific decisions, responsive to specific demands, in a political analogy to economic barter.

In a slightly more extended case, however, the government may assume a generalized leadership role—it assumes *responsibility*—far beyond this or that particular decision; and the population may give it general political *loyalty*—that is, generalized political support and trust—to some extent regardless of the greater or lesser popularity of any one of the government's policies. Far beyond the

former political logrolling, or trading of favors, generalized support is asked for and given in terms of an only partly quantifiable "currency" of responsibility and loyalty, where the political system bestows formal protection, citizenship, or permission to reside, on countable numbers of persons, and demands loyalty and allegiance from every one of them above the age of infancy. (The currencies of citizenship and loyalty are not immune to disorders resembling those of severe inflation. Examples may be found in the wholesale grant of citizenship by the Roman Emperor Caracalla in A.D. 212, or in the "loyalty oath" campaigns characteristic of both ancient and recent periods of political anxiety.) More accurately, countable currencies, such as votes, may appear in additional interchanges of transactions, where political systems give the franchise to certain groups of citizens, and votes are then put by the citizens back into the working of the system.

In the interchange between the economy and the political system, the economy provides the political system with a stock of disposable resources, skills, and expectable levels of productivity and capabilities (for example, the economic "war potential") while the political system provides the economy with guarantees of certain dependable expectations (for example, the protection of the prevailing patterns of private or collective property). Using once again the monetary currency, the economy furnishes taxes to the political systems, and every householder may have to pay an income tax in his secondary role as a member of the economy. The political system, and particularly the government, in turn, furnishes the economy with the regulation and preservation of contracts and credit, and thus with an important part of the control of interest rates and high or low levels and configurations of investment.

Credit and investment policies thus appear, from the viewpoint of this theory, largely as political inputs into the economy—a relationship illustrated by the close connection between public policy and central banking. To the extent that such central banking functions are carried out by wholly private and unsupervised banks, or other financial organizations—or to the extent that such organizations share in these policies—it might be surmised that banks and

carried on with the help of a smaller amount of monetary currency, and a relatively very small amount of gold. The last two mechanisms, gold and cash, thus appear as marginal aids or reserves rather than as the essence or substance of the wealth-producing process.

GOLD AND FORCE AS DAMAGE-CONTROL MECHANISMS

The political system, like the economic system, thus depends to a large extent on the fabric of coordinated expectations. Where and when this fabric is injured, particularly under conditions where strains and tensions are already present, coordination may break down in a spreading pattern, unless the damage is controlled. Gold bullion is the simple damage-control mechanism to stop a financial panic. If enough gold is available, all depositors could be paid off in full; and by making this prospect highly visible the timely arrival and public display of gold shipments may stop an incipient panic in its early stages.

Much in the same way, physical force with its instrumentalities—men with tanks and guns—is a damage-control mechanism of society. It can function as such a mechanism in situations where compliance with legal or political commands has broken down and where non-compliance or open resistance might start to spread. If a government, or an occupying army, commands enough physical force to kill or to coerce all persons disobeying its orders, it enjoys the most primitive kind of political solvency; and by displaying its possession of raw power it may forestall or stop in its early stages a possible chain reaction of disobedience or defiance. The tanks that President de Gaulle ordered into the streets of Paris on a critical day in early 1962 were thus analogous to the gold trucks conspicuously arriving at a bank threatened by a panic among its depositors. Each type of vehicle served to display its own kind of liquidity.

In less extreme crises of financial confidence, paper money or checks may suffice to preserve liquidity; and in less extreme political crises, demonstrations of political support through votes, public

declarations, or demonstrations may be enough to preserve the fabric of confidence in the political process.

Politics, like economic life, depends on human cooperation, based on coordinated expectations. The essence of economic activity consists in the coordinated, flexible, and dependable division of labor—in the productive combination of human skill and effort with technological machinery and with resources drawn from the environment. It is this fabric of flexible, productive cooperation that produces wealth and the capacity to produce more wealth, rather than any amount of paper money, coins, or gold bars. To mistake gold for wealth was the mistake of a primitive school of economic thinkers in the seventeenth century, the bullionists.[14] A less crude error consisted in mistaking money—such as foreign currencies—for wealth, and caused some financial writers to predict, as they did in 1940, that Hitler's Germany was bound to suffer economic collapse within a few months because its gold and foreign-exchange reserves were close to exhaustion. In fact the history of our century has shown that when a large country lacked gold, other damage-control mechanisms often were used to keep its economy working.

Something similar is true of politics. The coordination of human efforts toward the attainment of some goal or goals, set by the society or by any of its subsystems, can be greatly accelerated or facilitated by the use of power, and at particular times and places to some extent by the use of force. This is particularly true on occasions when some more normal machinery of social control has broken down. Troops or police may have to control riots where communal habits of compliance with the law have broken down, or policemen may have to arrest the driver who has shed the habit of stopping at a red traffic light or the juvenile delinquent who is no longer amenable to the controls of school and family.

Yet in the main it is coordinated habits, rather than threats, that keep things moving. It is the traffic engineers and the drivers' habits, aided by the streets, signals, and cloverleaf crossings, that keep the traffic flowing; and the role of fines and arrests is marginal, though it may be by no means negligible. It is the families and schools that educate our children, and the role of truant officers and

policemen, though not negligible, again is marginal. At intersections where there are no traffic policemen, traffic still moves, though sometimes with more difficulty. Where there are no truancy laws, most children still go to school where this is expected in their culture.

A PERSPECTIVE ON POLITICAL THEORY

Power is thus neither the center nor the essence of politics. It is one of the currencies of politics, one of the important mechanisms of acceleration or of damage control where influence, habit, or voluntary coordination may have failed, or where these may have failed to serve adequately the function of goal attainment. Force is another and narrower currency and damage-control mechanism of this kind. Influence and the trading of lightly desired favors—the traditional "playing politics" of American colloquial speech—are still others. All these are important, but each is replaceable by the others, and all are secondary to what now appears from this perspective as the essence of politics: the dependable coordination of human efforts and expectations for the attainment of the goals of the society.

This line of thought might have some implications for the long-standing argument whether the coercive aspects of government may be expected to recede in the long run with the increase in wealth, education, and perhaps in cultural and social integration, so that these aspects of the state might eventually "wither away," as envisaged by Marx and Engels, and more recently and vividly by H. G. Wells; or whether coercive government will have to be with us in all eternity, or as long as mankind lives.[15] The perennial vision of an eventual noncoercive world, so attractive to many early radicals and revolutionists, has become somewhat embarrassing to the bureaucratic rulers of the Communist states.[16] Yet it may not be impractical in principle; and it might be ironic if the most appreciable advances in this direction should come to be taken in the constitutional democracies of the West.

SOME QUANTITATIVE IMPLICATIONS

The interchange model of the political subsystem, and of its relations to the other main subsystems of society, has quantitative implications. The volume of effective demands made on a government can be estimated, and so can the range of matters for which the government assumes overt responsibility, as well as the amount of relevant resources and capabilities at its disposal. For these and similar interchanges, limiting conditions or equilibrium relations may be found to hold as a first approximation. Specific decisions of the government may require a commensurate backing by specific interest groups; and general demands for loyalty may have to be balanced by generalized political support extended by corresponding numbers of the populace.

A more refined analysis might trace the cumulative changes in these reciprocal transaction flows. If households consistently put more labor into the economy than that which corresponds to the consumer goods they take from it, savings and investments may result. If the government assumes responsibility for more matters than are effectively demanded from it, we may have an easily recognizable type of traditional paternalism. If the government on the contrary refuses responsibility for matters that are urgently demanded—such as demands for a tolerable level of living, for health, education, social security, or employment—political alienation may result, with eventual effects on political stability. If most of the specific demands made on a political system are met with specific and adequately rewarding responses, a process of reinforcement learning may increase the number of persons feeling a generalized loyalty to that political community—such as a nation or a federation—or to its form of government, or even to the particular party, group, or person in power.

The rates of many of these processes could be measured or estimated, at least in principle. Savings rates and investment rates are familiar to economists. It might be possible to estimate very roughly the rise in the volume of political demands made upon the govern-

ment from the rate of social mobilization, that is, the rate at which people leave the seclusion of subsistence agriculture and village life, their control of tradition, and the isolation of illiteracy and lack of contact with mass communications. The rates at which people leave these conditions and enter the ambit of the money economy, wage labor, urban life, literacy, exposure to mass media, and partial acculturation to modernity have been in part measured, and average estimates for the over-all process have been derived from them.[17] Similarly, the rate at which the responsibilities of the government are expanding might be estimated from the expansion of the share of the gross national product passing through the government sector, as measured perhaps by the ratio of government revenue to gross national product in different years, or between different countries, ranging in non-Communist countries in the mid-1950's from 9 per cent for India to 47 per cent for Western Germany.[18]

More directly, from voting data and perhaps from sample surveys, estimates could be derived for the political integration ratio, that is, the proportion of persons extending generalized political support to the government or to political parties pledged to such support—including "loyal opposition" parties. Similarly, a political ratio could be estimated as the proportion of people denying generalized support to the country and its type of government, or supporting opposition parties repudiating any such generalized allegiance to the state and its regime.

The rates of political integration and alienation are, then, the ones at which the respective ratios change over time; and a rate and ratio of political neutralization for the indifferent, the apathetic, and perhaps for those paralyzed by cross-pressures, might also be estimated in order to complete this part of the picture.

A POSSIBLE PROGRAM

These few examples of ratios, and of their rates of change, must suffice here to indicate the possibility of eventually making the inter-

change model in its entirety quantitative and in part predictive. A part of this development would be the application of cybernetic concepts to the system, making larger and more explicit use of time variables as well as of probabilistic and statistical considerations. This would mean, among other things, the measurement or estimation of the extent and probable distribution of imbalances in the transaction flows; of the corresponding loads upon the equilibrating or adjusting mechanisms in the subsystems; of the lags, gains, and leads in their responses; and hence of the probable stability and future states of the entire system and its parts.

What has been sketched here implies a very large intellectual program. First of all, it will be necessary to complete the qualitative sketch for all six interchanges among the four major subsystems. This would include the explorations of such matters as legitimacy and its symbols, which may turn out to be currency in some of the interchanges between politics and culture, that is, between the goal-attaining and the integrative subsystems. After this qualitative development of the model, the quantitative steps just sketched would follow. Altogether this adds up to a large enterprise. Parts of it may be carried out soon, while others may take a long time. At this stage, however, even this bare sketch may have some use for purposes of orientation and perspective. It may suggest ways of looking at some problems of political theory here and now; and it may serve to outline at least some of the possibilities of political theory in the future.

Whether the next steps in the evolution of political theory will indeed come from the development of some form of dynamic interchange model no one can know. All we can say is that there is some promise in the prospect and that it seems very much worth trying.

CHAPTER

8

Autonomy, Integrity, and Meaning

By now we have suggested, perhaps, the main elements of a theory of self-determination. All self-steering networks have three basic elements: receptors, effectors, and feedback controls. What "selfhood," or autonomy, such simple networks have is in their feedback controls. If these are destroyed or impaired, steering ceases and is replaced by drifting with external influences, or by coasting on momentum, or by some combination of these two. On the other hand, considerable damage to receptors or effectors may still permit steering as long as the feedback controls remain unimpaired. Even in simple systems of the type of self-steering automata, then, autonomy and integrity have operational meaning: they refer to their main feedback circuits and to their specific location and vulnerability.

AUTONOMY AND THE LOCATION OF CONTROL

More complex systems can change their goals, or "reset" their feedbacks, by interaction with information from their past, stored in particular memory devices. There, autonomy in the long run

depends on memory. Where all memory is lost, where all past information and preferences have ceased to be effective, we are no longer dealing with a self-determining individual or social group, but with a self-steering automaton. The facilities for memory storage, and particularly the circuits of channels for recall, recombination, new storage, and reapplication of memory data are critical here. There is no will, no conation, without some operating memory. The will of individuals or groups can be paralyzed by destroying their stored past information or by disrupting its flow into the system.

Still more complex networks may include processes of "consciousness," of internal monitoring of certain states of the net. Where consciousness exists to a sufficient extent, there it becomes a determining element in the over-all behavior of the system. The critical locations for autonomy are then the monitoring channels and the pools of information fed by them, which together carry the function of consciousness. The autonomy of an individual, a corporation, a social group, a party, or a government can be destroyed, without impairing its memory, by depriving it of consciousness, that is, by cutting the flow of information about the state of its different parts, and disrupting those controls over its own parts that depended on such internal information. Where this has occurred, gross automatic feedback controls may still function, as may the recall and feedback of some remembered data from the past, but the over-all effect might at best resemble that of a man "punch-drunk," or walking in his sleep, or a victim of severe brain injury.

A society or community that is to steer itself must continue to receive a full flow of three kinds of information: first, information about the world outside; second, information from the past, with a wide range of recall and recombination; and third, information about itself and its own parts. Let any one of these three streams be long interrupted, such as by oppression or secrecy, and the society becomes an automaton, a walking corpse. It loses control over its own behavior, not only for some of its parts, but also eventually at its very top.

Or let the monitoring of internal data be unimpaired, let consciousness exist, but inhibit its feedback into the behavior of the system—create a consciousness at once informed and powerless—and you have the pattern of a man who feels himself "possessed," who watches his own behavior in helpless surprise, unable to change it. On the level of a society, this is the experience of Cassandra watching her city rush to its doom. In both cases, the crucial element is not the presence of a new intruder, but the absence or breakdown of an inner channel of control. The patterns of internal feedback channels—the flow charts of information—may be used to evaluate the performance of particular communication systems. They could well be used in the investigation of social groups; of business enterprises; of large research organizations; of parties or governments. In all such cases, they could help to reveal the location of control, the critical connections or configurations of the channels of information and decision that keep the system behaving as it does.[1]

In self-determination, the location of the "self" can be sought at the location of the feedback circuits of the relatively highest hierarchical type. These feedback circuits themselves are never located at a single point; rather they may have a broad topological basis. When the channels of the highest type are lost or disrupted, the self-controlling behavior of the system is lowered to the next lower level, and its more primitive remaining "self" must now be sought at the location of the relatively highest type of feedback circuits that continue to function. Thus, if there is consciousness, selfhood is effectively located at the system of self-monitoring feedback circuits that carry it. If consciousness is lost, selfhood may rest on memory and the feedback of its data into the making of decisions, which we have called "will." If memory is lost, what selfhood remains would have to be sought in the system of feedback channels that connect and control receptors and effectors; and if even this were lost, the last poor remnants of "self-determination" might have to be sought in the reflex arcs. Evidence of what happens in the progressive destruction of the central nervous system would seem to bear out this view.

One test of "functioning" on all these levels would be the capacity to learn, that is, to produce internal rearrangements in the system so as to bring about changes in its behavior. The "self" under this aspect lies in the channels carrying the highest type of learning process of which the system at each stage is capable.

Selfhood, in this view, appears not as a static property but as the functioning of particular sets of channels in a communications system. Self-determination may increase with the increase in the number, effectiveness of organization, and level of type of these sets of channels. This view might well fit in with the suggestion of A. J. Toynbee that growth in organisms and civilizations should be measured by their increase in self-determination rather than by gains in size or complexity.[2]

INTEGRITY AND DIGNITY

Integrity, then, means the unimpaired functioning of the *facilities* that carry the processes of self-determination. The "integrity" of any self-steering system can be impaired either by imposing a non-autonomous change on some of its channels—such as by cutting a channel or disrupting its connections—or by leaving its channels intact, but forcing such traffic loads on some of them as to disrupt their functioning. The familiar case of "overpersuasion" or "high-pressure methods," and the frequent resentment against them, may illustrate the point.

When we defend our "integrity," we are defending our autonomous learning equipment, the structure of the personality we have acquired. When we defend a man's *dignity*, we defend his ability to use his personality: we defend him against the imposition of an intolerably high speed of learning, an intolerable speed of changing his behavior—intolerable, that is, because incompatible with the continuous functioning of his self-determination, his autonomous learning. A man's dignity is impaired, in this view, if he is forced to do something that deprives him of autonomous control over his own behavior, and makes him instead into the

object of another process, a "means" instead of an "end," in Kant's terminology. The more sweeping the loss of self-control, the deeper the loss of dignity.

Dignity can be lost or impaired temporarily. It is the suspension of a process of self-determination that later may be resumed again. Integrity when lost or impaired implies a permanent change in inner structure. Autonomy then may be regained, but most likely it will not be quite the same as before.

Concepts of "human dignity," "integrity," and "worth of the human personality" have had an important political and emotional appeal. They have been incorporated in the United Nations Charter and in the United Nations Declaration on Human Rights. They have been criticized sometimes as being vague, but it appears now that they can be given explicit and operational meaning: respect for every man's right to learn at his own speed and with his own inner equipment, in an unbroken sequence of autonomous acts of learning, in which his own unique stored past and his own acquired preferences at every single step have at least some share in the outcome. Dignity is nondisruptive learning. Integrity is undisrupted or unimpaired inner learning equipment. In catastrophes or catastrophic emergencies these may not always be compatible with survival. Where the power of events forces men to learn at breakneck speed, two types of casualties may result: the proud, stiff-necked men of the "old school" who "nail their colors to the mast" and kill and die; and the spineless opportunists who survive at the price of living for nothing but survival.

Perhaps the solution of the problem rests in an increase of inner facilities for continuous undisrupted adjustment at high speeds. To permit a high speed of learning and wide changes of behavior, without loss of inner structure and of an effective past, is perhaps best accomplished by increasing the inner communication channels in variety, flexibility, and numbers.[3] To ensure continued self-determination, integrity and dignity are not enough. In the language of religion, pride may mean death where a change of heart may mean survival. In less exalted terms, the best way to strengthen a communications system against the impact of large external changes may

well consist in enriching its internal structure and its range of possible new configurations.

A CONCEPT OF MIND

Perhaps a concept of mind may be tentatively suggested from this very sketchy survey of patterns of communications. *Mind* might be provisionally defined as any self-sustaining physical process that includes the nine operations of selecting, abstracting, communicating, storing, subdividing, recalling, recombining, critically recognizing, and reapplying items of information.

Discrete items of information moving through a net may be called "messages." Physically, each message is a reproducible pattern of changes of state of parts of the net, regularly followed by determinate processes depending on that pattern.[4] Any message that has acquired a relatively stable association with an event outside the net, or with another message within it, may function as a symbol; and a mind may be considered a self-preserving physical process of communicating and manipulating symbols.

Such a physical process may include the production of novelty and of initiative. If information from events in the outside world is abstracted and stored analytically, that is, if information about some of its different parts or aspects is stored separately (such as remembering not only a whole bird but also separately its wings), then these may be separately recalled and recombined into new patterns that did not exist in the outside world (such as imagining Daedalus, a man with wings, three thousand years before the first airplane). To complete the production of *novelty*, this new combination of old elements must itself be matched or abstracted in the mind: a new image or symbol must be stored, *pertaining to the new pattern as a whole, regardless of its earlier combinatorial origin.*

Creativity is consummated in this second step. Both creativity and eclecticism recombine old elements, but while eclecticism does no more than that, creativity abstracts and reapplies new patterns from the combination. After this secondary abstraction and storage, the

new pattern may be applied to new recombinations within the mind, or more directly to new patterns of action by the effectors of the net. This latter result of novelty, the beginning of some new item of behavior, we may call initiative. More important instances of initiative may concern whole sequences of behavior, where novelty would lead to changes in more general patterns of preference and value, and thus change the probabilities of whole classes of future actions.

The combinatorial richness of possibilities for novelty may already be vast within a single mind, vaster still in any mind that is in communication with others, vastest perhaps in a mind that is *open*, that is, that applies *initiative to the widening of its range of intake* of information from what we may still believe to be an infinite universe. The more capable a mind is of such *creative learning*, the more new kinds of information can be reached by its self-steering, abstracting, and combinatorial powers, the more properly it may be considered inexhaustible. Indeed, the only way a mind has to keep itself inexhaustible may well be to keep itself open in this sense.[5]

As a physical process of communication, mind depends on a set of physical facilities, circular communications channels, and the like, which are themselves physical processes of a simpler order. In this sense, there is no mind without "body." Processes constituting mind may occur in the body of a single human individual, or in the communication among several individuals, and their stored records of past information, in a social group, organization, culture, or society. There are mental processes that go beyond the capacity of any single individual, discoveries that may be the result of mental teamwork extending over several generations.

If we are to ascribe tentatively any meaning to such concepts as "group mind," "collective personality," "ideology," or "configuration of culture," then the concept of mind we have outlined so far may suggest observations and experiments to test it. Where and how in such groups is past information stored? How is it recalled, how applied to new decisions? Is there *group* learning, that is, does new information lead to significant changes in the structure of the group? If answers to these and similar questions should confirm

the notion of social mind, or group mind, as meaningful, it would follow that there can be plural membership of minds—that one individual or smaller group can participate in several self-steering communications networks that generate and process thoughts.

The process we have called "mind" depends on a particular configuration of physical facilities, but it does not depend on the preservation of any one of these facilities, so long as it is replaced by another without destroying the configuration. In this manner stones are replaced in repairing a cathedral; wires and relays are replaced in a telephone exchange; cells are replaced in a living body; men live and die and are replaced in their functions in a society—all without necessarily destroying the relevant configurations of these structures. If the last two structures named, the individual and the society, are also carriers of processes of mind, then mind may be found to persist over any number of material changes in their parts—including conceivably the replacement of every single part—provided only that the configuration of the whole remains preserved. Mind, in this view, depends among other things on a particular configuration of "body," within which every single part may be replaceable.

If this view is accepted, the greatest practicable speed and scope of safe replacements of the material parts that carry the processes of a mind become a question of empirical fact. Whether more radical transfers of mind are possible from one set of physical facilities to another is not known at this time, but this, too, appears now as a question of fact. As such, it can only be answered by observation and experiment. It may be noted, however, that there seems to be no deductive principle of science or philosophy incompatible with such a possibility. All such transfers would have to occur by physical processes, from one set of physical facilities to another.

Some fairly complex patterns of information are already capable of being transferred from one set of facilities to another. The sound of a symphony may be transferred from the score to the orchestra, to the electric recording device and, finally, to the phonograph record; thus recorded it may "survive" its composer and the musicians of the orchestra. We cannot say whether a similar preservation by an

agency in the universe would be possible for the vastly more complex patterns of a higher order, such as the patterns of living tissue or the patterns of mind, but in the light of today's science of communications there seem to be no deductive considerations opposing this possibility, as there would have been in the science of the nineteenth century. It is, of course, quite possible that such fundamental considerations will be discovered, precluding the possibility of preserving or transferring mind, or permitting it only within definite limits. All we can say now is that the question is again open, in a way in which it was not open thirty years ago.

Since mind depends on physical facilities, it necessarily depends on a process by which such facilities are preserved: it thus depends on *life*. Organic or social life can be called, from one point of view, a self-preserving, self-reproducing, and self-modifying autocatalyst, or more generally, a self-preserving, self-reproducing, and self-modifying structure of material processes.

If we imagine a complex machine, or set of machines, capable of self-repair and self-maintenance, capable of seeking out and utilizing sources of energy and materials in its surroundings, and even of producing and storing sufficient supplies of spare parts to build functioning replicas of itself, with lesser or greater modifications, then we should be imagining something close to life. It would be capable of metabolism, reproduction, and even under certain conditions—through chains of variously modified replicas more or less suited to their environments—capable of evolution guided in part by natural selection. Yet it would not have mind. Its incredibly complex processes would be wholly taken up with merely maintaining their own ensemble. It would use information from the outside world only for immediate purposes, to locate needed supplies. It would have no facilities for storing or recombining information, which are essential functions for mind.

No such machines could be built now—although parts of them could be designed, and trees in nature might come close to their performance—but our "experiment in imagination" may have served to clarify a basic relationship: life can occur without thought, because channel maintenance can continue without additional in-

formation feedbacks,[6] but mind cannot occur without life, because information feedbacks cannot continue without channel maintenance.[7]

Mind is not the configuration and maintenance of communication channels and storage devices; it is the pattern of information flow within them. Injury to channels may disrupt the information flow, much as bodily injury may damage mental health in people. Even with all physical facilities intact, information flow may be disrupted by the rise of "pathological" traffic patterns, much as traffic jams in a city may disrupt traffic flow without any change in the layout of the streets, or as certain undamaged electronic computers can be deadlocked or thrown into circular sequences of operations by feeding them problems involving paradoxes. In the latter case, traffic overloads may ultimately damage the channels: the disruption originates in the information flow and later spreads to the channels, resembling what has been called "psychosomatic illness" in man. Apart from such eventual "bodily" damage, it would appear that functional mental disturbances may occur in an individual in the course of his history without necessarily being preceded or accompanied by bodily changes.

Certain sets of channels, such as calculating machines, can be cleared of all information flow, and a new run can be started unaffected by anything that went before. Similarly, a city could be evacuated, and a new traffic pattern substituted for the old. Other communication systems, however, such as the human brain, are never wholly cleared; they function like "single-run machines." In fact, Norbert Wiener has suggested that the analogy to man should not be a calculating machine, but rather a single run on one.[8] This point may be put more generally: Mind is not a machine, but a run. It is a single-run pattern of information flow.

The nature of mind thus implies individuality. Two runs on two identical electronic calculators may go through different channels; or two calculators, exact copies of each other when fresh from the factory, may become different in operating characteristics if their memories acquired different data during their run, and remain different until they are cleared. If two minds operate over longer

periods in a nonstandardized environment, it would seem almost impossible for them not to become different, even if their "bodily equipment"—their channel facilities—had been exactly identical at the outset. The development of personality differences between identical twins would seem to confirm this view.

What has been said about mind thus far may suggest certain perspectives. If mind can be defined as a particular class of physical processes, then there seems to be no compelling reason for assuming a priori that processes of this type are limited to mankind and to this planet. The more we understand the characteristics of the process we call mind, the more plausible it may appear that processes of mind and thought may exist more widely in the universe, as do such simpler physical processes as crystals, storms, and flames, or as catalysis or cyclical nuclear reactions, and as biological and ecological processes.

We may perhaps infer from the study of entropy that the occurrence of mind will be relatively rare: processes of a high degree of order are statistically less frequent than those involving less order (that is, more entropy). The known facts likewise suggest that nuclear processes are more frequent in the universe than chemical processes, chemical processes more frequent than biological ones, and processes of life more frequent than processes of mind.

Yet by these same considerations, mind, though relatively rare, may be expected to exist widely in the universe. Centuries ago, the work of Copernicus and Giordano Bruno gradually brought home to men's imagination the fact that in astronomy the earth was not unique but was one world among many. Today the thought that man's mind is not alone in the universe, but represents one type of mind among many—a thought long vaguely held in many versions by religious or philosophic intuition—may well be on the verge of entering the fabric of rigorous and scientific thought.

A second perspective follows from the first. If mind is the name of a class of processes, united by significant common characteristics, then there should exist certain regularities or "laws" applying under suitable conditions to all members of the class. We may therefore expect the existence of some "laws of mind," as we expect the exist-

ence of laws of gravitation or combustion or biology, and expect them to hold good for any kind of mind regardless of the particular physical processes that make up their internal channels of communication. Any mind, for instance, must have memory, autonomy, and individuality. It must somehow balance present information from the world outside it against stored information from its own past, recalled from memory. It cannot operate without preferences or values. What we have said above about selfhood, integrity, and dignity will apply to it; so will the connection between creative and pathological learning.

It is well known that much in the ethical norms of the world's great philosophies and religions is similar, or even identical. Such norms of ethics may now be compared carefully and critically with what we may call "laws of mind": testable, predictive statements about the probable self-destruction of minds following upon certain patterns of overt or covert behavior, both on the level of the individual and of the community or social group. We have long had concepts of "mental health" and "moral health" and some intuitions about their connection. It should eventually become possible to fill these general concepts with more specific structural detail, testable by operations, and of more general validity.

From the foregoing, it will appear that cybernetics could offer social scientists not merely an approach toward a theory of self-determination but also an approach toward a theory of *growth*, and particularly of the growth of minds—of human personalities and organizations. The two concepts of self-determination and growth are connected, and a recent philosopher of history has considered an increase in self-determination to be the most essential criterion of growth.[9] Nevertheless, it may now be suggested that the concept of growth should go further.

Growth should mean not merely the highest degree of unity and self-determination within the existing limits of a system (which Parsons might call "integration"); nor should growth mean a mere enlargement of the system with no change in its characteristics of performance (that is, growth in terms of Parsons' "pattern maintenance"). Rather, growth also should mean an application of learning

capacity toward an increase in openness, that is, an increase in the range, diversity, and effectiveness of an organization's *channels of intake* of information from the outside world (similar to Parsons' "adaptation"). Still further, growth should mean an increase in an organization's ability to make effective responses to its environment and to change this environment in accordance with its needs, that is, to fulfill Parsons' function of "goal attainment." And, finally, growth should mean an increase in the range and diversity of goals the organization is able to follow, including the power to change goals and to add new ones. The third of these tests of growth—openness and adaptation—has long been known by philosophers and religious leaders under the name of humility. The last, the ability of an organization to change its goals and not to remain a prisoner of some temporary goal, has sometimes been spoken of as the ability of man to avoid the "idolization of ephemeral institutions."[10]

SPIRIT AND MEANING

A final note may be added. The concept of "mind" is frequently associated in some vague manner with the concepts of "spirit" and of "meaning." From the point of view of communications, this empirical association may be spelled out in more specific terms. While no mind can operate without values, *spirit* denotes second-order value. It is the set of preferences about sets of preferences. A man, a people, or an epoch are among other things also systems; the *spirit* of a man, or a people, or an epoch is the configuration of rules according to which their value systems are patterned and operated. Spirit is related to values as strategy to tactics or as policy to operations. A change in "spirit" means, therefore, a strategic change in the patterns of behavior. And, under suitable conditions, such a change can be communicated.

As "spirit" deals with the internal patterns of a mind, "meaning" deals with those of its wide context. When we ask for the *meaning* of a series of events, we are trying to abstract some pattern from

them that applies both to the continuation of this sequence itself and also to the continuation of another series of a higher logical type, that is, a series abstracted from a sequence of physical events on a larger scale in space or time, so that this second series comprises within itself the continuation of the first. More briefly, meaning is context. To recognize meaning is to recognize similarities in a series of one logical type, and to recognize further its extrapolation in another series at least one logical type higher.

Meaning, in this view, is physical position in a sequence of events. Meaning, therefore, is always relative. There are as many meanings as there are levels of logical type. And there are as many logical types as there are physical contexts, that is, objective sequences of physical events. In this sense "meaning" is real, whether perceived by an observer or not: the stone placed by a small boy on a railroad track may "mean" an accident in the context of the railroad, regardless of whether the boy was aware of this or not; at the same time it may mean a subsequent change in other contexts, such as in the education of the boy, or in his personality development, and these again need not depend on his knowledge, although they may be influenced by it.

Events may be "given meaning," therefore, in two ways: symbolically, by attaching symbols to them and to the context in which they already actually function; and physically, by putting an event physically into a context, that is, into some larger series of events. These two ways, of course, can, and often do, occur together. The first of the two, the imputing of context through symbols, is often useful for purposes of prediction. It is subject to error, and to verification, much as other statements about interactions among events in nature or society. The second—physical participation in a context—seems essential to the continued functioning of any mind: its stored information would become ultimately meaningless if it became closed to all further information from any context larger than itself.

It seems clear from our earlier description of minds that the imputation of meanings to events, and the verification of the imputed

meanings, are a major part of their activities; and, in the case of conscious minds, that this search for meaning will include their own position and their own activities in the context of their surroundings and in the even wider contexts of the universe they surmise or discover.

PART III

Communication Models and Political Decision Systems