The State of the Social Sciences

Papers presented at the 25th Anniversary of the Social Science Research Building The University of Chicago November 10–12, 1955

Edited by LEONARD D. WHITE

Simon, on computer programs [illegible]

© 1956

THE UNIVERSITY OF CHICAGO PRESS
CHICAGO AND LONDON

Table of Contents

HERBERT A. SIMON[1] *and* ALLEN NEWELL

Models: Their Uses and Limitations

In contemporary usage the term "model" is, I think, simply a synonym for "theory." I am to speak, then, on "Theories: Their Uses and Limitations." This is a topic I can handle very briefly: the uses of theories are obvious, and their only limitations are that they are often bad theories.

However, the persons who arranged this meeting did *not* presumably intend that "model" should mean simply "theory." I suspect—but it is only a suspicion—that by "model" they meant "mathematical theory," and they intended to exhibit in this arena another instalment of the prolonged guerrilla warfare between mathematics and language.

With respect to these hostilities, I have two comments. First, I stand with J. Willard Gibbs: "Mathematics *is* a language"—and, to my ear, the most dulcet of languages. Second, I do not believe that the form in which we clothe our thoughts is a matter of indifference —or even of taste, as my last comment may seem to imply. It may be true that words without thoughts never to heaven go; but the converse is equally true: wordless thoughts, too, are earthbound. The matter has been put very well by Roget, the author of the *Thesaurus.* In the Introduction to his work he has this (as well as many other wise and even profound things) to say:

> The use of language is not confined to its being the medium through which we communicate our ideas to one another; it fulfills a no less important function as an *instrument of thought;* not being merely its vehicle, but giving it wings for flight. Metaphysicians are agreed that scarcely any of our intellectual operations could be carried on to any considerable extent, without the agency of words. None but those who are conversant with the philosophy of mental phenomena can be aware of the immense influence that is exercised by language in promoting the development of our ideas, in fixing them in the mind, and in detaining them for steady contemplation. Into every process of reasoning language enters as an essential element. Words are the instruments by which we form all our abstractions, by which we fashion and embody our ideas, and by which we are enabled to glide along a series of premises and conclusions with a rapidity so great as to leave in the memory no trace of the successive steps of the process; and we remain unconscious how much we owe to this potent auxiliary of the reasoning faculty.

1. Since most of the notions discussed here are by-products of my collaboration over the past several years with Allen Newell, I have asked him to permit me to present this paper as a joint product—which it is. For infelicities of form and manner, and for downright errors, I alone am responsible.

If we interpret the term "word" literally, then Roget is probably wrong. But if he means that the form of our thought exercises a great control over the course of that thought, he is almost certainly correct.

To select a suitable language with which to wing our thoughts, we must understand what languages there are, and we must be able to compare them. In this paper I should like to discuss three main kinds of scientific languages or theories: the mathematical, the verbal, and the analogical. It will appear from our analysis that these three kinds of theory are really indistinguishable in their important *logical* characteristics; hence, that the choice among them must be based on certain *psychological* criteria. And since analogies, employed as theories, are somewhat less well understood than either verbal or mathematical theories, I shall devote the last part of the paper to two important current uses of analogies.

Before we can plunge into the comparison, however, we will need a clearer understanding of the nature of theory, and of these three types of theory in particular. The next two sections will be devoted to these preliminaries.

Models and the Modeled

It will be convenient for our purposes to define a theory simply as a set of statements or sentences. (They may, of course, be mathematical statements, or equations, instead of verbal statements.) It is important to observe that this definition refers to the form in which the propositions are clothed, that is, the actual explicit statements set forth. Thus, we distinguish the theory, so defined, from its *content,* to be defined next.

By the *content* (or logical content) of a theory I shall mean the totality of the empirical assertions that the theory makes, *explicitly or implicitly,* about the real world phenomena to which it refers.

That is, the content of a theory is comprised of all the assertions about the world, whether true or not, that are explicitly stated by the theory or that can logically be inferred from the statements of the theory.

Consider now some body of phenomena, and imagine that there is a theory whose content tells the truth, the whole truth, and nothing but the truth about these phenomena. By this I mean that any statement that is true of the phenomena is stated in or derivable from the theory and that any factual statement contained in or derivable from the theory is true of the phenomena. Then we may define the *content of the body of phenomena* as identical with the total content of this particular theory.

The particular theory I have just mentioned is, of course, nonexistent for any actual body of phenomena. The theories that actually occur do not have the same content as the phenomena to which they refer. They do not tell the truth—or at least they do not tell the whole truth and nothing but the truth.

The most conspicuous inadequacy of theories is that they do not tell the whole truth; they have a very much smaller content than the phenomena. Borrowing a term from statistics, we may call these errors of omission "Type I errors." But I think it can be shown that almost all theories also err in the other direction—they say things that are not so, as well as failing to say things that are so. Their errors of commission we may call "Type II errors." To the extent that theories commit errors of Type II—asserting some things besides the truth—they have, of course, a larger total content than the phenomena.[2]

The notion of content that I have introduced relates to the *logical* properties of a theory—to the facts that can be extracted from it by applying the laws of logic. Of at least equal importance to the scientist is its *psychological* or available content—the empirical propositions that the scientist is in fact able to derive from it. One theory can have exactly the same logical content as another but be

2. The ideas discussed in this section have been developed in a somewhat different manner by W. Ross Ashby in his book, *An Introduction to Cybernetics*, in which he discusses the relations among theories and between theories and phenomena by use of the concepts of isomorphism and homomorphism. We are indebted to Dr. Ashby for making his work available to us in preliminary mimeographed form. The printed edition of his book is to appear shortly.

infinitely more valuable than the other if it is stated in such a way as to be easily manipulated, so that its logical content is actually (psychologically) available to the inquirer.

For example, one theory (a trivial one, but one that will illustrate the point) tells me that the number of years from the birth of Christ to the Hegira is DCXXII; and from the birth of Christ to the present, MCMLV. A second theory tells me that the former interval is 622 years; the latter interval, 1,955 years. From the second theory, I deduce readily that it is 1,333 years from the Hegira to the present; from the first theory I also deduce, but much less readily, that the interval from the Hegira to the present is MCCCXXXIII years.

The distinction between the logical and the psychological content of a theory helps us to understand Roget's assertion that language gives thought its wings. Man is not an omniscient logician; he is an information-processing system—and a very limited one, at that. The logical content of a theory is of use to him only to the extent that he can make that content explicit by manipulation of the theory as stated. All mathematics (and verbal logic, to the extent it is rigorous) is one grand tautology. The surprise that is occasioned by the Pythagorean theorem derives from the psychological properties of mathematics—from the new information obtained by processing the explicit statements of the mathematical theory—not from its logic.

Three Kinds of Theories

In the preceding section I have introduced the notion of the content of a theory and the important distinction between logical and psychological content (i.e., between what is inferrable "in principle" and what we can actually succeed in inferring). I have pointed out that theories can and do make errors of omission (Type I errors) and errors of commission (Type II errors). I should now like to characterize several types of theories in the light of these distinctions. I shall use as an example certain phenomena that are of central importance in economics: national income, investment, saving, consumption, and similar variables that occur in "macroeconomics." I will distinguish three kinds of theories:

1. *Verbal theories.*—An example of a statement in such a theory is: "Consumption increases linearly with income, but less than proportionately."

2. *Mathematical theories.*—The approximately corresponding statement in the mathematical theory is: "$C = a + bY;\ a > 0;\ 0 < b < 1$."

3. *Analogies.*—The idea that the flows of goods and money in an economy are somehow analogical to liquid flows is an old one. There now exists a hydraulic mechanism, the Moniac, designed in England and available in this country through Professor Abba Lerner of Roosevelt College, one part of which is so arranged that, when the level of the colored water in one tube is made to rise, the level in a second tube rises (*ceteris paribus*), but less than proportionately. I cannot "state" this theory here, since its statement is not in words but in water. All I can give is a verbal (or mathematical) theory of the Moniac, which is, in turn, a hydraulic theory of the economy.

The three types of theory I have just illustrated by an economic example could have been equally well illustrated by psychological or sociological examples. Corresponding to Guthrie's verbal learning theory we have Estes' mathematical counterpart, and a number of robots have been constructed by Shannon, Grey Walter, and others incorporating Pavlovian conditioning and associational learning.[3] Homans (*The Human Group*) has constructed a verbal theory of group behavior which I have mathematized. As far as I know, no electromechanical analogue has been constructed, but it would be extremely simple to make one if the task struck anyone's fancy.

Verbal, mathematical, and analogical theories represent, I think, the main kinds of theories there are, but it is of interest to consider a few special cases to see where these fit into the classification.

Geometrical theories appear at first glance to be mathematical theories. Thus, we can represent income and consumption as the abscissa and ordinate, respectively, of a graph, and represent the postulated relation by a straight line with a positive slope of less than 45 degrees cutting the ordinate above the origin. However, if we look at the matter a little more closely, we see that this geometrical theory is really a mechanical analogue of a mathematical theory, for we do not usually employ geometry in a rigorous axiomatic way but instead draw diagrams—which are, of course, actual physical objects in a space that we hope is approximately Euclidean. (For the benefit of non-economists, I should observe that most so-called "non-mathematical" economists are, in fact, mathematical economists who prefer arithmetical and geometrical analogues to algebra, calculus, and set theory. There are few verbal economists in the strict sense.)

3. See Robert R. Bush and Frederick Mosteller, *Stochastic Models for Learning* (New York: John Wiley & Sons, 1955), and W. Grey Walter, *The Living Brain* (New York: W. W. Norton & Co., 1953).

Computing machines that have been programed to represent a particular theory constitute a slightly more complicated case.[4] In a so-called analogue computer there is generally a one-one correspondence between the circuits of the computer, on the one hand, and the equations of a mathematical theory of the phenomena, on the other. In the special case of a simulator there is a direct correspondence between the analogue and the phenomena. In addition to the Moniac, mentioned above, which can be considered a hydraulic simulator, Strotz and others have used electrical analogues to represent the theory of macroeconomics.[5]

In the case of the digital computer—of which most modern general-purpose electronic computers are examples—there is no direct correspondence between the computer circuits and particular features of the phenomena. First, a mathematical theory of the phenomena is constructed, and then the computer is programed to carry out the arithmetic computations called for in the mathematical theory. Thus, the computer is an analogue for the arithmetic process. *This is not, however, the only way of employing digital computing machines as theories*—an important point to which I shall return later.

Verbal and Mathematical Theories

We are now in a position to compare verbal and mathematical theories with respect both to their content and to the availability of that content to the theorist. At the very outset we are confronted with a paradox. It is usually argued that mathematics has certain logical advantages against which must be weighed its psychological disadvantages. A closer examination of the case shows that the truth is almost the exact opposite of this. In the arguments ordinarily

4. To program a computing machine is to instruct it as to what it is to do in sufficient detail (and in an appropriate language) so that it can execute its tasks.

5. See Arnold Tustin, *The Mechanism of Economic Systems* (Cambridge, Mass.: Harvard University Press, 1953).

used to compare the relative virtues of mathematics and words as languages, we find that the advantage claimed for mathematics is, in fact, largely psychological, while the advantage claimed for words is largely logical.

The mathematician is aware how difficult it is to squeeze more than an infinitesimal part of the logical content out of verbal theories, because of the awkwardness in their manipulation. On the other hand, the verbal theorist (assuming he knows enough mathematics to understand the issues) finds that the logical content of most mathematical theories is quite small compared with the logical content of verbal theories. I do not say that this is the only issue between the mathematician and the non-mathematician, but it is certainly one issue that is often stated explicitly.

Now, I am not a neutral in this particular dispute. I believe that the psychological advantage claimed by mathematics is real and vitally important and that the logical advantage claimed for words is often illusory. With respect to the psychological difference—the importance of ease of manipulation—my example of Roman and Arabic numerals will provide, perhaps, some food for thought.

The logical difference—the relative logical content of verbal and mathematical theories—requires additional comment. It can be verified, I think, by the examination of almost any verbal theorizing that makes claims of rigor that only a very small part of the logical content of the theory is or can be employed in the reasoning at any one time. It is almost impossible to handle more than two or three simultaneous relations in verbal logic. Hence verbal reasoning (i.e., manipulation of theories stated in verbal terms) is replete either with logical gaps, or with *ceteris paribus* assumptions, or with both. For this reason the potential advantage derivable from the rich logical content of verbal theories is almost entirely lost by their intractability. The incompatibility of the theory with the information-processing skills of the scientist makes most of this logical content inaccessible to him.

Let me illustrate. Suppose we wish to theorize about the lynx and rabbit population in Canada. Lynxes eat rabbits; hence, if the lynx population is very large relative to the rabbit population, the latter will presumably decrease. On the other hand, if the rabbit population is too small, the lynxes will have a hard time finding a square meal and will also decrease in number. Now I should like a verbal theorist to predict for me the outcome of this competition. Will the lynx population become extinct, will the rabbit population become extinct, or both? Or will both species increase in number? And, if a large number of squirrels is introduced (which lynxes also like to eat), will the rabbit population increase or decrease?

There is a perfectly good mathematical theory, due principally to Volterra and Lotka (*Elements of Physical Biology*), that answers all these questions in a definite manner. (Roughly, the answer is that under reasonable assumptions there will be cyclical fluctuations in both lynx and rabbit populations; neither will become extinct; and under most assumptions the introduction of squirrels will decrease the rabbit population.) This theory has also been fitted to the data and has been found to hold reasonably well.

Now, of course, the illustration I have used is biological and has nothing to do with social phenomena. But Lewis F. Richardson has produced a mathematical theory of armaments races that is closely analogous to the lynx-rabbit theory. And he has been able to show the conditions under which an armaments race is consistent with peace and the conditions under which it leads to war. Moreover, the theory has been tested, to a certain extent, against data.

Other examples can be supplied;[6] but I do not wish to appear more partisan than I feel. The construction of good theory is such an arduous task at best that it is foolish to tie our hands behind our backs by limiting the range of tools that we utilize.

Of all three types of tools—words, mathematics, and analogies—analogies are perhaps the least frequently used and certainly the most poorly understood. Instead of continuing a discussion of the more familiar verbal and mathematical theories, I should like to turn my attention to the problems and possibilities of making fruitful use of analogy in social science theory.

6. I invite comparison of my mathematical "Homans model" ("A Formal Theory of Interaction in Social Groups," *American Sociological Review*, April, 1952) with the verbal theorizing by Henry W. Riecken and George C. Homans on the same subject in the *Handbook of Social Psychology*, ed. Gardner Lindzey, Vol. II, chap. xxii. A similar comparison may be made between the model that Harold Guetzkow and I have constructed of Festinger's theories of social influence ("A Model of Short- and Long-Run Mechanisms Involved in Pressures toward Uniformity in Groups," *Psychological Review*, January, 1955), and the verbal theory of Festinger himself ("Interpersonal Communication in Small Groups," *Journal of Abnormal and Social Psychology*, January, 1951).

Analogy as Theory

Analogies are the object of considerable distrust. An important reason for this distrust is that there have been some prominent examples in the not-too-distant past of their gross misuse—for example, Spencer's analogy between society and an organism and his uncritical social Darwinism.

I believe that the usual reason given for distrusting analogies (as contrasted with other theories) refers to logical content: analogies cannot be depended on to tell "nothing but the truth," while theories, it is alleged, can. That is to say, theories may be lacking in content, and hence be guilty of making Type I errors; analogies, on the other hand, have a great deal of content that has no correspondence with the phenomena—their serious errors are of Type II.

It is undoubtedly true—and Spencer's theory is only one of many examples that could be cited—that analogies are particularly susceptible to Type II errors. *But I believe it can be shown that verbal and mathematical theories are also susceptible to such errors.* The exaggerated use of the concept of instinct, for example, that characterized one period in the history of psychology can be traced simply to difficulties in handling the nature-nurture distinction in a verbal theory. The tendency of Freudian theory to proliferate mental entities—the id, the ego, the superego—probably has something to do with the preference of our language for nouns over verbs.

Why should theories of all kinds make irrelevant statements—possess properties not shared by the situations they model? The reason is clearest in the case of electromechanical analogues. To operate at all, they have to obey electromechanical laws—they have to be made of something—and at a sufficiently microscopic level these laws will not mirror anything in the reality being pictured. If such analogies serve at all as theories of the phenomena, it is only at a sufficiently high level of aggregation.

A little reflection shows that the same is true of verbal and mathematical theories, but in a more subtle way. These theories must be fitted to a particular computing device—the human brain—and at a sufficiently microscopic level a theory will more closely mirror the neurological and psychological properties of that information-processing system than it will anything to be found in the outside world.

This was observed a long time ago by nominalistic philosophers, who noted that Aristotle's *Prior Analytics* bore a suspicious resemblance to Greek syntax. The same observation is the foundation of Kant's synthetic a priori and of modern phenomenology.

At this point you may wish to object. The theory, you will say, does not consist of the individual letters or words. It is the *meaning* of the statement or equation that contains the theory, not the mounds of ink or the neural circuits that are its physical embodiment.

Even if we were to change our definition of "theory" to agree, the same could then be said of the analogy. It would then not be the water and glass tubes of the Moniac that constituted the theory but rather the relations among variables that these exhibit. If propositions and equations live in the Platonic heaven of ideas, why cannot their earthly representatives be constructed of glass and water as well as of paper and ink?

The truth seems to be that we are accustomed to words and equations as analogies; consequently, we do not often mistake the paper and ink, or even the grammatical structure, for the meanings that are supposed to model the phenomena. Few of us are any longer convinced by the ontological argument—one of the classical Type II errors of verbal theory. Gradually over the centuries we have acquired the sophistication in handling words and equations that is essential to avoid errors of this kind.

We are not so accustomed to non-verbal analogies and particularly to electromechanical ones; hence, we do sometimes mistake the irrelevant properties of the analogy for parts of the theoretical model. But if analogies are intrinsically useful devices as vehicles of theory, this difficulty is certainly one we can learn to overcome.

In specific terms, the argument amounts to this: The content of the theory embodied in the Moniac is identical with the content of the theory embodied in the corresponding set of Keynesian equations or the corresponding set of verbal statements. All three are simplified aggregated theories of the economy, having virtually the same *logical* content. If we have a preference for one of these theories over the others, it must rest not on logical grounds but on information-processing considerations—the relative ease with which the theory can in fact be manipulated in order to extract from it the implicit logical content.

The relative power of words, equations, and computers to convey, psychologically, their logical content has to be determined case by case; and the answer, even in specific applications, may well depend on the time at which it is asked. For the ease with which a mathematical system or an electromechanical analogue can be manipulated will depend heavily upon the current state of the mathematical and computer arts, respectively. (For the last two thousand years no comparable progress has been made in the verbal art.)

I should like to devote the final portion of this paper to a discussion of the probable fruitfulness, as matters stand in the year 1955, of two analogical theories that have recently received considerable attention. These analogies present fresh and novel problems of methodology that should illuminate our general analysis. My first example will be the "natural" analogy—the organism analogy—that has been advanced by the proponents of general systems theory. My second example will be the digital computing machine as an analogy for human thought processes.

General Systems Theory: The Organism as Analogue

The premise that underlies the advocacy of "general systems theory" or "general behavioral systems theory" is that each of the classes of things designated a *system* is an analogy, in significant respects, of each of the other classes.[7] So an organism may be regarded as an analogue to a cell or to an organization, or vice versa. Our question is whether these analogies are likely to be useful, and whether some of the fallacies of earlier theorizing of this kind (e.g., Spencer) can be avoided.

First, let us specify the conditions of the problem. We suppose that there exists a theory, at some stage of development, for each of the classes of things we call "systems." There is a theory of cells, a theory of organisms, a theory of species (meaning by this term an organism and its descendants), a theory of groups, a theory of

7. James G. Miller, "Toward a General Theory for the Behavioral Sciences," *American Psychologist*, September, 1955.

organizations, and a theory of societies. Each of these theories may include verbal, mathematical, and even electromechanical components. Each theory has a definable logical content.

A general systems theory is feasible to the extent that these several theories, as they develop, have common content. This is obvious enough. If there are no statements about cells that are not true also (with appropriate changes in correspondences) of organizations, then a general theory that embraces both classes of systems is simply not feasible. Miller, in his paper at this session, mentions a number of propositions that apply, or might apply, to all classes of systems. As he points out, whether they do in fact apply is a question that has to be answered by empirical research.

But if a general systems theory were feasible, would it be useful? I think this is simply a question of economy of learning, specifically, a question of transfer of training. The relevant questions are these: (1) How easy or difficult is it to set up a correspondence between the elements of, say, an organism and the elements of an organization? How does the effort required compare with the conjectured common content of the two systems? (2) Has one or more of these bodies of theory evolved so much beyond the others that it would prove a cheap mode of discovery to borrow from the former in order to add to the latter?

On the whole, I suppose I am rather skeptical with respect to the first question and a bit more sanguine with respect to the second. This reaction probably reflects little more than my own habits of thought and my desire to appear a man of Aristotelian moderation. However that may be, I remain unconvinced that the common content of the several systems theories is sufficiently great to justify the investment of much effort in the construction of an elaborate formal structure.

Beyond my general doubts, my skepticism has a very specific basis. One of the analogies that general systems theory proposes to encompass is the human organism. The human organism contains as one of its parts the central nervous system, which appears to be a completely general computer capable of constructing any finite proof—hence of imitating any other computing program; hence of serving as analogue for *any* conceivable theory. (Technically, such a general-purpose computer is known as a Turing machine.)

The same thing may be said in another way. Because of the flexibility of the central nervous system, the human organism can in principle be programed to produce almost any physiologically possible output for almost any stimulus input. It seems unlikely that with this potential flexibility of behavior the analogy of this organism with a cell or even with an organism lacking such a central nervous system can have much content.

Having expressed these doubts as to how far the formal development of a general systems theory can be carried, I do not want to discourage the curiosity of the biologist who wishes to learn about social systems or the social scientist who wishes to study biology. It is probably useful for a scientist who wants to contribute to the theory of one of these systems to familiarize himself with the theories of the others. However incomplete, the analogy certainly has sufficient content to be of great heuristic value. I think I can cite a number of examples where this heuristic value has already been exploited in useful ways:

1. The lynx-rabbit cycle, already discussed, is a case in point. This biological theory was a major stimulus to Richardson's *Generalized Foreign Politics,* a theory of international competition; and it influenced also, directly or indirectly, Rashevsky's theories of social imitation and my own model of the Homans system.

2. W. Ross Ashby's theory of the central nervous system, set forth in his important book, *Design for a Brain,* might be regarded as a form of "neural Darwinism." Dr. Ashby has accomplished a transfer of the principle of natural selection from the theory of species to the theory of cerebral learning.

3. There is a broad class of frequency distributions, often encountered in biological and social data, that are highly skewed and may be regarded as the logarithmic counterparts of the normal, binomial, Poisson, and exponential distributions. The so-called Pareto income distribution is one instance; Fisher's log series distribution, which fits many biological "contagious" phenomena, is another; a third is the log-normal distribution; a fourth is a distribution, applicable for example to city sizes, that I have christened the "Yule distribution" in honor of the statistician who first provided it with a theory. The kinds of probability mechanisms that will generate distributions of these types and the reasons for their frequent occurrence in biology and sociology are beginning to be pretty well understood and will broaden the base of analogy among these phenomena.[8]

4. A final example is provided by information theory, first developed to handle certain problems of coding messages for electrical transmission, which has recently found exciting applications in genetics—specifically in contributing toward an understanding of how genes transmit the characteristics of the organism.

You will note that all the examples I have cited are at a relatively concrete level. They have not involved the construction of a common theory so much as an imaginative use of analogy to suggest special theories. It is perhaps also worth observing that what is transferred in these examples is largely the mathematical frameworks of the theories and only to a slight extent the more special content. I do not believe that there is between Miller and me any difference in principle on this point; there is, perhaps, a difference in strategy and tactics—a difference in the importance we attach to the construction of a formal general systems theory.

The Electronic Digital Computer as Analogue

As my final illustration of the relation of analogues to theories, I should like to talk about the fantastic modern toys that have been called "giant brains." Two supposedly fatal objections have been raised against regarding these systems as "brains." The first is that the anatomical structure of the central nervous system is demonstrably quite unlike the wiring diagram of a digital computer. The second is that the computers allegedly cannot do any "thinking" beyond simple arithmetic and that, like clerks and schoolboys, they must be instructed in detail what arithmetic to do.

The first objection is misdirected, and the second is not correct. The first objection rests on the common misconception about the nature of analogies that we have already discussed at length. Although the circuitry of a modern computer is clearly a very poor analogue to the anatomy of the brain, it does not follow at all that this disqualifies the functioning of a programed computer from serving as an analogue to the processes of human thought.[9]

8. See "On a Class of Skew Distribution Functions," *Biometrika,* December, 1955.

9. See John von Neumann, "The General and Logical Theory of Automata," in Lloyd A. Jeffress (ed.), *Cerebral Mechanisms in Behavior* (New York: John Wiley & Sons, 1951).

The usefulness, if there is one, in employing a digital computer as a theory of human thought processes rests not on any supposed similarities of gross anatomy but on the fact that the computer is a Turing machine. It is a general-purpose device that, subject to limits on its speed and memory, can be programed to imitate the behavior of *any* other system—and, in particular, to imitate human thought. (Lest this statement depress my listeners, let me observe again that a human being is a Turing machine too. In fact, we can assert with considerable conviction that there is nothing a digital computer can do that a human being, given time, patience, and plenty of paper, cannot do also.)

Whether the computer will in fact prove a useful tool for the study of thought depends on whether it is powerful enough for the task within the limits established by time, memory size, and the complexities of programing it. The question, to put it briefly, is whether a computer can learn to play a reasonably good game of chess or to become a geometer at, say, the level of a high-school sophomore; and whether it can acquire and execute these skills using, at least qualitatively, the same tricks and devices that humans use.[10]

The proposal to program a computer to play a game is not new. As a matter of fact, several reasonably powerful checkers-playing programs have already been constructed for digital computers. But in previous attempts of this kind the objective has been to get the machine to play a good game and not to simulate human problem-solving processes. Hence, the rational man of game theory and statistical decision theory (an entirely mythical being) has been taken as the model, instead of the problem-solving organism known to psychologists.[11]

10. For an extensive discussion of the problem of programing a computer to learn chess see Allen Newell, "The Chess Machine: An Example of Dealing with a Complex Task by Adaptation," *Proceedings of the 1955 Western Joint Computer Conference*, pp. 101–8. A remarkable analysis of the problems discussed in these concluding pages will be found in Edwin G. Boring, "Mind and Mechanism," *American Journal of Psychology*, April, 1946.

11. The important differences between these two creatures are discussed in "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, February, 1955; and "Rational Behavior and the Difficulty of the Environment," *Psychological Review*, January, 1956. For an example of a game theoretical approach to the chess machine see C. E. Shannon, "Programming a Computer To Play Chess," *Philosophical Magazine*, March, 1950.

I will not be tempted into a prediction as to how long it will be before we know how to teach these things to a computer. Nor do I wish to enter into the technical problems that are involved. This much I can say with confidence on the basis of some participation in such an undertaking. One cannot think seriously about the problem of programing a computer to learn and to solve problems without gaining very great insights into the ways in which humans learn and solve problems. Regardless of whether this analogy between machine and man can in fact be realized "in the metal," its heuristic value can hardly be exaggerated.

But apart from the heuristic value, what is the particular virtue of the computer analogy? Why not work directly toward a mathematical (or verbal) theory of human problem-solving processes without troubling about electronic computers? If we were sure that the construction of such a mathematical theory were within our powers, the question would have no answer. But it is at least possible, and perhaps even plausible, that we are dealing here with systems of such complexity that we have a greater chance of building a theory by way of the computer program than by a direct attempt at mathematical formulation. Let me indicate why I think this is so.

Remember, the proposal is not to program a computer to play chess but to program it to *learn* to play chess. It can be shown that to program something to learn means to program it to alter and modify its own program and to construct for itself new subprograms.[12] This means that, as the learning process progresses, the activity of the computer will be more and more self-programed activity. The scientist will be no more aware of the details of the program inside the computer than he is aware of the details of his own thought processes.

Suppose that we could achieve the goal of programing a computer to learn to play chess. How would we use the computer as a theory?[13] First, we would experiment with various modifications of the learning program to see how closely we could simulate in detail

12. Self-programing of digital computers has already been achieved at simple and elementary levels and is now a standard part of programing technology.

13. For a penetrating discussion of the use of a computer as theory see Walter Pitts, "Comments on Session on Learning Machines," *Proceedings of the 1955 Western Joint Computer Conference*, pp. 108–10.

the observable phenomena of human problem-solving. The program that achieved this simulation would provide, at a suitable level of aggregation, a theory explaining these observable phenomena and would without doubt suggest a number of crucial experiments.

Second, human beings have great difficulty in introspecting—and particularly in introspecting reliably and comprehensively. The computer, however complex its over-all program, could be programed to report, in accurate detail, a description of any part of its own computing processes in which we might be interested. Because of this, and because of our exact knowledge of the physical structure of the computer, we could find out directly a great deal more about what was going on in the computer than we are likely ever to find out directly, by introspective techniques at least, about what is going on in the human mind.

Third, it might prove easier to construct a mathematical theory of human problem-solving after we have constructed a mathematical theory of machine problem-solving. Ordinarily, we use a computer when we are confronted with a mathematical theory whose equations are too complicated to be solved explicitly. Then we program the computer as an analogue to the mathematical theory—which is, in turn, an analogue to the phenomena. The present proposal involves a quite different use of the digital computer in theory construction. If a computer were used to simulate human problem-solving activities, the analogy between computer program and the phenomena would be direct. Mathematical theory, as it first enters the picture, enters as a theory of the computer program and hence only indirectly as a theory of the phenomena.

Conclusion: Science as Analogy

The basic postulate underlying this discussion has been that, contrary to general belief, there is no fundamental, "in principle" difference between theories and analogies. All theories are analogies, and all analogies are theories. Two theories are not equivalent for the scientist simply because they have the same logical content. The choice between theories depends critically on the ease with which their logical content can be extracted by the manipulations of information-processing systems operating upon them and the ease with which errors of omission and commission can be detected and

avoided. This is the real core of the debate about the relative virtues of mathematical symbols and words as materials of theory.

We must not suppose, simply because verbal and mathematical theories have been with us a long time, that methodology is a static matter—an unchanging substratum for the changing substance of science. Methodology requires a re-examination today, both because of the novel substantive problems that the behavioral sciences face and because of the novel devices that are now available to help us solve these problems.

A theory of man that takes account of his characteristics as an information-processing system is just beginning to emerge. Already, the theory suggests a system exhibiting a degree of complexity with which the sciences—and certainly the behavioral sciences—have not hitherto dealt. Modern electronic computers have been, and continue to be, an important influence, by way of analogy, on the emergence of this theory. If the argument advanced here is correct, these same computing devices may provide us with the materials for a methodology powerful enough to cope with the complexity of the theory as it emerges.