Information Processing and Behavioral Science

Allen Newell*

The purpose of this essay is to develop the connection between the science of human behavior and a new area of research that is growing up around digital computers: the study of complex information processing systems. To the casual observer work in this area seems like an attempt to make machines "think" — that is, to accomplish intellectual tasks hitherto done only by humans. For example, the work in this area includes attempts to get machines to play chess, or to prove theorems in very abstract mathematics.

The field is less frivolous than the foregoing examples suggest. A rather complicated use of information is required in the solution of almost all difficult problems. For some problems — e.g., how to get out the payroll in a large factory — we understand moderately well what information is needed and just how it is to be used. For really difficult problems — e.g., how to discover a new scientific law — we know very little. This new area of research may be described as the scientific study of systems that process information in very complicated ways. At the present time its main methodological characteristic is the attempt to synthesize

---

*Since 1954 I have participated with J. C. Shaw of The RAND Corporation and H. A. Simon of Carnegie Institute of Technology, in the study of complex information processing systems. I am much indebted to them for all the ideas that underlie the present paper, although, of course, they are not responsible for the details of expression.

such systems by programming them for digital computers, and then to study their behavior empirically.

The implications of successful research in this area are clearly as broad as the range of tasks that require the complex use of information. In this paper I shall focus on the implications for predicting human behavior, and will disregard other implications. Thus, for example, I shall not explore the consequences of such research for business data-processing systems, although these are highly relevant to an applied science of organization.

Relations to Behavioral Science

The relation between the science of complex information processes and the science of human behavior can be summarized in three statements:

1. Work on complex information processes embodies a theory of human behavior centered around the concept of limited rationality. Although this is not a total theory of human behavior, it is particularly well adapted to the demands of organization theory.
2. Besides elements of a theory of human behavior, work on complex information processes contains a set of descriptive and analytical techniques adequate for developing such a theory.
3. Although the automatic digital computer is historically responsible for many of the ideas

involved, it is not used here as a model or analogy, but as an analytic device for deducing consequences of the theory.

We shall examine each of these assertions in turn.

Current Information-Processing Research

This field of research comprises much more than just the efforts of my colleagues and myself, but the propositions stated above do not represent a consensus of the attitudes of other investigators. Active groups exist in the field with almost every conceivable attitude toward these assertions. A brief indication of the composition of the field will help both in fixing its nature and in assessing these assertions.

1. There are practical workers who deny any connection between their work and a science of human behavior, but who take an engineering approach to some particular task that needs mechanizing. The work on mechanical translation of languages[1] and the work on machine literature searching,[2] both very lively and relevant research efforts at the present time, belong to this class.

---

[1] W. N. Locke and A. D. Boothe (eds.), Machine Translation of Languages, Wiley, 1955.

[2] J. W. Perry, A. Kent, and M. M. Berry, Machine Literature Searching, Western Reserve University Press, 1956.

The large amount of current practical research on machine programming belongs to this same class.[3] Programming large computers, either for engineering problems or for business applications, has proved very expensive. This has led to continuing efforts by both manufacturers and users to get the computer to help program itself. This work is closely allied to our second assertion, on techniques, since one direction of research in automatic programming is the development of languages that have some of the power and flexibility of natural languages and yet are precise enough to be understood by computers.

2. At the opposite extreme of practicality is a group concerned with pure artificial intelligence. They too prefer to disclaim any immediate relation to behavioral science, but are working directly to synthesize systems that will show as much of the higher human intellectual functions as possible — learning, the use of language, problem solving, etc. They are quite unconcerned whether the machines

---

[3] Proceedings of a Symposium on Automatic Programming for Digital Computers, May 1954. Office of Technical Services, U. S. Dept. of Commerce.

accomplish the tasks in the same manner as humans. This group is made up primarily of mathematicians and accounts for most of the abstract theorizing about such matters. A paper by Turing[4] on whether machines can think, and one by Shannon[5] on programming a computer to play chess are good examples of work in this area.

3. A third, rather diverse, group may fairly be called the cybernetic group.[6] Here the fundamental object is to construct a science of human behavior. The point of departure is physiology; and the source of new information and techniques with which to make progress is the possibility of designing physical systems that show feedback and purpose. For this group, the digital computer serves as a model of the nervous system — coordinating neurons with binary digital devices. Most of the work on actual automata (as opposed to

---

[4] Turing, A. M., "Computing Machines and Intelligence," Mind, vol. 59, 1950.

[5] Shannon, C. E., "Programming a computer for playing Chess", Phil. Mag., 41:256-75 (March 1950).

[6] For a rather comprehensive review with reference to most of the articles in the field, see W. Sluckin, Minds and Machines, Pelican, 1954.

speculation about automata, mentioned above) falls within this group; for example, the Homeostat of R. W. Ashby and the mechanical turtles of Grey Walter. The more directly physiological work of McCulloch and Pitts also belongs to the cybernetic tradition.

4. A small group can be separated from the cybernetics group by its attitudes towards digital computers. These investigators share in common with the cyberneticists a concern with the science of human behavior at the level of physiology and its first behavioral correlates. However, they use the digital computer as an analytic device for discovering the consequences of various theories, formulated as sets of interacting mechanisms. This group includes Rochester,[7] who has constructed models of the nervous system based on the theories of Hebb, and has examined whether they had the predicted properties (at first they didn't, but with modifications they did). It also includes

---

[7]N. Rochester, et al, "Tests in a cell assembly theory of the action of the brain, using a large digital computer," IRE Transactions on Information Theory, Vol. IT-2, No. 3, 1956.

the work of Selfridge and Dineen[8] on visual pattern recognition, which will be discussed later.

5. Finally we come to a group which I would call the information processing group, consisting mainly of ourselves.[9] We are concerned with the science of human behavior, but the point of departure is at the social and cognitive level. Also, the computer is viewed as a consequences-generating device, and not as a model of human behavior.

This rough mapping of the field shows the great diversity of attitudes and interests. The propositions in this paper represent only the attitudes and directions of a small segment of the total activity in the field. However, little enough is known so that a

---

[8] O. G. Selfridge, "Pattern recognition and modern computers," and G. P. Dineen, "Programming pattern recognition," both in Proceedings of the 1955 Western Joint Computer Conference, I.R.E., March, 1955.

[9] Besides the papers referred to explicitly in this article, see A. Newell and H. A. Simon, "Current Developments in Complex Information Processing," P-850, The RAND Corporation, 1956; H. A. Simon and A. Newell, "Models: their uses and limitations," in L. D. White (ed.), The State of the Social Sciences, U. of Chicago, 1956; and H. A. Simon: "A behavioral model of rational choice," Quarterly Journal of Economics, February, 1955.

considerable amount of mutual aid and benefit will occur among groups despite the wide distribution of motivations and research directions. In this sense all these undertakings belong to one field.

The Doctrine of Limited Rationality

The doctrine of limited rationality asserts that, within a broad range of behavior, man operates in a rational way given his available information and his limited intellectual powers. Rationality means that behavior is shaped to the attainment of ends. Thus, actual behavior falls short of objectively rational action primarily because of the limitations of the actor's cognitive powers in developing the consequences of the information he does possess, or because of external constraints on the information available to him. The doctrine does not imply that objectively rational behavior is a good approximation to actual behavior — the constraints are much too important for this.

The doctrine of limited rationality occupies a rather unenviable position in the behavioral sciences at the present time. Partly, it is a reaction to the doctrine of objective rationality, which has proved both popular and serviceable in economics. This reaction has been heightened by the advent of game theory which tremendously strengthened the doctrine of objective rationality and added the methodological doctrine that an analysis starts with an exhaustive listing of all possible alternative strategies, and deals exclusively in terms of this fixed set. This type of analysis simply highlights the gap between what objectively rational behavior demands and what reasonable men can be expected to know and think.

The doctrine is also partly a reaction to the anti-rationalism of psychology. Somewhere in the search for scientific objectivity, this objectivity was equated with behavior determined by fixed laws, while the concept of behavior determined by calculation on ends and means seemed mentalistic and was lost. In any event, psychology has come to talk about learning as if it just occurred, its rational nature being irrelevant. It thinks of Freud as revealing the ultimate, irrational nature of man, when in fact most Freudian theory is a large array of clever means to achieve ends that happen to be unconscious to the individual. Psychology has learned to analyse complex group situations as if they were simply pushes and pulls towards this or that, the outcome being determined by some simple resultant force.

The notion of limited rationality exists because so much of behavior is quite clearly at variance with the available extreme models. The doctrine has had strong influence in the study of organizational behavior, where both the rational, and the less-than-rational behavior of man are much in evidence. An organization is a vast machine for rational action, and the evidence of achievement is too clear to be ignored. It is difficult to explain how a steel factory or a television set gets built without some concept of rational action. On the other hand, observation of the processes that go on inside organizations makes it abundantly clear how far short of full rationality action always falls. It is difficult to maintain the objective rationality of decisions made around the conference table when relevant and available facts never get expressed.

Unfortunately, the doctrine of limited rationality starts as a negative idea. It really asserts that life is more complicated than either the rationalist or the "emotionalist" belief. It is in the same position as the doctrine that systems are nonlinear. Both doctrines are undoubtedly true, but there are essentially an infinity of ways of being either nonlinear or nonrational. What is needed is the specification of alternative sets of positive assumptions, and not merely a statement of the inadequacy of what exists. Both linearity and maximization derive as much popularity in scientific theories from their ability to provide unambiguous answers as from their ability to provide correct ones.

<u>The Information Processing Model</u>

The problem, then, is to select from all the possible ways of being partly rational a set that is both characteristic of human behavior and is describable in a scientifically useful way. The view of man as an information-processing system offers a solution. By finding a precise way of stating the processes that occur — how each consequence gets found in terms of the information available — it provides a positive alternative framework for the theory of human action.

The Chess paper[10] is an example of this kind of analysis. Since it was written early in the development of this research area, it helps to separate two major advances: the emergence of the positive alternative, which we are discussing here; and a technical advance, which will be discussed in the next section.

The Chess paper considers a total, continuing, complex decision situation and describes how an organism could arrive at decisions for each of the interdependent sequence of actions it must take. In the game of chess, the entire situation is precisely defined and the objectively rational way of behaving is known in principle. However, more calculation is required to play the optimal strategy than will ever be available (at least for the known algorithms). Hence, in chess the information-processing considerations become all important, and to solve this decision problem is to find at least one particular alternative to full rationality.

The Chess paper gives a complete scheme of processes for solving the decision problem. This is much more than a

---

[10] A. Newell, "The Chess machine: an example of dealing with a complex task by adaptation," Proceedings of the 1955 Western Joint Computer Conference, I.R.E., March, 1955.

list of the component mechanisms that make up the machine. Many of the component ideas existed prior to the paper: utilities, probabilities, feedback loops, levels of aspiration, and so on. In the chess machine, however, each is given in a particular form to solve a particular sub-problem, and they are all welded into a system sufficient to determine the behavior of the chess player at each choice point.

Of course, this is only the scientific equivalent of proving a mathematical existence theorem: we have exhibited an alternative to the model of man as a fully rational animal. The next question is how well the model fits the actual behavior of humans. This task is not undertaken seriously in the Chess paper; it has been started elsewhere[11] using a program for proving theorems in symbolic logic. Logic was chosen rather than chess for the technical reasons that will be discussed in the next section.

Techniques

In some sense the advance documented in the Chess paper is technical — the discovery of how to specify the kinds of systems required by a theory of man as a partially rational animal. The advance occurred because a way was found to use the technical and conceptual apparatus of the digital

---

[11] A. Newell and H. A. Simon, "The Logic Theory Machine," IRE Transactions on Information Theory, Vol. IT-2, No.3, 1956.

computation field to specify systems for performing tasks like chess, rather than systems for computing numbers. Oliver Selfridge and Gerald Dineen,[12] two mathematicians working at the MIT Lincoln Laboratories, were the first ones really to accomplish this. Taking up the problem of how organisms learn to recognize a visual pattern, given only some exemplars of the pattern under varying conditions, they succeeded in programming a computer for this task. Their program is truly non-numerical, for it performs its computations on two-dimensional arrays of patterned dots, and produces from them new arrays. It has two learning loops, one which selects the processes to be performed, and another which modifies the generator that provides the processes from which the first learning loop selects. This is a way of using a computer that is entirely divorced from numerical computations.

Once Selfridge and Dineen had shown the way, it was clear how to specify other systems. The Chess paper, which is an exercise in using this technique, was written as a direct result of my first contact with the work of Selfridge and Dineen in the fall of 1954.

However, both the chess machine and the pattern-recognition program are technically inadequate. The techniques used are the flow diagram, backed up by the machine code when

---

[12] Ibid.

the program is actually run on the computer. The flow diagram is, heuristically, a very powerful tool, but it is not precise enough to support detailed deductions about the nature of the system it represents. A glance at the flow diagram in the Chess paper should amply confirm this. The power of the flow diagram comes from the fact that the people who use it develop the ability to assess accurately whether it can be realized in terms of a precise program in machine code. On the other hand the terms of the machine code represent such minute processes that the only function a machine coded program serves, other than communicating instructions to the computer, is verification that given processes can in fact be realized.

The development of a theory of information processing requires much more powerful techniques for specifying systems and for analysing them. These techniques are in fact emerging, as a comparison of the Chess paper with the Logic Theory Machine paper shows. In the latter, instead of a schematic flow diagram, a precise language is used to specify the system. The terms of this language are meaningful processes for the task being performed, and although the program still looks formidable, it is smaller and simpler by several orders of magnitude than it would be if specified in machine code.

This technical development is a second distinct advance in the field, which can be separated both logically and historically from the first advance. It flows from the accumulated experience in programming and the efforts of the programmers themselves to do automatic coding. One important direction these efforts have taken is the construction of new languages (called pseudocodes) which are more convenient for specifying problems than are machine codes. Programs are written for the computer itself to interpret these pseudocodes — either producing a program in machine code, or doing the computation directly. These languages are concerned mostly with arithmetical calculations, but they provide a vision of general information-processing languages that can be understood by the computer.

The nature of these techniques can be illustrated by comparing them with the techniques available in a well developed scientific theory like analytical mechanics. The very foundation of mechanics is its ability to describe any particular system as a set of differential equations. The language of differential equations is independent of the laws of mechanics and provides a base for exploring the consequences of different laws. Most important, the description is <u>complete</u> in the sense that it contains all the relevant information about a system and its behavior over time. Thus the description becomes a perfect surrogate for the mechanical system being studied, and all further theoretical inquiry is based on it.

In our case a human (or an organization) is described as an information-processing system. The structure of the system is given by a set of memories, and the symbols they hold (corresponding to the variables and their values in mechanics). The behavior of the system is described by a program (corresponding to the set of differential equations). This program specifies which sequences of information processes will occur under what conditions. There is a set of primitive processes out of which all programs are composed, just as in the set of differential equations the effects on the rates of change of the variables are described by additions, multiplications, exponentiations, etc., and combinations of these. This description is also complete: given the initial symbols in the memories and the program, the entire behavior of the system is determined through time. It thus has the very strong virtue that it is capable of supporting theoretical inquiry — there are consequences that logically flow from the description.

Given descriptions that contain some information about a system, analytical techniques are needed for extracting the information. Given a program that describes a man playing chess, one wants to find out how the program plays and what its properties are. In what situations does it have difficulty? To what aspects of the chess position does it attend? Most of these properties cannot be ascertained simply by glancing at the program. In the case of differential equations there exist three ways of deducing the behavior of a particular system. First, there exist mathematical techniques, such as differentiation, for manipulating the expressions into more revealing form. Second, there exist theorems about the properties of special classes of systems, such as theorems about stability. Finally, there exist computational techniques for determining the time paths of particular systems. For example, there are finite difference approximations for differential equations that are adapted to computation on digital computers. As one passes from simple to complicated systems of differential equations, the relative use of the techniques shifts from the first two to the third.

In the case of information processing systems, only the technique of computation is currently available. No mathematics and no theorems of any importance exist to predict the behavior of programs. However, the high speed digital computer does give us the capability to compute the behavior of a program for any number of special cases we desire. Thus, it is the main analytic tool for developing a theory. Analysis proceeds by specifying programs, exploring empirically their properties on the computer, specifying new systems and so on.

The prospects for acquiring the other analytical techniques are uncertain. In principle there could exist a mathematics of programs, with powerful theorems about their behavior. Almost no attempt has been made yet to develop one. On the other hand, the programs needed to describe human behavior are already of considerable complexity, and may be beyond the reach of symbolic methods of analysis, just as large systems of differential equations are.

Computers and Analogies

The role of the digital computer as a consequence-generating device was described in the preceding section. An information processing system is specified by a set of symbols — the program. This specification contains all the implications that exist about the system, and the problem is to extract them. The computer lets us find some of these implications by being able to perform the program for any particular situation.

This is not an unfamiliar role for machines to occupy if we consider the use of computers in physical theory or statistics. However, machines have also played a different role *vis-à-vis* theories of man and his nature — the role of analogue. As mentioned earlier, one approach to the field of complex information processing is to construct thinking machines, and this leads naturally to the supposition that the thinking machine is to be made into a model of man. This idea is reinforced from the cybernetic movement, which in fact has used both electrical circuits with feedback and digital computers as analogies to the nervous system.

It should be evident by now that we are not using the digital computers as an analogy. It is true historically that many of the basic ideas for representing information processing systems were first formulated in connection with digital computers. It is also true that the large high speed computer is the only non-biological system currently known that can show anything like man's capability for complex behavior. But in fact we could write programs, investigate their properties, and use them to represent theories of human behavior even if computers never existed.

Summary

We have now examined in some detail the nature of the three assertions. I have claimed that if we have a program that behaves like a man, then we have an explanation of that behavior. I have indicated that techniques exist for specifying programs in a scientifically useful way. These share many of the virtues of the best techniques known for the description of dynamic systems — differential equations — although they are comparatively deficient in analytic power. Also, although more by indirection, I have indicated that theories of man centered around the concept of limited rationality are handled naturally by these concepts and techniques of information processing. Finally, I have shown that we are not simply drawing crude analogies between men and machines.

There are several ways in which the science of complex information processing may make direct contact with the theory of organizational behavior. Perhaps the clearest way is through the development of a better psychological theory of the individual as a decision maker with limited rationality. Given a good model of man, one can then consider the behavior arising from a system of men under various conditions governing their interaction. This is the slow and indirect route, for it means first working extensively in the area of individual psychology. However, some clear progress is already being made. In the last several months work has started on a theory of human problem solving, using the Logic Theory Machine as the theory.[13] Comparisons have been made between the properties exhibited by the program and the known characteristics of humans in problem solving situations. The Logic Theory Machine already exhibits qualitatively a large number of these characteristics. We have also begun to construct programs to describe the human memorization of nonsense syllables. This is a much-studied task in human psychology, for which large amounts of published data exist. This work is not yet published, but already very simple notions of the processing requirements yield some quantitative agreement with human data.

---

[13] A. Newell, J. C. Shaw, and H. A. Simon, "Elements of a Theory of Human Problem Solving," in draft.

The area of small group research may prove amenable to the approach indicated in this paper. In a Bales type discussion group the number of individuals is small enough (about five) to allow each to be represented as a separate information processing system. The group task proceeds entirely in terms of information processing: things happen in discussions because the participants direct a stream of facts, suggestions, agreements, etc., at each other. From the evidence to date, any theory of unstructured small group discussions will contain personality factors as a major constituent. In this respect it may differ from theories of more formal organizations. However, it is possible that a personality may be better represented as a particular system of complex processes than as a set of abstract traits. Some serious consideration is being given by Bales to using the techniques talked about in this paper to help develop a theory.

One need not focus on the individual as the unit of analysis. Instead one can consider directly the processes that an organization must go through to arrive at a course of action, a plan, or some specified information. These can be characterized abstractly as an information process made up of subprocesses, independent of the particular individuals involved. This characterization permits an aggregative analysis that may allow very large organizations to be described in relatively simple terms, compared to any model built up from individuals. Attempts are also beginning in this area, using the general concepts of program, search, etc., to help organize field data about business decisions.[14]

These points of contact with organizational science are only those most apparent at the present time. As techniques for specifying and analysing such systems develop there will arise also some quite different ways of applying the concepts of information processing.

---

[14] R. M. Cyert, H. A. Simon, and D. B. Trow, "Observation of a business decision," The Journal of Business of the University of Chicago, Vol. 29, 4, October, 1956.