Models of Man

SOCIAL and RATIONAL

Mathematical Essays
on Rational Human Behavior
in a Social Setting

HERBERT A. SIMON
Professor of Administration
in the
Graduate School of Industrial Administration
Carnegie Institute of Technology

JOHN WILEY & SONS, INC., NEW YORK
Chapman & Hall, Limited, London

PART IV

Rationality and Administrative Decision Making

> What is important here, however, is the superlative degree to which logical processes must and can characterize organization action as contrasted with individual action, and the degree to which decision is specialized in organization. It is the deliberate adoption of means to ends which is the essence of formal organization. This is not only required in order to make cooperation superior to the biological powers and senses of individuals, but it is possibly the chief superiority of cooperative to individual action in most of the important cases of enduring organizations.
>
> —Barnard, *The Functions of the Executive*

A theory of administration or of organization cannot exist without a theory of rational choice. Human behavior in organizations is best described as "intendedly rational"; and it merits that description more than does any other sector of human behavior.

It can be said with equal truth that a theory of rational choice can hardly exist without a theory of organization. Robinson Crusoe, it may be argued, proves the contrary. But an understanding of Robinson Crusoe, however important as a first step, is only a preliminary to an understanding of modern, urbanized man. The characteristic environment of man is constituted not of nature but of his fellows. His rational decision making —at least during most of his waking hours—takes place in social groups including organizations.

The essays of this part will show that our understanding of rational processes has not yet taken us much beyond the Robinson Crusoe stage. Almost all we shall have to say about decision could be said as truly of him as of the executive staff of a business concern. The task of incorporating organizational considerations into the theory of rational choice has still to be completed.

Chapters 13 through 15 contain the basic ingredients for a theory of human rational choice. Chapters 12 and 16 are included for purposes of contrast, for they represent approaches to the problem quite different from the middle three. In my introductory notes, I should like to comment a bit further on the role of the concept of rationality in the theory of organization; and then to discuss rather carefully the several points of view toward human rationality that are to be found in these essays, as well as in those we have already encountered in Part III.

The Principle of Bounded Rationality. There are two principal species of economic man: the consumer and the entrepreneur. Classical economics assumes the goals of both to be given: the former wishes to maximize his utility, which is a known function of the goods and services he consumes; the latter wishes to maximize his profit. The theory then assumes both of them to be rational. Confronted with a pair of alternatives, they will select that one which yields the larger utility or profit, respectively.

Beyond these postulates—that he is rational and that his goals are specified—the theory assumes nothing about the psychological characteristics of economic man. The factors that determine his behavior (apart from those already mentioned) are entirely external to him. The consumer is constrained by a fixed budget, say, and by the prices of goods and services; the entrepreneur is faced with determinate supply schedules for factors of production, demand schedules for his products, and a technologically determined production function. The economist predicts their behavior, to the extent that he is interested in it, without subjecting them to tests either of intelligence or of personality.

We have already encountered specimens of economic man in the essays of Part III. The actors in both "A Comparison of Organization Theories" and "A Formal Theory of the Employment Relationship" fit pretty well the description of the previous paragraph. (There are some deviations, upon which I shall comment in a moment.)

The first essay printed in the present part is included as an especially clear-cut illustration of the treatment of rationality in classical economics. The psychology of the people who inhabit the world described there is disposed of, once and for all, in three sentences. "In line with our original assumption of perfect mobility of labor, the wages of labor in the two types of production must be equal." "Consumer rationality requires the maximization of ϕ with prices given, and subject to the restriction that I is constant." "The producers will strive to maximize profits."

I do not intend to dispute the usefulness of the "ideal type" of economic man for many problems of economic analysis. But the specific problems with which organization theory is concerned are of a character that generally renders this particular idealization inappropriate. As soon as we turn from very broad macroeconomic problems and wish to examine in some detail the behaviors of the individual actors, difficulties begin to arise on all sides.

Two distinct types of difficulties that have attracted considerable attention in economics in recent years are illustrated by the essays of Part III. In "A Comparison of Organization Theories" (Chapter 10), we encountered the problems—made familiar by the theory of imperfect competition and the more recent theory of games—associated with "the rationality of more than one." When we examined the theory of the firm closely, as was done in this paper, we saw that the assumption of rationality did not lead to a unique determinate solution but to a whole set of

"viable" solutions. Additional assumptions (e.g., assumptions about which participants were "active" and which "passive") were needed to restore determinateness to the situation. Oligopoly theory today is embarrassed by its inability to choose among the rich assortment of alternative assumptions that are available to it. It is unable to choose because the alternatives represent conflicting *empirical* statements about the psychological characteristics of economic man, and economic theory is not accustomed to deriving his characteristics from empirical observation.

The second difficulty—this one represented by the paper on the employment relationship (Chapter 11)—is that the classical criterion of rationality is not easy to extend in an univocal fashion to situations where uncertainty is present. The whole argument of that paper was that the employment relationship is a "rational" adjustment to ignorance about the future. But if we admit that rational man may be ignorant about the future, that his rationality does not imply omniscience, what other limits may we place upon him? And if his rationality does not represent an *objective* orientation to the real world, but only a *subjective* orientation to his incomplete picture of it, how are the characteristics of this perceived world determined for him?

The reluctance of economic theory to relinquish its classical model of economic man is understandable. When even a small concession has been made in the direction of admitting the fallibility of economic man, his psychological properties are no longer irrelevant. Deductive reasoning then no longer suffices for the unique prediction of his behavior without constant assistance from empirical observation.

The essays of this part have been written with the conviction that traditional economic man, however attractive he is to the economic theorist, has little or no place in the theory of organization, or, for that matter, in most parts of the theories of imperfect markets and of economic development. All the valiant efforts of recent decades to save economic man by adding this characteristic to him, subtracting that, have only shown us how little is left of the predictive power of the classical theory once we allow the smallest deviations from the assumption of omniscient rationality. It is time, therefore, for a fundamental change in our approach. It is time to take account—and not merely as a residual category—of the empirical limits on human rationality, of its finiteness in comparison with the complexities of the world with which it must cope.

The alternative approach employed in these papers is based on what I shall call the *principle of bounded rationality*:

The capacity of the human mind for formulating and solving complex problems is very small compared with the size of the problems whose solution is required for objectively rational behavior in the real world—or even for a reasonable approximation to such objective rationality.

If the principle is correct, then the goal of classical economic theory—to predict the behavior of rational man without making an empirical investigation of his psychological properties—is unattainable. For the first consequence of the principle of bounded rationality is that the intended rationality of an actor requires him to construct a simplified model of the real situation in order to deal with it. He behaves rationally with respect to this model, and such behavior is not even approximately optimal with respect to the real world. To predict his behavior, we must understand the way in which this simplified model is constructed, and its construction will certainly be related to his psychological properties as a perceiving, thinking, and learning animal.

A second consequence of the principle of bounded rationality, which is a particular application of the one already stated, is described in *Administrative Behavior* (pp. 240–41) in the following terms:

> ... if there were no limits to human rationality administrative theory would be barren. It would consist of the single precept: Always select that alternative, among those available, which will lead to the most complete achievement of your goals. The need for an administrative theory resides in the fact that there *are* practical limits to human rationality, and that these limits are not static, but depend upon the organizational environment in which the individual's decision takes place.

Organizations are the least "natural," most rationally contrived units of human association. But paradoxically, the theory of an organization whose members are "perfectly rational" human beings (capable of unlimited adaptation) is very nearly a perfectly vacuous theory. It is only because individual human beings are limited in knowledge, foresight, skill, and time that organizations are useful instruments for the achievement of human purpose; and it is only because organized groups of human beings are limited in ability to agree on goals, to communicate, and to cooperate that organizing becomes for them a "problem."

Organization theory is centrally concerned with identifying and studying those limits to the achievement of goals that are, in fact, limits on the flexibility and adaptability of the goal-striving individuals and groups of individuals themselves. The entrepreneur of economic theory is limited only by constraints that are external to himself and his organization—the technology—and by the goal striving of individuals whose interests are not identical with his. Administrative man is limited also by constraints that are part of his own psychological make-up—limited by the number of persons with whom he can communicate, the amount of information he can acquire and retain, and so forth. The fact that these limits are not physiological and fixed, but are instead largely determined by social and even organization forces, creates problems of theory construction of great subtlety; and the fact that the possibilities of modifying and relaxing these limits may themselves become objects of rational calculation compounds the difficulties.

We see, then, that the principle of bounded rationality lies at the very core of organization theory, and at the core, as well, of any "theory of action" that purports to treat of human behavior in complex situations.

Bounded Rationality Contrasted with "Irrationality." It is important to distinguish between the principle of bounded rationality, just described, and the contemporary emphasis in social psychology upon the affective, nonrational factors in human behavior. Fashion in the scientific explanation of man's behavior oscillates between theories that assign supremacy to his reason and those that give predominance to his passions. The synchronized push that Freud and Pareto gave to this pendulum has, for the past generation, kept it far over on the side of passion. Economics alone among the social sciences has kept alive the belief in reason, a belief that even Veblen's challenge shook only slightly.

Clearly, a mature social science will have to accommodate both intellect and affect; but that is not our immediate concern here. One of the difficulties—perhaps the most serious—in incorporating cognitive processes in the theory of social behavior is that we have not had a good description of those processes. As we have pointed out in the previous section, the received theory of rational choice is a theory that almost completely ignores the limits of humans as mechanisms for computation and choice—what we have called the principle of bounded rationality.

The central task of these essays, then, is not to substitute the irrational for the rational in the explanation of human behavior but to reconstruct the theory of the rational, making of it a theory that can, with some pretense of realism, be applied to the behavior of human beings. When we have made some progress with this reconstruction, I believe that the return swing of the pendulum will begin, that we will begin to interpret as rational and reasonable many facets of human behavior that we now explain in terms of affect. It is this belief that leads me to characterize behavior in organizations as "intendedly rational."

Rational Components in Role Behavior. The comments of the last section have an important bearing on the sociological concept of "role." The increased use of "role" in sociology and social psychology has not been accompanied by a corresponding clarification of what is meant by the term or of how social roles are to be specified. When an actor plays the leading role of *Hamlet*, the notion of rational conduct is largely irrelevant. His task is to recite, as meaningfully as possible, the lines that were written for that role by Shakespeare. He cannot, in particular, improve upon the role by changing the lines.

But surely this cannot be what we mean by "role" in sociology. Nor is the difficulty removed when we say that the role is not completely determined, but that some freedom is left for the expression of the personality of the actor (just as Barrymore's Hamlet and Olivier's Hamlet are different, although they speak the same lines). With trivial exceptions, social roles do not specify at all the lines that are to be spoken. *The specification of a role consists in the specification of some subset of the premises that are to guide the decisions of the actor as to his course of behavior.*

The crucial point is that we define roles in terms of *decision premises* rather than in terms of the decisions compounded from such premises. If we take the decision premise—rather than more global concepts like the "decision" or the "role"—as our unit for the description of human choice, then it is easy to place the rational and the nonrational aspects of behavior in proper relation to each other.

> Individual choice takes place in an environment of 'givens'—premises that are accepted by the subject as bases for his choice; and behavior is adaptive only within the limits set by these 'givens.'[1]

In particular, the term "role" applies to that subset of the premises of choice that is derived from the social definition of the situation. The principle of bounded rationality tells us that many or most of the premises of rational choice will be determined by the social and psychological, rather than the technological, environment of the choosing subject. The role hypothesis asserts that a large fraction of these premises is obtained from the socially defined role appropriate to the situation in which the actor is placed. Hence, if we take the premise as the unit for role description, there is no antithesis between the notion that behavior is rational and the notion that it is to a considerable extent role-determined.

The principle of bounded rationality does contain one directive for role specification; roles, to be enacted, must be specified in such a way as to bring them within the computational capacities of the actor. Hence, our criticism of the classical model of the entrepreneur was not directed at the idea that such a role should be specified in a particular society but, rather, at the particular form of the specification that made it impossible for flesh-and-blood humans to enact the role. In the next section, I shall direct the same criticism against the role of the rational decision maker as this has been defined in game theory and statistical decision theory.

Contemporary Theories of Rational Choice. I am not alone in attaching great importance to the rational component of human behavior. As I have already shown, economics has never surrendered its faith—and an exaggerated faith at that—in the powers of human reason. But beyond this, in the past ten years there has actually been a vigorous revival of interest in rational behavior associated especially with the theory of games and modern statistical decision theory.

---

[1] *Administrative Behavior*, p. 79.

The publication in 1945 of von Neumann and Morgenstern's *Theory of Games and Economic Behavior* has attracted enormous attention to the theory of rational choice. This has been reinforced and amplified by parallel developments in mathematical statistics—associated particularly with the names of Neyman, Egon Pearson, and Wald—which have reinterpreted the theory of statistical tests as a theory of rational decision.[2]

These have been contributions of the greatest importance to a number of issues: for example, the construction of a cardinal scale of utility, the relation between utility and subjective probability, the definition of rationality in competitive situations, and the decisional implications of statistical tests. In addition to its content, the work of von Neumann and Wald has further significance in having attracted to the social sciences a considerable number of men of high talent and strong mathematical training who would otherwise, in all likelihood, have worked in other fields. For these reasons the end of World War II, when these developments occurred, marks the beginning of a new era in the social sciences.

Having said this, I must record my judgment, which is at the present time very nearly the judgment of a minority of two, that the approach taken in the theory of games and in statistical decision theory to the problem of rational choice is fundamentally wrongheaded.[3] It is wrong in precisely the same way that classical economic theory is wrong—in assuming that rational choice is choice among objectively given alternatives with objectively given consequences that reflect accurately all the complexities of the real world. It is wrong, in short, in ignoring the principle of bounded rationality, in seeking to erect a theory of human choice on the unrealistic assumptions of virtual omniscience and unlimited computational power.

In the introductory pages of *The Theory of Games and Economic Behavior* (p. 2), the authors state their objective: "to establish satisfactorily... that the typical problems of economic behavior become strictly identical with the mathematical notions of suitable games of strategy."

[2] For more particulars, the reader is referred to the review article by Ward Edwards, "The Theory of Decision Making," *Psychological Bulletin*, (July 1954), and to the recent book by L. J. Savage, *Foundations of Statistics* (New York: Wiley, 1954). I must add that I agree with neither Edwards nor Savage in their assessment of the import and implications of these developments.

[3] Perhaps I am unduly pessimistic about the number of my companions in heresy, but I will let them speak for themselves. As to the term "wrongheaded" that I apply to those who disagree—I utter it in the friendliest tone of voice possible. But I cannot think of a more appropriate word to describe the cheerful and buoyant obstinacy with which my friends in economics and statistics defend their myth of omniscient man. I hope that L. J. Savage will not object if I single out his book, mentioned in the previous footnote, as an extraordinarily consistent and competent exposition of the point of view with which I disagree.

But where does the theory lead, and what questions does it answer? I cannot do better than quote from *The Theory of Games* (p. 125):

> This shows that if the theory of Chess were really fully known there would be nothing left to play. The theory would show which of the three possibilities ["white wins," "tie," or "black wins"] actually holds, and accordingly the play would be decided before it starts.... But our proof, which guarantees the validity of one (and only one) of these three alternatives, gives no practically usable method to determine the true one. This relative, human difficulty necessitates the use of those incomplete, heuristic methods of playing, which constitute "good" Chess; and without it there would be no element of "struggle" and "surprise" in that game.

If the proximate goal of economic and administrative theory is to describe and explain actual human behavior—"intendedly rational" or not, as the case may be—a theory that leaves out the "relative, human difficulty" and consequently finds itself unable to account for "struggle" and "surprise" cannot be of much help. I can only repeat a sentiment attributed to Poincaré: "Such solutions are very little solved."

More will be said below (and more is said also in Chapter 14) about the theory of chess and its implications for rational behavior. Meanwhile, I should like to conclude the present discussion by clarifying what I mean by the "assumption of omniscience" in the theory of games and statistical decision theory. Both theories, of course, encompass decision making in the face of uncertainty; but the uncertainty that is admitted is of a very restricted sort: (1) uncertainty about random events that have a joint probability distribution; (2) uncertainty about the future behavior of another player (this other player may be "nature," in case we do not wish to specify the probability distribution of "states of nature"). In the first case, the goal of maximizing utility is replaced by the goal of maximizing expected utility; in the second case a maximization strategy is replaced with the now-famous "minimax" strategy. Neither of these devices greatly simplifies the computational problem that faces the decision maker, and hence they cannot be expected, by themselves, to lead to a satisfactory theory of rational choice.

It may be said in defense of the theory of games and statistical decision theory that they are to be regarded not as descriptions of human choice but as normative theories for the guidance of rational decision. Even this defense seems to me untenable, but I shall not pursue the issue here. We are interested here in the prediction and description of human behavior, rather than in normative rules of conduct.

Rational Choice in the Face of Uncertainty. I have already mentioned the difficulties that the classical theory of rational choice encountered when the attempt was made to extend it to situations involving the rationality of more than one or to those involving uncertain knowledge about the future. The essay reprinted here as Chapter 13 is concerned with the second of these two difficulties.

The usual procedure for introducing uncertainty into the theory of choice is to assume that knowledge about the future values of one or more variables is given in the form of a probability distribution. An example of this method of conceptualizing uncertainty will be found in Sections 5 and 6 of Chapter 11 above.

It is highly doubtful, however, whether this is very often the way in which humans formulate their estimates of an uncertain future. A sales manager can usually be persuaded to reply to the question: "What do you estimate that your sales will be for each of the next twelve months?" Whether his estimate will be very reliable or not is another matter, but at least the question will seem meaningful to him. I doubt whether very many sales managers can be induced to respond to the inquiry: "Please estimate for me the joint probability distribution of sales over the next twelve months." I have tried this out a couple of times; fortunately my behavior was interpreted as attemptedly humorous rather than insane.

It has occurred to me that a house thermostat is confronted with the same problem as a sales manager; for both to perform optimally, the latter would have to predict sales correctly and the former would have to predict the weather correctly. But a house thermostat has no illusions about its powers of prediction. It regulates the house temperature not by predicting but by taking relatively prompt corrective action to eliminate deviations between the actual temperature and the desired temperature.

Chapter 13 is motivated by the idea of applying this same notion of feedback correction to the control of industrial inventories and production rates. In point of fact it has laid the basis for subsequent work that has within the past year reached the stage of application to actual problems of production and inventory control. (See also the "Note on Statics and Dynamics" at the end of Part III, and the reference given there.) Its main interest for our present purpose, however, is in indicating one way in which, at a certain cost, a decision maker can avoid reliance upon predictions about the future (and avoid entirely making estimates of joint probability distributions), and can base his decisions instead on the principle of feedback.

The same problem is raised in a somewhat different setting, and a solution suggested along the same general lines, in the Appendix to Chapter 14.

Rationality and Maximization. I regard Chapters 14 and 15 as the central core of the theory of choice I am advancing here. In these two essays the focus is upon ways of simplifying the choice problem to bring it within the powers of human computation. In the former chapter, the emphasis is upon the properties of the choosing mechanism; in the latter, upon the properties of its environment.

The key to the simplification of the choice process in both cases is the replacement of the goal of *maximizing* with the goal of *satisficing*,

of finding a course of action that is "good enough." I have tried, in these two essays, to show why this substitution is an essential step in the application of the principle of bounded rationality. It will be seen that an organism that satisfices has no need of estimates of joint probability distributions, or of complete and consistent preference orderings of all possible alternatives of action.

I had originally appended to Chapter 14 (although it has not been published and is omitted here) a very crude scheme for a chess-playing machine that would make use of the principles developed in the chapter. The idea of such a machine is an old one, revived by the theory of games and the rapid development of digital computers. It has received attention during the past few years from von Neumann, Wiener, Shannon, and others. Shannon actually worked out in some detail a chess-playing program for a general-purpose computer. All of these efforts were based on game-theoretical notions of what constituted rational behavior, and all of them were rendered futile by the fact that a program built on these principles would either play weak chess or place demands on the computer of a magnitude that could not be met with present or prospective designs.

Again, the key to an effective solution appeared to lie in substituting the goal of satisficing, of finding a good enough move, for the goal of minimaxing, of finding the best move. I carried my analysis far enough to convince myself that this substitution would reduce by a very large factor the magnitude of the computational task. Meanwhile, Allen Newell, pursuing an investigation along parallel lines, has advanced the design of a chess machine a very great distance—to the point, in fact, where there are strong reasons for believing that his scheme will be reduced to practicality in the rather near future.[4]

Within the past few months even stronger evidence has been obtained for the hypothesis that the higher mental processes of humans can be explained in terms of this kind of model of rationality. Allen Newell and I, in a joint undertaking, succeeded in December 1955 in describing a program that will enable an ordinary digital computer to discover proofs (and in a manner that is very "human" in its appearance) for theorems in symbolic logic. The program has worked successfully in a hand simulation; it appears to be adequate to develop proofs for all the theorems in the propositional calculus that are found in Chapters 2 through 5 of Whitehead and Russell; and it has now been coded in detail for a digital computer. This program, which is based squarely on the idea of satisficing rather than maximizing, provides a striking demonstration of the

[4] A first description of his proposal will be found in "The Chess Machine: an Example of Dealing with a Complex Task by Adaptation," *Proceedings of the 1955 Western Joint Computer Conference*, published by the Institute of Radio Engineers, 1955, pp. 101–108.

power of this approach to the theory of rational human decision making and problem solving.[5]

Rational Choice and Learning. The final essay, Chapter 16, is somewhat more compatible with the prevailing *Zeitgeist* than are the others. It demonstrates a connection between game theory, on the one hand, and the stochastic learning theories of Estes, Bush, and Mosteller, on the other. As my concluding paragraph indicates, I am in a state of some puzzlement whether the connection thus revealed is to be regarded as pure coincidence or whether it has some deeper meaning. This note of inquiry is perhaps a proper one on which to conclude this volume, for it hints at the vast exploration still to be undertaken before we shall have even a tolerable understanding of the complex and Protean phenomena of boundedly rational human choice.

[5]The program for discovering proofs is described in detail in A. Newell and H. A. Simon, "The Logic Theory Machine," *Proceedings, 1956 Joint Symposium on Information Theory*, Institute of Radio Engineers, Cambridge, Mass., September 10–12, 1956.

Productivity and the Urban-Rural Population Balance

A WIDE range of economic forces have cooperated to produce a steady increase in the ratio of urban to rural population in the United States and most Western European countries during the past several centuries. It is the purpose of this paper to enumerate briefly these various forces, and then to explore formally the theory of one of them—increasing economic productivity.

While the conclusions that will be reached confirm existing economic theories as to the reasons for the increase in urban population, it will be of interest to show how this result can be rigorously derived from rather simple economic models. The models, in turn, will help to reveal the anatomy of the mechanisms that are responsible for the shift.

The urban-rural population ratio depends upon the proportion of the occupied population in nonagricultural and agricultural occupations, respectively. The relation between these two ratios is not quite direct, however. The rural population includes, in addition to those engaged in agriculture and their families, a large number of other persons who are located in rural areas either because they are engaged in occupations oriented toward farm markets (distributors and those occupied in market-oriented industries), or because they are engaged in occupations oriented toward the source of raw materials (e.g., cotton gins and cheese factories). Moreover a considerable part of the time of the farmer may be devoted to nonagricultural activities, particularly in a subsistence economy.

Throughout the analysis it will be assumed that disparities between incomes of those engaged in agriculture and those in nonagricultural pursuits will be removed by migration of occupied persons from the one type of occupation to the other. As is well known, there are in fact substantial lags in this migration, and of these the analysis will take no account. It will be strictly an "equilibrium" analysis. It will also be assumed that the length of the average work week remains unchanged.

I

In studying the urban-rural population ratio for any country or area two aspects of the problem must be considered: the international and the internal. International trade may affect the ratio in two distinct ways:

1. Changes in comparative advantage between agricultural and industrial production. Whether a nation will have net imports or net