



Human Nature in Politics: The Dialogue of Psychology with Political Science

Herbert A. Simon

The American Political Science Review, Vol. 79, No. 2 (Jun., 1985), 293-304.

Stable URL:

<http://links.jstor.org/sici?sici=0003-0554%28198506%2979%3A2%3C293%3AHNIPTD%3E2.0.CO%3B2-N>

The American Political Science Review is currently published by American Political Science Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/apsa.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

Human Nature in Politics: The Dialogue of Psychology with Political Science

HERBERT A. SIMON
Carnegie-Mellon University

This article compares two theories of human rationality that have found application in political science: procedural, bounded rationality from contemporary cognitive psychology, and global, substantive rationality from economics. Using examples drawn from the recent literature of political science, it examines the relative roles played by the rationality principle and by auxiliary assumptions (e.g., assumptions about the content of actors' goals) in explaining human behavior in political contexts, and concludes that the model predictions rest primarily on the auxiliary assumptions rather than deriving from the rationality principle.

The analysis implies that the principle of rationality, unless accompanied by extensive empirical research to identify the correct auxiliary assumptions, has little power to make valid predictions about political phenomena.

This article is concerned with the nature of human reason and the implications of contemporary cognitive psychology for political science research that employs the concept of rational behavior. I shall begin with a bit of history, written from a rather personal viewpoint, to provide a setting for the discussion.

The older and/or more scholarly among you will recognize the essay's title as having been plagiarized from Graham Wallas, whose seminal book, *Human Nature in Politics*, appeared in 1908. When I began graduate study, in the middle 1930s, that book, along with Walter Lippmann's *Public Opinion*, was still wholly fresh, and both stood out as harbingers of the "behavioral revolution" that was then just getting under way at the University of Chicago.

Not that we graduate students thought of ourselves as participants in a scientific revolution. The realities of the political process had long since replaced the formal legal structure of political institutions as the main subject for study in political science—at least at the University of Chicago. Merriam's studies of power, Gosnell's quantitative methods, Lasswell's psychoanalytic probes seemed to us merely (paraphrasing Clausewitz) "the continuation of political realism by other means."¹

I was little prepared, therefore, for the violence of the polemic pro and con "behavioralism" that echoed over the land in the first two decades after World War II. Nowadays, my periodic soundings in *The American Political Science Review* reassure me that this civil strife in the profession is largely over, and that the behavioral revolution is now seen as continuity rather than discontinuity in the development of political science. I am not sure it would even qualify, in today's revisionist view, as one of Thomas Kuhn's major paradigm shifts. Perhaps what we were doing was not revolutionary science at all, but just everyday normal science.

This is probably the right moment, while I am alluding to behavioralism, to record a *culpa mea* for my part in popularizing that awkward and somewhat misleading term. It appeared, of course, in the title of *Administrative Behavior* (Simon, 1947/1976a), and also in the title of my chief epistle to the economists, "A Behavioral Model of Rational Choice," published in the *Quarterly Journal of Economics* in 1955. However, I doubt that I was the main culprit. That honor belongs to the Ford Foundation, which at that same time introduced and diligently popularized the phrase "behavioral sciences."

Whatever its origins, the term was picked up with enthusiasm—as an epithet—by the opponents of behavioralism, who frequently employed it as though it were synonymous with the Behaviorism then rampant in the discipline of psychology. In fact, there was never any substantive connection between the two labels, and much of what went on in political science, sociology, economics, and anthropology under the heading of behavioralism would have been anathematized

The present essay is a slightly revised version of the James Madison lecture presented by the author at the Annual Meeting of the American Political Science Association in Washington, D.C., 1984.

¹See David Eastman's perceptive account of this history in his article on political science in the *International Encyclopedia of the Social Sciences* (1968).

by the psychological Behaviorists if they had been aware of it—which they mainly weren't.

However, my aim here is not to reminisce about old battles. We should rejoice that political scientists are devoting all their efforts to advancing the science, and we should do nothing to encourage a renewal of the *Methodenstreit*. Instead, I shall offer a commentary on the role of the rationality principle in recent political science research.

I emphasize that this is a commentary and not a new piece of substantive research. The basic values for political science to which I and my contemporaries were and are committed include sound empirical data as the foundation for theory and for normative recommendations; new sources of data including polls, structured interviews, and systematic samples; the use of statistics, mathematics, and computer simulation where appropriate as tools for data analysis and theory construction; and the analysis of phenomena in terms of basic categories like power, decision making, rationality, and systems.

The research on which I shall comment exemplifies those values: it is empirically based, employing many different kinds of data-gathering methods, often uses mathematical and other formal techniques, and is sophisticated in its use of theory. My commentary will not touch on any of those aspects of the work except the last, and in particular its employment of ideas derived from the theory of human rationality.

The commentary will take us through three main topics. First, I shall have to say something about the two main forms of theories of human rationality that prevail in social science today—the one of them having its center in cognitive psychology, the other in economics. Next, I shall consider the implications, for the balance in political science between rationalism (or a priorism) and empiricism, of adopting one or the other of these two paradigms of rationality. In particular, I will argue that there is a natural alliance between empiricism and the psychological version of rationality, on the one hand, and an alliance between rationalism and the economic version of rationality, on the other. Finally, I will comment on the balance between reason and passion—"radical" irrationality—in political affairs.

The Forms of Rationality

The term "rational" denotes behavior that is appropriate to specified goals in the context of a given situation.² If the characteristics of the

²For a more extensive discussion of the concepts of substantive and procedural rationality, see Simon (1976b), reprinted as chap. 8.3 in Simon (1982).

choosing organism are ignored, and we consider only those constraints that arise from the external situation, then we may speak of substantive or objective rationality, that is, behavior that can be adjudged objectively to be optimally adapted to the situation.

On the other hand, if we take into account the limitations of knowledge and computing power of the choosing organism, then we may find it incapable of making objectively optimal choices. If, however, it uses methods of choice that are as effective as its decision-making and problem-solving means permit, we may speak of procedural or bounded rationality, that is, behavior that is adaptive within the constraints imposed *both* by the external situation and by the capacities of the decision maker.

The terms "procedural" and "substantive" were, of course, borrowed from constitutional law, in analogy with the concepts of procedural and substantive due process, the former judging fairness by the procedure used to reach a result, the latter by the substance of the result itself. In the same way, we can judge a person to be rational who uses a reasonable process for choosing; or, alternatively, we can judge a person to be rational who arrives at a reasonable choice.

There is a fundamental difference between substantive and procedural rationality. To deduce the substantively, or objectively, rational choice in a given situation, we need to know only the choosing organism's goals and the objective characteristics of the situation. We need to know absolutely nothing else about the organism, nor would such additional knowledge be of any use to us, for it could not affect the objectively rational behavior in any way.

To deduce the procedurally or boundedly rational choice in a situation, we must know the choosing organism's goals, the information and conceptualization it has of the situation, and its abilities to draw inferences from the information it possesses. We need know nothing about the objective situation in which the organism finds itself, except insofar as that situation influences the subjective representation.

If we review the history of political science over the past 40 years, I believe we will see that it was mainly the procedural view of rationality that was embraced by behavioralism, but that during the past two decades this view has received growing competition from the substantive view. Anthony Downs's *Economic Theory of Democracy*, published in 1957, may be used to date the first nudgings of this new camel into the tent.

I should now like to develop a little further the fundamental characteristics and theoretical structures of the two views of rationality, and then consider the implications of employing them,

separately or jointly, in the study of political behavior.

Procedural Rationality and Cognitive Psychology

A central theme for Graham Wallas in *Human Nature in Politics* was the interplay of the rational and nonrational components of human behavior in politics. That, of course, was also a central theme for Harold Lasswell in *Psychopathology and Politics* (1934) and *World Politics and Personal Insecurity* (1935). But while Lasswell's psychological apparatus comes largely from Freud, Wallas acknowledges as his principal mentor William James. Although Lasswell was concerned with borderline and not-so-borderline pathology, Wallas was interested in the ubiquitous workings of instinct, ignorance, and emotion in normal behavior. Wallas, like his mentor William James, is the more closely attuned to the contemporary orientation in psychology.

What is that orientation? I expressed skepticism, earlier, that political science has experienced, since World War II, any change that deserves being called a revolution. I have no such doubts about the field of psychology. Cognitive psychology, in the past 30 years, has undergone a radical restructuring, from a severe Behaviorism (no relation, I remind you, to behavioralism) to a framework that views thinking as information processing.

In psychology, Behaviorism carefully avoided speaking about what went on inside the head—it preferred to stick to the observable facts of stimuli and responses. It preferred rats to humans as subjects in its experiments, presumably because rats could not be induced to give unacceptable introspective accounts of their mental experiences. Even the term “cognitive” was eschewed, as implying an illicit mentalism.

Today, all of these barriers are down. The term “cognition” is uttered openly and proudly to refer to the human thought processes and to distinguish them from the processes of sensation and emotion. Most experiments use human subjects, and many instruct the subjects to speak aloud as they perform the experimental tasks, the tape-recorded protocols from such sessions being now regarded as wholly objective and analyzable data.³ Theories, in modern cognitive psychology, are expected to provide detailed descriptions of the information processes that go on in the human head when it is performing problem solving and other tasks in the laboratory.⁴

Within this new paradigm, cognitive psychology has made great strides toward understanding how an information processing system like the human brain solves problems, makes decisions, remembers, and learns. That understanding has advanced so far that psychology is no longer limited to dealing with “toy” tasks—puzzles and nonsense syllables—in the laboratory, but can give rather impressive accounts of adult performance in professional-level tasks: making medical diagnoses, solving physics and mathematics problems at high school and college level, learning new mathematics and chemistry, and even making new scientific discoveries, to mention just a few examples.

As examples of explicit applications of the new theories to political science, I can mention the models of public budget-making behavior constructed by Crecine (1969) and Gerwin (1969) and their students, and Carbonell's (1979) ingenious “Goldwater machine,” which predicts the response of an appropriately specified political figure to a situation or set of events. Later, I will cite a number of other accounts of procedural rationality at work in the political process, but in most of these the appeal to cognitive theory and research is only implicit.

The human capabilities for rational behavior that are described by contemporary cognitive psychology are very congenial to the paradigm of bounded rationality as that is described in *Administrative Behavior*. The models of problem solving describe a person who is limited in computational capacity, and who searches very selectively through large realms of possibilities in order to discover what alternatives of action are available, and what the consequences of each of these alternatives are. The search is incomplete, often inadequate, based on uncertain information and partial ignorance, and usually terminated with the discovery of satisfactory, not optimal, courses of action.

To understand the behavior of this kind of problem solver, who is provided in advance with a knowledge of neither alternatives nor consequences—and who may even discover what his or her goals are in the course of the problem-solving process—it is necessary to specify what the problem solver wants, knows, and can compute. Within the framework of these conditionalities, the mere assumption of rationality provides little basis for the prediction of behavior. To be of much use, that assumption must be supplemented by considerable empirical knowledge about the decision maker.

Substantive Rationality and Economics

Just as procedural, bounded rationality is most

³See Ericsson and Simon (1984).

⁴See, for example, Newell and Simon (1972), Simon (1979) and Anderson (1983).

extensively developed in modern cognitive psychology, so substantive, objective rationality finds its principal base in neoclassical economics and statistical decision theory.⁵ The two conceptions of rationality are radically different. The foundation for the theory of objective rationality is the assumption that every actor possesses a utility function that induces a consistent ordering among all alternative choices that the actor faces, and, indeed, that he or she always chooses the alternative with the highest utility.

If the choice situation involves uncertainties, the theory further assumes that the actor will choose the alternative for which the expected utility is the highest. By expected utility of an alternative is meant the average of the utilities of the different possible outcomes, each weighted by the probability that the outcome will ensue if the alternative in question is chosen.

The theory of objective rationality assumes nothing about the actor's goals. The utility function can take any form that defines a consistent ordering of preferences. Nor does the theory postulate anything about the way in which the actor makes probability estimates of uncertain events; in fact one version of the theory, the so-called subjective expected utility, or SEU, theory, explicitly denies that these probabilities are to be identified with objective probabilities of the events, determined by some outside observer. In this one respect, the label "objective" for this version of the theory must be qualified.

In principle (i.e., in a wholly idealized laboratory setting), it should be possible to obtain independent evidence about the nature and shape of any particular person's utility function, as well as evidence of the probabilities that person assigns to events. In practice, this is completely infeasible. In fact, when such experiments have been run, it has generally been found that human subjects do not possess consistent utility functions or probability assignments.⁶

In application, therefore, auxiliary assumptions about utility and expectations must usually be supplied before the theory of objective rationality can be applied to real situations. In economic applications, for example, it is customary to identify the utility function of a firm with its profit, and to assume that actors generally are trying to maximize economic well-being—perhaps some weighted average of income and leisure. In applications to political science, it may be assumed that the goal is to maximize power, or to maximize economic well-being as a function of the policies

pursued by the government. (I will have more to say later about the assumptions that are made regarding political "utility" in applying the principle of rationality to problems in political science.)

In the same way, in applying the theory of objective rationality to real-world behavior, either uncertainty must be ignored, or auxiliary postulates must be provided to define the expectation-forming process. In contemporary economics, for example, the very lively "rational expectations" school, whose leaders include such figures as Robert Lucas and Thomas Sargent, assumes that each economic actor has a more or less accurate model of the economic system, and expects that system to proceed toward its equilibrium in the near future. Of course, there is much doubt whether this particular assumption about the formation of expectations bears any close resemblance to the reality, and a majority of neoclassical economists have different, and simpler, beliefs about how economic actors cope with uncertainty.

When neoclassical economics in its purest form addresses itself exclusively to questions of the existence, stability, and Pareto optimality of equilibrium, it can generally get along without introducing auxiliary assumptions about the utility function or the nature of the expectation-forming processes. In fact, it usually finesses the latter by ignoring uncertainty. The price that is paid is that the conclusions reached by this kind of analysis are extremely general and abstract: roughly, that under conditions of perfect competition, the economic system has a stable equilibrium, and that this equilibrium is, indeed, Pareto optimal (not everyone can simultaneously be made better off than the equilibrium).

When economists want to draw conclusions about nonequilibrium phenomena, matters get stickier. The theory of business cycles provides an important illustration of the difficulties.⁷ The economic theory of Keynes and that of neoclassical economists like Friedman or Lucas are only inches, not miles, apart. Most of Keynes's general theory can be (and has been) interpreted as an exercise in quite orthodox neoclassical reasoning—except at one or two critical points, the most important being the supply of labor. At these points economic actors depart from objective rationality and suffer from persistent illusions or confusions. The assumption in Keynes's theory that produces a business cycle and the possibility of long-continuing unemployment is that labor mistakes its money wage for its real (purchasing power) wage. It is not human rationality, but the

⁵A classical treatment is Savage (1954).

⁶For a number of examples and references to the literature, see Kahneman, Slovic, and Tversky (1982).

⁷This account is based on Simon (1984).

limits on that rationality and its breakdown, that accounts for Keynes's important predictions.

But the same thing can be said of the other, non-Keynesian, theories of the business cycle. (I must except Milton Friedman (1968), who essentially denies that there is such a phenomenon as real unemployment.) For example, Lucas (1981), among the most orthodox of neoclassical economists, attributes the business cycle to a different limit on human rationality. In his theory, it is not labor but businessmen who behave irrationally. When general price changes occur (e.g., inflation), they mistake these changes for *relative* changes affecting only prices in their own industry. It is this departure from objective rationality that produces the cycle in Lucas's model.

I have developed this example at some length because it is perhaps the most dramatic illustration of a widespread phenomenon that is not well understood outside the profession of economics, and perhaps not even within the profession: A large part of the "action" of economic models—the strong conclusions they support—does not derive from the assumptions of objective rationality at all, but depends on auxiliary assumptions that are introduced to provide limits to that rationality, assumptions about the process of decision.

This being the case, one would suppose that a great deal of attention would be devoted to the empirical validity or plausibility of the auxiliary assumptions—in the examples just cited, the assumptions that labor or business, as the case may be, suffers from a money illusion. However, this is not the way the practices and traditions of economics have developed. Instead, there is a tradition that is often referred to, within economics itself, as "casual empiricism." Assumptions about the shape of the utility function or the limits on the rationality of economic actors are commonly made in an armchair, on the basis of feelings of "plausibility" or "reasonableness," and without systematic support from empirical evidence. The assumptions are never tested directly, but only in the context of the models in which they are embedded. The goodness of fit of a model, usually to aggregate data, is regarded as the best justification for the assumptions embedded in that model, whatever their source.⁸

⁸Friedman's well-known methodological essay transforms these methodological practices into a strongly defended doctrine. Friedman argues that direct tests of the behavioral assumptions underlying an economic model are superfluous at best, and positively misleading at the worst.

Bounded Rationality Is Not Irrationality

Skepticism about substituting a priori postulates about rationality for factual knowledge of human behavior should not be mistaken for a claim that people are generally "irrational." On the contrary, I think there is plenty of evidence that people are generally quite rational; that is to say, they usually have reasons for what they do. Even in madness, there is almost always method, as Freud was at great pains to point out. And putting madness aside for a moment, almost all human behavior consists of sequences of goal-oriented actions.

When, in spite of the evidence for this goal-oriented character of human behavior, we call some of that behavior "irrational," we may mean any one of several things. We may deem behavior irrational because, although it serves some particular impulse, it is inconsistent with other goals that seem to us more important. We may deem it irrational because the actor is proceeding on incorrect facts or ignoring whole areas of relevant fact. We may deem it irrational because the actor has not drawn the correct conclusions from the facts. We may deem it irrational because the actor has failed to consider important alternative courses of action. If the action involves the future, as most action does, we may deem it irrational because we don't think the actor uses the best methods for forming expectations or for adapting to uncertainty. All of these forms of "irrationality" play important roles in the lives of every one of us, but I think it is misleading to call them "irrationality." They are better viewed as forms of bounded rationality.

To understand and predict human behavior, we have to deal with the realities of human rationality, that is, with bounded rationality. There is nothing obvious about these boundaries; there is no way to predict, a priori, just where they lie.

The Rationality Principle in Politics

After this long excursion into the views of human rationality that are commonly held in psychology and economics, let me come back now to the subject of political science. What kind of rationality does *Homo politicus* exhibit? Is he or she a creature of objective, substantive rationality; or instead, one of subjective, procedural rationality? But I am afraid that I have already tipped my hand and made it quite clear that I believe the latter to be the case.

If that is true, the rationality principle, as it is incorporated in theories of substantive rationality, will provide us with only limited help in understanding political phenomena. Before we apply

the methods of economic reasoning to political behavior, we must characterize the political situation, not as it appears "objectively" to the analyst, but as it appears subjectively to the actors. We can only select the appropriate model of adaptation after we undertake the requisite empirical study to determine this subjective representation both of goals and of the situation or draw upon research in cognitive psychology to tell us about the nature of that representation. A few examples drawn from the political science literature will show what is involved.

An Example: Duverger's Law

Recently, William Riker (1982) provided us with an instructive account of a descriptive generalization that usually goes by the name of Duverger's Law. In its roughest form, the law asserts that plurality election rules bring about and maintain two-party, rather than multi-party, competition. In an informative way, Riker takes us through the history of the empirical research that has been done to test, to confirm, refute, or amend, this law. He also shows that political scientists have not been content simply to assert the law, or to test it empirically; they have also sought to "explain" it. He says

From the first enunciation by Droop, the law has been implicitly embedded in a rational choice theory about the behavior of politicians and voters. This theory has been rendered more and more explicit, especially in the last two decades, so that recent empirical work consciously invokes the rational choice model. (1982, p. 766)

The so-called rational choice argument for Duverger's Law goes something like this. If a number of candidates are running for office under a plurality election rule, and if candidates A and B are well ahead of the pack so that it is unreasonable to suppose that any other candidate will win, then it is rational to limit your vote to your preference between A and B. The argument has to be elaborated somewhat to account for two-party configurations that are stable over time, but I think that I have conveyed the general idea.

What assumptions does this argument make about you, the voter. First, it assumes that you have a preference ranking among candidates and wish to vote so as to secure the election of a candidate who is as high as possible on your ranking. Second, it assumes that you believe that one vote may decide the election (otherwise it is indifferent, in terms of the stated goal to whom the vote goes). Third, it assumes that you have an assessment of the relative prospects of the candidates, and a

considerable confidence in that assessment (e.g., you do *not* believe that one more vote could bring success to any but one of the two candidates judged to have the most support). Fourth, it assumes that you do not attach a large value to providing public evidence that your most preferred candidate has extensive, even if not pluralistic, public support.

Since I have not tried to construct a formal axiomization of this choice, perhaps there are other assumptions that must be made, in addition to those listed above. For the purposes of the present argument, however, my inventory of assumptions will suffice. What the assumptions show is that only a small part of the work of explaining Duverger's Law is being done by the rationality principle. Most of the work is being done by propositions that characterize the utility function of the voter and his or her beliefs, expectations, and calculations—that is to say, the limits of rationality. These propositions are subject to empirical test.

Perhaps the key assumption here is the postulate of "sophisticated voting," that a rational voter believes "his vote should be expended as part of a selection process, not as an expression of preference" (Downs, 1957, p. 48). But this postulate is wholly independent of the usual definition of objective rationality. There is no irrationality in a utility function that regards a vote as an expression of preference rather than an attempt to influence the selection. In fact, it is realistic to believe that one can express a preference (i.e., change the numerical result of the vote, if only by a unit), but seldom realistic to believe that one can affect the outcome of an election. Moreover, a voter might correctly (or incorrectly, but certainly not irrationally) believe that expression of preference for a party could increase the chances of that party's succeeding in subsequent elections.

There are many more changes we can ring on the possible beliefs of voters without impugning their (subjective) rationality. With these alternative sets of beliefs are associated different voting behaviors. It is not at all hard to build a rational model of the voter who stays home from the polls and does not vote at all. Hence, we get very little understanding or explanation of voting behavior simply from invoking the principle of utility maximization. That principle does not exempt us from the arduous task of testing all the auxiliary empirical assumptions about voters' values, beliefs, and expectations. And, as Riker shows us, when we subject an auxiliary assumption like the postulate of sophisticated voting to empirical test, we discover that the actual pattern of human response can be very complex indeed. We are then constructing and testing theories of bounded rationality, not theories of substantive rationality.

Additional Examples

It should not be thought that Duverger's Law is an isolated case and that rational choice theories derived from the assumption of utility maximization and unalloyed with auxiliary assumptions about preferences and beliefs have much more predictive and explanatory power in most other cases. Recent issues of the *American Political Science Review* provide a rich mine of examples that support our analysis of the respective roles of reason and fact. One can stumble upon such examples by opening the pages almost at random, and it appears to make little difference whether the author is a behavioralist or an economic rationalist by persuasion. (Or if there is a difference, it is that the behavioralist makes fewer explicit claims for rationality as the source of his or her conclusions than does the rational choice theorist.)

My next example is a study by Hibbs (1982) of "Economic Outcomes and Political Support for British Governments among Occupational Classes." Hibbs demonstrates that various indicators of the health of the British economy are related to voting preferences. Score one for the objective rationality principle. Presumably voters vote for the party that they think will enhance their economic well-being. But how do we get from that general proposition to a prediction of their vote? We can make the leap only if we can discover how voters *judge* which party will do the better job of managing the economy. There are many ways in which that judgment could be made, none of them, probably, having high objective validity.

Hence, the interesting and significant finding of Hibbs's study is not that people employ a rationality principle. The interesting finding, which does not follow from such a principle, is that "voters evaluate the cumulative performance of the governing party relative to the prior performance of the current opposition," weighting current performance more heavily than past performance (p. 259).

Now I don't know if Hibbs's model will hold up under further analysis or will apply equally well to other times and places. However valid or invalid the model, its powerful motor is not a theory of objectively rational choice but a very specific empirical assumption, based on notions of bounded rationality, about how voters form their beliefs regarding the connections between the economy and government. If Hibbs's model is correct, voters do this not by solving a maximization problem but by setting an aspiration level (the opposition's past performance) against which to measure the performance of the incumbents. This is what modern cognitive theory would lead us to

expect, but not what would be predicted by a theory of utility maximization.

A third example has to do with the application of rationality principles to a game resembling the prisoners' dilemma, but allowing the players the additional alternative of exiting from the situation (Orbell, Schwartz-Shea, & Simmons, 1984). In their abstract, the authors, using the usual distinction between defectors and cooperators in the prisoners' dilemma, sum up the matter very well:

We derive the prediction that the exit option will drain the community or group more of cooperators than of defectors.

But experimental data do not support this prediction; cooperators do not leave more frequently than defectors. . . . [We] present data supporting the hypothesis that cooperators often stay when their personal interest is with exiting because of the same ethical or group-regarding impulse that (presumably) led them to cooperate in the first place.

In this experiment, again, the principle of objective rationality contributes little to predicting or explaining the findings. Everything rests, instead, on the assumptions that are made about the utility functions of two classes of players, those who are prepared to cooperate with the other players and those who are prepared to betray them. What is more, to explain the behavior of the cooperators, a strong component of altruism must be introduced into their utility functions.⁹

Other research within a game-theoretical framework shares many of the characteristics of this study. The predicted outcome depends sensitively upon assumptions not derivable from the principle of objective rationality, about participants' beliefs and values. For example, in a study involving the conditions under which subjects would contribute to the provision of public goods, the authors summarize their findings thus (van de Kragt, Orbell, & Dawes, 1983, p. 112):

We present hypotheses about why designating a minimal contributing set works. . . . The essential property of the minimal contributing set . . . is *criticalness*: the contributions of the members of the minimal contributing set are each critical to obtaining the public good the members desire, and they know it. It is reasonable (albeit not a dominant strategy) to contribute because reasonable behavior can be expected from other minimal contributing set members who are in the same situation.

⁹For a discussion of the problems of reconciling altruism with rationality in systems subject to evolutionary selection, see Simon (1983, chap. 2).

What is called reasonable behavior here is clearly the behavior we might expect of a creature of bounded rationality. And its reasonableness depends on expectations about the behavior of others.

Perhaps the major contribution of game theory to political science has been to demonstrate how rare and unusual the situations are where a game has a stable equilibrium solution consistent with the principle of objectively rational choice. Under these circumstances, the task of determining how people actually do behave in situations having game-like characteristics must be turned over to empirical research: research that seeks to determine what values people actually act on, and how they form their expectations and beliefs.

My final example concerns considerations of economic advantage in voting decisions. Weatherford (1983) points out that the concept of economic voting is ambiguous. It may mean voting in response to perceptions of one's own economic well-being, or voting in response to perceptions of the health of the economy. But this distinction is itself ambiguous, for it may refer to differences in utility functions or to differences in the voter's model of reality.

You, the voter, may want to vote for the candidate who will do best for you (for example, support the "right" kinds of tax laws, impose or remove the "right" kinds of regulations), or for the candidate who will best foster the vigor of the whole economy, even if it costs you, personally, a loss of income or of a job. Put in these terms, the difference lies in the structure of your utility function.

But we can look at the matter in a different way. How do you judge the state of the economy or your well-being? You can use the immediate evidence of your personal situation—your employment or unemployment, your salary, your taxes. Or you can look at published economic indexes. And, because the question before you is not the current state of the economy, but how it is likely to be affected if one candidate or another is elected, there are still other kinds of evidence that may influence you. You may consider the candidates' past voting records or the economic predispositions of the parties to which they belong.

Differences in the kinds of evidence you respond to may have nothing to do with your utility function. Instead, they may reflect the model you have of the world, the beliefs you have formed about the meanings and predictive value of different kinds of available information, and what information has come to your attention.

All of these examples teach us the same lesson: the actors in the political drama do appear to behave in a rational manner—they have reasons for what they do, and a clever researcher can

usually obtain data that give good clues as to what those reasons are. But this is very different from claiming that we can predict the behavior of these rational actors by application of the objective rationality principle to the situations in which they find themselves. Such prediction is impossible, both because, even within the framework of the SEU theory of substantive rationality, behavior depends on the structure of the actors' utility functions, and because it depends on their representation of the world in which they live, what they attend to in that world, and what beliefs they have about its nature.

The obvious corollary is that rationalism can carry us only a little way in political analysis, even in the analysis of the behavior of boundedly rational people. The rest of the path requires continuing, painstaking empirical investigation within the framework of modern cognitive theories of human behavior.

Rationalism and Empiricism

I should not like my comments to be interpreted as a complaint that political science worships at the altar of rational choice theory. On the contrary, I think we political scientists have generally been behaving quite well in this respect. If I take the pages of the *American Political Science Review* as representing the attitudes and methods of our discipline, then I observe that there is a healthy respect for sophisticated empirical research. Assumptions of rationality are used to provide a framework for analyzing behavior, but they are generally used tentatively, and with a sensitivity to the assumptions of value, expectation, and belief that have to be added to the models before they can yield predictions of behavior.

Authors who use rational choice models are not always conscious of the extent to which their conclusions are independent of the assumptions of those models, but depend, instead, mainly upon auxiliary assumptions. Nor is advantage taken as often as it could be of the knowledge of cognitive mechanisms to be found in the psychological literature. But these defects, if defects they be, are easily remedied.

It is also a good omen for the future of our science that empirical work means both the study of social aggregates, whose behavior is recorded in public statistics, and the study of the individual actors at the microscopic and face-to-face level of the interview and the poll. The graduate training we provide our students gives them opportunities to acquire skill in both kinds of empirical methodology, and others (e.g., historical inquiry) as well. In this respect, we are better off than our brethren in economics, who are seldom trained in the skills of observing economic phenomena at first hand.

We sometimes, perhaps, experience a mild malaise in that our research does not seem to be taking us in the direction of a few sweeping generalizations that encompass the whole of political behavior. A hope of finding our "three laws of motion" was probably a major part of the appeal of rational choice theory in its purer forms. But a more careful look at the natural sciences would show us that they, too, get only a little mileage from their general laws. Those laws have to be fleshed out by a myriad of facts, all of which must be harvested by laborious empirical research. Perhaps our aspirations for lawfulness should be modeled upon the complexities of molecular biology—surely a successful science, but hardly a neat one—rather than upon the simplicities of classical mechanics.

Radical Irrationality

Thus far, I have dealt with the picture of procedural rationality that emerges from modern cognitive psychology and the relation between that picture and the economist's notion of substantive rationality. My main conclusion is that the key premises in any theory that purports to explain the real phenomena of politics are the empirical assumptions about goals and, even more important, about the ways in which people characterize the choice situations that face them. These goals and characterizations do not rest on immutable first principles, but are functions of time and place that can only be ascertained by empirical inquiry. In this sense, political science is necessarily a historical science, in the same way and for the same reason that astronomy is. What will happen next is not independent of where the system is right now. And a description of where it is right now must include a description of the subjective view of the situation that informs the choices of the actors.

But you may feel that I have not gone far enough in my skepticism about reason in political behavior. Surely even the concept of bounded rationality does not capture the whole role of passion and unreason in human affairs. Don't we need to listen to Lasswell and Freud as well as to Wallas and James?

Assuredly we do. From the earliest times it has been seen that human behavior is not always the result of deliberate calculation, even of a boundedly rational kind. Sometimes it must be attributed to passion, to the capture of the decision process by powerful impulses that do not permit the mediation of thought. The criminal law takes explicit account of passion in assigning different penalties to deliberate and impulsive acts.

In psychoanalytic theory, passion takes mainly the form of unconscious drives, largely unknown

to the actor, that provide the "real" wellsprings of action. This approach, whether it be correct or false, has always been troublesome for empirical research, because it makes suspect human testimony about motives.¹⁰ If we don't know why we act, if our motives are unconscious, then we can't report them, no matter how much we wish to cooperate with the researcher.

Let me take a more conservative approach, which accords well with what we know about the mechanisms that link emotions to reason (Simon, 1978, chap. 1.3). People are endowed with very large long-term memories, but with very narrow capacities for simultaneous attention to different pieces of information. At any given moment, only a little information, drawn from the senses and from long-term memory, can be held in the focus of attention. This information is not static; it is continuously being processed and transformed, with one item being replaced by another as new aspects of a stimulus are sensed, new inferences drawn, or new bits of information retrieved from long-term memory. Nevertheless, of all the things we know, or can see or hear around us, only a tiny fraction influences our behavior over any short interval of time.

If a particular strong drive takes control of our attention, determining not only our goals of the moment but also selecting out the sensory and memory facts that we will consider, then behavior can be determined by that drive or passion as long as its control persists. But passionate behavior in this extreme form is exceptional and not common in human behavior. The control process is usually more complex.

Even in the case of a person like Hitler, whose behavior might be interpreted by some clinicians as a pure instance of an all-consuming hatred or self-hatred, a large cognitive element intrudes into the behavior. Hitler was not just angry; he directed his hatred toward a particular group of people, Jews, and he made decisions that were arguably rational on the premise that the Jewish people were to be extirpated to satisfy that hatred. For some purposes of political analysis, it may be enough to postulate the overtly expressed values and goals without seeking their deeper roots in the unconscious, or at least without trying to explain how they arrived there.

The methodological lesson I would draw is that we need to understand passion and to provide for

¹⁰For a review of some reasons why we should suspect testimony about motives, see Nisbett and Wilson (1977). The authors of that study draw conclusions that are rather too broad for their evidence, but their main point about reports of motivation are well taken. See also Ericsson and Simon (1984).

it in our political models, but we need particularly to provide in those models for the limited span of attention that governs *what* considerations, out of a whole host of possible ones, will actually influence the deliberations that precede action. In particular, we need to understand the conditions that predispose human beings to impulsive action that disregards much of the potentially relevant reality. I would like to comment on three aspects of this question: the nature of the attention mechanism, the role of uncertainty, and the process whereby novel ways of viewing situations are evoked or generated.

Attention

The human eye and ear are highly parallel devices, capable of extracting many pieces of information simultaneously from the environment and decoding them into their significant features. Before this information can be used by the deliberative mind, however, it must proceed through the bottleneck of attention—a serial, not parallel, process whose information capacity is exceedingly small. Psychologists usually call this bottleneck short-term memory, and measurements show reliably that it can hold only about six chunks (that is to say, six familiar items) of information.

The details of short-term memory and the bottleneck of attention are not important for our purposes. What is important is that only one or a very few things can be attended to simultaneously. The limits can be broadened a bit, but only modestly, by “time-sharing”—switching attention periodically. The narrowness of the span of attention accounts for a great deal of human unreason that considers only one facet of a multifaceted matter before a decision is reached.

For example, it has been hypothesized that the art of campaign oratory is much more an art of directing attention (to the issues on which the candidate believes himself or herself to have the broadest support) than an art of persuading people to change their minds on issues.¹¹ Similarly, shifts in expressed voting intentions during the course of an election campaign have been explained as caused by evocation of beliefs and attitudes already latent in voters’ minds (e.g., party loyalties) (Lazarsfeld et al., 1948, chap. 9). Another example, highly characteristic of the political process, was the shift of attention from environmental problems to problems of energy supply that took place immediately after the Oil Shock, and that greatly altered public priorities for a number of years.

The unreason associated with attention focusing has no necessary connection with passion—cold reasoning can be as narrow and one-sided as hot reasoning. But the existence of these narrow limits on the span of human attention is a principal reason why we must distinguish between the “real” situation and the situation as perceived by the political actors when we try to apply the rationality principle to make predictions of behavior. People are, at best, rational in terms of what they are aware of, and they can be aware of only tiny, disjointed facets of reality.

Uncertainty

Lack of reliable knowledge and information is a major factor in almost all real-life decision making. In our soberer moments, we realize how little we know and can predict about the decision-making premises and processes of the rulers of the USSR. Yet the content of a rational foreign policy is highly sensitive to our hypotheses about these matters. The effects of the policies of the president upon the well-being of the American economy are only slightly less uncertain. At least there is often little consensus in the economics profession about these effects.

Wherever such uncertainties are present, an enhanced opportunity is provided for unconscious, or only partly conscious, drives and wishes to influence deliberation. Where the facts are clear (to the actors as well as to us), we have some chance, by application of the principles of reason, to calculate what the choice will be. Where evidence is weak and conflicting, a rationality principle has little independent predictive power.

Evocation

Finally, to understand political choices, we need to understand where the frame of reference for the actors’ thinking comes from—how it is evoked. An important component of the frame of reference is the set of alternatives that are given consideration in the choice process. We need to understand not only how people reason about alternatives, but where the alternatives come from in the first place. The process whereby alternatives are generated has been somewhat ignored as an object of research.

But not wholly ignored! Turning again to my favorite source of information on the state of the profession, I find in a recent issue of *The American Political Science Review* another imaginative paper by William Riker, in fact his 1983 Presidential Address to the Association, on precisely this issue. (I could wish that he had not invented the word “heresthetics” to conceal the heresies he is propagating.) Riker traces the history of pro-

¹¹For a classic statement of this hypothesis, see Lazarsfeld, Berelson, and Gaudet (1948, chap. 8).

posals in the Constitutional Convention for electing the President, with particular concern for the generation of new alternatives, and for the shifts in attention and emphasis on issues that accompanied their introduction.

Riker speaks of these matters in terms of "artistry within the rational choice context." I think that the generation of alternatives is much more than that: that it is an integral component of any veridical account of human decision making, or of human bounded rationality generally. The theory of the generation of alternatives deserves, and requires, a treatment that is just as definitive and thorough as the treatment we give to the theory of choice among prespecified alternatives.

But is such a treatment possible? Are we not treading upon the sacred precincts of creativity? Indeed we are; but I think the precincts are no longer sacrosanct. The same cognitive psychology that has been elaborating the theory of human bounded rationality has made considerable progress toward constructing models of the processes of discovery and creativity that can account for these processes in terms quite akin to those it uses to account for ordinary problem solving. Again, I cannot tell that story here but must limit myself to pointers to the literature (Bradshaw, Langley, & Simon, 1983; Lenat, 1983).

Conclusion

In this essay I have tried to provide an overview—a very general one—of our current knowledge of human nature in politics. I first undertook to compare the two principal theories of human rationality that have found application in political research: the procedural bounded rationality theory that has its origins in contemporary cognitive psychology, and the substantive global rationality theory that has been nurtured chiefly in economics. Then, by means of a series of examples, I examined the relative roles played by rationality principles and by the auxiliary assumptions that accompany them, respectively, in predicting and explaining human behavior in political contexts. Finally, I commented on the more extreme deviations from the objective rationality model that exhibit themselves in political affairs, and showed how they could be explained, in considerable measure, in terms of the mechanisms of attention and the severe limits that the architecture of the mind places on the span of human attention.

My overview, if it is even partly valid, carries a number of implications for research in political science. First, it dissipates the illusion, if anyone holds it, that an application of principles of rationality can discharge us, to any considerable degree, from the need to carry on painstaking em-

pirical research at both macro and micro levels. It is far easier (for the political scientist and for the political actor) to calculate the rational response to a fully specified situation than it is to arrive at a reasonable specification of the situation. And there is no way, without empirical study, to predict which of innumerable reasonable specifications the actors will adopt.

Second, my overview suggests that the study of the mechanisms of attention directing, situation defining, and evoking are among the most promising targets of political research. In particular, the question of where political ideas come from is not only highly deserving of study, but also within the competence of our contemporary research techniques. I join Bill Riker in commending it to you as one of the truly exciting and significant areas of investigation in our field.

Nothing is more fundamental in setting our research agenda and informing our research methods than our view of the nature of the human beings whose behavior we are studying. It makes a difference, a very large difference, to our research strategy whether we are studying the nearly omniscient *Homo economicus* of rational choice theory or the boundedly rational *Homo psychologicus* of cognitive psychology. It makes a difference to research, but it also makes a difference for the proper design of political institutions. James Madison¹² was well aware of that, and in the pages of the *Federalist Papers* he opted for this view of the human condition (*Federalist*, No. 55):

As there is a degree of depravity in mankind which requires a certain degree of circumspection and distrust, so there are other qualities in human nature which justify a certain portion of esteem and confidence.

—a balanced and realistic view, we may concede, of bounded human rationality and its accompanying frailties of motive and reason.

References

- Anderson, J. *The architecture of cognition*. Cambridge, Mass.: Harvard University Press, 1983.
 Bradshaw, G. F., Langley, P. W., & Simon, H. A. Studying scientific discovery by computer simulation. *Science*, 1983, 222 (Dec. 2), 971-975.
 Carbonell, J. G. *Subjective understanding: Computer*

¹²Or was it Hamilton? Mosteller and Wallace (1964) attribute No. 55 of *The Federalist* to Madison, but it is the least certain of their attributions of the numbers whose authorship has been disputed. Since the sentiment quoted here is certainly consistent with the beliefs of both authors, we need not be too concerned with the uncertainty of authorship.

- models of belief systems*. New Haven, Conn.: Yale University Dept. of Computer Science, 1979.
- Crecine, J. P. *Governmental problem solving: A computer simulation of municipal budgeting*. Chicago: Rand McNally, 1969.
- Downs, A. *An economic theory of democracy*. New York: Harper & Row, 1957.
- Eastman, D. Political science. In D. L. Sills (Ed.), *International encyclopedia of the social sciences* (vol. 12). New York: Macmillan, 1968, pp. 282-298.
- Ericsson, K., & Simon, H. A. *Protocol analysis: Verbal reports as data*. Cambridge, Mass.: MIT Press, 1984.
- Friedman, M. The methodology of positive economics. In *Essays in positive economics*. Chicago: University of Chicago Press, 1953.
- Friedman, M. The role of monetary policy. *American Economic Review*, 1968, 58, 1-17.
- Gerwin, B. *Budgeting public funds: The decision process in an urban school district*. Madison: University of Wisconsin Press, 1969.
- Hibbs, D. A., Jr. Economic outcomes and political support for British governments among occupational classes: A dynamic analysis. *American Political Science Review*, 1982, 76, 259-279.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press, 1982.
- Keynes, J. M. *The general theory of employment, interest and money*. London: Macmillan, 1936.
- van de Kragt, A. J. C., Orbell, J. M., & Dawes, R. M. The minimal contributing set as a solution to public goods problems. *American Political Science Review*, 1983, 77, 112-122.
- Lasswell, H. *Psychopathology and politics*. Chicago: University of Chicago Press, 1934.
- Lasswell, H. *World politics and personal insecurity*. New York: McGraw-Hill, 1935.
- Lazarsfeld, P. S., Berelson, B., & Gaudet, H. *The people's choice* (2nd ed.). New York: Columbia University Press, 1948.
- Lenat, D. B. EURISKO. A program that learns new heuristics and domain concepts. *Artificial Intelligence*, 1983, 21, 61-98.
- Lippmann, W. *Public opinion*. New York: Macmillan (1922/1944).
- Lucas, R. E., Jr. *Studies in business cycle theory*. Cambridge, Mass.: MIT Press, 1981.
- Lucas, R. E., Jr., & Sargent, T. J. (Eds.). *Rational expectation and economic practice*. Minneapolis: University of Minnesota Press, 1981.
- Mosteller, F., & Wallace, D. W. *Influence and disputed authorship: The Federalist*. Reading, Mass.: Addison-Wesley, 1964.
- Newell, A., & Simon, H. A. *Human problem solving*. Englewood Cliffs, N.J.: Prentice Hall, 1972.
- Nisbett, R. E., & Wilson, T. D. Telling more than we know: Verbal reports on mental processes. *Psychological Review*, 1977, 84, 231-259.
- Orbell, J. M., Schwartz-Shea, P., & Simmons, R. T. Do cooperators exit more readily than defectors. *American Political Science Review*, 1984, 76, 753-766.
- Riker, W. H. The heresthetics of constitution-making. The presidency in 1787, with comments on determinism and rational choice. *American Political Science Review*, 1984, 78, 1-16.
- Riker, W. H. The two-party system and Duverger's Law: An essay on the history of political science. *American Political Science Review*, 1982, 76, 753-766.
- Savage, L. J. *The foundation of statistics*. New York: Wiley, 1954.
- Simon, H. A. *Administrative behavior*. New York: Free Press, 1947/1976. (a)
- Simon, H. A. A behavioral model of rational choice. *Quarterly Journal of Economics*, 1955, 69, 99-118. Reprinted in H. A. Simon. *Models of bounded rationality* (vol. 2, chap. 7.2). Cambridge, Mass.: MIT Press, 1982.
- Simon, H. A. From substantive to procedural rationality. In S. J. Latsis (Ed.), *Method and appraisal in Economics*. Cambridge: Cambridge University Press, 1976, pp. 129-148. (b)
- Simon, H. A. *Models of bounded rationality*. Cambridge, Mass.: MIT Press, 1982.
- Simon, H. A. *Models of thought*. New Haven, Conn.: Yale University Press, 1979.
- Simon, H. A. Motivational and emotional controls of cognition. In *Models of thought*. New Haven, Conn.: Yale University Press, 1979, chap. 1.3.
- Simon, H. A. On the behavioral and rational foundation of economic dynamics. *Journal of Economic Behavior and Organization*, 1984, 5, 35-55.
- Simon, H. A. *Reason in human affairs*. Stanford, Calif.: Stanford University Press, 1983.
- Wallas, G. *Human nature in politics* (4th ed.). Gloucester, Mass.: Smith, 1906/1944.
- Weatherford, M. S. Economic voting and the "symbolic politics" argument: A reinterpretation and synthesis. *American Political Science Review*, 1983, 77, 158-174.