Rationality in Political Behavior

Herbert A. Simon
Carnegie Mellon University

Public choice theory borrows the basic assumptions of neoclassical economics about the nature of human rationality and applies them to the explanation and prediction of behavior in the political domain (Downs, 1957). An attractive, and seductive, feature of the theory is a very strong rationality assumption (maximization of subjective expected utility) that appears to permit a great deal of explanation and prediction without the painful necessity of first constructing an empirically based theory of human behavior, in particular, a theory of the nature and limits of human rationality. To an important extent, deductive reasoning from the theory's basic postulates of rationality substitutes for a great deal of costly empirical inquiry.

The frequent use of public choice theory today in political science calls for an examination of the assumptions of rationality that the theory employs. This paper carries out such an examination. The analysis focuses on three issues: the nature of the rationality assumption, the orthogonality of rationality and selfishness, and the possibility of altruism. Our inquiry will lead us to the conclusion that human rationality is much more complex than it is portrayed to be in neoclassical and public choice theories, and that much less can be derived by deductive means from the (amended) assumptions of rationality than has been supposed by the exponents of those theories. An important corollary to this conclusion is that a veridical theory of public choice requires a solid foundation of empirical fact about the nature of human goals and about the processes that people use in reasoning from their actions to their values.

People are Rational

By a weak definition of rationality, virtually all human behavior is rational. People usually have reasons for what they do, and if asked, can opine what these reasons are. Of course, as Freud (and laboratory experiments) have taught us, people may deceive themselves; the real reasons may be different from what they suppose they are.
When we do not find people's reasons for their actions to be credible, we do not thereby judge them to be irrational. We simply conclude that they were deceiving us (or themselves), and assign other, more plausible, reasons to the behavior. When we judge people to be irrational, we do not mean that they do not have reasons for what they do, but rather (a) that their reasons are quite different from the reasons people normally give (as in various forms of insanity), (b) that the reasons are based on incorrect assumptions of fact, (c) that there are other, and stronger, unstated reasons for not taking different actions, or (d) that their values are not ours (Simon, 1983).

**Slips twixt Lip and Cup.** To say that people have reasons for their actions means that there is a connection between the actions and the goals (values, utility functions) the actors possess. The actions increase the likelihood that some of these goals will be achieved. However, even in behavior that we would call rational, there may be serious gaps between action and goal achievement. First, the actor may have (and usually will have) incomplete or erroneous information about the situation and the potential changes in the situation with the passage of time. The action may fail to achieve its goal because of lack of information or wrong information.

Second, even if the information is complete, the actor may be unable (and usually will be unable) to compute all of the consequences of an action. Computational limits may cause a large gap between intended goal and actual results. Actions will have different consequences than expected, and often side effects will ensue that were not contemplated at all.

Third, actors generally have more than one goal, and there may be incompatibilities among goals, the realization of one interfering with the realization of others. Such incompatibility often arises when efforts aimed at one goal consume scarce resources, so that these are then no longer available for realizing the other goals. Or the realization of a goal may entail side effects upon others, or may block goals that would otherwise have been reached. Seldom are more than a few of the goals and consequences that might be affected considered before actions are undertaken. Attention is usually, and of necessity focused on just the most salient of them.

Fourth, the actor may fail to reach a goal through ignorance of possible courses of action; or might reach it more slowly and with greater use of resources than if another (unknown) action had been known and available. In general, actions are not given, but
must be discovered or devised. The whole sets of human activities we call invention and
design are aimed at generating new means for reaching our ends. These means only
become available after investment of substantial design effort, and decisions may be
reached before search for effective actions is adequate.

In short, people almost always have reasons for what they do, but seldom the
"best" reasons. That is to say, consequences of actions are ignored or misjudged, either
because information is lacking or erroneous or because computational power (thinking
power) is insufficient for estimating the consequences; tradeoffs among goals are
handled inadequately or not at all; and finally, potential effective actions may be
unknown (and even unknowable) or ignored. As a general label for these departures
from the global rationality postulated in economic and public choice theory, we speak of
human bounded rationality.

The framers of the United States Constitution were (boundedly) rational: they had
reasons for the decisions they made, and for the votes on these decisions that they cast in
the Convention. This is not to say that these were the "best" reasons. The founding
fathers frequently misjudged the consequences of structural provisions: for example, the
designation as Vice President of the candidate scoring the second largest electoral vote for
the Presidency. They aimed at certain goals (equality of representation of the States) at
the expense of other goals (equality of representation of voters). In the case just cited,
they invested in the Senate the attainment of one of these goals, and in the House, the
attainment of the other, thus avoiding the need for any kind of weighting function for
comparing them.

Nor did the framers have before them a menu of all possible constitutions or all
possible constitutional provisions. Instead, they invented some provisions as they went
along: the Virginia Compromise, the time limit on the continuation of the slave trade, and
so on (Riker, 1984). Another Convention might have invented quite different
provisions, and might have framed a quite different document.

With all of these qualifications and provisos, we must still say that the framers
of the Constitution were rational. They had reasons (and gave reasons, real or spurious)
for their preferences and proposals. They gave arguments, sometimes elaborate, with
evidence, occasionally extensive, for the connections they claimed between provisions
proposed and the goals these provisions would achieve. They were rational enough that
their structure has endured more than 200 years, with only modest repairs along the way. It can be claimed that it has achieved to a large degree many of the goals it was aimed at, and the American population today, and even a sizeable part of the world's population, give considerable credence to these claims.

**Rationality in Economics.** However, the behavior of the framers of the Constitution was not rational by the standards of modern neoclassical economics, where rationality of a decision implies more than that reasons can be given for it. It is only rational if it is supported by the *best* reasons and achieves the best possible outcome in terms of all the goals. Appealing to all the goals requires that there be a way of adjudicating among them, and that means some sort of weighting function. Economists postulate such a function, which they call the utility function and require to be consistent. In particular, if A is preferred to B, and B is preferred to C, then A must be preferred to C.

Beyond requiring that the utility function be consistent, no specific content is specified for it. As far as economics is concerned, a utility function can assign as much or more utility to giving away goods as to consuming them; it can award as much utility (to me) for an increase in the living standard of an Indian peasant as it does for the same increase in my own living standard. Rationality is orthogonal to selfishness.

Utility can reside in the possession of wealth, the consumption of wealth, the amassing of power, the accumulation of the good will of other people, the acquisition of glory, the suffering of masochistic pain. Name your favorite tangible or intangible good; there is no objection to including it in your utility function -- provided that consistency is preserved. If we define an altruist as someone whose utility lies in giving pleasure to other people, altruism is entirely compatible with the economist's definition of utility.

The theory does permit uncertainty, but only if probabilities are somehow attached to the possible different outcomes, so that the action having the best expected value of the outcome, that is, the highest expected utility, will be chosen.

Moreover, the theory assumes that all of the alternatives of action are given. In one variant of the theory, alternatives may be searched for, at a cost. Then the problem for the rational actor is to stop searching exactly when the marginal cost of continuing would just equal the expected marginal increase in the value of the best alternative discovered to date. However, this extension of the theory has had only a few applications.
(e.g., consumer product choice, or search for employment), for it usually imposes an impossible task of estimating the marginal costs and returns of additional search. It complicates the whole decision process even further beyond human capabilities than when the alternatives are assumed to be given in advance (Simon, 1983, Chapter 1).

The economist's definition of rationality is the definition used in public choice theory (with one qualification that I will introduce later). The definition imposes a crushing weight of knowledge acquisition, computation, and global consistency on the decision maker at the same time that it allows him or her nearly complete freedom in the selection of the goals of action. The former property of the theory might appear to provide it with strong predictive power: given the actor's goals as defined by a utility function, and the (known) alternative actions, and the known (at least up to a probability distribution) connections between action and utility, only one action, in general, can be rational. The actor's freedom to define the utility function destroys almost all of the predictive power so arduously obtained, because what action is rational will depend critically on what goals are assumed to have utility, and this is not specified by the theory.

For example, rational choice theorists have puzzled at great length about the "voting paradox." Everyone knows that his or her single vote will almost never change the outcome of an election. Why, then, does anyone go to the trouble of voting? Ordeschook cut the Gordian knot by simply including the pleasure of voting in the voter's utility function. This solution -- apart from the fact that it is completely ad hoc, fails to predict how many people will vote and in what elections, and does not tell us anything we did not already know. It has been criticised for violating the "principle" of consequentialism (See Riker, this volume), which rules out actions that have no consequences.

But what are consequences? Since rational choice theory leaves the content of the utility function entirely free, why not stuff it full of pleasures of voting, daydreams, and what not, as well as desires to gain and spend wealth? Assuming a pleasure in voting is no more or no less grounded than any of the other auxiliary assumptions that are necessary before the theory can predict behavior. They are all assumptions of fact, and in science, matters of fact are to be settled by observation and experiment, not by posit.
Then there is the matter of knowledge and computation. Of course adherents to the rational choice theory of economics are not so naive as to believe that decision-makers will have all the knowledge and will carry out successfully all the calculations called for by the theory. They believe that actors choose "as if" they had made such calculations. Or they believe that small departures of actual behavior from that predicted by the rational model will be averaged out in aggregating the behavior of many people, and will have no serious effect upon decisions at the level of the polity.

Unfortunately, no empirical evidence is available to show that this "as if" assumption is valid. At the micro level of individual decision makers there is a great mountain of evidence that actual behavior is quite unlike the neoclassical rational model. The very foundation stone -- the assumption of consistency of the utility function -- has been subjected to extremely careful study in the laboratory (Kahneman, Slovic and Tversky, 1982) and has been found to be almost always false.

When the detailed behavior of businessmen and consumers has been examined in a wide range of situations, it has been found to be very different from the behavior postulated in the model (for one example, Cyert and March, 1963). In particular, a large part of the effort in practical decision making goes into designing and constructing likely alternatives. No computation appears to be made to decide exactly when the search for alternatives should stop and a choice be made. Moreover, even in the case where it is reasonable to characterize the goal as maximizing profit, the evidence usually reveals that the decision makers are trying to achieve a satisfactory profit, satisfaction being measured by past experience and aspirations.

At the macroscopic level, certain kinds of predictions can often be made. If there is a freeze in the Florida citrus groves, the price of oranges will rise. These kinds of relations between supply and demand can be used successfully, and often are, in prediction. But it has been shown (e.g., Becker, 1962) that such predictions can be made without using the stifling assumptions of neoclassical theory. For example, if prices rise, consumers will buy less by reason of limits on their total budgets, quite apart from any utility calculations. There is little or no evidence that accepting the apparatus of neoclassical rationality, instead of simply assuming that people have reasons for what they do, provides additional predictive power.
The Goals that Guide Choice

I said earlier that one qualification must be introduced into my account: there is one vital respect in which analyses of neoclassical economists and public choice theorists depart from pure utility-maximization premises. A scrutiny of the way in which public choice theory is actually applied to political phenomena shows that strong auxiliary assumptions, going well beyond the postulate of maximization of expected utility, are always required. These auxiliary assumptions often amount to specifying the content of the utility function. In neoclassical analysis of business enterprises, the most common auxiliary assumption is that firms maximize profit (short-term profit or long-term expected profit). In public choice analyses of political phenomena the corresponding assumption is that governments maximize political power; elected officials and bureaucrats, their chances of survival in office; and voters, their economic welfare. The question of time horizon, and how choices would be affected by it, is seldom discussed in the public choice literature.

Sometimes, the auxiliary assumptions about motives and goals are supported by another layer of inference that derives them from quasi-Darwinian survival goals. Thus, it is argued that those candidates for office who do not maximize their election chances will be replaced by others who do.

Conflicting Goals. Characteristic of all of these assumed goals is that they are "selfish," based on the axiom that "what's in it for me" is the sovereign driving force of human action. Most of us would agree that holding and exercising power, remaining in office, and improving one's economic situation are prevalent human goals. This assumption is generally consonant with our everyday experience, with folk (and media) wisdom, with our reading of history, and with many findings of empirical research in political science.

However, even if selfish goals are important to decision makers in politics, knowledge of that fact does not permit easy prediction of political behavior. Because of bounded rationality, in any given situation many different choices might be consistent with such goals. A Congressman may vote for a raise in congressional salaries even if he knows that his vote will be an issue in the next campaign. Another Congressman will vote against the same raise because he knows it will be an issue.
If we know how close each Congressman believes the next contest will be in his or her district, perhaps that will help predict the vote. But what is the basis of the Congressman's belief, and how valid is it? Applying the principle that people have reasons for what they do, at best we might predict that there will be a correlation between the votes on this bill by different congressmen and the closeness of the previous contests in their respective districts. That prediction, of course, does not assume that Congressmen are maximizing anything. It derives its force from the auxiliary specific assumptions that wealth and power are important, not from the general assumption that utility is maximized.

**Salience of Goals.** In the same way, we may sometimes predict the behavior of voters from some knowledge of their values, but we seldom know any more than that certain goals are very important to certain voters (or to the "average" voter); we certainly do not know what their utility functions are, or whether they are maximizing utility. When we say that a certain Representative's reelection is endangered by his or her stand on the abortion question, we mean simply that many voters in the district have strong views on abortion that conflict with the Representative's. In such situations, swaddling our prediction in the clothes of maximization simply gives us an illusion of prescience.

It is well known that in particular political campaigns certain issues are salient, while others are outside the focus of public attention. In fact, political wisdom suggests that a main purpose of campaigning is not to change people's values or beliefs but to turn their attention to those issues on which majority views favor the campaigner. This wisdom entered the political science literature at least as early as the Lazarsfeld, Berelson and Gaudet (1948) voting studies, if not before, without the help of maximization assumptions. (Advertising wisdom has reached a similar conclusion: that the main function of ads is to direct attention, hence to change the structure of the goals that influence choices.)

**Measures of Goal Attainment.** Predicting political behavior from the assumption that people wish to advance their economic welfare encounters additional difficulties (Simon, 1985). How do voters judge how well off they are and which candidate or party will improve their economic situation? Voters have direct information of their own economic situation, and how it compares with their situations a year ago or ten years ago. But how do they diagnose the reasons for a satisfactory or unsatisfactory situation, and how do they judge what course of action will improve conditions?
Various economic indicators are available and are widely known: the level of unemployment and the rate of inflation are two of them. The national debt, the Federal deficit, and the balance of foreign trade are others. How do voters weigh these different measures; or do they weigh them? In the period from 1930 to 1960, there was much public concern about unemployment, little about inflation. Since 1970, there has been much concern about inflation and, until quite recently, little about unemployment. Now the emphasis appears to have shifted sharply back again. Does this mean that utility functions have changed, or theories of economics, or means of diagnosis? (Simon, 1985 discusses examples of these ambiguities in the recent political science literature.)

**Focus of Attention.** Such shifts are best explained as shifts in attention (Simon, 1987). One set of symptoms replaces another in the public awareness, providing a new basis for voting decisions. To predict the shift, it is important to know what level of employment or inflation will attract attention; it is of no use to introduce a utility function to be maximized. Again, the auxiliary assumptions are doing all of the work, and these are empirical assumptions that cannot be deduced from the theory.

The same role of attention can be seen in the rapid shift in priorities toward environmental concerns in the late 1960s, the shift back to competing energy concerns after the Oil Shocks, then the return to environmental concerns when gasoline prices dropped sharply and such phenomena as El Niño and global warming became prominent in the media. Over the thirty years during which these shifts in priority have occurred, there has been no comparable change in scientific knowledge about the nature and magnitude of the problems. Virtually all of the problems of environmental pollution and energy exhaustion that trouble us today were in the literature in 1965.

**Group Identifications.** But there is one further gap between our knowledge of voters' goals (especially such broad goals as economic welfare) and political behavior. People often identify their economic (or other) welfare with the welfare of one or more of the groups to which they belong: for example, the nation, and most obviously today, ethnic groups defined in terms of race, language, religion, and shared history. To these we must add (at least) social and economic class and gender.

That voters are economically motivated (if that is the claim) helps little to predict how they will vote or behave until we know empirically with what economic group they associate themselves. Are they Roman Catholics, or South Slavs, or
Croatians, or proletarians, or women, or farmers who grow Merino goats? The answer to that question will determine how the sides will line up and who will join them. And the answer is among the auxiliary assumptions that have to be made, not a part of the general theory of rationality. Again, to the extent that we can predict, the auxiliary assumptions are doing all of the work of prediction, and they are empirical assumptions.

This same process of identification, of measuring one's welfare (or utility, if you prefer) by the welfare of a group, constitutes a major aspect of behavior in large organizations. No theory of bureaucracy can make predictions without taking account of the group identifications of administrators and employees. Are they loyal to the Department, to the Division, to the Section? Is the Department of the Interior a focus of identification or the National Park Service? The answers to these questions are the major predictors of behavior, and the answers cannot be derived from postulates of rationality, weak or strong, without auxiliary empirical assumptions (Kaufman, 1960; Simon, 1991).

**Auxiliary Assumptions are Empirical Assumptions.** At every step in our analysis we encounter the need to make auxiliary assumptions about the content (and strength) of peoples' goals, and we find that our power to explain and predict depends almost wholly on these assumptions (Simon, 1986). Moreover, there is no way to deduce the correct assumptions from the postulates of rationality: alternative goals, alternative loyalties are equally rational or irrational. Hence a major task of political science is to establish the facts: to understand, under what particular circumstances and at what particular points of time, people will hold and act upon particular goals and values.

A large part of the research effort is necessarily devoted to observation and experiment; simply to finding out how things are in fact. A part of the effort can be devoted to discovering the processes that govern the formation and alteration of values, the processes that determine how the focus of public attention will shift from time to time, the processes that determine group and organizational loyalties. It is at this level, and in this intimate relation with the empirical auxiliary assumptions, that we can expect to find theory in political science. Of course research of these kinds has constituted for a long time (especially since the "behavioralist revolution") the major occupation of political scientists. The advent of public choice theory and neoclassical economics has not made it obsolete. It is and will remain the main stream.
Those of us who live outside the "hard" sciences (e.g., physics) often have a romantic view of formal theory as Queen of those sciences. Newton's laws of motion say it all! (Or quantum theory, or relativity, or the unified field theory that Einstein dreamed of, but which does not yet exist.) Nothing could be further from the truth. The application of physical theory to concrete situations (especially, but not exclusively, the complex situations with which engineers deal) depends crucially upon the auxiliary assumptions that are based, in turn, upon direct physical observation and measurement. Plank discovered the quantum when a theory he had previously proposed (on the basis of earlier data) to explain blackbody radiation was blown out of the water by observations taken with new instruments for measuring infra-red radiation (Langley, et al., 1988).

In physics, there are twenty experimenters for every theorist: peasants, perhaps, but essential peasants, without whom the theorists would have no food for thought. The parameters that define the strength of a particular kind of steel are as important to the strength of a bridge as the laws of mechanics. The controversy about cold fusion was settled, not by debates about theory, but by turning back to the laboratory. And when we move a bit away from physics, into such realms as molecular biology, theory plays a muted role in the face of validated facts. In political science, we might do well to inquire into the ways of biologists before we follow the ways in which we (mistakenly) imagine that physicists do their work.
Selfishness and Altruism.

The main thrust of my argument does not depend upon the content of human goals, but on the general principle that this content must be known in order to predict behavior. However, it has sometimes been claimed that the content of human values and goals can be deduced from Darwinian theory. In particular, it has been claimed that Darwinian selection assures priority for selfish motives, and that we can therefore assume such motives for purposes of prediction, without the painful necessity of making empirical observations to see what goals are actually operative.

There is, indeed, a sense in which the principle of survival of the fittest is almost a tautology. If we define as fittest those who have the most progeny, and if heritable traits have something to do with the probability of having progeny, then over time (and in a stable environment) those genes that are fittest will gradually replace those that are less fit. It is less clear what predictions about the real world follow from this tautology.

Altruism and Fitness. The current neo-Darwinian argument, a quite rigorous one that can often be given precise mathematical form, leaves only a little room for altruism: behavior that contributes to the fitness of others at the sacrifice of self's fitness. Altruism towards close kin can be explained (but "close" means children and siblings, hardly anyone more distant than that). Under certain systems of intermarriage among tribes or villages (so-called "structured demes"), altruists may be fitter than egoists. But few other exceptions, and these very narrow, have been found, until recently, to the Darwinian prediction that altruism cannot survive.

Of course the limitation does not hold for reciprocal altruism: acts of altruism premised on the assumption that they will be reciprocated in the long run. Much social behavior involves reciprocal altruism, the bestowal of gifts and favors in most societies being an example. And yet it has been argued with great cogency that reciprocal altruism is not sufficient to explain a great deal of very common human behavior. As examples of particular importance, I would cite the frequent willingness of people to sacrifice personal welfare (and even their lives) in support of the groups with which they identify. One does not have to look very far around the world today to find striking instances of such behavior.
Meanings of Selfishness. The prediction of selfishness from Darwinian assumptions runs into another serious difficulty in our world today, because selfishness, in everyday speech, means something different from selfishness in the Darwinian theory. In common speech, we use the word "selfish" to denote behavior aimed at material gain, particularly of economic wealth and secondarily of power. (Of course, power can also be explained as a means for acquiring wealth, or vice versa, so we need not trouble to distinguish too carefully between them.)

Often, the term selfish is broadened to include desiderata like glory or attention. And we can even view the martyr as selfishly seeking martyrdom. Upon reaching this point, the term "selfishness" loses any usefulness it may have had when defined more narrowly. We might as well supplant it with "utility": People seek whatever increases their utility, i.e., whatever it is they seek. So let us return for a moment to special wants: wealth, and/or power.

It is easy to see how in times past wealth and power were highly correlated with fitness. Since genes do not distinguish between legitimacy and bastardy, the king, the duke and the brigand could have many progeny. Matters are different nowadays. There is little reason to believe today that, at least in large parts of the world, there is any relation between wealth or power, on the one hand, and number of progeny, on the other. It is another question how long such a lack of correlation will last. A Darwinian might even argue that if the particular form of selfishness that desires wealth or power makes no contribution to fitness, it will in time cease to be an important motive in the society.

This line of inquiry, while fascinating, would, if pursued further, take us too far afield. Let me leave it after drawing one conclusion: the Darwinian argument against altruism, even if valid, cannot be transformed into an argument that people desire wealth and power above most other things. If they do, it must be due to something other than the current forces of selection upon their genes. Therefore, the assumptions that people do desire wealth and power above all else, essential auxiliary assumptions in most contemporary applications of public choice theory to politics, derive no special credence from Darwinism. They are just empirical assumptions, to be tested like all other assumptions against empirical evidence.
The Case for Altruism. In fact, there is a serious flaw in the usual neo-Darwinian
demonstration that altruism (in its Darwinian meaning) is inconsistent with fitness, except toward close kin. With the help of some simple algebra, we can demonstrate a quite robust and general mechanism for altruism that is consistent with Darwinism (Simon, 1990).

We consider two types of individuals, (S) and (A). The fitness (F) of both types of individuals has a general component, G. In addition, the individuals of type A have a component that we will call docility (D): by virtue of both temperament and intellectual ability, they learn from their social environment more effectively than do the individuals of type S. This social environment is benign, in the sense that most of the information, instruction, and guidance they receive from it (from parents, schools, peers, books) enhances their fitness. Their total fitness is then:

$$F_A = (G + D) > G = F_S,$$

so that $$F_A > F_S.$$  

Thus, individuals of type A will have greater fitness than those of type S. The society can now impose a "tax" upon these individuals. It can include in its influence on them some guidance that will lead them to behave altruistically towards others, even at the expense of their own fitness. It may inculcate in them, for example, tendencies to aid others even when there is danger for themselves. The amount by which this tax reduces their fitness is T.

The altruistic behavior, whose cost is the "tax," bestows a benefit that increases the fitness of all members of society, those of type S and A alike, by an amount, B. Now the net fitness of the altruists will be $$F_A = G + D - T + B,$$ which will still be larger than $$F_S = G + B$$ as long as $$T < D,$$ that is, as long as the tax is not so large as to negate the contribution to fitness of docility. Under these assumptions, the number of individuals of type A will continue to grow relative to the number of type S.

Why do some individuals of type A not learn to avoid the altruistic behaviors and accept only that instruction which contributes to their fitness? Here we must evoke another very important property of human beings, the very property, in fact, that makes it impossible for them to optimize utility in any meaningful sense. Their rationality, that is to say, our rationality, is exceedingly bounded in relation to the complexity of the environment in which we live. Most of of our knowledge and skill is
not acquired by direct evaluation of our experiences with the environment (by touching hot stoves), but by instruction under the guidance of others.

We reduce our cholesterol intake because we have been told, by "reliable" (i.e., socially legitimated) medical authorities, that it is bad for our health. We "know" that the Sun, not the Earth, is at the center of the solar system because our textbooks say so and argue the case, not because we observed it ourselves. Even if we know the reasons, we are unlikely to have invented them ourselves after rejecting schooling. In few cases have we ourselves evaluated the evidence thoroughly, and in most cases we are not even capable of doing so.

All over our globe, we find millions and billions of human beings holding beliefs about the origins of the cosmos, the nature of matter, the existence (or not) of life after death, morality, and a plethora of other matters largely because these are the beliefs of their reference groups in the society around them. It is not a question of whether the beliefs are true or false (of course, they cannot all be literally true), but of how one accounts empirically for their widespread acceptance.

To refuse to accept beliefs or advice until they can be verified personally -- to be from Missouri -- is to reject learning most of the things that society can provide to enhance fitness. The child who persistently rejects social influence, or who is unable to learn the skills society offers to teach it is severely disadvantaged in the competition of life. Docility, even if taxed, pays.

Finally, members of a society that taxes docility to induce altruism will be fitter than members of a society that does not, as long as B > T, that is, as long as the benefits to fitness obtained from the altruism of others are greater than the cost to fitness of one's own altruistic behavior. A society with altruism will grow more rapidly than one without altruism. On Darwinian grounds, then, we would expect to find altruism in human societies (but not to the exclusion of "selfish" behaviors). We might even find that in societies where wealth and power do not lead directly to greater genetic fitness, the drive to acquire wealth and power are inculcated through social influence as part of the mechanism of docility (and altruism?). In this way a society can be supplied with resources and leadership. I would not like to endorse a too rosily optimistic view of these two particular mechanisms, but merely to suggest that the processes of society interact in ways that are not always evident to the naked eye.
In sum, the argument for the possibility (better, the probability) of altruism is this: docile individuals, individuals who are intelligent, and who are dispositionally inclined to respond positively to the influence of their social environment, will attain higher levels of fitness, on average, than others who are less docile. A society can use this fact to influence its members toward altruistic behaviors that are of net advantage to all, although immediately disadvantageous to the altruist. By reason of bounded rationality, the influence cannot be avoided without losing the benefits of docility. The "meek," more accurately, the docile (as I have defined them), will inherit the earth. But before we conclude that this mechanism would lead to the best of all possible worlds, we need to inquire into the behavior that altruism induces.

**Altruism and Behavior**

In interpreting the social implications of altruism in human behavior, including political behavior, we must pay careful attention to the way in which altruism is defined. Altruism is behavior that sacrifices individual fitness for the fitness of others. What behavior fits this definition depends on whom we include among the "others." Social injunctions to behave altruistically commonly urge behavior that benefits "us" rather than behavior that only benefits the individual. The most important form that human altruism takes is loyalty to groups, with the consequent tendency to evaluate actions in terms of their consequences for the group and without consideration of their consequences for people outside the group.

The "we" that gains the benefits of altruism may be one's immediate family, extended family (as, for example, in the Chinese culture), town or state, nation, religious sect, ethnic or language group, political party, industry, the organization in which one earns one's livelihood. The possibilities are legion, although those I have mentioned appear historically to have been the most important.

Actions that benefit us frequently harm them. What is good for the Guelphs is often bad for the Ghibellines. Although not all human activity is zero-sum in character, one large class of actions that benefit a group are actions that gain control of resources at the expense of another group. The willingness of a Nazi soldier to die outside Stalingrad qualifies under the definition of altruism as fully as does the willingness of someone to
jump in the water to save a drowning person. Individual altruism is frequently recruited to the service of group selfishness.

The rational altruist finds and carries out actions that are advantageous to the group with which he or she identifies. To predict what actions this rationality will produce, we must know with what group or groups the altruist identifies. Hence understanding economic and political phenomena requires a veridical theory that can predict the formation and strengths of group loyalties. Among the groups that attract loyalties are organizations, including business corporations, military units, political parties, government agencies, universities.

In economic literature, organizations have been assumed to be held together mainly by monetary rewards -- to owners, managers and other employees, vendors, customers and bondholders. But it is quite unclear that monetary rewards can account for the observed effectiveness of large organizations. There are great difficulties in measuring the results of organizational efforts, both in nonprofit organizations and in components of business organizations that are not self contained. There are even greater difficulties in measuring the marginal contributions of managers and employees to organizational output. Under these circumstances, it is hard to make a convincing case that monetary rewards can be targeted with sufficient accuracy to support the large corporation as the dominant organizational form in industrialized societies. Introducing organizational identification as another source of motivation in organizations provides a new and powerful mechanism to explain how large groups of people can be brought to collaborate in advancing organizational goals.

In politics, political parties and single-issue advocacy groups provide alternative kinds of organizations to which people can attach their loyalties, with quite different consequences in the two cases for focus of attention and energy. Each kind of attachment has its reasons, but there is no indication that such loyalties maximize or optimize anything. Instead, they provide further evidence of the boundedness of human rationality, and the need people experience to simplify and focus in order to be able to reach decisions. Only extensive empirical research will enable us to understand the circumstances under which people form attachments to one or the other of these forms of political organization.
The same thing can be said about the larger social groups, especially ethnic and religious groups, that are playing such an enormous role in today's national and international political scene. The fact that this great resurgence of ethnicity and religious fundamentalism was not foreseen by most students of politics (any more than the international student movements of the 1960s were foreseen) illustrates how much more we must learn before we can build an adequate theory of such developments.

Conclusion

My thesis is that human behavior is generally rational, and that it cannot be understood without finding the connections between its actions and its goals. However, finding these connections is not a simple matter of reasoning from an abstract model of optimizing rationality. There is an enormous variety of human values and goals; they are present in different people and in the same people at different times with greatly varying strengths. Selfish goals are no more "rational" than altruistic ones, or vice versa. Rationality can only go to work after final goals are specified; it does not determine them. Moreover, the connections between goals and behavior are mediated by factual knowledge and beliefs about the relations of means and ends.

Only a few of the values that people hold can influence their decisions at any given time. Much depends upon their focus of attention, upon what values and knowledge are evoked while their decisions are being reached. Focus of attention is a variable of particular importance for political phenomena because there is considerable synchrony in the direction of attention of different members of a society, the more so, the stronger the communication network among them and the speed with which messages are transmitted.

To understand human behavior, especially in a social context, one must have a large store of knowledge about the minds of the actors, what they know and believe, what they want, what they are currently attending to, how they structure the world around them. Included in their beliefs are convictions about the legitimacy and reliability of various channels of information that seek to inform or influence them. In making predictions of human behavior, the auxiliary assumptions (based, it is hoped, on sound empirical studies) that postulate the states of mind of the actors are far more important than consequences deduced from a formal definition of rationality.
There is a proud tradition in political science, going back to Aristotle's collection of the constitutions of the Greek cities, of empirical study of human behavior. It was continued by the historians (e.g., Thucydides, Gibbon), the more worldly of the philosophers (e.g., Machiavelli, Montesquieu), and traveler-commentators (de Tocqueville, Bryce). It was picked up again in this century by pioneers like Pendleton Herring and Paul Lazarsfeld, and vastly extended in recent decades with valuable additions of modern quantitative methodology. This is the bedrock of fact on which rests our ability to predict behavior, and such generalizations and theories as we use to organize our knowledge. All of the students of politics who built these foundations used freely the assumption that people are rational, that they have reasons for what they do. None of them used the assumption of utility maximization.

Optimizing assumptions can rarely be applied literally (the information is simply not there to permit sophisticated calculation, nor is the calculation usually computationally feasible). The human condition requires both the people whose behavior is being studied and the person studying them to satisfice. Fortunately, the unworkable tools of optimization are not required for making effective predictions of behavior -- provided that the auxiliary assumptions are empirically sound. This basic principle provides an important guideline for the direction and nature of scientific research in politics.

Nor can we hope that the picture of human motivation can be much simplified by placing it within a Darwinian framework. Virtually none of the vast gamut of goals and values that have been attributed to human beings can be ruled out by the mechanisms of natural selection. In particular, altruistic behavior, especially in the form of identification with groups, plays a major role in human social systems. On the other hand, the relation of such pervasive motives as acquisition of wealth and power to genetic forces is quite unknown.

References


