

CHAPTER TWO

THE NATURE OF RATIONAL CHOICE THEORY

Before examining how rational choice theories are applied, it is necessary to say something about what rational choice theories are and how they are distinctive. The rational choice approach to the study of politics, as Jackman (1993, 281), Grofman (1993b), and others have noted, is often caricatured by being reduced to one or two of its characteristic assumptions and presented as a monolithic theory that all practitioners are presumed to accept. A more accurate depiction is that most practitioners agree on some, but not all, features of the definition of rational choice. As a result, there is no single rational choice theory or unambiguous standard for assigning the label “rational choice” to a theory. Our purpose in this chapter is to describe the rational choice approach to the study of politics in a way that takes these complexities into account. It is not part of our objective here to adjudicate disputes among rational choice theorists as to how rationality is best understood. Rather, our aim is to supply the reader with an understanding of the principal ways that rationality has been characterized in the rational choice tradition and a sense of what is at stake in the different characterizations from the standpoint of empirical testing.

We begin with an account of the less controversial assumptions that are generally shared by rational choice theorists. These assumptions concern utility maximization, the structure of preferences, decision making under conditions of uncertainty, and, more broadly, the centrality of individuals in the explanation of collective outcomes. Next we turn to the issues on which there is disagreement, notably the nature and content of human goals and the amount of information rational agents are presumed to possess and use. We conclude with some remarks about the different views of explanation that seem to guide rational choice theorizing, noting the implications of these views for empirical research. This overview sets the stage for our discussion of rational choice models of politics in subsequent chapters.

GENERALLY ACCEPTED ASSUMPTIONS

The first assumption about which there is widespread agreement among rational choice theorists is that rational action involves *utility maximization*. To say that a person maximizes utility is to say that when confronted with an array of options, she picks the one she believes best serves her objectives. As Olson (1965, 65) puts it, an individual's actions are rational when her objectives are "pursued by means that are efficient and effective for achieving these objectives," given her beliefs.

Rational behavior is typically identified with "maximization of some sort," as Arrow (1951, 3) puts it, even if there is scant agreement among rational choice theorists on just what is maximized and how, if at all, this utility should be measured.¹ Individuals are assumed to be uninterested in others' fortunes (or, indeed, their own) except insofar as those fortunes impinge on their particular maximizing strategies. Rational choice theorists need not dissent from the proposition that the welfare of others might affect an individual's conception of his or her preferences, as when an egalitarian wants the income of the poor to increase or an elitist wants the income of those around him to be comparatively lower. The maximizing assumption requires only that *some* schedule of preferences is maximized; it "does not specify any particular goal" (Riker 1990, 173).

Rational choice theorists agree, second, that certain *consistency* requirements must be part of the definition of rationality. These requirements are seen as essential to a science of rational action. "Unless economic units act in conformity with some rational pattern no general theory about what would follow from certain premises would be possible" (Rothschild 1946, 50). Following the lead of microeconomists, rational choice theorists of politics have sought to keep their consistency requirements minimal, but two appear to be widely accepted. First, it must be possible for all of an agent's available options to be rank-ordered. This is sometimes called the assumption of connectedness. It requires that an agent regard any two available outcomes as either unequal (that is, she prefers one to the other) or equal (she is indifferent). Connectedness does not require that numerical values attach to preferences for different options, that comparisons can be made across individuals, or that arithmetic functions can be performed on an individual's preference ranking. But it does assume the possibility of rank-ordered preferences over all available outcomes for every individual.

Rational choice theorists also assume that preference orderings are transitive. If A is preferred to B, and B is preferred to C, then this consistency rule requires that

1. There can be forms of strategic behavior that are not maximizing, as is illustrated in Herbert Simon's contention that people do not seek the best alternative in any feasible set (1955, 1956). Instead, he argued, they "satisfice"; they limit themselves to what seems to be "good enough" or satisfactory. See also Eckstein 1991.

A be preferred to C. Transitivity assumes nothing about the intensity of preferences or the amount by which the different outcomes are valued in comparison with one another. It does not even require that these amounts be known to the individual, much less that they be measurable by a third party. Transitivity requires only minimal consistency within preference orderings. When the connectedness and transitivity requirements are both met, we have what Arrow (1951, 13) described as a weak ordering of preferences. This is generally assumed by rational choice theorists to be axiomatic of rationality.²

Third, rational choice theorists routinely assume that each individual maximizes the *expected value* of his own payoff, measured on some utility scale (Luce and Raiffa 1957, 50). The focus on expected rather than actual utility is required by the fact that decision making often takes place under conditions of uncertainty. A farmer who chooses to plant one crop rather than another has to make assumptions about future weather, which he cannot forecast with certainty. It is usually assumed by rational choice theorists that numerical probabilities can be attached to different eventualities—for example, the value of an outcome for an agent weighted by the probability of achieving it (Elster 1986b; Harsanyi 1986).

If, for example, a person were neutral with respect to risk (though nothing in rational choice theory requires that he or she must be), the idea of expected utility maximization implies that the individual would be indifferent between having \$5 or having a 50 percent chance of having \$10. The assumption of expected utility maximization is usually justified by reference to von Neumann and Morgenstern's theorem (1947). Using weak assumptions about rational behavior, they demonstrated that for a decision maker whose choices among outcomes and gambles follow certain assumptions of consistency, there is a way to assign utility numbers to the various outcomes so that he or she would always select an option that maximizes expected utility (Myerson 1991, 2). "Expected utility," observes Fishburn (1988, 1), "has served for more than a generation as the preeminent model of rational preferences in decision making under conditions of risk."

A fourth assumption that commands widespread agreement among rational choice theorists is that the relevant maximizing agents are *individuals*. Unlike evolutionary biologists, for example, who have debated for decades over whether the basic unit of survival is the species, the group, the individual, the gene, or some other entity (see Gould 1992), rational choice theorists of politics generally agree that it is by reference to the maximizing actions of individual persons that collective outcomes must be explained. Buchanan and Tullock (1962, 13) declare that collective action is nothing

2. As Arrow (1951, 12–13) puts it, if we define R as a single relation "preferred or indifferent to," then the idea of a weak ordering requires that we accept the following two axioms: *For all x and y, either xRy or yRx, and For all x, y, and z, xRy and yRz imply xRz.*

more than “the action of individuals when they choose to accomplish purposes collectively rather than individually,” so that for them the state “is seen as nothing more than the set of processes, the machine, which allows such collective action to take place.” Riker and Ordeshook (1973, 78–79) put it thus: “Society, not being human, cannot have preferences in any proper sense of ‘have,’ nor indeed can it order the preferences that it does not have.” Consequently, they argue, we are bound to assume the existence of individual preference orderings and individual choices among alternatives as our basic theoretical building blocks in the study of politics. Likewise, Elster (1986b, 3) contends that because the mechanisms through which rational choice explanations operate are the preferences and beliefs of individuals, rational choice explanations cannot be predicated upon entities other than individuals. “A family may, after some discussion, decide on a way of spending its income,” he notes, “but the decision is not based on ‘its’ goals and ‘its’ beliefs, since there are no such things.” Riker (1990, 171) goes so far as to suggest that consistent generalization in the social sciences is possible only when “the central propositions are about rational decisions by individuals.”³

The task for rational choice theorists, then, is to explain collective outcomes by reference to the maximizing actions of individuals. Indeed, Olson’s original thesis about the logic of collective action stemmed from his observation that the then orthodox group theory of politics associated with the writings of Bentley, Truman, Latham, and others offered no account of why rational individuals would coalesce into groups to pursue their objectives. The group theorists had failed, in Olson’s view, to see that rational individuals will “*not* voluntarily make any sacrifices to help their group attain its political (public or collective) objectives” (Olson 1965, 126). Even when the individual in question greatly values the good that a group provides, there will always be an incentive for that individual to free-ride, to avoid participating,

3. It is sometimes said that rational choice theory is not necessarily individualist in its assumptions, that political parties are assumed, for example, to be maximizing agents in theories of electoral competition, and that nation-states are treated as the basic maximizing units by game theorists of international relations. There are also versions of rational choice Marxism in which classes are regarded as the basic maximizing units (compare Przeworski 1991). But in such circumstances parties, nation-states, or classes are assumed to be unitary actors of whom preferences, goals, and strategies may meaningfully be predicated. Even in such applications, therefore, rational choice remains individualist in its basic ontology. For this reason, rational choice theories are sometimes criticized as misleading just because they ignore pertinent complexities that are internal to their primitives. For criticism of the Downsian model of party competition along these lines see Budge and Farlie 1977, 115. For criticism of the assumption that nation-states are monolithic “rational” individuals in many rational choice models of international relations see Maoz 1990, Bueno de Mesquita and Lalman 1992, and Russett 1994. For criticism of Przeworski’s Marxism on analogous grounds see Swenson 1991a, 1991b and Shapiro 1993.

secure in the knowledge that this course of action will likely have no effect on whether the good is provided. The rational choice literature has generally followed Olson's individualist account; indeed, many of its central research questions would not arise were it not assumed that individuals are the basic maximizing units.

Finally, rational choice theorists generally assume that their models apply equally to all persons under study—that decisions, rules, and tastes are “stable over time and similar among people” (Stigler and Becker 1977, 76). Although nothing in the core assumptions of rational choice theory requires that the content or even strategic character of agents' preferences necessarily be identical for all agents, in practice to allow interpersonal variation may generate insuperable problems of tractability (Strom 1990, 126). “If utility functions and perceptions differ widely,” Goetze and Galderisi note, “and if people have very different combinations of altruistic and self-interested motives then the construction of adequate explanatory models might be frustrated. Patterns of universal behavior may not [in that case] be discoverable” (1989, 38). To avoid this result, rational choice theorists generally assume away such differences, at least when constructing empirical applications. This *homogeneity* assumption is usually justified on grounds of theoretical parsimony. If an outcome can be accounted for only by assuming that in deciding how to vote some voters make sophisticated calculations about the likely votes of others while other voters do not, for example, “we are forced to say that *two* different models—the models of sincere and sophisticated voting—must be used simultaneously to explain what we observe, a decision that is, to say the least, scientifically unparsimonious and one which would call for an explanation as to why the behavior of some voters must be explained with one model and the behavior of other voters with another” (Enelow 1981, 1077–78; see, however, Denzau, Riker, and Shepsle 1985).

In sum, rational choice theorists generally agree on an instrumental conception of individual rationality, by reference to which people are thought to maximize their expected utilities in formally predictable ways. In empirical applications, the further assumption is generally shared that rationality is homogeneous across the individuals under study.

COMPETING VIEWS OF RATIONAL CHOICE

Yet there is more to any theory of rational behavior than the features just described. As we delve more deeply into the meaning of rationality, we find several areas of disagreement among rational choice theorists.

The first such area of dissensus concerns the robustness of assumptions about human goals. In what Ferejohn (1991, 282) dubs the “thin-rational” account, agents are assumed to be rational only in the sense that “they efficiently employ the means available to pursue their ends.” In “thick-rational” accounts, by contrast, “the analyst

posits not only rationality but some additional description of agent preferences and beliefs." Adherents to this view "generally assume that agents in a wide variety of situations value the same sorts of things: for example, wealth, income, power, or the perquisites of office." Utilitarianism and classical economics rested on thick-rational accounts for most of their histories, as did the embryonic rational choice arguments of Hobbes (who assumed that individuals maximize power) and Bentham (who assumed that they maximize pleasure). Neoclassical economics is, by contrast, thin-rational in its assumptions about consumers: they are presumed to maximize their utilities, but the content of those utilities is not specified. On the other hand, the neoclassical theory of the firm is thick-rational in Ferejohn's sense, since all firms are assumed to be maximizers of profits.

Some rational choice theorists of politics claim to assume only thin rationality. Riker (1990, 173) argues, for example, that so long as the consistency requirements of an Arrovian weak ordering are met, any choice—including suicide—can be interpreted as rational. He concedes that this makes the sense in which individuals are self-interested tautological, arguing that it is the formal structure of preferences, not their content, that does rational choice theory's explanatory work. Other rational choice theorists embrace models that assume more robust conceptions of self-interest that are incompatible with altruistic and consciously self-defeating behavior (compare Klosko 1987).

Riker is correct that some rational choice literatures in political science, notably the literatures on cycling and instability, depend almost entirely on thin rationality. As a result, these literatures keep controversial assumptions about human goals and motivation to a minimum. It will become plain, however, that what is gained by avoiding controversial assumptions about human nature can come at a considerable cost from the standpoint of measurement and empirical testing of rational choice hypotheses. If the content of preferences is not specified, it becomes enormously difficult to determine, for example, whether a changed outcome in the majority vote of a committee reflects the presence of stable but cyclical preferences among the voting members, changes in their preferences over time, or some other phenomenon.

In addition, it is sometimes unclear whether an account is thin-rational or thick-rational. Even if nothing is specified about the content of preferences, the researcher may make certain assumptions about the stability of preference orderings that are more robust than what mere thin rationality requires. For instance, an otherwise thin account may assume that people do not change their preferences toward the same set of available outcomes over time, or that the actors' tastes are not directly influenced by the choices offered them or by the behavior of others. In principle, theories range from thick to thin, but empirical applications seldom approximate the latter ideal type.

Much of the rational choice literature rests on unambiguously thick-rational assumptions. For instance, the literature on party competition typically assumes that parties try to maximize votes and, in so doing, maximize power; the rent-seeking literature assumes that interest groups try to maximize a variety of goals, from profits to environmental conservation; much of the law-and-economics literature assumes that judicial decisions maximize the production of wealth; and the literature on legislators and bureaucrats assumes that they try in various ways to maximize career advancement. These assumptions may be more controversial than thin-rational accounts, but *prima facie* they should be expected to present fewer difficulties from the standpoint of empirical testing, because there is less room for ambiguity in the definition and measurement of what allegedly is being maximized. However, we show in subsequent chapters that thick-rational accounts have often proved to be just as slippery as thin-rational accounts when tested empirically.

A second area of disagreement among rational choice theorists concerns the amount of relevant information that agents can normally be presumed to possess and act on. Conventional neoclassical models of market behavior assume both perfect information and consumers' ability to understand and use that information. These assumptions are unrealistic, all the more so in politics, where voters are reputed to be ill-informed about the leaders and policies among which they are presumed to choose. As a result, many rational choice theorists of politics have moved away from the assumption of perfect information, though they retain the assumption that actors make the most of the imperfect information they possess (see McKelvey and Ordeshook 1987).

Imperfect information arguably reflects the fact that acquiring information is often time-consuming and costly. Taking the view that information-gathering resembles other economic investments, Downs (1957, 215) reasons that any seeker of information "continues to invest resources in procuring data until the marginal return from information equals its marginal cost." As Elster (1986, 19–20) notes, however, such logic leads to a conundrum: the agent has to assess the value of information that she does not yet have in order to determine whether it is worth taking the trouble to gather that information. A variant of this conundrum arises when rational choice theorists debate the rationality of "myopic" behavior, in which actors pursue immediate rewards without regard for the possibility that this strategy may lead to undesired outcomes (Krehbiel and Rivers 1990; Austen-Smith 1991). If strategic foresight and planning are assumed to be costless, myopic action cannot be characterized as rational. But if one allows for cognitive costs (or distractions arising from the pursuit of objectives in other aspects of life), myopic strategies may be construed as rational, given an actor's shortsighted beliefs. Rational theories, in sum, encompass a range of assumptions about the knowledge actors have concerning the strategic choices before them.

RATIONAL CHOICE CONCEPTIONS OF EXPLANATION

Rational choice theorists see themselves as engaged in a common explanatory enterprise. As with the definition of rationality, however, they do not all characterize this enterprise in quite the same way. Yet most accounts of it share two basic features. One relates to the type of causal arguments that are considered; the other concerns the universalism to which rational choice theorists aspire.

Intentions as Causes Because rational choice theorists assume that social outcomes are the by-products of choices made by individuals, rational choice explanations are typically formulated by reference to individual intentions. According to Satz and Ferejohn (1993, 1–2), the most common philosophical interpretation of rational choice theory “conceives of it as a psychological theory wedded to a reductionist program in the social sciences, where the behavior of a social aggregation is explained in terms of the mental states (i.e., the desires and beliefs) of its component individuals and their interactions.” Elster (1986b, 12) also argues that rational choice explanation is a “variety of intentional explanation.” It requires not only that agents’ “reasons be causes of the action which they rationalize,” but also that agents’ beliefs and desires, on which those reasons are based, be both rationally held and internally consistent. As Elster elaborates:

Ideally, then, a rational-choice explanation of an action would satisfy three sets of requirements. First, there are three optimality conditions. The action is the best way for the agent to satisfy his desire, given his belief; the belief is the best he could form, given the evidence; the amount of evidence collected is itself optimal, given his desire. Next, there is a set of consistency conditions. Both the belief and the desire must be free of internal contradictions. The agent must not act on a desire that, in his own opinion, is less weighty than other desires which are reasons for not performing the action. Finally, there are [*sic*] a set of causal conditions. The action must not only be rationalized by the desire and the belief; it must also be caused by them and, moreover, caused ‘in the right way’ [it must have been intended by the agent to produce the effect it in fact produced]. Two similar causal conditions are imposed on the relation between belief and evidence. (16)

Not surprisingly, some rational choice theorists do not want to commit themselves to every aspect of so demanding an account. In the real world of politics, coming up with explanations that can be shown to meet the relevant optimality, consistency, and intentional conditions would be a tall order. It is evident, however, that neither the optimality requirement nor the consistency conditions (which guarantee an Arrovian weak ordering) can be relaxed without abandoning the entire rational choice venture. This leaves the rationality of the agent’s beliefs and the intentional account of causa-

tion as the obvious candidates for less demanding kinds of formulation. Rational choice theorists have explored both possibilities.

There are good reasons for seeking to relax Elster's strong requirements about the rationality of an agent's beliefs, if only because of the enormous burdens that are otherwise placed on the researcher. One way to do this is to take the agent's beliefs as given. In effect this means that beliefs are subjected to requirements that are no more epistemologically demanding than are preferences in the model, thereby obviating any need for the researcher to broach questions that move beyond the subjective world of the agent under study. Another way is to remain agnostic about whether the content of a belief is true or false, requiring only that an agent's beliefs be rationally updated as he or she encounters new information.

Moves of this kind create problems of their own, however. As Downs (1957, 8) notes, "How can we distinguish between the mistakes of rational men and the normal behavior of irrational ones?" If one takes the agent's beliefs as given, it may be impossible to distinguish these two cases. If, on the other hand, the researcher does try to distinguish them empirically, substantial measurement problems have to be confronted. It is often difficult to know whether a person's beliefs are rationally held in Elster's sense, or even whether they have been rationally updated in the light of new information.

Demonstrating empirically the existence and causal efficacy of intentions is difficult in the best of circumstances, and some rational choice theorists have flirted with abandoning the intentionality requirement entirely. Thus McKelvey and Ordeshook (1982, 312) argue that political candidates employ complex strategic decision rules, even if the "substantial numerical complexities" required by these strategies make it doubtful that the candidates "could ever compute and abide by such solutions." Likewise, Posner (1972, 1979, 1980) abandons the intentionality requirement when he argues that common law judges make decisions that maximize the efficient production of wealth, but he thinks that the judges are typically unaware of this result and that often they do not intend to produce it.⁴

This theoretical move obviates the need to identify intentional causal mechanisms, but at a considerable cost from the standpoint of empirical testing. It then becomes exceedingly difficult for the researcher to pin down what the causal mechanism involved is or to know what would count as evidence in support of its existence. Satz and Ferejohn (1993) try to sidestep these difficulties by distinguishing "internalist" from "externalist" rational choice explanations. Describing the conventional requirement of intentional causal agency as an internalist interpretation, they contend that it is unnecessarily demanding for many of the questions social scientists study. This is an

4. Indeed, he goes so far as to excoriate one judge for trying to apply wealth-maximization theory in a particular case (Posner 1979, 298–99).

important move because Simon (1955, 1956), Kahneman and Tversky (1979, 1984), Abelson and Levi (1985), and others have shown that the psychology of choice often differs significantly from the deliberative processes of rational choosing. If rational choice theory “is taken to specify a psychological mechanism,” Satz and Ferejohn (1993, 6) concede, “then these criticisms may be fatal.” Accordingly, they propose that the theory should be thought of as illuminating “structures of social interaction in markets, governments, and other institutions.” On an externalist understanding, rational choice theorists “are not interested in explaining a particular agent’s behavior, but in the general regularities which govern the behavior of all agents.” Satz and Ferejohn contend that these regularities reflect the fact that “it is not the agents’ psychologies which primarily explain their behavior, but the environmental constraints they face.” On this view, rational choice explanations are best thought of as accounting for environmental constraints and their effects; as such they “do not necessarily depend on psychological foundations” (7).

Satz and Ferejohn recognize that there are difficulties with unqualified forms of externalism. Although they want to insist that good rational choice explanations need not be derivable from postulates about psychological states of individuals, they do not deny the causal influence of mental states, and they concede that predicting an action is not the same as explaining it. Consequently, they argue that rational choice explanations must be compatible with, though not necessarily deducible from, maximizing assumptions about the intentions of individuals. The relevant agents’ actions must be explicable *as if* they were maximizing utility.

Theorists who employ hypothetical assumptions of this kind often think by reference to evolutionary metaphors, since evolutionary theory is in substantial part a theory about the structural constraints within which organisms exist. As Satz and Ferejohn (1993, 17–18) put it, “Evolutionary biology views nature as a selective structure. The structure of nature selects types with certain properties: those who lack those properties do not reproduce. However, those properties are not necessarily the consequence of the intentional states of the organism. Nonetheless, these properties themselves can often be described in a decision theoretic way; we can predict the behavior of an organism by assuming that, within constraints, it will behave in ways that will maximize its expected reproductive output.” Just as evolutionary theory is not a theory about the intentions of organisms, so rational choice models in the social sciences are best understood as a models of “powerful selective mechanisms.”

The Satz-Ferejohn strategy for avoiding robust assumptions about the causal efficacy of intentions is intuitively appealing; yet it runs into difficulties that are especially troublesome from the standpoint of evaluating the theory’s empirical power. It is notoriously difficult to test evolutionary theories empirically because they are compatible with so many outcomes. Granted, evolutionary theory does yield certain types of testable predictions. For instance, a version of evolutionary theory might

generate the hypothesis that there was a gradual expansion in the cranial capacities of a particular species over a specified period of time. This, then, would lead to the prediction that newly discovered skulls from later in the specified period should be larger than those from earlier in the period. Such a prediction could be falsified by the discovery of larger earlier skulls, or smaller later skulls, of the relevant type (assuming that independent procedures for dating skulls were available). Typically, however, it is not possible to predict the development of a given organism or even species on the basis of an evolutionary hypothesis. More generally, it is often difficult to come up with testable predictions from evolutionary theory because the workings of natural selection are compatible with an organism's evolving in a myriad of directions, with its surviving or dying out or its evolution being critically shaped by random external events.

On the Satz-Ferejohn interpretation of individual maximization, evolutionary theory would predict, presumably, that lemmings will not jump off cliffs to their deaths or that human beings will not choose to go to certain death in war. Satz and Ferejohn might respond that they are not in the business of predicting particular outcomes or events, but once the move is made to abandon the "internal" reading of the micro-foundations of rational choice hypotheses, it is difficult to see how such hypotheses *can* "illuminate structural relations and causes" (Satz and Ferejohn 1993, 26), except via testable predictions. Yet it will become plain in subsequent chapters that on an "external" reading rational choice hypotheses are compatible with so many divergent empirical outcomes that testing becomes problematic.

The differences between rational choice internalists and externalists, in the Satz-Ferejohn senses of these terms, should not be overstated. Both are methodological individualists in that they posit maximizing propensities of individuals in explanations of political behavior. Both accept the standard "thin" definition of strategic rationality by reference to an Arrovian weak ordering of preferences, and each can also embrace "thick" assumptions about rationality defined in terms of self-interest or some other variable that agents might be thought to maximize. What divides internalists from externalists is that internalists assume that something like Elster's intentionalist account of the psychological microfoundations of political action is true, whereas externalists say that one should proceed as if it were true and see what predictive success can be achieved.

Universalism and the Search for Equilibria A second assumption about explanation that commands widespread agreement among rational choice theorists concerns their universalist aspirations. Rational choice theorists "are committed to a principle of universality," Ferejohn (1991, 281) observes, according to which "[all] agents act always to maximize their well-being as they understand it, based on their beliefs, preferences, and strategic opportunities." Rational actor theory, Noll and Weingast

(1991, 239) note in a similar vein, "should seek consistency and universality." The rational choice commitment to universality results from its proponents' conception of scientific advance, which is thought to occur when generalizable results can be shown to follow from analytic propositions derived from axioms.

Rational choice theorists are skeptical that universal theories of politics can be developed through the inductive methods that have characterized political science through most of its history. "Deductive theoretical propositions are of interest," Achen and Snidal (1989, 168) contend, because they "interconnect with one another." Use of such theories prevents "arbitrary multiplication of explanatory categories" and ensures that surprises and insights flow from the theory rather than mercurial inventions that arise to cope with the idiosyncrasies of particular cases. Riker (1990, 177) insists that the failure of the social sciences to advance reflects the fact that they have "not been based on rational choice models." Bueno de Mesquita (1985, 129) links the scientific status of rational choice models to their lawlike character. "We must not be lulled by apparent empirical successes," he warns, "into believing that scientific knowledge can be attained without the abstract, rigorous exercise of logical proof." In the same spirit Achen and Snidal (1989, 168) argue that social scientists who work from the analysis of particular cases toward empirical generalizations (in their view commonly but mistakenly called "middle level theory") fail to see that this method "makes decisive theory-verification well-nigh impossible." Whatever the merits of inductive generalizations, they "are not a substitute for theorizing; empirical laws should not be mistaken for theoretical propositions."

For many rational choice theorists, the search for theoretical propositions is a search for equilibria. Ordeshook (1982, 25) notes, for example, that despite other methodological differences, rational choice theorists "share, knowingly or unknowingly, a common goal: to search for political equilibria." With characteristic decisiveness Riker (1980, 443) declares that in "the absence of such equilibria we cannot know much about the future at all." Though they interpret the concept of equilibrium in competing ways (see Ordeshook and Shepsle 1982), rational choice theorists insist that unless equilibria can be discovered, lawlike statements—from which predictive hypotheses are derived—cannot be developed. Ordeshook (1986, xiii) explicates the relationship between equilibria and lawlike statements as follows: "An equilibrium is a prediction, for a prespecified circumstance, about the choices of people and the corresponding outcomes. This prediction generally takes the form 'if the institutional context of choice is . . . and if people's preferences are . . . then the only choices and outcomes that can endure are . . . ' Thus, equilibria replace both journalistic interpretations of events and statistical correlations between environmental factors and political outcomes as explanations. In the deepest meaning of the word, the study of equilibria, in game theory, combined with substantive applications, is an attempt to provide *causal* explanations."

Concepts of equilibrium “are the link between our abstract models and the empirical world that we are trying to understand.” But what does equilibrium mean in the study of politics? It is borrowed, with modification, from the natural sciences. “Physical equilibria occur when forces balance one another so that a process repeats itself (such as orbits) or comes to rest (as in completed reactions). The scientist explains such equilibria by showing that, in an equilibrium, the forces must in fact balance; or that, in a disequilibrium, the forces must fail to balance” (Riker 1990, 177). Whereas the physicist’s equilibrium is the product of mechanical forces, that of the rational choice theorist stems from the purposive behavior of individuals: “What must be balanced is choices of actions—that is, intentions, which are thus analogous to physical forces. Social equilibria occur when actors choose in the most advantageous way, given the choices of others, and reach an outcome they would not wish to depart from. That is, they would not wish to have chosen differently because the outcome reached is the best they can achieve under the circumstances” (Riker 1990, 177).

The rational choice conception of equilibrium was influenced greatly by the work of John Nash (1950). A Nash equilibrium occurs if there is a potentially self-reinforcing agreement whereby each actor “does what is best for her given what others [would] do” (Przeworski 1991, 20). It can be understood intuitively as an agreement from which no party has an incentive to defect. Harsanyi (1986, 92) defines it more exactly: “A given strategy of a certain player is called a *best reply* to the other players’ strategies if it maximizes this player’s payoff so long as the other players’ strategies are kept constant. A given combination of strategies (containing exactly one strategy for each player) is called an *equilibrium point* if every player’s strategy is a best reply to all other players’ strategies.” When people can enter into binding agreements with others, “an equilibrium corresponds to an outcome in which no coalition has the incentive or the means for unilaterally insuring an improvement in the welfare of all of its members. In game-theory terms such an equilibrium is called a core and corresponds in simple voting games to a Condorcet winner” (Ordeshook 1982, 26).⁵

If a single equilibrium point exists for a given configuration of actors’ preferences and set of institutional rules, then it is possible to derive predictive hypotheses about

5. The term *game* refers to a formal representation of a choice situation. This formal representation specifies the set of players, the strategic options available to them, the outcomes associated with each combination of players’ moves, and the way the players rank the possible outcomes in terms of their preferences. See Luce and Raiffa 1957, chaps. 1 and 3. Noncooperative game theory concerns social interaction in which agreements and promises are not enforced by a third party, as distinct from cooperative game theory, in which there is usually communication among players, the opportunity to make binding agreements, and third-party enforcement. The presence of third-party enforcement means that there is no need for self-enforcing contracts in cooperative games. See Nash 1950; Harsanyi 1986, 92–93.

what actual agents will do, assuming that people behave rationally. If there are many possible equilibria, then rational choice models become more indeterminate; if there are no equilibria, then the political world threatens to be chaotic and inherently unpredictable in its basic structure. This is why so much of the theoretical rational choice literature revolves around trying to identify the necessary and sufficient conditions for the existence of equilibria.

The dominant view among rational choice theorists is that in politics unique equilibria can seldom be identified, though theorists differ on the significance of this fact. For those like Riker (1980, 443) it means that political science is “*the dismal science*.” On his view, if determinate predictions cannot be derived from the laws in which equilibrium models are embedded, then the claim that rational choice models amount to anything more than mere empirical generalization has to be abandoned. Other rational choice theorists take less than an all-or-nothing view. Elster (1986b, 19) notes, for example, that when a model predicts multiple equilibria “it can still help us to eliminate some alternatives from consideration, even if it does not conform to the ideal of eliminating all options but one.” Ordeshook (1986, 98) points out that the discovery that no equilibrium exists can be “a clue to what actions and outcomes we can anticipate,” and a considerable rational choice literature has developed in an effort to model strategic behavior in such settings.⁶

Rational choice theorists who resist Riker’s pure, all-or-nothing universalism do not entirely abandon universalist ambitions. The qualified forms of universalism that they adopt do vary, however. One account, advocated by Elster and Ferejohn, may be described as *partial universalism*. This is the view that rational individual maximization explains part, but not all, of what goes on in every domain of politics. There is disagreement, among those who adopt this view, over just how much explanatory work rationality should be expected to do in different circumstances. For Elster (1986, 27) rationality should play a “privileged, but not exclusive role” in explaining political outcomes. Ferejohn (1991, 284) makes a weaker claim, based on the acknowledgment that multiple equilibria are ubiquitous: “In a very wide class of situations of strategic interaction—indeed, in virtually any game that takes place over time or in which there is a nontrivial informational structure—almost any outcome can occur in some game-theoretical equilibrium. This indeterminacy, often called the ‘folk theorem’ by game theorists, suggests that unless we substantially enrich the concept of rationality itself, or supplement it with extra assumptions about human nature, rationality by itself cannot fully account for the selection of one outcome rather than another.”

This leads Ferejohn to argue that rational choice theory should be complemented

6. See reviews in Ordeshook 1986 and McKelvey 1991. This issue is taken up again in Chapter 6.

by other partial theories, such as cultural theories. Because human actions are located at the boundaries of the sphere of action, which is constrained by the logic of rational calculation, and the sphere of meaning, which is constrained by “subtler ideational logics,” they cannot be explained “without taking both spheres into account” (1991, 283–86). For Ferejohn, then, cultural theory supplements rational choice theory by enabling the researcher to discover which of the many possible equilibria rational choice theory predicts will actually occur. Just how much remains that is genuinely universal on this type of account is debatable. In subsequent chapters we note that rational choice theorists who have advocated partial universalism have left unexplored the extent to which a phenomenon is explained by individual maximization as opposed to habit, blunder, and the like. Nor have they devoted much attention to how individual maximization interacts with other independent variables, preferring instead to focus on the rationality components of partial universalist explanations.

A more radically revisionist approach is *segmented universalism*, the view that rational choice explanations are successful only in certain domains of political life. On this view, the systematic failure of rational choice theories in certain domains—such as in the explanation of voter turnout—suggests that rational choice theorists should try to develop accounts of the circumstances in which rational choice explanations will succeed. Satz and Ferejohn (1993, 3) suggest, for example, that “rational choice explanations are most plausible in settings in which individual action is severely constrained.” Just as rational actor models do a better job of explaining the behavior of firms than of consumers in the economy, they contend, so these models should be expected to do a better job of explaining the behavior of parties than voters in politics. Another possibility is that rational choice models tend to be more successful in domains of politics that are comparatively similar to economics, so it would be reasonable to expect more success in accounting for bureaucratic capture than for ethnic riots (Schumpeter 1942; Green 1992). A third view, advanced by Maoz (1990, 318–21), is that rational choice models will be more successful in situations that do not involve extremely high or extremely low levels of stress. Low stress “can imply both low motivational drives and low practical constraints,” suggesting that agents are “likely to resort to routine mechanisms for problem solving.” Conversely, in high-stress circumstances typically “the motivational drive is extremely strong, and time pressure is acute.” This increases the likelihood that “emotional factors and practical constraints inhibit analytic procedures” and prevent rational decision making. On this view, rational choice models should be expected to do best in situations of moderate stress. A fourth hypothesis, suggested by Elster (1986, 19–20), is that rational choice explanations are more likely to be correct when the options confronting an agent are fixed rather than when they hinge on the possible actions of others, and in less urgent decisions than in more urgent ones. Brennan and Buchanan (1984) embrace yet another logic when they argue that since the act of voting is not plausibly regarded as

an investment—because of the infinitesimal chance that any one vote will affect the outcome—spatial models of voter preference do not fall within the domain of the theory. Aldrich (1993) and Fiorina (personal correspondence, 1993) generalize this argument, suggesting that rational choice models will be useful when the stakes are substantial and when the actions of the individual have a significant impact on the disposition of the stakes. If the outcome does not matter much or if the agent is unlikely to be able to influence it, then it would not be worthwhile to be strategically rational.

Defenders of segmented universalism might be thought to have abandoned universalism entirely; this would be a misperception. First, they often have in mind what Ferejohn describes as thick-rational interpretations when they acknowledge rational choice theory's limits. Thin-rational accounts, which are conceived of by Ferejohn, Riker, and others who defend them as tautologies applicable to all human action, continue to be maintained across the board. Second, most rational choice theorists who affirm segmented universalism back into this affirmation, domain by domain, as a result of empirical setbacks. They want to defend the most universal a variant of the theory possible, and they typically search for a variant that can explain what goes on in a particular recalcitrant domain rather than give up on explaining that domain. Third, some of the arguments for segmented universalism are themselves rooted in rational choice logic, most obviously Fiorina's conjecture that in certain circumstances it will not be strategically worthwhile to behave strategically, and Satz and Ferejohn's claim that relatively constrained conditions are more likely to prompt strategic action than relatively unconstrained conditions. It will become clear in subsequent chapters that conjectures of this sort have yet to be tested empirically and that the theoretical arguments that underpin them remain relatively underdeveloped.

Finally, some rational choice theorists have tempered their universalism by depicting rational choice as a *family of theories* rather than as a single theory. Different versions of the theory involve various claims about what is maximized, as we have seen. Apart from the distinction between thin and thick accounts of rationality, there are different thick-rational accounts in the literature. Votes, wealth, profits, power, influence, or some other entity can be maximized, depending on the stipulations of the theorist. Different rational choice theorists also work with various assumptions about what instrumental rationality entails. Some theorists try to account for agents' anticipation of the strategic behavior of other actors. This aspect of the picture is further complicated by the range of rational choice views concerning the attitudes of maximizing individuals toward one another—attitudes that range from mutual indifference to various kinds of interdependent utilities.

It is sometimes said, therefore, that rational choice theory is really a family of theories that share in common a commitment to the idea that the maximizing behavior of individuals explains political outcomes (see Becker 1976; Laver 1981; Elster

1986a; Riker 1990, 172–77). Indeed, some theorists appear to suggest that even the maximization hypothesis can be modified or abandoned in certain circumstances. Ferejohn and Fiorina (1993, 1) characterize their theory in which voters go to the polls in order to minimize the chances of experiencing that outcome they most regret (failing to cast the decisive ballot in an election) as a rational choice model, although Schwartz (1987) and others disagree. Both Lowi (1992) and Monroe (1991) regard Simon's model of satisficing behavior as a type of rational choice explanation, a characterization that Simon (1993) contests.

Depending on one's interpretation of the relevant basis for family resemblance, adopting the family-of-theories view might amount to anything from a mild qualification of universalism to complete abandonment of it. At the latter extreme, note that Wittgenstein (1963, 31–32) coined the term *family resemblance* as part of his attack on universals, pointing out that words like *game* refer to classes of related phenomena that share no single defining feature. It is doubtful, however, that any rational choice theorist would want to go that far; such a move would be at odds with the frequently trumpeted aspiration to come up with a theory that could credibly be described as more systematic and less ad hoc than the going alternatives in political science. Our impression is that rational choice theorists tend to prefer some members of the family to others, creating a hierarchy that ranges from versions of the theory that are implausible but interesting to versions that are plausible but banal.

The versions of rational choice theory that are most arresting, and usually most coveted by rational choice theorists, are thin-rational accounts that produce counter-intuitive results regardless of agents' tastes and preferences or their knowledge about one another's likely behavior (Stigler and Becker 1977). Arrow's impossibility theorem rests on such an account, but results like his are few and far between. More common are thick-rational accounts that posit self-interest as the basic political motivator. Among these, rational choice theorists most often try to vindicate those that posit the self-conscious maximization of money, power, or influence under conditions of full information. It was because Olson offered a theory of this kind that *The Logic of Collective Action* attracted so much attention. When such explanations fail, rational choice theorists typically move to imperfect information models as the first line of defense (Harsanyi 1986; Coughlin 1992). If this fails, the next step often involves turning to thick-rational accounts that appeal to motives other than narrow self-interest, as in Riker and Ordeshook's contention (1968) that voters go to the polls to maximize the psychic benefits of fulfilling their civic obligations.

If these levies around the definition of rationality do not hold, other characterizations of human motivation wait in reserve, as we point out in Chapters 4 through 7. Although results that vindicate the strategic capacities of utility-maximizing individuals are generally preferred, when these do not pan out theorists turn to more realistic decision-theoretic models that make less taxing assumptions about people's cognitive

capacities. When agents apparently fail to maximize even under these conditions, satisficing and minimax regret strategies wait in the wings. And when the possibility of agents' demonstrating even quasi-rational behavior no longer seems viable, there is always the last-resort expedient of turning from "internalist" to "externalist" accounts of causation, thereby opening up the porous world of evolutionary metaphors. Rational choice theorists do not explicitly defend this hierarchy among rational choice family members. Rather, their preferences among the different variants may be inferred from the pattern according to which empirical literatures develop: a research program is founded upon an arresting proposition at or near the top of this hierarchy, but as anomalies arise subsequent work gravitates downward.

ASSUMPTIONS ABOUT SCIENTIFIC METHOD

The rational choice penchant for holding onto some form of universalism, no matter how qualified, is linked to a particular view of the scientific method. The aspiration to deduce true explanations from axioms makes sense by reference to the deductive nomological conception of explanation, which requires that empirical laws include only general terms, referring to "general kinds, not to individuals," and that, "taken together, the laws must entail that when initial conditions of the general kinds described are realized, an event of the kind to be explained *always* occurs" (Miller 1987, 19).

In spite of the commitment to develop general laws, much rational choice research does not fit comfortably within the strictures of this model. The reason is that much rational choice research is based on what Moe (1979, 215–16) describes as "core statements"—axioms, postulates, and assumptions about people and the contexts in which they act—that are concededly unrealistic. These usually include several of the following: that people always act rationally (according to the specified definition); that people base their actions on certain types of information, sometimes "perfect information"; that people update their beliefs in accordance with Bayes' Rule; that people evaluate their options on the basis of values specified in the theory (usually nonaltruistic values or utility schedules that exhibit such mathematical properties as transitivity, ordinality, etc.); that the relevant political "commodities" are homogeneous and infinitely divisible; and that preferences remain fixed for the duration of the time frame in question. Although some rational choice models are more unrealistic than others, they are, as Moe notes, usually "not even close to descriptive accuracy."

The use of explanations based on unrealistic assumptions is usually justified by reference to a model of explanation different from the deductive nomological one. On this view, which Milton Friedman (1953, 3–43) developed partly as a critique of the covering-law model, science does not necessarily advance via the development of lawlike generalizations; valid theories may just as reasonably emanate from a super-

stitution or a scientist's dream as from a theorem. Indeed, noting that wildly implausible and contradictory theories (measured against the existing stock of knowledge of the day) have sometimes survived and become accepted, Friedman argued that paying too much attention either to realism or to consistent generalization of a theory was likely to produce an illicit conservative bias in theory building. On the Friedman-instrumental view, the decisive test of a theory is its predictive or explanatory power, not its internal structure or its concurrence with received wisdom.

The Friedman-instrumental view can justify building rational choice hypotheses on unrealistic assumptions, but at the cost of undermining the claims of Achen and Snidal, Bueno de Mesquita, Noll and Weingast, and others who insist that scientific advance comes only with developing theory—that is, establishing the existence of covering-laws. As Moe (1979, 215–39) and Miller (1987, 18–19) have noted, the covering-law model gets its distinctiveness and power from its requirement that covering-laws be both general and empirical—subject, that is, to disconfirmation through observation. This reality check is essential to ensuring that covering-laws are not mere flights of intellectual fancy; if they turn out to be at variance with the observed data, they must be abandoned, or modified and then subjected to new empirical tests.

Given the competing rationales behind the covering-law and instrumental views, it is not legitimate both to justify the unrealism of rational choice explanations on instrumental grounds and to appeal to the covering-law model in defense of axiomatic proofs. Either the development of general theory is justified on covering-law grounds (in which case it cannot legitimately be based on unrealistic assumptions), or the unrealism is justified on instrumental grounds (in which case the particular mode of theory building is beside the point; testable predictions are what matter).

This leaves two options for those who want to pursue rational choice theory as part of the endeavor of advancing the empirical study of politics. One option is to modify the assumptions on which the theory rests to make them more realistic (thereby remaining within the strictures of the covering-law model); this is the tack advocated by Austen-Smith (1984), Krehbiel (1988), Ferejohn (1991), Noll and Weingast (1991), and Johnson (1991). Striving for realism raises the demand for empirical testing and the standards that a theory must meet to be successful. The other possibility is to abandon the covering-law model, justifying the unrealism of the theory on instrumental grounds. This is a view from which rational choice theorists often shy away in their explicit pronouncements on method, since it calls into question the value of much of their work, the bulk of which is not empirical at all. Often, however, rational choice work derives its intellectual appeal from its putative ability to make predictions about politics from a small handful of unrealistic assumptions about motivations, information, and incentive structures. On this description, it is difficult to make sense of the rational choice venture in anything other than Friedman-instrumental terms.

In the end, whether rational choice theory is thought of in covering-law or Friedman-instrumental terms, empirical testing cannot be escaped. On either view, a theory of politics has no payoff if its hypotheses do not survive empirical scrutiny. In this light, it is surprising that both defenders and critics of rational choice theory have paid so little attention to empirical testing. It is to that subject that we now turn.

CHAPTER THREE

METHODOLOGICAL PATHOLOGIES

Whatever may be said on behalf of the analytic elegance or heuristic value of rational choice theories, empirical applications have tended to suffer from two classes of methodological infirmities. The first encompasses what may be described as pedestrian methodological defects. Scholars working within the rational choice tradition from time to time misapply statistical techniques, overlook problems of measurement error, or rely excessively on inferences drawn from a small number of case studies. Although potentially serious, methodological shortcomings of this kind come with the territory in political science and are not the main focus of our critique.

More interesting is the syndrome of fundamental and recurrent methodological failings rooted in the universalist aspirations that motivate so much rational choice theorizing. These concern the ways hypotheses are conceptualized, the manner in which they are transformed into testable propositions, and the interpretation of empirical results when tests are conducted. We contend that these (often mutually reinforcing) mistakes stem from a method-driven rather than problem-driven approach to research, in which practitioners are more eager to vindicate one or another universalist model than to understand and explain actual political outcomes. More than anything else, it is this aspiration that leads to the errors that we describe here as the pathologies of rational choice theory. We make good on the claim that these are *characteristic* failings in Chapters 4 through 7, where we review in systematic fashion rational choice literatures on turnout, collective action, legislative behavior, and electoral competition. In this chapter we describe and illustrate these methodological failings, explaining why they are at odds with basic requirements of sound empirical research.¹

1. It is not our position that every attempt to test rational choice models empirically goes

POST HOC THEORY DEVELOPMENT

Many of the methodological failings of applied rational choice scholarship are traceable to a style of theorizing that places great emphasis on the development of post hoc accounts of known facts. Can a rational choice hypothesis explain the existence of seniority systems in Congress? The growth of deficit spending by governments? Why people vote for third parties? To answer such questions the theorist engages in a thought experiment designed to generate an explanation of a given phenomenon that is consistent with rational choice assumptions, somehow specified. Fiorina and Shepsle (1982, 63) offer a lucid description of this approach:

Our position is that scientific progress reflects (1) the scholarly *choice* of models that (2) possess equilibria that (3) correspond to observed regularities. This entails neither constructing equilibrium models *ex ante*, generalizing and refining subject to the constraint that equilibrium be preserved . . . nor retaining disequilibrium models only to be tongue-tied when asked to say something positive about the world. . . . To travel the first path is to say little that applies to the world of phenomena, and to travel the second is to say little, period. Instead, we recommend a third path, one termed “retroduction.” . . . Put simply, the retroductive process begins with an empirical regularity X and poses the question, “How might the world be structured so that X is an anticipated feature of that world?” The answers (and there should be several) are models, all of which have in common the regularity X as a logical implication.

To be sure, striving to explain observed empirical regularities is preferable to fashioning theories according to the dictates of “neatness, or other aesthetic criteria” that otherwise guide rational choice theorizing in both political science and economics (Fiorina and Shepsle 1982, 63). But given the lack of specificity about what it means to be a rational actor, it is not obvious what sorts of behaviors, in principle, could fail to be explained by some variant of rational choice theory. Rational choice theorists have at their disposal a variety of assumptions about actors’ objectives (wealth, power, moral satisfaction, etc.) and the extent to which individuals derive utility from the well-being of others, as well as the sorts of information and beliefs actors possess, their tastes for risk, the rate at which they discount future rewards, whether their decisions are informed by reasoning about strategic behavior of others, and, if so, the decision rules used when actors face conditions of uncertainty.² As

awry. But as we point out in the chapters that follow, in those rare cases when appropriate tests are appropriately conducted, the results seldom sustain any novel or counterintuitive propositions.

2. Although rational choice theory is often advertised as a unified approach to the study of social, economic, and political behavior, we saw in Chapter 2 that there seem to be few constraints on the assumptions that underlie empirical accounts, and sometimes quite

Ordeshook (1993, 95) points out, those who craft post hoc explanations have not necessarily achieved much: "Even if such models fit the data up to an acceptable level of statistical accuracy, we must contend with the fact that we can establish nearly any reasonable outcome as an equilibrium to some model, provided only that model is sufficiently complex. . . . Designing assumptions so that a model's predictions fit the data is, in fact, little more than an exercise in curve fitting, albeit of a slightly more complicated sort than the type we generally hold in disrepute."

One indication of the ease with which post hoc accounts may be generated is that a great many sufficient explanations arise to explain phenomena such as nonzero voter turnout or differences between the platforms of the two American parties. Another indication is that sufficient explanations pop up to explain certain "stylized facts" that, on reflection, are not facts at all. McKelvey and Riezman (1992, 951), for example, set for themselves the task of explaining both why incumbent legislators tend to be reelected by wide margins and why legislatures have seniority systems. But neither of these premises holds for legislators or legislatures generally. The reelection rates of U.S. senators and representatives contrast sharply, and the strength of the seniority system in Congress has varied over time. Furthermore, statistical studies of Congressional elections (Feldman and Jondrow 1984; Ragsdale and Cook 1987) detect no evidence of the putative causal connection between seniority and incumbent electoral fortunes. Under these circumstances, it is difficult to know what to make of McKelvey and Riezman's analytic result that in equilibrium legislators adopt a seniority system and voters unanimously reelect all incumbents.³

One might at this point object that what we are calling post hoc theorizing might well be characterized as puzzle-solving, a legitimate scientific activity. It could be argued, for example, that the fact that voters go to the polls in large numbers despite the theoretical prediction that rational citizens abstain leads to the discovery of civic-mindedness. Our reservation about such "discoveries" (if they may be described as such) is that retrodution merely establishes the proposition that it is not impossible that some rational choice hypotheses might be true. Often rational choice theorists seem to regard this as the end of the exercise; that the post hoc account they propose indeed vindicates the approach of looking at politics as though it were populated by actors who approach "every situation with one eye on the gains to be had, the other eye on costs, a delicate ability to balance them, and a strong desire to follow wherever rationality leads" (Downs 1957, 7-8). Data that inspire a theory cannot, however, properly be used to test it, particularly when many post hoc accounts furnish the same prediction. Unless a given retroductive account is used to generate hypotheses that

contradictory motives are imputed to agents, depending on the domain of application (Mueller 1979).

3. McKelvey and Riezman (1992, 958) caution that their model implies more than one equilibrium. An alternative equilibrium is one in which "seniority is rejected by the legislature and all legislators are defeated for reelection."

survive when tested against other phenomena, little of empirical significance has been established.

For example, many rational choice theorists have sought to explain why, as Schumpeter (1942, 261) put it, “normally, the great political questions take their place in the psychic economy of the typical citizen with those leisure-hour interests that have not attained the rank of hobbies.” The hypothesis of “rational ignorance” (Downs 1957) holds that citizens know little beyond what they can learn costlessly, because they have no incentive to expend resources to become knowledgeable about political affairs. In light of the small probability that any voter’s ballot will prove decisive in an election, the rational citizen reasons that the benefits of casting a well-informed vote will not offset the expenditure of time and money spent gathering information. As we note in Chapter 5, this argument is widely touted as a successful explanation of what is taken to be widespread voter ignorance. But since other post hoc explanations for voter ignorance are imaginable, one must ask: Why should we put stock in *this* explanation? What else does this account tell us about the conditions under which voters will or will not invest in costly information?

Post hoc theories are not only tested inadequately, the manner in which they are developed tends to be in tension with the enterprise of empirical testing. To the extent that theorists exploit the ambiguity in the meaning of rationality to transform successive disconfirming instances into data consistent with a newly recast theory, one must question whether the succession of theories is susceptible to empirical evaluation in any meaningful sense. As we will see in subsequent chapters, rational choice theorists seldom set forth a clear statement of what datum or data, if observed, would warrant rejection of the particular hypotheses they set forth or, more generally, their conviction that politics flows from the maximizing behavior of rational actors.⁴

These problems of empirical evaluation are compounded by the fact that rational choice models of a given phenomenon are difficult to evaluate vis-à-vis alternative theoretical perspectives that are not rooted in the assumption of utility maximization. In principle as well as in practice, rational choice models may be constructed from a wide assortment of assumptions about beliefs, tastes, and environmental constraints. Not surprisingly, rational choice models may generate diametrically opposing predictions. Some rational choice accounts, for example, predict that collective political action will collapse under the weight of the free-rider problem; others suggest that such movements may be sustained by solidary incentives. Some variants of rational

4. It is not hard to understand why rational choice theorists might be reluctant to relinquish the propositions that they advance. Leaving aside rare instances in which theorems rest on flawed proofs (e.g., Austen-Smith and Riker 1987), these propositions *are* true as analytic statements. Rational choice theorists, therefore, often regard empirical setbacks as indicating a given theorem’s limited range of application. As we point out in Chapters 5 to 7, theorists in this position often cling to the notion that the forms identified in a theorem are fundamental and operative, even if they are offset in specific applications.

choice theory predict that candidates in a two-party system will adopt identical platforms, while others assert that candidates will adopt divergent political stances. That constructions of rational choice theory predict X and Not-X creates vexing problems for those seeking to compare the performance of rational choice models against competing perspectives. The predictions of one rational choice model will invariably overlap with those derived from another kind of theory.

Alternative theoretical accounts, it should be noted, occupy a small pedestal in the rational choice pantheon. The drive for sufficient accounts of political phenomena often impels rational choice theorists to focus instead on what the theory *does* seem to explain. As Russell (1979, 11) notes, this style of analysis is often accompanied by a striking disregard for alternative explanations, leaving open the question of whether the data conform equally well to the predictions of competing theoretical accounts. Sometimes the failure to consider the relative strength of rational choice versus alternative explanations stems from mere sloppiness or parochialism. More often, however, it results from a faulty approach to theorizing that stresses the formulation of sufficient explanations. Ironically, the insistence on pressing one form of explanation to the exclusion of others has the effect of diminishing the persuasiveness of rational choice accounts.⁵

Because of the lack of interest in competing explanations, research is seldom designed with an eye toward rejecting a credible null hypothesis, a conjecture accorded presumption of truth by the researcher, in favor of a rational choice-derived alternative. The null hypothesis that the researcher seeks to reject is frequently rather prosaic—for example, the hypothesis that experimental electors vote randomly (McKelvey and Ordeshook 1984) or that behavior is unresponsive to changes in price (Wittman 1975).⁶ Just as overcoming an adversary like Grenada does little to attest to the military might of the United States, one's views of politics are not much influenced by the fact that a rational choice proposition vanquishes a trivial or implausible null hypothesis. This is not a critical failing, but we should accord explanatory power to rational choice theories in proportion to the credibility of the null hypotheses over which they triumph. More often than not, rational choice scholars consider either untenable alternative explanations or none at all.

In sum, when post hoc theorizing is used to come up with possible rational choice

5. Olson's rational choice explanation for the economic decline of Britain (1982), for example, surely would have been more compelling had he compared (or even mentioned) any of the more than half-dozen competing explanations (see Cameron 1988). Much the same may be said of the large literature that places the blame for inflation and the growth of government at the doorstep of democratic institutions and the incentives they engender (see Barry 1984; Mueller 1989, chap. 17).

6. Wittman (1975, 738) offers (though does not test) the hypothesis that those given paid time off work in order to vote will be more likely to do so. He also suggests that turnout will be higher, all things being equal, among citizens in good health.

explanations of observed phenomena or to reformulate rational choice hypotheses in ways that evade or appear to account for anomalous instances, the rational choice theorist may believe that the theoretical approach has in some significant way been “saved.” In reality, the specific hypotheses in question have yet to be tested.

This critique of post hoc theorizing is not meant to foreclose the possibility of genuine theoretical innovation. Our point is not that theoretical predictions can never be changed to accommodate new evidence. Rather it is that the “innovations” that typically emerge do not involve new predictions as such; they involve mere redescription of the processes by which a previously known outcome obtains. Having recast their hypotheses to encompass known facts—and, in particular, anomalies—rational choice theorists typically fail to take the next step: proposing a coherent test to gauge the empirical adequacy of the newly revised hypothesis. Even less often do they take the step after that: gauging the empirical power of their preferred theoretical formulation against that of alternative explanatory accounts.

FORMULATING TESTS

To test a theory, one needs to know in advance what the theory predicts. From time to time, certain rational choice theorists have expressed discomfort with the lack of attention devoted to this aspect of applied rational choice scholarship. For instance, in 1978 Fiorina and Plott observed that “game theoretic and social choice—theoretic models . . . are developed and advocated without a hint of possible operational definitions—one can find proof upon proof, but one searches in vain for a detailed discussion of exactly how and when a model should be applied” (575–76). Concerns of this kind, however, have had surprisingly little impact on the evolution of rational choice scholarship, and the imbalance between analytic exposition and application remains marked.

Those who seek to derive testable propositions from rational choice models frequently find, moreover, that these theories are constructed in ways that insulate them against untoward encounters with evidence. This problem turns up in various forms. Those who advance models so parsimonious or abstract that recognizable features of politics are all but absent (for example, models of policy making that omit mention of political parties and treat each branch of government as unitary actors [Banks 1989; Spiller and Spitzer 1992]) deflect empirical scrutiny by describing their theories as simplifications or first cuts at thorny theoretical issues. Others assert that their models capture general truths that need not coincide with specific applications, as when Calvert (1985, 87) defends a model of candidate strategy “because it reveals the properties that underlie all electoral competition, even though these properties may be counteracted by the particular conditions of a real world situation” (see also Strom 1990, 11).

Arguably the most important source of slipperiness in model building is the multiplication of unobservable terms, which causes the complexity of a theory to outstrip the capacity of the data to render an informative test. This general problem is compounded by the specific difficulties that attend the ambiguous translation from equilibrium models to empirical tests.

Slippery Predictions Rational choice explanations typically comprise an array of unobservable entities. Tastes, beliefs, decision rules, and, at a higher order of abstraction, equilibria, form the essential ingredients of most rational choice models. The problem is not the positing of unobservable terms per se, but rather the ratio of latent constructs to observable measures in rational choice accounts.⁷ As this ratio grows, it becomes increasingly difficult to establish whether a set of data confirms or disconfirms a rational choice explanation.

Consider, by way of illustration, a game in which two players must divide \$14 between them. If the players can agree on how to allocate the money, then that agreement becomes binding; if no agreement is reached, then player 1 receives \$12, and player 2 receives nothing. “Cooperative game theory,” note Hoffman and Spitzer, “predicts that the subjects will cooperate and divide the rewards \$13 to \$1 (the Nash bargaining solution: an even division of the \$2 gain from trade). Under no circumstances should [player 1] settle for less than \$12, according to game theory” (1985, 259). Suppose that after repeated observations of this game actually being played, one encountered a substantial number of resolutions in which the players divided the \$14 evenly.⁸ What may be inferred from this pattern of results? That the dollar amounts were too small to induce preferences over and above preexisting tastes for fairness? That despite the proscription of threats, player 1 feared physical retaliation from player 2? Mistaken understanding of the game? A temporary departure from equilibrium that would be rectified through greater exposure to the real world of cutthroat negotiations?

As this example indicates, rational choice hypotheses that meet with unanticipated facts may be resuscitated by recourse to a variety of unobservable thought processes

7. The problem is exacerbated to some degree by the skepticism with which rational choice scholars regard “psychological” measures of tastes and beliefs. Although tastes and beliefs figure prominently in rational choice explanations, many scholars working within this tradition question the validity of measures other than behavior—actual choices—as indications of preference. As we note in the chapters that follow, however, this skepticism about soft data has not prevented rational choice theorists from voicing speculations about psychological processes based on no data.

8. Indeed, Hoffman and Spitzer (1985, 260) report that all of their experimental subjects do precisely that when the roles of players 1 and 2 are assigned by coin flip. Under these conditions, the subject in the role of player 1 always “agreed to take \$5 *less* than the \$12 that he could have obtained *without* the other subject’s cooperation.” See also Hoffman and Spitzer 1982.

for which there are insufficient direct or indirect measures. When faced with discordant results, it may be difficult, therefore, to distinguish empirically among three different claims about the principal unobservable term, equilibrium:

- The preferences assumed by the model are accurately represented in the setting one observes, but some or all of the actors lack the strategic acumen to play the game as rational choice recommends, and hence predicts.
- The model accurately captures the actors' objectives, but, perhaps owing to the particular characteristics of the equilibrium itself, there is a temporary departure from this predicted outcome.⁹
- The model does not capture one or more features of the observed game, and the outcomes conform to the equilibria (or lack thereof) associated with some other game.

The propagation of theoretical terms that are either unmeasurable or difficult to measure creates a situation akin to *underidentification in statistical models involving latent variables* (Bollen 1989). Under these circumstances, data cannot furnish a convincing test. When any hypothesis fails, the researcher is always in a position to argue that a successful prediction was thwarted by an offsetting tendency or temporary aberration. In this respect, empirical discussions in rational choice scholarship are reminiscent of debates about the declining rate of profit that once preoccupied Marxists. Having convinced themselves by analytic argument that the rate of profit in capitalism must fall over time but failing to find evidence to support this contention, Marxists for decades devoted their energies to identifying masking, fleeting, and countervailing tendencies that obscure this alleged phenomenon. Declining profitability was believed to be going on just beneath the surface on the strength of a theory that insisted that this must be so (compare Roemer 1979a; Van Parijs 1980).

The underidentification problem may be addressed in two ways. One is to set limits on the range of theoretical arguments that may be used in the construction or resuscitation of a theory. This kind of restriction, however, proves difficult to maintain against the impulse to defend the universal applicability of the rational choice approach. Often these restrictions are endorsed by such figures as Downs (1957) and Olson (1965), who introduce rational choice inquiry into a given domain of politics. But over time these constraints are relaxed by subsequent authors seeking to preserve a model in the face of discordant evidence. Another approach is to gather additional data so as to give the number of measures a sporting chance to catch up with the

9. Fiorina and Shepsle (1982) offer a lucid typology for various kinds of equilibria. Some, like "black holes," attract and retain outcomes in a social system. Others are retentive but not attractive, or vice versa. In the latter cases, it may be impossible to determine empirically whether a system is temporarily or permanently out of equilibrium.

number of theoretical terms. Rational choice scholars tend to shy away from this approach, perhaps a tacit admission that the formal precision of rational choice models greatly outstrips political scientists' capacity to measure.

Vaguely Operationalized Predictions A second common pathology related to hypothesis testing concerns the fit between the hypotheses advanced and the empirical tests used to evaluate them. Since equilibrium analysis is at the heart of much rational choice scholarship, many rational choice propositions are stated in the form of point predictions. Sometimes that point prediction is a rate or proportion, as in the case of Olson's conjecture that in the absence of selective incentives or coercion, members of large groups will not engage in collective action to advance their joint interests (1965). In other cases, the point prediction involves a particular outcome, as in the case of a specific majority rule equilibrium point in a cooperative bargaining game. Such propositions are invariably false to some degree; strategic blunders sometimes occur, producing nonequilibrium outcomes. The argument then shifts to the often expressed "hope that enough people act rationally enough of the time in their political behavior for economic theories of politics to yield descriptions, explanations, and predictions which are frequently useful approximations to the truth" (Kavka 1991, 372).¹⁰

It is unclear whether a rigorous test of a point prediction can be constructed in the form of an approximation. If several millions of dollars in small contributions are collected by referendum campaigns, is that evidence in support of the free-rider hypothesis (Lowenstein 1982, 572–73), given the paltry ratio of contributions to public concern over the outcome of these elections, or against it (Tillock and Morrison 1979), given the presumed irrationality of absorbing personal costs on behalf of a broadly diffused public good?

The match between theory and evidence becomes more ambiguous when rational choice hypotheses move seamlessly between point predictions and marginal predictions. The former concerns the location of an equilibrium under static conditions; the latter—derived from "comparative statics" analysis—concerns the direction in which an equilibrium is expected to move in response to exogenous changes in goals, beliefs, or environmental constraints. It is logically possible that only one sort of prediction will survive empirical testing, but the availability of two standards of evaluation affords defenders of a model more opportunity to claim support for its predictions. In particular, predictions at the margin are often hailed when static predictions fall into trouble. Whatever the defects of rational choice explanations of why citizens bother to go to the polls, Grofman (1993a) argues, rational choice theory does predict correctly that people are less inclined to vote in bad weather.

10. As we note in Chapter 5, when empirical failures occur, this "approximation" notion accompanies attribution of anomalies to the behavior of an irrational few.

We have no objection to the use of comparative statics to generate hypotheses. To the contrary, we find tests that focus on change at the margin much more amenable to traditional quasi-experimental methodology than those involving point predictions. Our concern is with the notion that the rationality of certain actions can be rescued on the grounds that the actors are to some degree responsive to changes in costs or benefits. Take, for example, the study of why politically inexperienced candidates challenge incumbent members of the House of Representatives. The behavior of these challengers is something of a mystery, since their chances of defeating an incumbent are nothing short of dismal. Like most puzzles of this sort, the behavior of challengers may be explained by reference to such ancillary factors as self-delusion, eagerness to promote legal practices while on the campaign trail, belief that *somebody* should contest the incumbent, and so forth. Banks and Kiewiet (1989, 1007) try to salvage the notion that rational, election-seeking behavior accounts for the decisions of weak challengers by arguing that “weak challengers can maximize their probability of getting elected to Congress by running now against the incumbent” rather than waiting for an open-seat contest in which they may have to defeat other strong opponents in both the primary and general elections. As the authors note dryly, “This probability may not be very high, but they are maximizing it.” The study of whether weak challengers are *more* attracted to races against incumbents or to open-seat contests may be a worthy endeavor in its own right, but it is unclear how the results speak to the question of whether weak challengers are rational to oppose House incumbents, so long as rationality requires that the benefits of doing so outweigh the costs (1000).

SELECTING AND INTERPRETING EVIDENCE

Another set of characteristic pathologies concerns the manner in which rational choice hypotheses are tested. The first has to do with the biased fashion in which evidence is selected. The second deals with subtler ways in which evidence is projected from theory rather than gathered independently from it. The last involves the strategic retreat from domains in which the theory is found to perform poorly. All three methodological defects undermine the theoretical claims they are intended to support, as it is the structured search for disconfirming evidence that is essential to scientific testing.

Searching for Confirming Evidence When reading applied rational choice scholarship, one is struck by the extent to which advocates of rational choice models permit their theoretical commitments to contaminate the sampling of evidence. The procedure of adducing instances that confirm a hypothesis is perhaps most transparent in such domains as regulation and bureaucratic politics, where the ideological stakes are

high. This practice, reminiscent of advertisements that show one brand's achievements while mentioning neither its failings nor the comparable achievements of its competitors, is not limited to these ideologically charged domains, however. In its more qualitative manifestations, rational choice scholarship tends to ruminate over confirming illustrations combed from the political landscape, memorable moments in history, and biblical texts (Brams 1980, 1993; Riker 1982, 1986). Elsewhere, this pathology leads researchers to dwell on instances of successful prediction, be they the phenomena of strategic counteramendments by committee leaders on the House floor (Weingast 1989, 810) or the suboptimal provision of collective goods (Olson 1965). The tendency to adduce confirming instances also manifests itself, though in subtler form, in quantitative research that goes through the motions of contrasting treatment and control conditions en route to a conclusion that follows trivially from the research design. McCubbins (1991, 107), for example, finds that time-series analyses of federal data for the period 1929 to 1988 "strongly support" his game theoretic account of how divided party control of Congress leads to budget deficits. Granted, his statistical estimates suggest that "since 1929, divided government has yielded sizable increases in the national debt" (102), but the period studied contains just two such episodes: the advent of supply-side economics under Ronald Reagan, and the drought of federal revenues during the waning days of the Hoover administration.

A variant of this methodological problem surfaces in studies that use laboratory behavior to support rational choice propositions but fail to build a control group into the experimental design. As we argue in detail in Chapter 6, successful experiments of this sort at most suggest that a laboratory setting can be constructed to approximate the conditions presupposed by a theorem; a researcher seeking to defend a rational choice hypothesis need only engineer a confirming illustration. Generated without a control group, the results give no indication of whether the observed outcome would have obtained anyway for reasons unrelated to the theory in question, nor does the experiment tell us whether this theory predicts successfully under other circumstances. Experiments crafted in this way illustrate rather than test.

Projecting Evidence from Theory A profound desire to establish rational choice theory's breadth of application from time to time opens the door to a tendentious reading of the empirical record. Sometimes this is a simple matter of authors imagining a datum consistent with economic logic (for example, bad weather depresses voter turnout) and assuming this datum to be empirically verified. At other times, one finds rational choice theorists asserting almost by way of afterthought that some eccentric feature of a model mirrors reality. For example, McKelvey and Riezman's legislative model (1992) hinges on the assumption that those with seniority are more likely to be recognized on the floor in the initial round of voting but not in subsequent rounds. The authors insist that this characterization provides a "realistic description

of the seniority system for the U.S. Congress” because seniority-influenced committees get first crack at making proposals, and “once the bills go to the floor, the committees lose most of their power” (958). Suffice it to say that this is a rather sparse depiction of the process by which legislation is proposed and amended in Congress (Weingast 1989).

Even when a full-blown empirical study is undertaken, the theoretical convictions of the authors may guide what they infer from a set of observations and how they reconstruct the data for presentation. For example, an obscure set of House votes on the Powell Amendment to a 1956 measure authorizing school construction has been offered up time and again as evidence of how legislators vote to strengthen a proposal they dislike in an effort to make the amended bill unpalatable (Riker 1965, 1982, 1986; Denzau, Riker, and Shepsle 1985). A dispassionate examination of the historical record, however, shows that the facts surrounding the Powell Amendment are at best ambiguous with respect to the phenomenon of strategic voting (Krehbiel and Rivers 1990). Indeed, the omissions and factual distortions that Krehbiel and Rivers detect in previous accounts (556–60, 574) suggest that earlier writers were unable to assimilate data that did not conform to their theoretical expectations.

Arbitrary Domain Restriction On occasion, rational choice theorists will concede that there are domains—such as voter turnout and organized collective action—in which no plausible variant of the theory appears to work. Some theorists are then inclined to withdraw, choosing to concentrate on applications in which these theories appear to have better success. For instance, in trying to make the case that his wealth-maximization hypothesis explains the evolution of the criminal law, Posner (1985) is forced to come to accept that he cannot explain the existence of laws against such “victimless crimes” as prostitution and drug abuse. He therefore abandons this domain, insisting nonetheless that wealth-maximization provides a powerful explanation of the rest of the criminal law.

Such a move might at first sight seem reasonable, even modest, but there is more at stake here than meets the eye. Suppose it transpired one day that red apples did not fall to the ground as other heavy bodies do. One would not be much impressed by the physicist who said of the theory of gravity that, though it seems not to work for red apples, it does a good job of explaining why other things fall to the ground and that consequently from now on he was going to restrict his attention to those other things when using the theory.

What we are calling arbitrary restriction to domains where a theory seems to work is not to be confused with two nonarbitrary forms of domain restriction that scientists engage in routinely. First, as Moe points out (1979, 235), testing of all scientific theories involves the insertion of *ceteris paribus* clauses designed to exclude omitted factors, so that a proper test of the hypothesis that objects of unequal mass fall to earth

at the same rate presupposes wind resistance to be held constant.¹¹ Second, theories may properly include an account of what are conventionally termed “interaction effects,” factors that limit or enhance the influence of the independent variables of theoretical interest. Indeed, the value of a theory in the eyes of those who wish to understand and influence politics may hinge on a clear account of the conditions under which it is held to apply. Arbitrary domain restriction occurs when an empirically testable set of limiting conditions is lacking but retreat is sounded anyway. There is, in other words, a critical difference between specifying the relevant domain in advance by reference to limiting conditions and specifying as the relevant domain: “wherever the theory seems to work.”¹²

The problem of arbitrary domain restriction is thus the obverse of the tendency to adduce confirming instances. The latter involves fishing for supportive evidence; the former, draining lakes that contain problematic data. While the practice of adducing confirming instances produces misleading tests, arbitrary domain restriction renders problematic the enterprise of testing. If the appropriate domain within which a theory is to be tested is defined by reference to whether the theory works in that domain, testing becomes pointless.

Posner, in our example, pushes the case for wealth-maximization as far as he can and cuts and runs when he has to. Yet he neither considers any alternative explanation nor sees the need to offer an account of why the theory breaks down in the domain of victimless crimes. For domain restriction to be adequate, the relevant domain must be

11. It is important to note that *ceteris paribus* provisos must refer to confounding factors, such as wind resistance, whose effects are in principle testable. One cannot take the position that only when all the logical assumptions of a theorem are satisfied empirically do the theorem’s empirical predictions follow.

12. In much the same vein, arguments about when and where to apply a theory must be advanced in a consistent fashion. For example, in an effort to bolster their claim that House “leaders will be chosen in such a fashion that their personal reelection is not too incompatible with the duties of office,” Cox and McCubbins (1993, 130) point out that one rational choice argument, based on the idea of the “uncovered set” (see Chapters 6 and 7), predicts “definite limits to the policy platforms that those seeking leadership positions will adopt” and, in particular, rules out successful bids by noncentrist candidates (130). Although Cox and McCubbins wish to embrace this prediction, they note that it is open to the objection that decisions enacted by majority rule are inherently unstable, that “there will always be some majority, all of [w]hose members could be made better off if its policies were changed” (131). Cox and McCubbins respond that this objection about the inherent vulnerability of House speakers rests on the assumption that actors incur no transaction costs when identifying or forming new majority coalitions. When these costs are taken into account, they contend, the instability problem no longer applies to the choice of speaker. They neglect to mention, however, that their preferred predictions based on the uncovered set also presuppose the absence of transaction costs.

specified independently of whether the theory explains the phenomenon within it. Furthermore, the hypothesis about the limiting conditions of rational choice explanations must itself stand up to empirical testing. As we noted in Chapter 2, rational choice theorists such as Brennan and Buchanan, Fiorina, and Satz and Ferejohn have suggested some hypotheses about the conditions under which rational choice explanations are likely to apply. It will become plain in subsequent chapters, however, that these recommendations have not yet had much impact on the design and application of rational choice models.

CONCLUDING COMMENTS

Although widespread among empirical applications, the methodological problems identified in this chapter are not inextricable features of rational choice theorizing. Indeed, the larger message of this book is not that rational choice models of politics should simply be abandoned. Rather, the rational choice approach must be rethought fundamentally, and its relations with the existing stock of knowledge and theory in the social sciences should be reevaluated. It is therefore necessary to understand what the recurrent methodological problems are, why they turn up, and how they might be remedied. In this spirit, we turn to the literatures on turnout, collective action, legislative behavior, and electoral competition.