2

Rationality: The Model of Choice

In analyzing politics I shall take what has come to be known as the *rational choice approach*. It also goes by other names: formal political theory, positive political theory, political economy. Indeed, in a genuine (but failed) attempt at intellectual imperialism, some economists like to think of it as the economic approach to politics.[1] They are right in one sense. The rationality assumption has been used most extensively and has seen its fullest flowering in economics. But there is nothing distinctly economic about rational behavior, as we shall see.[2]

The term *rationality* has a long history and, in ordinary language, often means something entirely different from what I have in mind. If a friend of yours does something that you would not have done were you in the friend's shoes—say, go to the movies the night before a final exam—you might say, "Jeez, that's really irrational." By that you might mean: Given what *your friend* wants, that is not the best way to go about getting it. Or perhaps you mean something different: Given

[1] For an excellent essay on positive political theory as a failure of economic imperialism, see Peter C. Ordeshook, "The Emerging Discipline of Political Economy," in James E. Alt and Kenneth A. Shepsle, eds., *Perspectives on Positive Political Economy* (New York: Cambridge University Press, 1990), pp. 9–31.

[2] Even in economics, rationality is undergoing revision under the rubric of "behavioral economics." For a fine review of this revisionst interpretation, see Norman Frohlich and Joe A. Oppenheimer, "Skating on Thin Ice: Cracks in the Public Choice Foundation," *Journal of Theoretical Politics* 18 (2006): 235–66.

what *I* want, I would not do what she is doing (and she ought to want what I want). In either case, you are claiming that what your friend is doing is crazy. Crazy it may well be, but I shall reserve *irrationality* for something quite specific.

The term *rationality* as I shall use it does not mean brilliant or all-knowing. The men and women whose behavior we wish to understand are not gods, so we certainly do not want to characterize any deviation from omniscient, godlike behavior as irrational (for then nearly all behavior would fall in this category). The people we model are neither all-knowing nor worldly wise; they are ordinary folks. As such they have *wants* and *beliefs*, both of which affect their behavior.

Preliminaries

Individual wants, which I refer to as *preferences*, can be inspired by any number of different sources. Clearly we humans come hardwired with a number of wants related to survival and reproduction: food, protection from the elements, sexual desires. Other wants may be socially acquired and only indirectly related to such large and weighty matters as survival of the species—a preference for the latest fashion in jeans or the most recent jazz CD. Modern man and woman are economic and social animals. While one cannot deny the strong influence of material, economic wants on individual preferences, additional important sources of preference include religious values, moral precepts, ideological dispositions, altruistic impulses, and a sense of common destiny with a family, clan, tribe, ethnic group, or other community.

The individuals who populate our model world are assumed to have preferences derived from any and all of these various sources. We do not pretend to know why people want what they want—we leave that to evolutionary biologists, psychologists, and sociologists. Nor do we need to know why in

order to proceed. For us, preferences are one of the givens of a situation and, for purposes of analysis, we assume that they don't change much in the short run. In short, we take people as we find them.[3]

I shall occasionally say that people who act in accord with their preferences are *self-interested.* Hindmoor has cleverly noted that "most of us have no difficulty in accepting that some people are self-interested all the time and that everyone is self-interested some of the time, but we balk at the notion that everyone is self-interested all of the time."[4] As already noted, however, I do not require a pinched view that people are selfish in the ordinary sense of that word, but rather that people are selfish only in a less self-absorbed sense: People pursue the things they regard as important, to be sure, but this may include empathy for family, friends, whales, trees, or random strangers. An individual's conception of self is reflected in his or her preferences and priorities. Pursuit of those preferences and priorities is self-interest in this weaker sense at work.

The world of preferences and priorities is an *interior* world. Indeed, because a person does not wear her preferences on her forehead, and sometimes, for subtle reasons, may not be all that she seems, we often have to make *assumptions* about her preferences. That is, in trying to figure out what someone might do, we have to start somewhere, and entertaining hunches and intuitions about that person's motives is often a useful point of departure.

[3] Methodologically, this is very similar to the approach of economists, who take tastes for goods, services, labor, and leisure as fixed in the short run and determined outside the boundaries of their inquiry. However, let us reemphasize that preferences in our discussion are construed more broadly than in conventional economic models—they should not be equated with material well-being. And we should leave room for preference change arising from learning, experience, persuasion, or deliberation.

[4] Andrew Hindmoor, *Rational Choice* (London: Palgrave-Macmillan, 2006), p. 5.

But preferences, tastes, and values are not all there is to rational behavior. Complementing this interior world is an *external environment* in which people find themselves. This environment is filled with uncertainty—about how things work, about the preferences of others, about random events over which individuals have neither control nor sometimes even knowledge. This uncertainty is of interest to us because it affects the way people express their preferences. Individuals have preferences, as already stated, and I assume they have a behavioral repertoire or behavioral portfolio available as well. They may do any of a number of things in pursuit of whatever it is they want (things like going to the movies or studying on the night before a final). They often cannot choose the thing they want directly (like getting an A on the final), but instead must choose an instrument—something available in their portfolio of behaviors. If each instrument leads directly to some distinct outcome, then the job of the rational person is simple: choose the instrument that leads to the outcome preferred the most. If you want an A on the final, and studying the night before produces it while going to the movies does not, then by all means study.

Enter uncertainty. More often than not, individuals may not have an exact sense of how an instrument or behavior they might adopt relates to the outcomes they value. That is, they may have only the vaguest sense of "how the world works," may not quite appreciate how the choices of others influence the final outcome, and may not be able to anticipate random events (like the virus that arbitrarily picks you on the morning of the exam). Consequently, the effectiveness of behavioral instruments for the things an individual wants is only imperfectly known. Personal knowledge and wisdom take one only so far. But one must use what he or she has available. We describe the hunches an individual has concerning the efficacy of a given instrument or behavior for obtaining something he or she wants as that person's *beliefs*. Beliefs con-

nect instruments to outcomes. Acting in accord both with one's preferences and one's beliefs is called *instrumental rationality*.

Beliefs, like preferences, come from a variety of sources, and we need not resolve their origins in order to take them as part of what defines an individual at any moment in time. Indeed, beliefs may change as the individual acquires experience in his or her external environment. Learning takes place, causing the individual to revise initial opinions about the effectiveness of a specific instrument for achieving some particular objective. To be on the "steep" part of a learning curve means for you to be in a relatively novel situation of high uncertainty in which each new bit of experience causes you to revise your views about how the world works in this situation. As bits of experience accumulate, your beliefs begin to settle down, your opinions begin to firm up, and you revise your opinions less frequently and dramatically; you are in the "flat" part of the learning curve—you've learned most of what there is to know and have squeezed out most of the uncertainty (that is squeezable).

I've done a fair bit of throat-clearing to this point. To sum up this preliminary discussion, the conception of rationality employed in this book incorporates both preferences and beliefs. A rational individual is one who combines his or her beliefs about the external environment and preferences about things in that environment in a consistent manner. Since we have no time to spare, I can only note in passing that the rational choice approach is a form of *methodological individualism*.[5] The individual is taken as the basic unit of analysis. In contrast, many sociological theories take the group as the basic building block. Marxist approaches begin with economic classes as the actors in their models. Most theories of interna-

[5] For readers interested in pursuing this theme further, see Geoffrey Brennan and Michael Gillespie, eds., "Special Issue: Homo Economicus and Homo Politicus," *Public Choice* 137 (2008): 429–524.

tional relations aggregate all the way up to the nation-state as the unit of analysis. Indeed, even some economic theories treat aggregates like a firm or an entire industry as the unit of analysis. The most important thing to know about methodological individualism is that it is taken as fundamental that individuals have beliefs and preferences. These things are the stuff of human cognition and motivation. Groups, classes, firms, and nation-states do not have minds, and thus cannot be said to have preferences or hold beliefs.

Now it is time to make these ideas more precise.

Motivation

To motivate a rational model of political behavior, let's begin with a glimpse of how economists practice their craft. I simplify shamelessly in advancing the view that economics is concerned mainly with how four different classes of actors choose to allocate what's theirs. For the *consumer* the choice is one of how to spend his or her monetary endowment so as to achieve a maximum of contentment (or utility, as the economist likes to say). *Producers*, on the other hand, possess various productive inputs and must determine how best to combine them so as to maximize their profits. The endowment of a *worker* consists of time. To keep things simple, suppose that workers toil at a fixed wage rate so that once they decide how much time to spend at work, both their total wages (and hence monetary endowment from which they derive contentment when they transform themselves into consumers) and the amount of time left over for leisure are determined. Workers, then, pick an amount of time to work in order to acquire purchasing power and leisure time, each of which contributes to their contentment. Finally, *investors* are providers of capital. They allocate their wealth across alternative investment opportunities with

DISPLAY 2.1

Actor	Endowment	Objective
consumers	budget	contentment
producers	inputs	profits
workers	time	purchasing power/leisure
investors	wealth	long-run return

an eye on the overall long-term financial return. This is laid out in Display 2.1.

Now, there is surely ambiguity in each of these ideas, but it is fair to say that economists, in one fashion or another, firmly commit themselves to what it is that animates various economic actors. This is not because they think their assumptions—and Display 2.1 is one very simple set of assumptions—are verifiable as descriptive statements. Start asking some obvious questions and you will quickly determine that these assumptions are seriously flawed as descriptive statements. Don't consumers care about anything except their own consumption? Are producers driven entirely by the profit motive, or do they give some weight to other things, like the welfare of their workers or the quality of their products, even if these sometimes come at the expense of profits? Don't workers get any satisfaction from the job itself, or do they only care about how much effort they must expend (and indirectly about their wages)? Is investing for long-run return the only thing an investor can do with her wealth? Can't she devote some of it to finance a tropical vacation in the dead of winter, or a new wing on the local children's hospital?

It is evident that descriptive accuracy is not the point or purpose of the economist's assumptions. The reason is *scientific*, not substantive. The idea is this: Can we explain variations and regularities in economic performance, outcomes, and

behavior with a simple set of assumptions? The modern theory of economics is a grand intellectual edifice precisely because it has succeeded, as no other social science has, in constructing explanations logically, rigorously, and in empirically meaningful ways. At the foundation of this edifice is a scientific commitment to explanation, not description.[6]

This does not mean there is no controversy in economics. What it does mean is that over the past two centuries, a *corpus of scientific knowledge* has accumulated, a corpus different from either an encyclopedia of descriptive detail or even a body of wisdom (by which we mean serviceable commonsense notions). It is, instead, a logically integrated collection of principles, a set of tools of inquiry—a methodology, if you will—for prediction and explanation. Of great importance is the fact that this scientific knowledge is *cumulative*, something that distinguishes it from wisdom, which is intuitive, implicit, and often nontransferable (it dies with its possessor).

Is it possible, in a manner precisely analogous to what has occurred in economics, to create a science of politics? That is, is it possible to begin with a simple set of premises or assumptions and, from these, derive principles of political performance, outcomes, and behavior? This is a daunting challenge, but it is the objective that has motivated the body of work in positive political theory that is the focus of this book.

The Simple Logic of Preference and Choice

Our first building block is the notion of preference. We must begin by defining terms, explaining notation, and making assumptions. Since the machine we are building must serve in a

[6] For an elaboration of this issue as a philosophical matter, see Frank Lovett, "Rational Choice Theory and Explanation," *Rationality and Society* 18 (2006): 237–72.

variety of contexts, our building blocks must be developed in an abstract and general fashion (for which I beg the reader's indulgence). To give the reader something concrete to hold onto, however, consider the dilemma that Claire McCaskill, state auditor of Missouri, confronted after the 2004 election.

CASE 2.1
CLAIRE MCCASKILL'S ELECTORAL OPTIONS

Claire McCaskill graduated from law school in 1978 and went on to become a very successful state politician in Missouri. After several years of judicial clerking and private practice, she was a local county prosecutor, a county representative, a state representative, and, in 1998, was elected statewide as auditor, a position to which she was reelected in 2002. Along the way she broke down barriers for women in politically conservative Missouri. In 2004 she defeated Governor Bob Holden in the Democratic gubernatorial primary, becoming the first person in Missouri history to defeat a sitting governor in a primary election. She lost the general election to Secretary of State Matt Blunt, a Republican, with 48 percent of the vote to his 51 percent. This was her first losing effort in a twenty-year political career.

What would her next career step be? Talking heads and political insiders assumed that her close loss for governor in a year when circumstances favored Republicans would propel her candidacy for governor four years later—a rematch against Blunt, the only politician who had ever beaten her. However, there was the possibility of 2006. Jim Talent, the incumbent Republican senator was up for reelection. He had come to the Senate in the 2000 election cycle in a most peculiar way. Mel Carnahan, then Democratic governor of Missouri, died in an airplane crash while campaigning for the Senate seat only days before the election. His name re-

mained on the ballot and he actually *won* the election! A special election was then called, and Talent defeated Carnahan's widow, Jean.

McCaskill had a choice to make. She could challenge the incumbent Talent for the Senate seat in 2006, or wait and challenge the incumbent Blunt for the governorship in 2008. If she ran and won the Senate seat in 2006, she would not enter the governor's contest in 2008. If she ran and lost the Senate seat in 2006, her prospects, as a two-time loser, of succeeding in 2008 would have been rather dim.

Thus, the career outcomes facing McCaskill were these:

x: a term as senator
y: a term as governor
z: out of politics for the near term

We can be reasonably confident that McCaskill preferred x to y and y to z (and, as will be shown later in this chapter, if she weren't "incoherent" she would also certainly have preferred x to z). But she could not literally choose from $\{x,y,z\}$. Her behavioral options were "run for senator," "run for governor," and "run for both." Since I've assumed she would not choose to run for governor if she won the Senate seat, and she could not expect success if she ran for governor after *losing* the Senate race, it is reasonable to suppose that McCaskill focused only on {"run for senator in 2006," "run for governor in 2008"}.

We can think of her decision problem as that of choosing between two *lotteries*. If McCaskill chose to run for the Senate in 2006, she would obtain outcome x with some probability p and outcome z with probability $1-p$. If she held off to run for governor in 2008 on the other hand, she could obtain outcome y with probability q and outcome z with probability $1-q$. The keys to her decision are how much she prefers x to y and how good are her chances for the former (p) as opposed to the latter (q).

McCaskill ended up running for (and winning, as it happens) a Senate seat in 2006 because her outright preference for x over y was reinforced by a belief about her chances of victory. In either case she would be running against an incumbent, so that factor was more or less a wash. The reason she thought 2006 would be a better year to run than 2008 was her fear that Hillary Clinton would win the Democratic presidential nomination in 2008. (Recall that back in 2005 and 2006, people believed that Clinton as the Democratic nominee was a *fait accompli*.) In her view, no Democrat would win statewide office in relatively conservative Missouri with Ms. Clinton at the head of the ticket.*

*For a wonderful essay on Claire McCaskill's decision making, see Jeffrey Goldberg, "Letter from Washington—Central Casting: The Democrats Think About Who Can Win in the Midterms—and in 2008," *The New Yorker* (May 29, 2006).

In the remainder of this section I abstract from the specific features of this case in order to develop a general logic of rational choice. We begin with a situation in which there are three objects over which a typical actor, named Mr. i, has preferences. We call the objects *alternatives*, and label them x, y, and z. Mr. i, in a manner we make precise below, has the capacity to make statements like, "I prefer x to y," or "I am indifferent between y and z." The alternatives may be career paths (as in McCaskill's choice problem), or political candidates, or potential marriage partners, or laptop computers. It does not matter, for our purposes, what comprises the choice situation or the set of alternatives. Nor does it matter how Mr. i arrived at his preferences. What does matter is that Mr. i is rational in the sense that his preferences have coherence and that his ultimate choice bears a logical relationship to his preferences.

Symbolically, "xP_iy" means Mr. i (whose name appears as a subscript) prefers x to y. In words, the symbols in quotation marks state that "x is better than y according to Mr. i's prefer-

ences." Similarly, "xI_iy" means Mr. i is indifferent between x and y. Thus P_i is i's *strict preference relation* and I_i is i's *indifference relation.*[7]

If Mr. i is given the opportunity to choose among x, y, and z, then we say that his choice is *rational* if it is in accord with his preferences. Thus, a choice is rational if the object chosen is at least as good as any other available object according to the chooser's preferences. Put differently but equivalently, an object is a rational choice if no other available object is better according to the chooser's preferences.

So far this is pretty straightforward and, once you get used to the notation, pretty commonsensical. Now we must determine what must be true about the preference and indifference relations just described so that choosing in conformity with them accords with our intuitions about rational choice. What we are seeking, in effect, are properties of preference relations that allow the chooser to *order* the alternatives in terms of preference (and enable him, being a rational soul, to choose the top-ranked alternative in the ordering). It turns out that two underlying properties capture the commonsensical notion of rationality as ordering things in terms of preference:

> *Property 1: Comparability (Completeness).* Alternatives are said to be *comparable* in terms of preference (and the preference relation *complete*) if, for any two possible alternatives (say, x and y), either xP_iy, yP_ix, or xI_iy. That is, the alternatives are comparable if, for any pair of them, the chooser either prefers the first to the second, the second to the first, or is indifferent between them.[8]

[7] Putting strict preference and indifference together yields i's *weak preference relation*, R_i, so that "$x\ R_i y$" means that Mr. i either strictly prefers x to y or is indifferent between them. In words, "x is at least as good as y according to Mr. i's preferences."

[8] Equivalently, the alternatives are comparable if, for any pair of alternatives like x and y, either $x\ R_i y$ or $y\ R_i x$ or both. In words, a person has complete preferences if either x is at least as good as y, or y is at least as good as x, or both (that is, each is as good as the other). If the latter, then $x\ I_i y$.

Property 2: Transitivity. The strict preference relation is said to be *transitive* if, for any three possible alternatives (say, x, y, and z), if xP_iy and yP_iz, then xP_iz. That is, if Mr. i strictly prefers x to y and y to z, then he strictly prefers x to z. Likewise, the indifference relation is transitive if xI_iy and yI_iz imply xI_iz (if i is indifferent between x and y and between y and z, then he is indifferent between x and z, too).[9]

As Case 2.1 makes clear, Senator McCaskill possessed complete and transitive preferences over the alternatives $\{x, y, z\}$. She preferred a Senate seat (x) to the governorship (y); a term as governor (y) to being out of politics altogether (z); and, of course, the Senate seat (x) to the political wilderness (z).

If i's preferences satisfy comparability and transitivity, then i is said to possess a *preference ordering.* As noted, the rational choice is the alternative at the top of the ordering. Note that P_i and I_i are exactly like $>$ (greater than) and $=$ (equal to), respectively, as applied to real numbers. For real numbers x and y, either $x > y$, or $y > x$, or $x = y$; hence, they are comparable. Similarly, for any three numbers, x, y, and z, if $x > y$ and $y > z$, then $x > z$; and if $x = y$ and $y = z$, then $x = z$; hence the relations are transitive. In consequence, real numbers can be ordered in terms of magnitude. (The reader can check for himself or herself that the weak preference relation, R_i, is analogous to $\geq$ [greater than or equal to] as applied to real numbers.)

This is all pretty simple. Preferences that permit rational choices are, in effect, *ordering principles.* They are personal—P_i is Mr. i's particular way of ordering alternatives, which may differ from P_j, Ms. j's way of ordering the alternatives. They allow comparisons of alternatives a pair at a time (compara-

[9] Finally, the weak preference relation is transitive if $x\ R_iy$ and $y\ R_iz$ imply $x\ R_iz$.

bility). And the comparison they permit are internally consistent (transitivity).

Before concluding that all is well and moving on, however, we must satisfy ourselves about exactly what we are assuming. We need to ask if *all* relations satisfy properties 1 and 2. If so, then we haven't made very hard demands at all. If not, then we need to know precisely what is excluded from consideration by our assumptions.

In fact, not all relations are complete or transitive (or both). Some relations satisfy transitivity but not completeness.[10] Others satisfy completeness but not transitivity.[11] And still others satisfy neither.[12]

[10] The relation "is the brother of " applied to the set of all males satisfies transitivity but not completeness. It violates completeness since neither "John is the brother of Bob" nor "Bob is the brother of John" is true if they are not brothers! However, if John is the brother of Bob, and Bob is the brother of Charles, then John and Charles are also brothers, so the relation is transitive.

[11] Suppose Ms. i prefers Bill Clinton to George H. W. Bush ($C \, P_i \, B$), Bush to Ross Perot ($B \, P_i \, P$), and Perot to Clinton ($P \, P_i \, C$) in the 1992 presidential election. The alternatives clearly satisfy comparability, but they violate transitivity. You may think Ms. i quite daffy in this case, but we know people like her and expect you may, too. For example, whenever i thinks about the Clinton-Bush comparison, domestic policy issues are triggered in her mind ("It's the economy, stupid!" was the Clinton campaign war chant in 1992, after all), and she prefers the Democratic candidate on these issues. Whenever she thinks about the Bush-Perot comparison, foreign policy issues loom large and she worries about the ship of state in the hands of a businessman with no diplomatic or political experience, like Perot. Finally, whenever Ms. i makes the Perot-Clinton comparison, she can't help thinking about the character issue on which the businessman with no political skeletons in his closet dominates someone who has been nothing but a politician his entire adult life. Intransitivity or inconsistency may arise when different criteria are used for different pairings. When this happens, it is not possible to order all three alternatives in terms of preference. Ms. i ranks Clinton ahead of Bush, Bush ahead of Perot, and Perot ahead of Clinton.

[12] The relation "is the father of " satisfies neither completeness nor transitivity. Suppose we take the population of males and draw two at random. It is entirely possible, indeed highly probable, that neither one is the father of the other; thus, not all pairs of alternatives are comparable according to this relation. On the other hand, even if, for three selected males, the first is the father of the second and the second is the father of the third, it is obvious to any five-year-old that the first is not the father of the third, but rather is the *grandfather*. That eliminates transitivity.

So I have actually said something of substance when I assume properties 1 and 2. The issue now is whether I can defend what's been said. Regarding comparability, clearly you could push things far enough so that making a comparison in terms of preferences would be absurd. *Sophie's Choice* is author William Styron's literary invention for this absurdity. In his novel, a concentration camp prisoner in Poland is permitted to save one of her two children from the gas chamber, but she must choose whom to save; in the absence of a choice, both will die. It is a horrible, inhuman choice. Nevertheless, Sophie does indeed choose (for not to do so is far worse), even though she does not regard her children as comparable. Horrible choices may be painful to make, and some of us may ultimately lack the courage to do what we must. But, as in Sophie's case, even the failure to choose is a choice with its own consequences.

The real problem for the comparability property comes in situations in which the comparison doesn't make sense to the chooser. If objects do not connect up in the mind of the chooser as competing alternatives, then you are likely to get shrugs, puzzled looks, and, if given the option, a response of "don't know." If pollsters, in late 2007 or early 2008, were to have asked a random sample of voters whether they preferred John McCain or Barack Obama in the 2008 presidential contest, they would have obtained many don't-know responses, for that far in advance of an election, most candidate pairings really don't connect in the mind of the average voter. This is not a critique of rationality-based models so much as a warning label advising appropriate use. Choices must have meaning to the choosers if they are to be guided by principled considerations such as those associated with rationality.[13]

[13] Although peripheral to the main line of argument, it is nevertheless interesting to ask what it means for someone to say "don't know" when confronted by a pollster with one of these puzzling choices. It could either mean "this comparison is loony and I cannot make a choice," or "the alternatives

Transitivity requires that the chooser not be confused in a different sense. It requires consistency, something in short supply at times. Psychology professors have, since time immemorial, imposed saline-solution and shades-of-gray experiments on captive sophomores in introductory psychology classes. The typical experiment begins with ten bottles of water of varying salt content (or ten pictures of a triangle colored white, black, or some shade of gray). A student is asked, when presented with two bottles (or triangles), which tastes saltier (or is darker). Her answer is recorded and then a different pair is presented. This continues for some time as alternative pairs are presented and answers recorded (there are forty-five distinct pairs). Because the different saline solutions shade into one another (as do the gray triangles), invariably the student, sometime during the experiment, answers that bottle 2 is saltier than bottle 9, that bottle 9 is saltier than bottle 7, but that bottle 7 is saltier than bottle 2—a clear violation of transitivity. It is hard to be consistent in the manner property 2 requires when the comparisons are so difficult, when there are potentially consequential random events for which the experimenter does not control (such as how thirsty the subject is or how much sunlight is coming into the room), when so little is at stake, and when the answers of a particular subject aren't likely to make much difference.

This, too, is less a critique of rationality than a warning about the domain over which it is likely to be more or less relevant and useful. When the stakes are low, uncertainty is high, and individual choices are of little consequence to the chooser, then inconsistencies are likely to be common. Behavior is likely to be more random than rational, more arbitrary

---

are so close in my mind in terms of preference that it is a matter of indifference to me." Which one it is is a judgment call that the researcher needs to make. In terms of predicting behavior, however, it may not make any difference. Whether a person is indifferent or confused, if a choice is forced, his or her behavior is likely to be random.

than principled. But when the choices *matter* to the chooser, he or she is likely to be more intent on being consistent. As in the case of comparability, whether transitivity is appropriate or not is a judgment call to be made by the investigator. The kind of consistency required by this property is demanding, to be sure, even in more significant situations. But we need it to get on with our business and must content ourselves with the knowledge that, as in other sciences, simplifying assumptions are necessary in order to make progress.[14]

The Maximization Paradigm

The assumptions of comparability (completeness) and transitivity yield an "ordering principle"—they permit an individual to take a set of objects and place them in an order, from highest to lowest (with ties permitted), that reflects personal tastes and values. Rationality is associated with both this capacity to order *and* an aptitude to choose from the top of the order.

The very existence of a "top" to a preference ordering, and individuals with sufficient sense to choose it if given half the chance, is the reason that most of us working in this tradition think of rationality as consisting of *maximizing* behavior. Individuals in social situations are thought to be seeking some goal, pursuing some objective, aiming to do the best they can according to their own lights. Indeed, instead of describing an individual in terms of his or her preferences, we may write down the principle that led the individual to order alterna-

[14] Transitivity strikes me as an assumption like that of perfectly spherical atomic particles in particle physics, perfectly spherical planets in astronomy, and frictionless planes in mechanics. All were known to be contrary to fact, even as they continued to be used; all nevertheless proved essential to move the science forward; all ultimately were relaxed as later generations of scientists subsequently saw how to strip away the offending parts.

tives as he or she did. We may, in other words, state what it is that the person is seeking to achieve or trying to maximize.

The earlier economic example (Display 2.1), in fact, did this. Consumers are interested in maximizing contentment, producers want to maximize profits, workers want the best division of their time between labor and leisure, and investors want the highest long-term return on investment. In the various political models that are examined in the next three parts of the book, political actors are similarly intent on maximizing. Elected politicians, for example, are interested in maximizing their votes at the next election. Legislators seek to maximize the amount of pork and other policy satisfaction they can deliver to the folks back home. Bureaucrats are interested in maximizing their budgets or their turf. The language in the remainder of this book will often reflect this maximizing perspective.

Environmental Uncertainty and Beliefs

Rational individuals choose from the top of a set well ordered according to preference. In many circumstances, however, the individual doesn't get to choose outcomes directly, but rather chooses an instrument that affects what outcome actually occurs. Claire McCaskill (Case 2.1), for example, could not simply *choose* to be the senator from Missouri in 2006. All she could choose was the option to run for that office. So we should revise our idea of rationality, saying now that a rational individual chooses the instrument or action he or she believes will lead to the best outcome.

I slipped the word "believes" into the reformulated definition of rationality. Just as I was precise about preferences earlier, I need now to be precise about beliefs. A belief is a probability statement relating the effectiveness of a specific action (or instrument) for achieving various outcomes. If an

individual is highly confident that he knows what will happen if he does some particular thing (for example, if I turn the handle the door will open), then he is operating under conditions of *certainty*. An incumbent politician's choice to seek reelection against a "sacrificial" opponent is made under conditions of (near) certainty. If, on the other hand, a person is not so confident that she knows what will happen, but nevertheless has a pretty clear sense of the possibilities and their likelihoods (if I turn the handle, there is a fifty-fifty chance that the door will open or be locked), she is operating under conditions of *risk*. Thus, McCaskill's choice to oppose the incumbent Jim Talent was a gamble made under conditions of risk. Finally, if in the mind of the chooser the relationship between actions and outcomes is so imprecise that it is not possible to assign likelihoods, then she is operating under conditions of *uncertainty*.

To see what is meant by certainty, risk, and uncertainty, consider the following example in which there are three possible outcomes—x, y, and z—and three actions—A, B, and C. Our chooser has preferences over the outcomes; suppose she ranks x first, then y, then z—written xyz, but she must make a choice from among the actions. If she knew for certain that C led to y, that B led to z, and that A led to x, then her decision is one of certainty (and, as the reader can ascertain, a pretty simple one—choose A). If, on the other hand, she knew that A led to a fifty-fifty chance of x or z, that B led to a fifty-fifty chance of y or z, and that C led to an even chance of x, y, or z, then the choice involves risk (and is a bit more complicated). Finally, if she weren't sure how to put probabilities on the odds of various outcomes from specific actions, then she would be uncertain (and, without further analysis, the appropriate choice would be quite illusive).

When there is certainty, rational behavior is pretty apparent: Simply pick the action or instrument that leads to your highest-ranked alternative. When beliefs about action-outcome

relationships are more complex, the principle of rational behavior requires more explanation. You need to assign a numerical value to each outcome, called a *utility number*. The utility numbers for x, y, and z, respectively, are $u(x)$, $u(y)$, and $u(z)$; they reflect the relative value you associate with each outcome. If you like x a whole lot better than y and z, and there is not much difference between the latter two in your mind, then $u(x)$ will be a much larger number than $u(y)$ and $u(z)$, and the latter two numbers will be close in magnitude—for example, $u(x) = 1$, $u(y) = 0.2$, $u(z) = 0$. On the other hand, if x is only barely your first choice, with z trailing badly, then the utility numbers would be on the order of $u(x) = 1$, $u(y) = 0.9$, and $u(z) = 0$.

In effect, we have "quantified" preferences by moving from ordered preference information to numerical preference information. There is nothing magical about the particular numbers we wrote down—they are gauged, by you, to best reflect your relative valuations of the alternatives.[15] Now let's do the same for beliefs. For each action or instrument, we can write down the probability that it will lead to one of the final outcomes. In the example above, action A led to a fifty-fifty chance of x or z; that is, $Pr_A(x) = 1/2$, $Pr_A(y) = 0$, and $Pr_A(z) = 1/2$. The probability numbers must all be between zero and one, and they must add up to one. As you can see, these beliefs about action A effectively make A a *lottery*—one in which y is

[15] It is the relative numerical values, not their absolute values, that convey this kind of information. Consequently, it is typical to "normalize" the utility numbers, setting your most-preferred alternative to a utility value of one, your least-preferred to a value of zero, and intermediate alternatives at utility levels between zero and one. It would have done just as well to set most-preferred and least-preferred alternatives at 100 and 0, respectively, or 1000 and −1000, respectively. The normalization values are arbitrary. We report on all this only for the rare reader who wishes to delve more deeply. A standard, accessible reference for further details is Howard Raiffa, *Decision Analysis* (Reading, Mass.: Addison-Wesley, 1968). Readers will not need very much detail to digest the materials in the remainder of our book, so breathe easy!

an impossibility and x and z are equiprobable. We can write $A = (1/2\ x,\ 0\ y,\ 1/2\ z)$. Each of the other actions is a different lottery over final outcomes.

Making a decision under conditions of risk involves choosing from among alternative lotteries. A rational choice entails choosing the "best" lottery. The rule of rational choice is known as the *Principle of Expected Utility*. It provides a method for assigning a single number to each action-lottery and then choosing the one with the largest number. The expected utility of action A of the previous few paragraphs is

$$EU(A) = Pr_A(x) \cdot u(x) + Pr_A(y) \cdot u(y) + Pr_A(z) \cdot u(z)$$

That is, the expected utility of action A is simply the sum of the utilities of all the outcomes that could result from A, weighted by the likelihood that each outcome will happen. If we make the same calculation for actions B and C, then we have a basis for comparing them. Rationality requires a chooser to select the action that *maximizes expected utility*.

Under conditions of uncertainty, a chooser is sufficiently confused that he or she cannot even figure out the likelihoods of various outcomes associated with each action. Needless to say, it is hard to be rational, however you might define it, if you are utterly confused. It turns out, however, that many people do have hunches about likelihoods that they can associate with various actions. So, if pushed a bit, they can give some quantitative precision to their beliefs. They, too, can be treated as if the expected utility principle covered their behavior.[16]

[16] There are many theories of decision under uncertainty that cover the circumstances in which choosers cannot assign probabilities of outcomes to alternative actions. We do not review them here. Still one of the best presentations of this material is to be found in R. Duncan Luce and Howard Raiffa, *Games and Decisions* (New York: Wiley, 1957), Chapter 13.

Conclusion

I have covered quite a bit of ground in this chapter and the last. But everything can be summarized with a few simply stated ideas. First, our general enterprise is that of explaining social and political events and phenomena. Second, the individual is our basic explanatory building block. Third, because we are interested in prediction and explanation rather than description, we characterize individuals in a very abbreviated form, namely in terms of their preferences and beliefs. Fourth, the individuals in our analysis are rational. This means that they act in accord with their preferences for final outcomes and their beliefs about the effectiveness of various actions available to them. The cause-and-effect relationships between actions and outcomes may be well defined (certainty), probabilistic (risk), or only crudely known (uncertainty). Fifth, acting rationally requires ranking final outcomes, assigning utility numbers to them if necessary, determining the expected utility of actions by weighing outcome utilities by action probabilities, and then selecting the action that has the highest expected utility. Sixth, and perhaps most controversial of all, rational political choices—whether career paths chosen by politicans, candidates chosen by voters, decisions to go to war made by kings or presidents, or something as banal as pizza toppings chosen by a group of friends—all are premised on the same comparability-and-transitivity foundation. Aristotle, Hobbes, Rousseau, and other great political thinkers and philosophers have suggested that there is something special about politics—that the collective choices for a nation, for example, are altogether different from choosing pizza toppings. Perhaps. Indeed, certainly this must be the case. But the process of choosing *rationally* bears characteristic markings in all these contexts and so may be analyzed with the same intellectual framework.

In what follows, this rationality machinery is used repeatedly while keeping technical matters to an absolute minimum. So, having covered the preliminaries, let's move on to the study of groups and their politics.

Problems and Discussion Questions[a]

1. How is rationality defined in this chapter? Answer with reference to both preferences and behavior, and then concoct an example of a violation of each of these aspects of the rational actor model, explaining carefully which assumption has been violated.

2. Rational choice models generally start with a well-defined set of actors ($N = \{1, 2, \ldots, n\}$), a number outcomes over which actors hold preferences ($X = \{x, y, z, \ldots\}$), a set of behaviors or instruments with which to achieve preferred outcomes ($I = \{A, B, C, \ldots\}$), and some rule which links actors' instrumental choices to outcomes (R). For example, each of n voters may vote for A, vote for B, or abstain. A candidate wins if he or she gets more votes than any other. Thus, $N = \{1, 2, \ldots, n\}$, $X = \{A$ wins, B wins, tie$\}$, and $I = \{$vote A, vote B, abstain$\}$. The rule, R, is plurality rule, implying that a vote for A (or B) increases the likelihood A (B, respectively) wins. Give simple characterization of each of these model foundations for the following political actors: campaign contributors, political activists, and candidates. In what forms do these actors confront uncertainty in making their behavioral choices?

[a] In this and succeeding chapters, I provide some problems and discussion questions to elaborate ideas in the chapter and to allow the student to test his or her mastery. Difficult questions are marked with an asterisk.

3. Rational choice is a methodology defined by instrumental action toward a goal, where the goal itself is determined by individual values. Given this definition, is it possible for rational individuals to undertake altruistic acts? Provide an affirmative response that a dyed-in-the-wool rat choicer would offer, as well as the perspective of a critic.

4. Mr. *i* holds the following preferences over outcomes *w*, *x*, *y*, and *z*: *xPw*, *xPy*, *zPx*, *yPz*, *wPy*, and *wPz*. When presented with a choice over any subset of these outcomes (e.g., *x*, *y*, and *z*; or all four outcomes), for which subsets can Mr. *i* identify his most-preferred choice? Do any of those subsets contain a preference intransitivity among all outcomes in the subset? Now consider Ms. *j*, who holds preferences: *xIy*, *xPz*, *xPw*, *yPz*, *yPw*, *wIz*. Answer the same questions as before. What does this exercise suggest about the relationship between transitive preferences and maximizing behavior?

5. In November 2008, a couple of weeks after the election of Barack Obama, Hillary Clinton was offered the job of Secretary of State of the United States. It was generally assumed that she faced the following trade-off: joining the new administration, in perhaps the highest-profile cabinet position, which offered the chance of enhanced prestige and policy-making clout in the executive branch, or continuing in the Senate, an option that promised less power (she would still be only one of a hundred) but greater autonomy. The other wrinkle was that most commentators assumed that taking an administration job would preclude a primary challenge against Barack Obama in 2012, and thus meant giving up on a lifelong dream to be president of the United States. Thus, Hillary Clinton faced three possibilities: Remain in the Congress and not win the presidency in 2012 (*C*), remain in the Congress and win the presidency in 2012 (*P*), or join the administration as secretary of state (*S*). State what you think Hillary Clin-

ton's preference ranking was at that time. If the probability of winning the White House in 2012 if she had remained in the Senate is p, then use an expected utility argument to determine the smallest p that would have induced Clinton to remain in the Senate in order to run in 2012. In your opinion, did Hillary Clinton's decision make sense?

*6. Imagine that you are confronted with two pairs of lotteries over the following three outcomes: x = \$2.5 million, y = \$.5 million, and z = \$0. The first pair pits P_1 against P_2, where $P_1 = (p_1(x),p_1(y),p_1(z)) = (0, 1, 0)$ (i.e., you are certain to win \$500,000) and $P_2 = (p_2(x),p_2(y),p_2(z)) = (.10, .89, .01)$. The second pair is a choice between $P_3 = (p_3(x),p_3(y),p_3(z)) = (0, .11, .89)$ and $P_4 = (p_4(x),p_4(y),p_4(z)) = (.10, 0, .90)$. Empirically, most individuals express a strict preference for P_1 to P_2, and P_4 to P_3. Is this behavior consistent with the theory of expected utility? In order to solve this problem, rephrase each of the expressed opinions in terms of expected utility (e.g., $EU(P_3) = .11u(x) + .89u(z)$) and then use basic operations on the resulting inequalities to see if a contradiction emerges. No knowledge of the actual utility function is necessary to solve this problem.