CHAPTER ONE

Bounded Rationality and Elections

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

—Herbert Simon (1957, p. 198; original emphasis)

One may speak of grand campaign strategy, rationally formulated and executed with precision, but a great deal of campaign management rests on the hunches that guide day-to-day decisions. The lore of politics includes rules of thumb that are supposed to embody the wisdom of political experience as guides to action.

—V. O. Key (1964, p. 468)

An intellectual revolution has occurred in political science: the diffusion of rational choice theories. The study of elections has been one of the most receptive subfields. All of its major components—party competition (Downs 1957), turnout (e.g., Riker and Ordeshook 1968), and voters' choices (Downs's spatial-proximity theory; see Merrill and Grofman 1999)—have been strongly influenced by rational choice models.

We think this has been a salutary development for both the discipline in general and the study of elections in particular. The rational choice program has given political science a much-needed degree of intellectual coherence. This new-found coherence connects subfields both by causal claims—we can now more easily see the connections between foreign and domestic politics via, e.g., models of interest groups on trade policy (Grossman and Helpmann 1994)—and by giving us ideas that unify previously disconnected subfields—e.g., problems of credible commitment in governmental borrowing (North and Weingast 1989) and in fights over succession (Powell 2004). Rational choice theories have generated some predictions that have stood up rather well to empirical tests: delegation to congressional committees (Krehbiel 1991), macroeconomic effects of partisan elections (Alesina and Rosenthal 1995), bureaucratic independence (Huber and Shipan 2002), fiscal effects of constitutions (Persson and Tabellini 2003), and cabinet formation and stability in parliamentary democracies (Diermeier, Eraslan, and Merlo 2003; Ansolabehere et al. 2005). Some rational choice predictions, however, have been spectacularly falsified—famously so regarding turnout. Nevertheless, these theories have been wrong in interesting ways and so have stimulated much research.

Further, rational choice theorizing is now flourishing in subfields which once had been terra incognita to rigorous theories of decision making: e.g., the study of democratization (Acemoglu and Robinson 2006) and politics in violence-prone systems (Dal Bó, Dal Bó, and Di Tella 2006). All in all, no research program in political science has been more productive.

But nobody is perfect. Not even a research program.[1] The major weakness of the rational choice program is well known: virtually all models in this program assume that human beings are fully rational, and of course we are not. Some of our cognitive constraints are obvious. For example, our attention is sharply limited: we can consciously think about only one topic at a time. Some are more subtle: e.g., we are sometimes sensitive to small differences in how problems are described (framing effects). But their existence is indisputable.[2] And there is also considerable evidence (e.g., Rabin 1998; Gilovich, Griffin, and Kahneman 2002) that these constraints can significantly affect judgment and choice.

Rational choice theorists have tried a variety of responses to these criticisms of bounded rationality. For a long time they tended to be dismissive (famously, Friedman 1953), but as experimental evidence about cognitive constraints accumulated, a certain unease set in.[3] Most scholars working in the rational choice program know about the obvious cognitive constraints, and many have read the critiques of Simon and of Tversky and Kahneman and their coauthors. Indeed, today enough scholars in the home disciplines of the rational choice program, economics and game theory, take bounded

---

[1] We are using the term "research program" in Lakatos's sense (1970): roughly speaking, it is a sequence of theories united by a few core premises, e.g., about the rationality of decision makers. Thus, in Lakatos's view there is a hierarchy of symbolic formulations: a single research program contains multiple theories, and a single (often verbal) theory can generate multiple (often formal) models. Hence, competition between a specific rational choice and a specific bounded rationality model and competition between their parent research programs are not equivalent, though they are related.

[2] Psychologists even have accurate quantitative estimates of certain cognitive constraints: e.g., ordinary untrained working memory can only handle four to seven bits of information before getting overloaded (Miller 1956; Cowan 2000).

[3] See also Green and Shapiro (1994) for a critique of rational choice theories that is not confined to the cognitive foundations of the research program, and see Friedman (1996) for rebuttals.

rationality sufficiently seriously so that new subfields—behavioral economics (Camerer, Loewenstein, and Rabin 2004), behavioral game theory (Camerer 2003), and behavioral finance (Barberis and Thaler 2003)—are now flourishing.[4] (As evidence for this claim, one needs only to sample a few mainstream journals in economics and game theory and count the number of papers presenting behavioral models.) Things have heated up quite a bit in the home disciplines of rational choice theory—more so, it seems, than in political science. This is ironic, given that two of the most important behavioral theorists, Herbert Simon and James March, were trained in political science and, as indicated by the many disciplinary awards they have won, we still claim them as ours. As it is said, colonials can be more royalist than the king.

A change is overdue. The issues raised by the bounded rationality program—the impact of cognitive constraints on behavior—are as pertinent to politics as they are to markets, perhaps even more so. This is evident in the subfield of elections. Indeed, it is in this domain that the rational choice program has encountered one of its most spectacular anomalies: turnout. The problem is well known: as Fiorina put it, "Is turnout the paradox that ate rational choice theory?" (1990, p. 334). Canonical rational choice models of turnout, whether decision-theoretic or game-theoretic, predict very low turnout in equilibrium: if participation were intense, then the chance of being pivotal would be very small, so voting would be suboptimal for most people. Yet, of course, many citizens *do* vote: even in the largest electorates, participation rates are at least 50 percent in national elections. The difference between prediction and observation passes the ocular test: one needs only to eyeball the data to see the anomaly. Of course, as is often the case with anomalies, eminent scholars have tried to solve the problem. The best-known attempts

---

[4] Although it would be interesting to explore the relation between behavioral economics and bounded rationality, we have a tighter focus in this book: to develop a behavioral theory of elections. Thus, two short points must suffice. First, behavioral economics and bounded rationality are similar in significant ways. Both emphasize the cognitive foundations of social science theories; both rely on evidence and theories drawn from psychology. Second, they exhibit some subtle differences. Behavioral economics—especially work based on the heuristics-and-biases approach (Gilovich, Griffin, and Kahneman 2002)—often focuses on how decision makers make mistakes even in simple task environments. (See, however, Gilovich and Griffin (2002) and Griffin and Kahneman (2003) for a different perspective.) In contrast, work in the bounded rationality program is more likely to emphasize the adaptive qualities of human judgment and choice. This is especially true of the fast-and-frugal approach to the study of heuristics; see Gigerenzer and Goldstein (1996) and Gigerenzer (2004). For detailed discussions of these issues, see Samuels, Stich, and Bishop (2002), Samuels, Stich, and Faucher (2004), and Bendor (forthcoming).

(e.g., Riker and Ordeshook 1968) focus on voters' utility functions, positing that the costs of voting are negative either because of an internalized duty to vote or the pleasures of the process. Doubtless there is something to these claims. But as both rational choice modelers and their critics (e.g., Green and Shapiro 1994) have noted, one can "explain" virtually any behavior if one can freely make ad hoc assumptions about agents' utility functions. The victory—the purported solution to the anomaly—then seems hollow. Accordingly, there are craft norms that impose a high burden of proof against such approaches. Hence many scholars, rational choice theorists and others, are dissatisfied by such explanations and believe that a major anomaly persists regarding turnout.

The scientific situation is somewhat different for the two other components of elections. The study of party competition is probably in the best shape, empirically speaking, of the three components. Although the most famous prediction of rational choice models—that in two-party competition the unique equilibrium is for both parties to espouse the median voter's ideal point—has met with empirical difficulties (Levitt 1996; Stokes 1999; Ansolabehere, Snyder, and Stewart 2001), the gap between prediction and evidence is much smaller than it is in turnout. Moreover, the rational choice program has generated quite a few models in which the parties differ in equilibrium (Wittman 1983; Calvert 1985; Roemer 2001). Further, the Downsian tradition has been remarkably fruitful in the study of party competition. Even scholars (e.g., Wittman 1983) who develop models based on different premises[5] acknowledge the impact of Downs's formulation. The study of party competition clearly owes a great deal to *An Electoral Theory of Democracy* and other work in that tradition.

Rational choice models of voters' decision making are in between turnout and party competition. On the one hand, there's no 800-pound gorilla of an anomaly dominating the picture. But there is a sharp tension between the premises of most rational choice models of voting and the empirical findings of political psychologists. The former typically presume that voters have coherent ideologies in their heads and know a lot about politics: e.g., they know where

[5] Calvert-Wittman–type models presume that candidates are not merely seeking office but also have policy preferences, just as ordinary citizens do. Since one of Downs's central ideas is that parties compete in order to win office, this is a nontrivial departure. For a thorough analysis of the differences between models of opportunist versus ideological candidates, see Roemer (2001). See also the literature on citizen candidate models, e.g., Osborne and Slivinski (1996) and Besley and Coate (1997).

candidates stand in the (commonly constructed) ideological space[6] or at least have unbiased estimates of these positions[7]—claims that are vigorously disputed by scholars studying voter behavior (e.g., Delli Carpini and Keeter 1996; Kinder 1998).

Thus, rational choice theories of elections exhibit a mixed scientific picture: a big anomaly regarding turnout, a qualified success regarding party competition, and some serious issues about voters' decision making.

For the most part, political scientists have criticized rational choice electoral models only on empirical grounds. Although verisimilitude is tremendously important, the failure to construct alternative formulations has allowed rational choice scholars to use the defense "you can't beat something with nothing" (e.g., Shepsle 1996, p. 217). This defense has some merit: it describes a sociopsychological tendency of scholars and arguably makes sense as a normative decision rule. Our goal is to facilitate debate about theories by providing such an alternative formulation.

But because bounded rationality is a research program, it contains a set of alternative formulations, not a single theory or model. Indeed, the program now offers quite a few approaches that address a wide array of topics (Conlisk 1996; Rabin 1998; Mullainathan and Thaler 2000; Camerer 2003). To situate our approach in this collective endeavor, we briefly discuss two major topics: framing and heuristics (e.g., satisficing). As we will see, both topics are central to our theory.

1.1 Framing and Representations

A decision maker's *frame* is his or her mental representation of the choice problem he or she faces.[8] Tversky and Kahneman (1986) pioneered the study of framing in behavioral decision theory. In their work, framing has mainly been associated with just one approach:

---

[6] For an early pointed criticism of the Downsian assumption that voters in an electorate share the same mental model of campaigns and locate parties in this homogeneous cognitive construction, see Stokes (1963).

[7] And because processing probabilistic information is cognitively more difficult than processing deterministic data, models which assume that voters know party platforms only probabilistically trade greater realism in one respect (what voters know) in exchange for less realism in another (how they process information).

[8] In the cognitive sciences the term "representation" is much more common than the term "frame." This terminological difference may have inhibited theoretical unification—a pity, given the surprisingly weak connections between behavioral decision theory and cognitive psychology (Weber and Johnson 2009).

Prospect Theory. But cognitive psychologists use the notion of representation much more widely: "Virtually all theories about cognition are based on hypotheses that posit mental representations as carriers of information" (Markman and Dietrich 2000, p. 138–139; see also Stufflebeam 1999, p. 636–637). Indeed, in standard computational theories of mind, thinking is seen as operations performed on a sequence of representations (Billman 1999; Tversky 2005). In particular, a computational theory—as opposed to an "as if" formulation (Friedman 1953)—of optimal choice posits that a decision maker constructs a mental representation of her choice problem, which includes her feasible alternatives and their payoffs, and executes an operation of value maximization on this representation.[9]

Prospect Theory assumes that decision makers represent choice problems in a way that differs sharply from the representation implied by a computational version of classical decision theory. Whereas the latter assumes that alternatives and their payoffs are compared only to each other, the former posits that agents compare alternatives to a *reference point*—an agent's internal standard. (Most applications of Prospect Theory presume that the reference point is the decision maker's status quo endowment. However, a close reading of Kahneman and Tversky (1979) reveals that this is not part of the theory's axiomatic core; it is an auxiliary hypothesis.) This difference in hypothesized mental representations is fundamental: indeed, Prospect Theory's other two hypotheses about preferences—that people are risk-averse regarding gains and risk seeking regarding losses and that they are loss-averse—would not make sense without the first axiom and its central concept of a reference point.

More generally, one of Tversky and Kahneman's main findings, that people often violate the classical principle of *descriptive invariance*, follows almost immediately from the centrality of mental representations in most theories of information processing in cognitive psychology. It would be astonishing if agents covered by this class of theories satisfied descriptive invariance. These formulations (e.g.,

---

[9] It is no accident that classical utility theory posits (in effect) that alternatives are represented by a preference ordering. This representation makes the critical operation—select the optimal option—relatively easy. As cognitive psychologists have argued for a few decades, a specific type of representation facilitates certain operations while hindering others (Novick and Bassok 2005). And effects on mental operations can impact behavior. In particular, there is strong experimental evidence that the representation of options—whether "multiple options are presented simultaneously and evaluated comparatively, or … options are presented in isolation and evaluated separately" (Hsee et al. 1999, p. 576)—significantly influences choice behavior. Because pure retrospective voting involves separate evaluation, whereas classical Downsian voting involves joint evaluations, this finding bears directly on the study of elections.

Simon 1999) usually presume that people solve problems by transforming one representation (e.g., the initial state) into another one (e.g., the goal state) by a sequence of operations. Although there are many computational theories of mind which allow for many different kinds of representations (Markman 1999; Markman and Dietrich 2000), this perspective is not vacuous: in particular, any theory in this class assumes that people perform operations on representations. Hence it follows, for example, that all else equal, the more operations that are required in order to solve a problem, the more time it takes to do the job (Tversky 2005). This point is familiar to us in our capacities as teachers: when we write up exam questions, we know that we can vary a problem's difficulty by describing it in different ways, so that solving it requires different numbers of operations.[10] Thus, such theories imply that for humans the representation of 492 × 137 is *not* cognitively equivalent to the representation of 67,404, even though the former implies the latter, and both of these are significantly different from the Latin numeral representation $\overline{\text{LXVII}}$CDIV.

In contrast, an agent who is logically omniscient (Stalnaker 1999) would immediately grasp *all* the information implied by a representation. Hence, such an entity would not be subject to framing effects. Of course, positing that any human is logically omniscient directly contradicts the principle of bounded rationality articulated in the Simon quote that began this chapter.

Prospect Theory is usually discussed as an alternative to rational choice modeling, but it is worthwhile pausing for a moment in order to note three ways in which Prospect Theory and classical decision theory overlap. First, both are forward-looking: e.g., in Prospect Theory, it is *anticipated* payoffs that are compared to the agent's reference point. Second, choices based on reference points involve value maximization. By now, however, this should cause no confusion. Maximization is an operation in the context of a representation. As framing experiments (Kahneman and Tversky 2000, passim) have repeatedly shown, if two decision makers use sufficiently different representations, their behavior will differ in some choice contexts even if they are using similar operations, i.e., both are maximizing some kind of objective function. Third, Prospect Theory assumes that the agent is following an *algorithm* that completely specifies

---

[10] The now-classic experiment by Shepard and Metzler (1971) beautifully demonstrated this property. It showed that the time it takes subjects to figure out whether a pair of three-dimensional objects, depicted by pictures, are equivalent is linear *in the degrees of angular rotation required to make the two pictures look the same.*

his or her behavior for the specified choice problem. This use of a complete plan is another feature shared by classical decision theory.

So, while the idea of framing is alien to the rational choice program, the concept of selecting complete plans of action by solving a forward-looking maximization problem is common, even though what is maximized, the objective function, differs between the approaches. In this sense Prospect Theory is quite similar to rational choice formulations. But not all behavioral theories presume that agents have complete action plans in their minds, especially when they confront complex problems. This takes us to the second topic: the notion of heuristics.

1.2 Heuristics

Although there is no unique agreed-upon definition in the cognitive sciences, a *heuristic* is generally presumed to have at least one of the following properties. (1) In general, heuristics are not complete plans: in all but the simplest problems they do not specify what to do under all possible contingencies.[11] (2) There is no guarantee that a heuristic will find a solution. Instead, heuristics promise only that they will increase the chance of discovering a solution (Pólya 1945). The two properties are related: if a plan is silent about what action to take if a certain contingency arises, then obviously it cannot guarantee optimality. (For example, the dramatic collapse of the economy in the middle of the 2008 presidential race made both parties improvise; John McCain's surprising move—reinjecting himself into the Senate's debate on bailouts—was widely seen as a blunder.) As we will see in chapter 2, however, incompleteness is only one cause of suboptimality.

More recently, it has been argued that heuristic reasoning is often rapid and automatic, per dual-process theories of mind (Stanovich and West 2000; see, however, Moshman 2000 and Evans 2008). This claim is now prominent in the heuristic-and-biases approach pioneered by Kahneman and Tversky (e.g., Kahneman and Frederick 2002). But automaticity is not universally regarded as a defining feature of heuristics. Some psychologists who study heuristics and biases argue that some rules of thumb can be deliberately used (Gilovich and Griffin 2002; Kahneman 2002). Scholars adhering to

---

[11] A *set* of heuristics may form a complete plan in somewhat complex situations. For example, in the repeated Prisoner's Dilemma the heuristics of niceness (don't defect first) and reciprocity (do today what your partner did yesterday) together are a complete strategy: tit for tat (Axelrod 1984).

the fast-and-frugal approach to heuristics not infrequently see adaptive rules as consciously deployed (Gigerenzer and Goldstein 1996, p. 666; Gigerenzer 2004). And when heuristics are socially transmitted, whether in professional communities, e.g., of mathematicians (Pólya 1945) or computer scientists (Silver 2004), or within lay communities, then these rules of thumb are obviously explicit. (For example, the common saying "If it ain't broke, don't fix it" is an explicit—and quite accurate—statement of satisficing.) Hence, in this volume we do not require heuristics to be implicit, though we do not exclude this possibility.

The most famous heuristic in political science is satisficing. In the original context of search problems (Simon 1955, 1956), satisficing was a stopping rule: terminate search once an acceptable solution is found. As is evident from the statement of the rule, the notion of an acceptable alternative is crucial. Simon's stipulation was straightforward: an option is acceptable if its payoff is at least as good as the agent's *aspiration level* (Lewin et al., 1944). In turn, an aspiration level is a quality threshold or cutoff which partitions options into two mutually exclusive subsets: satisfactory versus unsatisfactory alternatives. Thus, it is evident that the satisficing heuristic shares an important aspect of Prospect Theory's premise about the representation of choice problems: each posits quality or payoff thresholds that partition options into two subsets—essentially, good ones and bad ones.[12] The similarity is so strong that the terms "aspiration level" and "reference point" appear to be synonyms (Bendor, forthcoming). The only significant difference arises from their original use in different kinds of choice contexts: static for Prospect Theory and dynamic for satisficing.

In general, satisficing is not a complete plan. The reason is simple. Although the heuristic pins down what to do if the solution in hand is satisfactory (keep it), it says little about what to do if the option in hand isn't okay, beyond "look for new ones." This injunction may suffice for simple problems—e.g., new options are tickets in an urn and search simply involves randomly drawing from the urn—but in realistically complicated choice contexts "look for new ones" is merely the start of a complex process of policy design (Kingdon 1984).

There is one important class of problems for which satisficing constitutes a complete plan: the agent has only two options. In such situations, dissatisfaction with the current option can lead in only one direction: the decision maker must take up the other alternative.

[12] Note, however, that under satisficing, agents do not solve maximization problems, unlike Prospect Theory.

Because these problems—now called two-armed bandits—are simple enough to help us gain valuable intuition about satisficing and related forms of imperfectly rational adaptation, we will use them as a running example throughout the book.

Outside of this class, however, positing that a decision maker satisfices typically does not yield precise predictions when outcomes are unsatisfactory. Obviously, this is methodologically undesirable. Hence, in order to make our predictions more precise, in models in which agents have more than two options, we usually add auxiliary assumptions about how a dissatisfied agent searches.

In sum, satisficing is an often incomplete heuristic that encodes two key properties: (1) a choice representation in which an aspiration level is central; (2) an operation or decision rule—keep (or keep doing) satisfactory alternatives—that follows naturally from the representation. By generalizing property (2), we construct the class of heuristics that form the core of all the models in this book. The generalization is based on the observation that in the context of repeated choices, satisficing is simply the extreme case of psychological learning theory's *Law of Effect* (Thorndike 1898). This law—probably "the most important principle in learning theory" (Hilgard and Bower 1966, p. 481)—states that organisms tend to become more disposed to try alternatives that generate positive feedback and less likely to try those associated with negative feedback.[13]

The satisficing heuristic is a special case, as it assumes that the reaction to positive feedback is deterministic: if one tries an option today and finds it satisfactory, then one will try it again tomorrow

[13] Originally, many learning theorists in psychology thought that reinforcements could be identified objectively; observers would not need to make claims about subjects' mental states. This view was important to early learning theorists because it was integral to behaviorism, the dominant methodological posture of the time. But the search for objectively defined reinforcers failed: too much evidence conflicted with the idea (Flaherty 1996, pp. 11–16). Hilgard and Bower explain how the thinking of learning theorists evolved: "... in studies of learned performance, a given reward for a response may have either an incremental or decremental effect upon performance depending on what reward the subject expects or on the range of alternative rewards the subject has been receiving in similar contexts. If a person is expecting a one cent payoff, getting ten cents is going to be positively rewarding; if he is expecting a dollar payoff, the ten cents is frustrating and may have the effect of a punishment. Effects such as these have been observed with animals as well as men. . . . They can all be interpreted in terms of Helson's concept of adaptation level. The rewards obtained over the past trials in a given context determine, by some averaging process, an internal standard or norm called the adaptation level. *Each new reward is evaluated in relation to this adaptation level, having a positive influence on behavior if it is above the norm, a negative influence if it is below*" (1966, p. 486; emphasis added).

with certainty. Since relaxing an assumption is an armchair method of increasing the empirical plausibility of a theory, whenever possible we use the more general assumption described by the Law of Effect. (Sometimes we fall back on the specific case of satisficing in order to make a model more tractable.)

Note also how satisficing and related types of adaptation relate to the literature on framing and mental representations. We take from this literature the importance of two points: one general; the other, specific. The general point is that our theories of choice should take decision makers' problem representations into account; the specific, that reference points are often a significant part of these representations. Thus, because the notion of reference point is central to Prospect Theory, there are important similarities between that approach and formulations based on satisficing. However, we must remember that there are also significant differences: in particular, in contrast to Prospect Theory, agents in satisficing theories do not solve maximization problems; they adapt their behavior according the Law of Effect. Moreover, although their adjustment is shaped by reference points, these internal standards themselves adapt to experience in ways that are closer in spirit to Simon's original formulation than they are to canonical Prospect Theory. It is now time to examine this family of adaptive rules more closely.

1.2.1 Aspiration-based Adaptation

As we define them, *aspiration-based adaptive rules* (ABARs) have two key components: a representational property and an operation linked closely to that mental representation. The representational feature is a threshold, variously called an aspiration level or a reference point, that partitions all possible payoffs into two subsets: good and bad. Since we are examining mental representations, an agent's threshold is not directly observable. Instead, we infer its existence based on observed behavior: e.g., preference reversals (Slovic and Lichtenstein 1983; Hsee et al. 1999).[14]

The procedural property, which is probabilistic, is based on the Law of Effect: the organism compares payoffs to its reference

[14] That aspiration levels are not directly observable would be regarded as a methodological defect by behaviorists, who thought that theoretical concepts were unscientific. Today, however, virtually all social and cognitive scientists freely use beliefs, preferences, attitudes, and other unobservable mentalistic constructs in their theories.

point and becomes more inclined to try alternatives associated with good payoffs and less likely to try those associated with bad ones. Together, these properties identify a form of trial-and-error learning.[15]

Chapter 2 gives a precise mathematical definition of these properties and the corresponding class of adaptive rules. For now the main point to keep in mind is that both properties identify *classes* of representations and operations, respectively. Hence, our subsequent results about ABARs include many adaptive rules in addition to satisficing. Many of these rules are considered by cognitive scientists to be general heuristics that can be deployed in a wide array of situations—including, importantly, ill-defined problems. However, domain-general heuristics tend to be weak procedures (Newell 1969) that can often fall well short of optimality, even when they constitute complete plans.[16]

For the most part we focus on pure ABARs, i.e., those in which decision makers compare the payoffs of alternatives only to their reference points. It is possible to construct coherent hybrid rules, which compare alternatives to each other as well to aspirations. Indeed, one such rule already exists: Prospect Theory's value maximization procedure, which operates on a representation built on a reference point, makes it a hybrid. In a few places we show that our results are robust, i.e., hold for hybrid rules.

1.3 Aspiration-based Adaptation and Bounded Rationality

ABARs do not exhaust the set of bounded rationality theories. They are just one family of theories in the broader research program.[17] Needless to say, we believe that it is a particularly important class of theories.

That said, to ensure that readers grasp the difference between ABARs and the larger research program, we now spell out what we see as the heart of the latter, and how this programmatic core relates to ABARs. Our starting point is the sentence from Simon's *Mod-*

[15] We use the terms "adaptation" and "learning" interchangeably throughout this book.

[16] For an in-depth examination of the performance profile of the satisficing heuristic, see Bendor, Kumar, and Siegel (2009).

[17] An even more egregious error is to conflate *one* member of this family, satisficing, with the entire bounded rationality program. This is, unfortunately, a common mistake.

els of Man that opened this chapter. We unpack his statement into three claims or programmatic premises (Bendor 2001). First, humans are cognitively constrained in many ways, e.g., there are limits on short-term memory and on attention. Second, these mental properties significantly affect decision making. Third, the impact of these information-processing constraints increases with the difficulty of the choice problem that a person faces.

These premises can help us understand the relation between the research programs of bounded rationality and rational choice. To use Simon's metaphor, together "the structure of task environments and the computational capabilities of the actor" act as "a scissors [with] two blades" (1990, p. 7): theories of bounded rationality have cutting power, compared to theories of complete rationality, only when both blades operate (Bendor 2003, p. 435). Cognitive constraints "show through" only when a problem is sufficiently difficult (Simon 1996). If a problem is simple enough so that a person can easily maximize expected utility, then we expect this to happen. Consider, for example, a wealthy citizen repeatedly facing a choice between two parties in a general election. Suppose further that the citizen cares only about which party offers him the lower top tax rate, nothing else. In this case voting is easy and citizens are very likely to vote optimally given their objectives.

If a reader thinks that even such problems might be too difficult for some citizens, consider the proverbial $20 bill lying on the sidewalk right in front of a pedestrian. The natural problem representation and the corresponding operation—"There is $20 on the sidewalk. Either I bend over and pick it up or I don't bother"—together imply that the pedestrian will maximize expected utility.

Thus, the claim that people will optimize when the choice problem is sufficiently easy is very plausible. We might disagree about what *is* sufficiently easy, just as might a group of teachers who are designing a test that should contain problems simple enough for even weak students to solve. [Hence, figuring out what makes certain problems hard for most humans (e.g., Kotovsky, Hayes, and Simon 1985) is an important topic in the research program of bounded rationality.] But the general point is clear.

Since theories of aspiration-based adaptation belong to the bounded rationality research program, it follows that the same point holds for ABARs: when choice problems are sufficiently easy, we do not expect people to use these rules or their associated mental representations. Indeed, thinking about cognitive constraints can help us generate hypotheses about situations in which aspirations will *not* be part of an agent's activated problem representation.[18] Because constructing new representations is effortful and because relevant cognitive resources are constrained, people sometimes accept problem representations offered by others: politicians, advertisers, experimenters, and the like.

How one person influences another person's choice representation might be viewed as manipulation—some critics of framing studies adopt this perspective—but some choice representations may be offered to decision makers quite innocuously yet can still have an impact. Further, that impact may vary based on the context of the choice problem. Consider a diner who enters a restaurant. A typical diner perusing a long menu in a fancy restaurant might entertain a wide variety of options and might consider issues of nutrition, taste, recent dining history, and so on. This is a hard decision; hence all diners often use heuristics (I'll have the usual, I'll have what she's having, I always order seafood in a restaurant on the coast, etc.) to deal with this problem. A waiter recommending both the New York steak and the sole Florentine provides useful information that may inform the diner's heuristics but likely nothing more. For example, a diner that uses some version of an ABAR may usually select the same dish as long as it is prepared to his satisfaction but may occasionally experiment with a new dish.

Now consider a diner who needs to make a theater performance an hour later. For that diner the choice problem is much simpler: appeasing her hunger as quickly as possible is her overriding goal; all other considerations are secondary. When the waiter tells her about, say, the steak and the sole—the only two dishes which can be prepared within her time frame—this diner is perfectly willing to act on the waiter's recommendations, as doing so simplifies her choice problem: she no longer needs to create a problem representation—one is provided by the waiter—and most options have been eliminated to boot! For this much-simplified problem, heuristics are

---

[18] If aspirations are not part of an evoked representation, it does not follow that the activated representation will lend itself naturally to an operation of value maximization. However, scholars who study problem solving do offer a related (ordinal) hypothesis. As noted, many aspiration-based heuristics are general—hence weak—rules. Newell (1969) and Simon (1999) have argued that people use such heuristics when they lack more powerful domain-specific knowledge and problem-solving methods or when they do not recognize the problem as belonging to such a domain. But if a decision maker is an expert in the given domain, she or he may have recourse to a domain-specific heuristic that is more powerful than a domain-general ABAR. Nevertheless, if the problem is of chess like complexity, it may not be known whether the powerful domain-specific heuristic is optimal (Silver 2008), and its associated representation may not be the canonical one of classical decision theory.

unnecessary. She can easily maximize her utility by explicitly comparing steak versus fish. Note that, given this problem representation, our diner will probably be able to rank-order the steak versus the sole in terms of her preferences (suppose she prefers a lighter dish before attending the theater)—exactly the procedure mandated by classical decision theory.[19] Once again we see Simon's scissors at work: the cutting power of theories of bounded rationality depend on the relation between the difficulty of the choice problem and the decision maker's cognitive capacities.[20]

Note, however, that this dinner example involves *myopic* utility maximization. As represented, the problem turns on an immediate choice between two entrees; implicitly, the consequences are limited to the immediate future, ignoring considerations such as the long-term health effects of diet. If the future were seriously engaged, as in the choice of a mate or a career or whether to challenge a well-entrenched incumbent, the problem would be much harder.

In sum, people can represent complex problems in simple ways. This process, called "editing" by March (1994), is a major way of cutting a complicated situation down to cognitively manageable proportions. Because this point is both important and somewhat subtle, it is worthwhile examining it carefully in contexts that are substantively relevant to this book. We will look at voters first and politicians second.

---

[19] Whether this is analogous to a voter's choice between two candidates is an interesting but open question. But Sniderman (2000) is surely right in asserting that choice in a two-party system is cognitively simpler than it is in a multiparty one.

[20] For this reason some behavioral decision theorists (e.g., Bettman, Luce, and Payne 1998, p. 191) have argued that people facing a complex choice might initially screen alternatives by testing them against a set of independent constraints. This process is brutal and swift partly because it is *noncompensatory*: if an alternative does not satisfy a constraint (e.g., does the job candidate have a master's degree in statistics?), then reject it out of hand. Trade-offs are not considered at this point. Having winnowed the number of options down to a manageable number, the decision makers would then engage in the more challenging task of comparing alternatives against each other, weighing their pluses and minuses. This involves *compensatory* decision rules, including classical ones such as maximizing utility in multiattribute choice problems. Hence, behavioral decision theorists are arguing that the same person, in the same (extended) choice process, may in one phase compare alternatives to an aspiration level and later compare them directly to each other. Although this book focuses on the former process, we believe that most of our results would continue to hold if our agents used both choice modes, i.e., compared options both to an aspiration point and directly to each other. (To buttress this claim, we show that two propositions—2.3 and 5.7—are robust in exactly this sense.)

1.3.1 Voters and Turnout

Scholarly debates on the choice-theoretic foundations of turnout often focus on whether voters are rational. This, in turn, typically turns into a discussion of whether they are optimizing in light of anticipated consequences of their actions. This, however, overlooks a key difference *within* the rational choice program between decision-theoretic (e.g., Riker and Ordeshook 1968) and game-theoretic formulations (e.g., Feddersen and Pesendorfer 1996). Under both types of theories citizens maximize utility, but they make significantly different assumptions about people's mental representations.

In standard decision theories (of two-party races) such as Riker and Ordeshook's, a focal citizen estimates a probability of being pivotal and votes if and only if that probability times the value of his or her preferred party winning exceeds his or her private cost of voting. This is not a very difficult problem, and we suspect that it is within the cognitive capacities of most adults. Perhaps the most difficult part is estimating the probability of being pivotal, but in the United States and many other affluent democracies, polls usually provide a reasonable approximation. And once a citizen has estimates of the three parameters, applying the optimization rule itself is straightforward, involving only one multiplication and then a comparison of two magnitudes.

The cognitive operations are easy because the initial representation has a crucial feature: *the probability of being pivotal is treated as an exogenously fixed parameter*. But as game theorists have recognized for some time, if citizens were fully rational and this was common knowledge,[21] then pivot probabilities would be endogenously determined, i.e., in equilibrium (e.g., Palfrey and Rosenthal 1983, 1985). Hence in order to implement the internal logic of the rational choice research program, most scholars working inside this program have turned to game-theoretic models (Feddersen 2004).

The significance of this shift, ignored by most critics of rational choice theories of voting, is huge. As noted, decision-theoretic models assume that voters use a mental representation that imposes quite modest cognitive demands. In contrast, in a game-theoretic model a focal citizen represents the situation as involving thousands or even millions of rational peers who are simultaneously making participation decisions. Hence, here optimization entails the choice of an alternative that is only conditionally best: it depends on the actions of other agents who are simultaneously trying to solve

[21] In a two-person game, a property of player A is common knowledge if player B knows it, A knows B knows it, and so on ad infinitum.

the same complicated problem. To paraphrase Converse, the game-theoretic representation is "orders of magnitudes of orders of magnitudes" (Converse 1990, p. 373) harder than the decision-theoretic one. Indeed, the variation in the difficulty of these problem representations probably swamps the variation in voters' sophistication: we believe that almost no one, apart from a handful of brilliant game theorists, can use the game-theoretic representation to think about the situation; still less can they use it as a basis for optimization.[22] Note that, from the perspective of the rational choice program, the game-theoretic representation is "correct"; it is necessitated by that program's premises. But, outside the rational choice framework, one may very well want to posit that the actual representation in voters' minds may be closer to the simple decision-theoretic frame which would simplify the decision-making problem significantly. Which representation to use would then be an empirical question.

Similar issues occur in the debate over voters' motivations. This is normally discussed in terms of whether citizens are purely egoistic or partly altruistic. Although this issue is obviously important, it is orthogonal to the competition between the research programs of rational choice and bounded rationality. As in the case of purely self-interested voters, moral motivations can be represented in very different ways, some much more cognitively demanding than others. As some moral philosophers (e.g., Braybrooke 2004) have pointed out, utilitarianism is mentally demanding, a claim that has been supported experimentally (Greene et al. 2008). In contrast, certain kinds of nonconsequentialist moral considerations can greatly simplify the choice problem. For example, if voters simply follow a strict moral norm of participating in elections, perhaps out of a sense of duty, then one needn't worry about the chance that one will be pivotal.[23] Similarly, the cognitively simple decision-theoretic framework presented by Riker and Ordeshook can

---

[22] Indeed, many rational choice theorists (see the contributions in Friedman 1996) argue that game-theoretic models make no claims about how agents optimize. They are to be interpreted "as if" and judged only by their empirical implications. A key problem for this approach is that many of these empirical implications are not well supported by the data, whether at the aggregate level of, e.g., turnout levels or in laboratory studies, e.g., in the context of framing effects or the inability of the subject to reason probabilistically. (A second issue concerns the explanatory power of theories given an "as if" interpretation. This, however, involves an unsettled debate—between instrumentalist and realist philosophies of science—that we cannot examine here.)

[23] Students of modernization have long understood that obeying unconditional religious rules because they are prescribed by a sacred text is cognitively simple. Indeed, the critique of "the oversocialized conception of man" (Wrong 1961) was partly a rejection of the Parsonian view of people as rule-following automata.

be easily extended by adding the (in)famous "d-term" representing the utility component of doing one's duty. Feddersen and Sandroni (2006), on the other hand, provide a (nonstandard) game-theoretic treatment of the duty to vote derived from a rule-utilitarian framework. Their analysis reveals that if citizens were game-theoretically rational, doing one's civic duty involves significant complexities.[24] More generally, whether a person's mental representation of turnout is morally charged has no direct bearing on the fundamental differences between the two big research programs considered here. Both moral and selfish motivations are consistent with both simple and complex representations.

Thus, the main claim of the bounded rationality program is not that people satisfice instead of optimize. Rather, it is that cognitive constraints will bind *somewhere* in choice processes when tasks are sufficiently complex—in mental representations if not in operations on those representations (Simon 1979a, p. 498; Bendor 2003).

1.3.2 Voters Okay, but Politicians…?

Because most voters are political amateurs, hypothesizing that they use adaptive heuristics such as satisficing is quite plausible. Candidates and their staffs, on the other hand, are professionals. Many have been politically active for decades. And whereas amateurs may satisfice, professionals optimize. Or so it is believed.[25]

But the slogan "Amateurs satisfice; professionals optimize" reflects the same serious misunderstanding of the core notion of bounded rationality discussed above.[26] We follow Simon in stressing that bounded rationality is a *relation* between a decision maker's mental abilities and the complexity of the problem she or he faces. It is *not* a claim about the brilliance or stupidity of human beings, independent of their task environments. It is common to miss this central

---

[24] In particular, Feddersen and Sandroni observe that, if going to the polls involves any costs—e.g., driving or lost time—a duty to vote has a strange property: it is inefficient and a fortiori inconsistent with utilitarian criteria. This holds because any election outcome can be reached by either exactly one citizen voting (producing a winning candidate) or exactly two doing so (one from each faction, generating a tie). Hence, mass turnout wastes effort.

[25] This is not just casual hallway talk. It is a standard claim made in economics: see, e.g., Rabin's description of the conventional wisdom (1998, p. 31). Sometimes the claim, appropriately and carefully stated, has been supported by data (Feng and Seasholes 2005); sometimes not. [See Barberis and Thaler (2003) for a review of the findings in financial economics.] Presently this is an open question and fundamentally an empirical one. To our knowledge it hasn't been subjected to the kind of careful empirical scrutiny applied to equivalent behavior of, say, investors.

[26] The rest of this paragraph is from Bendor (2003).

point and to reify the notion of bounded rationality into an assertion about the absolute capacities of human beings, whether novices or experts. The fundamental notion is cognitive limits, and as is true of any constraint, cognitive constraints matter—affect behavior—only if they bind. *And per Simon's scissors, whether they bind depends vitally on the demands placed on decision makers by the problem at hand.* Thus, any analysis that purports to fall into this research program yet focuses only on the agent's properties is incomplete.

Hence, whether a specialist optimizes—more precisely, whether cognitive constraints bind—depends vitally on the difficulty of the task at hand. Of course, all else equal, we expect a specialist to outperform an amateur. (If that usually didn't happen, then we would suspect that the profession in question was bogus.) But "all else equal" includes problem difficulty. If the task facing a professional is much harder than that facing an amateur, the former might be just as cognitively constrained as the latter. In other words, while children quickly learn to play tic-tac-toe optimally, even grandmasters do not play chess optimally (Simon and Schaeffer 1992).

Running effective campaigns in large jurisdictions is a hard problem. It is not merely a one-shot selection of a point (i.e., a platform) in a low-dimensional space, as Key's description makes clear.

> A presidential campaign … may be conducted in accord with a broad strategy or plan of action. That general plan may fix the principal propaganda themes to be emphasized in the campaign, define the chief targets within the electorate, schedule the peak output of effort, and set other broad features of the campaign. The strategic scheme then provides a framework to guide the detailed work of the party propagandists, the labors of the speech writers, the decisions of those who parcel out the campaign funds, the schedulers of the itineraries of the principal orators, and the day-to-day endeavors of all the subordinate units of campaign organization.
>
> *Often the outlines of a campaign strategy are scarcely visible amidst the confusion of the campaign* and, indeed, campaigns often rest on only the sketchiest of plans. The preparation of a reasoned and comprehensive strategy requires more of a disposition to think through the campaign in its broad outlines than often exists around a national headquarters. Once the plan is made, its execution requires organization sufficiently articulated to respond to general direction in accord with the plan, a requisite that is not always met. And even when a campaign is blueprinted in advance, *a flexibility must be built into it to take advantage of the breaks and to meet unexpected moves by the opposition.*
>
> (Key 1964, pp. 462–463; emphasis added)

Key was arguing that complete plans of actions—strategies, in the game-theoretic sense—*do not exist in major campaigns*. The task is far too complex and filled with too many uncertainties.[27] Campaigns are of chess like complexity—worse, probably: instead of a fixed board, campaigns are fought out on stages that can change over time, and new players can enter the game.[28] Hence, cognitive constraints (e.g., the inability to look far down the decision tree, to anticipate your opponent's response to your response to their response to your new ad) inevitably matter. Professionals are not immune to Simon's scissors.[29]

Thus, political campaigns, like military ones, are filled with trial and error. A theme is tried, goes badly (or seems to), and is dropped. The staff hurries to find a new one, which seems to work initially and then weakens. A third is tried, then a fourth. (Recall, for example, the many themes in Robert Dole's 1996 presidential race.) In short, there are good reasons for believing that the basic properties of experiential learning—becoming more likely to use something that has worked in the past and less likely to repeat something that failed—hold in political campaigns.[30]

In sum, we have much stronger theoretical commitments to the claim that cognitive constraints are causally important in many significant political situations than we do to the hypothesis that people adjust via aspiration-based adaptation. The first claim is a fundamental premise in the bounded rationality research program. The latter, though an important theoretical position in this program, is

[27] Key's point applies to some relatively well-bounded events that are only a part of campaigns, such as nomination battles. It is unlikely, for example, that many strategists at the Democratic convention of 1924 thought about what they should do if the 102nd ballot were reached without picking a winner. (John Davis was selected on the 103rd.) Nobody thinks that far ahead: not in 1924; not now.

[28] For example, in the 1996 Israeli elections, four suicide bombings at a crucial point in the campaign pushed Benjamin Netanyahu past the incumbent, Shimon Peres.

[29] Indeed, if institutions sort problems sensibly, giving the hardest ones to the top specialists, then the raw performance scores of the best professionals or organizations may be worse than those of less skilled agents or organizations. (For an example of this problem regarding the evaluation of hospitals, see Dranove et al. 2003).

[30] In our formal model of party competition (chapter 3), we make the conventional assumption that a campaign strategy is simply a platform: no more, no less. But *everything* is scaled down in these models: both the cognitive sophistication of the decision makers as well as the complexity of their problems. From the perspective of the bounded rationality program, a key question regarding a model's plausibility is whether the *relation* between the agents' mental capacities and the difficulty of the problems they face is reasonable. The absolute level of either parameter is less important (Bendor 2003).

not one of its fundamental premises. As the Simon-March point on problem-editing makes clear, what is central to the program is a claim that humans use *some* method of reducing a complex problem to a cognitively manageable one. Aspiration-based choice is one way of doing this; it is not the only way. In this book, of course, we are betting that this particular family of theories is a fruitful way to go.

1.4 Plan of This Book

Chapter 2 will discuss aspiration-based theories of adaptation in some detail. It will provide some general properties of ABARs that will be useful for specific models of voting and elections. Chapters 3–7, the heart of the book, present models of elections. The approach is modular: first we present partial models that focus intensely on specific topics. This one-at-a-time approach also enables us to generate some analytical solutions to the models, which helps us to understand the guts of aspiration-based adaptation in the context at hand.

The first partial model, in chapter 3, takes up the classic issue of Downsian party competition. In this model, incumbents do not change policy positions; only challengers search for alternatives. We find conditions under which sets of policies are ruled out by this process but also that platform convergence can only occur under some special circumstances. The second model, in chapter 4, deals with voter participation. Now campaign platforms are fixed and everything turns on electoral participation: whichever side mobilizes more voters wins the election. Contrary to the well-known "paradox of turnout" raised by game-theoretic models of turnout, our model consistently generates realistically high levels of turnout. Moreover, this model produces comparative statics that are intuitive, consistent with those of game-theoretic models, and empirically supported. Finally, chapter 5 brackets both party platform locations and turnout and considers the voter's choice between candidates. We find here, inter alia, that using simple retrospective voting rules, citizens can generate endogenous partisan affiliations. This creates ideological polarization when aggregated over the entire population.

Chapter 6 then assembles these modules into one large model of elections. This synthesizes the constituent models into one complex formulation in which everything is at play: citizens must decide whether to turn out, and if they do, whom to vote for, while parties figure out what platforms to espouse. The complexity of this synthetic model obliges us to turn to computation as the main way to generate results (predictions). We do this exclusively for pragmatic reasons: if a model is too complicated to solve by hand, computing is better than giving up. This integrated model yields a "general equilibrium" of the election game. Many of the elements of the partial models appear here: winning parties adopt relatively centrist platforms, and citizens vote in significant numbers. One consequence of the constantly shifting party platforms is that voters choose the ideologically more distant party with surprising frequency.

All of the above is confined to two-party races. Starting there is sensible, but ending there would not be; there are too many multiparty systems to do that. Thus, chapter 7 extends the model to multiparty democracies. This raises a host of interesting questions that do not arise in two-party races, such as the need for voters to coordinate on a preferred candidate. Voters in this environment often face the problem of coordinating their behavior in order to prevent their least-preferred candidate from winning. This has given rise to an extensive literature on Duverger's Law. We show that the model leads to the selection of Condorcet winners yet allows significant vote shares for all candidates. We also find that our model does a good job of accounting for the partial coordination seen in election data from the United Kingdom.

Chapter 8 summarizes our major findings and provides some concluding thoughts.