Behavioral Rationality and the Policy Processes: Toward A New Model of Organizational Information Processing

BRYAN D. JONES, GRAEME BOUSHEY, AND SAMUEL WORKMAN

INTRODUCTION

Behind any study of public policy processes is a theory of organizations. Policy is made by organizations, but organizations are made up of interacting human decision-makers. As a consequence, any theory of organizations harbors a theory of individual choice.

This chapter considers the analysis of organizations from the competing perspectives of rational choice theory and bounded rationality. Our essay is divided into four parts. The first section surveys the broad characteristics of these behavioral models in the study of public policy. Although both begin with a common goal of connecting individual choice to macropolitical outcomes, they disagree fundamentally over how individual behavior should be understood. Rational choice theory claims analytical rigor and theoretical parsimony, holding that individuals behave as if they were pure utility maximizers to deduce patterns of outputs from social systems (Friedman 1953). There is no scientific analysis to support this assumption; it is simply an assumption, presumably validated by its deductive power. Bounded rationality insists that any model of choice be based in scientific analysis of the cognitive architecture of humans, even if that scientific analysis implies less parsimony.

In the second part, we survey recent empirical assessments of behavioral rationality—how people actually behave in experimental and observational situations where comprehensive rationality makes precise predictions about outcomes. Findings across disciplines of decision making not only point to the theoretical inadequacies of the rational choice framework, but also provide the important tools for constructing more realistic models of human choice.

The third part of this essay examines policy implementation and administration in light of studies of principal-agent dilemmas in organizations to highlight key differences between rational choice theory and bounded rationality. While rational choice theory has evolved to provide a more accurate portrayal of human behavior and preferences (Ostrom 1999a; Ostrom 1999b; Levi 1997), research in this paradigm nonetheless continues to overemphasize
the problems of delegation and control in bureaucracy. The obsession with control is symptomatic of the larger pathologies of rational choice theory—the core assumptions of individual utility maximization simply do not capture the complexities of human decision-making. By focusing narrowly on questions generated from the control trap, rational choice theory ignores a wide range of intriguing questions about information processing and policy dynamics in bureaucratic decision-making.

In the final part of this essay, we point toward an information-processing model of the policy process that avoids the obsession with control that permeates rational theories of policymaking. We argue that it is both more scientifically sound and fundamentally more parsimonious than models based on rational choice.

**PART I: BOUNDED RATIONALITY AND RATIONAL CHOICE THEORY: AN OVERVIEW**

In the behavioral sciences, such as psychology and behavioral biology, the major topic of inquiry is individual behavior. In the social sciences, such as political science and economics, the aim is to understand social systems of interacting individuals. Social scientists need both a model of individual behavior and an understanding of how individuals interact to produce social outcomes, such as are produced from free markets or from policymaking processes in government. As a consequence, social scientists cannot dally with the nuances of human behavior that do not impact on the interactions of people in organizations. Any model of individual choice in policymaking processes must be parsimonious, and they must link individual level actions to policymaking outcomes.

The behavioral models of comprehensive rationality and bounded rationality, which today divide the theories of public policy, were driven by a common desire to improve the rigor of political analysis. Applications of both rational choice theory and bounded rationality in the study of public policy and administration evolved from the efforts of such diverse theorists as Herbert Simon, Mancur Olson, and Anthony Downs to link individual human decision-making with broader macropolitical outcomes (Simon 1947; Olson 1982; Downs 1957). Although the competing models of rationality disagreed about the fundamental motivations behind individual choice, they were unified in their belief that the processes and policy outcomes be most powerfully understood through exploring the role of individual behavior in collective decision-making. A theory of rationality which anticipated individual choices in the context of the larger political process would present a significant advancement in the study of government—such a set of assumptions would act as a theoretical tool-box that could be used, not only to cut through the complexity of public decision-making, but also to locate areas of conflict within organizations, and potentially predict future political outcomes. Both rational choice theory and bounded rationality pushed the study of public policy away from case studies and atheoretical descriptions of public administration toward a generalized theory of public policy.

Despite these common aims, the theories of rationality are deeply divided over the most basic assumptions of individual choice. Rational choice theory borrows heavily from economic assumptions of individual preferences, and believes that a sufficient behavioral model could be drawn from deductions of an individual’s self interested utility maximization. Bounded rationality began as a critique of comprehensive rationality, and grew from an effort to reconcile the reductionist economic assumptions of rational choice with observed psychological constraints on human decision-making. The following sections outline the distinguishing characteristics of both rational choice theory and bounded rationality as applied to the study of public policy. This critique intends not only to cast light on the major differences between the two theories, but also calls attention to exciting empirical findings which we believe are fueling increasing theoretical convergence around a single positive theory of human choice.
COMPREHENSIVE RATIONALITY AND THE THEORY OF POLICY PROCESSES

In recent decades rational choice theory has been widely applied to the study of government. This dominance extends in no small part from the model's theoretical parsimony. Rational choice theorists have explored some of the most complex aspects of politics by making relatively few core assumptions about individual behavior. As in neoclassical economic models, the theory of comprehensive rationality posits that political decision-makers are self-interested utility maximizers who hold stable preferences and objectives, and make strategic decisions to maximize the personal benefits of a given choice. To understand politics at the aggregate level, researchers need only to understand ordered preferences of individual actors who populate a specific institution or political sphere and the formal rules by which these fixed preferences are combined. In this approach, preferences + rules = policy outcomes.

Important distinctions can be drawn between rational choice models in public policy, political science, and economics; however, all rational choice models share common characteristics. First, decision-makers hold stable ranked and ordered preferences for outcomes. Given three possible alternatives – options A, B, and C, a rational chooser will form clear preferences between each of the three given alternatives. These preferences are transitive, meaning that if an individual prefers option A to B, and prefers option B to C, he also prefers option A to C. Second, a decision-maker possesses necessary information to connect choices to outcomes. With this information, individuals then optimize when making decisions – they make strategic choices in order to achieve their most preferred outcome. “Thick” rationality adds the assumption of individual self-interest in determining preferences. For example – preferences for a higher tax rate to fund public education over a lower overall tax rate suggest that an individual will receive greater personal utility for increased funding in public education than from lower taxes. Regardless of the underlying reasons for this preference, rational choice scholars reduce it to a question of self-interest. Their behaviors reflect only their effort to maximize the utility of their choice.

Even the staunchest proponents of rational choice theory regard this characterization of the model as more of an ideal type than a realistic portrayal of human behavior. Decisions made under conditions of complete certainty – when specific strategic choices are known to lead to explicit outcomes – are rarely found in political or economic life. More powerful insights have evolved from rational choice theory when theorists have examined decision-making under risk – where a strategy might lead to several different outcomes with known probability, or conditions of uncertainty – where outcomes are known but the probabilities associated with those outcomes are not, and must be estimated.

Rational choice theorists use expected utility theory to approximate how individuals make calculations that rank alternatives by their expected value (under risk or uncertainty) rather than their known value. Under conditions of risk, individuals form strategic preferences probabilistically – they compare the probability that their most preferred outcome will occur against the probability that their less preferred outcome will occur, and both against the cost of making a decision.

It is not enough that individuals calculate the probabilities that outcomes will occur if they take a particular action; they must also calculate the likelihood that their choice will yield the outcome against the probability that their choice will have no bearing on the outcome. Individuals must not only predict the probability that an event will occur, but also the chance that an event would occur without their participation, or that their less preferred outcome would occur in spite of their participation. A classic example is rational voter models, where people are predicted to vote only if their expectation that their vote will make a difference exceeds the marginal cost of voting.

Comprehensive rationality holds great appeal as a model of choice for three basic reasons. First, rational choice promises a parsimonious method for studying complex political
behavior. By assuming that actors are singularly motivated to maximize gains available in political decisions, a wide range of considerations that have long complicated the study of political science—social class, partisanship, or cultural values—become peripheral to a decision-maker’s preferences. The challenge of rational choice theory is to characterize accurately the payoffs available to political actors in a given arena. A classic example of this approach can be drawn from David Mayhew’s study of congressional voting behavior in *Congress: the Electoral Connection*. In his analysis, democratic representation is motivated singularly by a congressman’s desire to retain political power (Mayhew 1974). Ideology and policy-making matter only inasmuch as they instruct a politician as to where his source of political power is located, and how to make appropriate votes to retain electoral support.

The second advantage of the rational choice method is its broad theoretical generalizability. Unlike descriptive studies of political behavior, the baseline assumptions of rational choice methods are characterized by a ‘universalism that reveals generalizable implications beyond those under immediate investigation’ (Levi 1997, 20). Deductions from rational choice theory are not specific to a particular time and case, but should offer insights wherever similar conditions can be observed. Mancur Olson argues that ‘the persuasiveness of a theory depends not only on how many facts are explained, but also how diverse are the kinds of facts explained’. In Olson’s view, the rational choice approach has the double benefit of preserving the parsimony of a theory by removing ‘any inessential premises or complexities that ought to be removed from an argument’, while retaining a high degree of explanatory power (Olson 1982: 12–13). Drawing upon a relatively small set of assumptions about human behavior, Olson is able to advance a series of predictions that serve as the basis of a theory of how the formation of interest groups in democratic societies detract economic growth. That democratic nations have sustained high levels of growth, even as interest groups have proliferated, means that Olson’s theory was wrong, but that is not the point. It was clearly and cleanly stated in a generalized manner that facilitated testing against empirical observation.

This leads to the third advantage. The rational choice method is a comparably rigorous approach to the study of political processes. Rational choice researchers derive formal mathematical models from a set of assumptions about individual preferences. These models derive a set of hypothesis of anticipating a specific outcome in politics, and then test them against the self interested behavior of actors put forward in the theoretical assumptions of rational choice theory—in this regard, rational choice hypothesis are verified or falsified. This approach forces researchers to advance well-specified models of political choice. Rational choice models focus only on those most crucial elements of the political process that are necessary to explain outcomes.

Both the simplicity and the promise of theoretical generalizability have made rational choice theory a popular tool for the study of government. The rational choice approach has been applied to such subfields as organizational behavior (Bendor and Moe 1985; Moe 1984) congress (Mayhew 1974; Arnold 1990), parties and elections (Downs 1967) and collective action problems (Olson 1982). The breadth of this research is suggestive of the theoretical power of the model. By exploring the strategic behavior of self interested individuals, rational choice theorists have produced a rich and theoretically unified body of research in a discipline once marked by methodological eclecticism.

**BOUNDED RATIONALITY AND THE THEORY OF POLICY PROCESSES**

Perhaps because rational choice theory takes a decidedly reductionist approach to the study of government—one which is ‘willing to sacrifice nuance for generalizability, detail for logic’ (Levi 1997: 21)—dissenting researchers have long charged that these sparse assumptions of individual utility maximization distort
the complexity of both individual behavior and organizational decision-making. Field researchers complain that observations of individual behavior rarely match the calculating self-interested actor posited in rational choice theory (Brehm and Gates 1997; Lipsky 1980). The behavioral norm of individual utility maximization simply does not seem to reflect accurately the actions of politicians, bureaucrats and voters, whose choices so often seem to be motivated by risk aversion, sense of mission, identity, fairness, or altruism. For these critics, rational choice theory is at best idiosyncratic—applicable only to discrete institutions such as the US congress where political behavior can be safely matched to self-interest (Rockman 2000). At worst rational choice theory is misleading, as it reduces potentially interesting social behavior such as altruism to an individual’s self interest (Monroe 1996).

Researchers who actually study individual decision making find the rational choice approach curious. Cognitive psychologists find the economic assumptions of stable ordered preferences, transitivity, and utility maximization to be a strange abstraction of human decision-making. Findings in psychology indicate that people are poor at forming preferences, generating alternatives, and making decisions (Sniderman et al. 1991; Tetlock 2000; Jones 2001). Studies show that individuals often lack even the most basic tools with which to make informed rational decisions. Preferences and choices are bounded by emotion and environmental context than by rational analysis (Jones 2001). Research in cognitive psychology makes one wonder how people are able to form preferences or make decisions at all.

Herbert Simon (1947) developed bounded rationality as effort to reconcile the strict economic assumptions of comprehensive rationality with actual decision-making revealed by the empirical study of organizations. Preferences and choices seemed bounded by cognitive and emotional constraints that interfered with the process of purely rational decision-making. Much like comprehensive rationality, bounded rationality offers an efficient method for moving between individual decisions and organizational outcomes. Bounded rationality retains the hallmarks of a theoretical model—it captures only those aspects of human behavior needed to understand collective decision-making.

As with so many researchers who followed him, Simon noticed that the assumptions of expected utility analysis failed to match his own observations of real world economic decision-making (Simon 1999). Looking at budgeting in Milwaukee, Simon observed that relatively little individual behavior matched ‘substantive or objective rationality, that is, behavior that can be adjudged to be optimally adapted to the situation’ (Simon 1985, 294). He instead found that the processes of both individual and organizational decision-making were a good deal messier than rational expectations would have us believe. Bureaucratic budgets were often adjusted incrementally, using the prior year’s budgeting as a benchmark for future spending needs (Thompson and Green 2001; Simon 1991). Organizational decisions were made more through horse trading and bargaining than a process even remotely resembling fully rational decision-making (Simon 1999). Organizations and individuals proved poor at generating complex alternatives or making trade-offs. Environmental factors, such as issue salience, individual attention, and time constraints, shaped the depth of solution searches in organizations.

Simon (1985) believed the most glaring problem with comprehensive rationality was its focus on outcomes rather than the process of individual decision making. By ignoring the procedures of choice, rational choice theory blithely accepted that ‘rational’ outcome emerged from self-interested behavior. Such an approach is especially dangerous when attempting to understand the broader mechanisms of collective policy-making, where the motivations for preference formation and political behavior are at least as important as the outcome itself. Simon (1985: 294) explains:

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

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

Simon argued that a theoretical model oriented toward understanding procedurally or boundedly rational decision-making would provide a more realistic bridge between individual and collective choice. A ‘behavioral’ model of rationality would take a decidedly scientific and inductive approach to understanding decision-making. In order to make efficient generalizations connecting individual psychological processes to collective political and economic choice, research needed to follow strict scientific guidelines. Underlying theoretical assumptions of human behavior needed to be tested and retested. Those assumptions that were verified through scientific research would be preserved in the scientific model, while others would be modified or abandoned. What bounded rationality lost in parsimony it would gain in accuracy. Unlike the artificial behavioral assumptions of comprehensive rationality, bounded rationality would capture the biological, emotional, and environmental constraints that constrained the procedures of decision-making.

As with rational choice theory, bounded rationality has been widely applied in the study of public policy. While bounded rationalists might emphasize different elements of behavioral model in their research, virtually all research in bounded rationality draws from four core principles: the Principal of Intended Rationality, the Principal of Adaptation, the Principle of Uncertainty and the Principle of Trade-offs (Jones 2003). From these principles, modern researchers have advanced a vigorous research program that explores how both people and institutions behave. As we will see, even though experimental and empirical research have improved our understanding of the behavioral model of choice, these four central tenets remain largely valid.

The Principle of Intended Rationality

The principle of intentionality suggests that we look at the goal directed behavior of people, and investigate the manner in which their cognitive and emotional constitutions either help or hinder their goal-seeking. This distinguishes bounded rationality from psychological theories, which generally focus only on the limitations of individual choice-makers. While comprehensive rationality assumes single-minded maximization, the principle of intended rationality allows researchers to distinguish between careful cost benefit analysis that closely approximates utility maximizing decision-making, quick decisions based on heuristic cues, unthinking reliance on past strategies, or even spontaneous decisions that seem to make no reference to potential gains or losses. Given the time, costs, and demands of a specific decision, humans may rely on hardwired biological responses, generalized decision-making strategies, or full information searches. ‘Cognitive architecture is most obvious when action occurs at short time scales. As one moves toward actions that take longer times, cognitive architecture is less and less evident, and the nature of the task takes on more and more importance in explaining action’ (Jones 2001, 56). Humans may intend to be rational, but their decision-making capabilities break down under time constraints or very high information costs. Moreover, there is no good evidence that people are more rational when the stakes are high, as some rational choice theorists have maintained. In direct defiance of that claim, state lotteries sell more tickets when the pot is large, lowering the probabilities and the expected return.

Principle of Adaptation

The principle of adaptation is closely related to the notion of intentionality. Much of human
behavior is explained by the nature of the 'task environment' surrounding a decision. With time, human decision-making adapts to the specific nature of the problems they face in a specific circumstance. The more time and learning an individual invests in a specific problem, the less constrained they become by environmental or biological constraints. This notion of adaptation accounts changes in decision-making efficiency in a single problem space over time. When a problem is iterated over time, people learn or develop coping strategies. Even more intuitively, the principle of adaptation may explain why organizations encourage specialization in areas of complexity, and routinization in decisions under severe time constraints.

A rich tradition of research in political science and psychology explores the use of heuristics in political decision-making—attempting to identify cost-cutting cues individuals rely upon in order to limit investment costs from making complex decisions in low-information environments. The key issue is whether these heuristics are maximally adaptive (Lupia, McCubbins, Popkin, 2000; Gigerenzer et al. 1999). While some have argued that heuristics invariably follow what a fully rational individual would choose, that is quite clearly not the case. 'Buy a lottery ticket when the stakes are high' is a classic heuristic that leads to lower expected returns than not buying at all, or buying when the stakes are low. It is likely that some heuristics are adaptive calculational crutches and some are misleading and even mal-adaptive.

**The Principle of Uncertainty**

Individuals operate in an environment of almost constant risk and uncertainty. Because of human cognitive architecture, uncertainty is far more fundamental to choice than expected utility theory admits. Not only are individuals unaware of the outcomes that will result from strategic choices, but they are uncertain of the procedures of choice themselves and are even uncertain about their own preferences. Uncertain outcomes may produce dependence on procedures, and may explain instances of extremely conservative risk adverse behavior.

**The Principle of Trade-Offs**

The final notion central to models of bounded rationality is the principle of trade-offs. Unlike comprehensive rationality, which suggests that individuals are able to move seamlessly between ranked goals, bounded rationalists argue that people find it difficult to trade off one goal against another when forming preferences and making choices (Slovak 1990; Tetlock 2000). This critique is a major shift away from the transitivity assumption in rational choice theory; however it is important to capture the volatile shifts in preferences observed by both behavioral economists (Kahneman and Tversky 1986) and public opinion researchers (Zaller 1992). Preferences are determined by emotional and cognitive cues, and are rarely as stable as rational choice theorists have us believe. Because of these trade-off difficulties, Simon argued that individuals, and by extension organizations, 'satisfice,' quickly choosing an option that is 'good enough' rather than searching for one that weighs the payoff of every possible choice.

Because bounded rationality is concerned with the procedures of individual choice, the school of thought tends to approach the study of politics by looking to how individuals and organizations respond to changes in their problem environments. Because individual decision-makers have limited attention for problem solving, they must address problems serially, one-at-a-time, which means they are forever juggling inputs, prioritizing them via the allocation of attention and the sense of urgency that inputs generate. The salience of a particular problem is almost always generated by non-rational elements in politics—by scandal, by crisis, by the mobilization of critics—rather than calm decision to allocate the scarce resource of attentiveness.

In governments, as well as in all organizations, attention is allocated in a process political scientists call *agenda setting*. Organizations also suffer from limited attention spans, and
must process at least major problems serially. Routine problems may be delegated and handled according to rules, but fresh problems cannot be handled this way. It is common for many problems to press forward on the governmental agenda, and whether the most severe problems receive the most attention is an empirical question. In individuals, as well as in organizations, the allocation of decision-making attention is crucial for understanding the immediate and future behavior of an organization.

Because it is explicitly oriented toward understanding the processes of decision-making, bounded rationality has been widely applied to the study of public policy and public administration. Bounded rationalists have explored a range of policy problems, ranging from agenda setting (Kingdon 1995; Baumgartner and Jones 1993), congressional decision-making (Kingdon 1973), federal budgets (Padgett 1980), incrementalism (Wildavsky 1964; Davis, Dempster, Widalsky 1966) and risk aversion (Kahneman and Tversky 2000). As a model of choice, bounded rationality is at least as broadly applicable as comprehensive rationality.

PART II: BEHAVIORAL DECISION THEORY AND THE SHORTCOMINGS OF RATIONALITY

Concerns over competing models of political choice are by no means limited to researchers in political science. Some of the most hostile critiques of the rational choice model have come from pioneering work by behavioral decision theorists, who have struggled to reconcile the assumptions of the rational calculating economic man with research in cognitive psychology that insists that human decision-making is frequently neither optimizing nor rational. The most intriguing contributions from behavioral decision theory have come from efforts to answer two basic questions about decision-making. First—how do individual preferences respond to changes in the framing or emotional stimuli of the problem? Second—how do game theoretic strategies based on rational choice assumptions compare to actual human strategies in real world games?

FRAMING EFFECTS AND PUBLIC POLICY

Studies linking framing effects present an especially intriguing challenge to our understanding of rational political behavior. Framing studies have demonstrated that an individual’s understanding of a policy idea depends heavily on the context and heuristic cues surrounding a policy problem (Garrett 2003; Jones 1994; Kahneman and Tversky 2000). Rather than holding stable opinions about public policies, individual opinions are volatile, and shift according to the framing of the political issue (Zaller 1992). The relationship between a voter and a policy idea is thus at least partially contingent on the context in which the voter is presented the policy choice. By ‘directing attention to one attribute in a complex problem space,’ policy entrepreneurs can produce desired changes to public responses to policy problems (Jones 1994, 104). Public receptivity to public policies seems highly responsive to shifts in the framing of a policy problem.

PROSPECT THEORY AND RISKY CHOICE

Daniel Kahneman and Amos Tversky’s research in prospect theory provides one interesting framework for understanding how framing effects limit rational behavior. As a response to errors they saw in Bernoulli’s expected utility assessment of choice under risk (Kahneman 2002) prospect theory advances a boundedly rational model of risky choice, by focusing on the reference point and framing of decisions rather than the strict utility of ‘final asset positions’ put forward by rational choice theorists (Kahneman 2002, 460). The approach grew from a series of studies in decision-making that demonstrated how individual preferences were shaped by
environmental context. People would hold different preferences for virtually identical decision based on how the problem presented. Kahneman explains:

Preferences appeared to be determined by attitudes to gains and losses, defined relative to a reference point, but Bernoulli’s theory and its successors did not incorporate a reference point. We therefore proposed an alternative theory of risk, in which the carriers of utility are gains and losses—changes of wealth rather than states of wealth. Prospect theory (Kahneman and Tversky, 1979) embraces the idea that preferences are reference dependent. (461–462).

Approaching the study of risky choice from the perspective of changes in wealth rather than overall states of wealth, Kahneman and Tversky arrive at the intuitively appealing notion that humans cope with gains and losses differently—that people ‘are risk adverse in the domain of gains, and risk seeking in the domain of losses’ (Quattrone and Tversky 1988, 723). Prospect theory further demonstrates that people make different choices depending on how decisions are framed—the weights they give in decision-making shift in response to how their attention is directed. ‘An individual would prefer a sure gain of $80 over an 85% chance to win $100. The same individual would prefer a gamble offering an 85% chance of losing $100 to a certain loss of $80’ (Berejikian 2002, 762.) Although the chance to win $100 is the same as the chance of losing the $100, individuals generally make different choices. In the domain of gains people value certain gains over possible gains. In the domain of losses, people will make riskier decisions in the hopes of avoiding any loss. These findings clearly contradict the tenets of rational choice theory, which would assume that a rational chooser would prefer certain gains or certain losses equally.

In a similar vein, Richard Thaler’s research on mental accounting further suggests that individual preferences and expectations are shaped by emotional cues. Basic economic theory suggests that money is fungible, that money ‘has no labels’ and is treated the same way regardless of how it is earned (Thaler 1992, 108). However, Richard Thaler draws a series of hypothetical scenarios to show that the manner in which people receive and then psychologically allocate the use of money has an important bearing on how the money is perceived and spent. ‘The effect on current consumption … of winning the $300 football pool should be the same as having a stock in which I own 100 shares increase by 3 dollars a share, or having the value of my pension increase by $300. The marginal propensity to consume all types of wealth is supposed to be equal’ (Thaler 1992, 108–109).

Yet people do not approach the gain of $300 equally. The $300 won in lottery or betting is seen as a sudden windfall, and might be spent with clear conscience on celebration. A sudden $300 increase in the pension plan will be treated with miserly caution, an essential part of a savings account linked to concerns for future financial stability. Although the amount is the same, people perceive it differently because it was accounted for differently. Money can be thus better thought of as placed in various mental accounts. Money appropriated for one use is spent or saved differently than money appropriated for another (Thaler 1992).

Findings from prospect theory and mental accounts are compelling because they provide empirical evidence that individual decision-making is strongly constrained by framing cues which illicit emotional responses. Unlike the calculating maximizer posited by rational choice theory, ‘in real life, generating maximum utility is neither simple nor smooth, and is affected by the cognitive and emotional constitution of the decision maker’ (Jones 2001, 26). The strategic choices pursued by individuals may not reflect calculations of optimality, but rather their disposition towards loss aversion or risk acceptance, or even their emotional response to a sudden windfall.

EXPERIMENTAL GAME THEORY

Unlike studies of framing effects, which examine how individual preferences shift in response to external emotional cues, experimental game theory began by asking if people behave
in a manner consistent with game theoretic expectations of rational self-interest. To test accurately assumptions of economic choice, individuals must be given clear guidelines and all relevant information with which to form strategic preferences. In spite of this convention, the most popular experimental games have provided mixed evidence at best that individuals behave in a manner consistent with game theoretic assumptions of strategic preferences. Players with full information about the rules and objectives of the game frequently prove themselves to be inefficient in forming game theoretic strategies (Fudenberg and Levine 1997).

By far the most intriguing findings have come from research in the now famous 'divide-the-dollar' games, which captured the imagination of decision-theorists exactly because they have failed to substantiate the most basic hypothesis drawn from expected utility theory.¹ In the simplest form of the game, the ultimatum game, a proposer and a responder are asked to share a sum of money. The proposer must divide the sum however he sees fit with the responder. Because neither player receives any money if the responder rejects the offer, rational choice theory would lead us to predict that the smallest permissible offers would be made and accepted. The proposer would rationally seek to maximize the monetary gains available in game play, while the responder would prefer even an incremental gain over no gain at all.

Yet mounting evidence suggests that individuals rarely if ever play as pure income maximizers. Proposers frequently offer up to half of the sum being divided, while responders routinely reject offers of under one third of the total—even when neither player knows the other, and often simply play against a computer (Nowak, Page, and Sigmund 2000). Even in the dictator form of the divide-the-dollar game, where the responder is required to accept any offer made by the proposer, an impressive 76% of responders chose to divide the money equally (Kahneman, Knetsch, and Thaler 1986).

Ultimatum game experiments not only routinely fail to reproduce purely rational utility maximizing behavior in individuals in the lab, but they also provide strong evidence that human decision-making cannot be fit neatly into economic assumptions of comprehensive rationality. Individual trading behavior in ultimatum games has been shown to be highly context-specific. Repeated play games suggest that learning, social interaction, and norms of reciprocity strongly shape ultimatum trading (Nowak, Page, and Sigmund 2000; Goeree and Holt 2000). A sense of entitlement seems to shape how a proposer or responder plays the game. Players who earn the responsibility of proposer engage in distinctly different trading behavior than those who are simply designated the role of responder (Hoffman and Spitzer 1985).

The brunt of experimental research in divide-the-dollar games has remained narrowly focused, first on revealing the pathologies of comprehensive rationality, and then in revealing altruistic, fair or other 'anomalous' behavior in research subjects. The discussion and conclusion sections of most ultimatum game research almost invariably speculate that evolutionary norms of fairness, cooperation, or rational risk aversion constrain responders from making decidedly one-sided offers. This points to a common complaint of experimental research—it is a powerful critique of one model of rationality, but so far the tradition has done little to generate and test new and scientifically realistic alternatives of choice. Experiments are fantastic devices for testing particular behavioral assumptions, but they are less useful in generating new ones.

Despite these limitations, efforts to replicate assumptions of comprehensive rationality have succeeded in fueling a movement to integrate findings of actual behavior into a more realistic model of human choice. Findings from
cognitive psychology, cultural anthropology, political science, and behavioral economics, demonstrate how considerations outside of strategic rationality influence the procedures of individual decision-making. Experts across disciplines increasingly agree that a successful model of human choice must integrate empirical findings about the effects of culture, attention streams and issue salience, emotional cues, framing effects, and identity and social perspective into a parsimonious model of decision-making. This research agenda has led to renewed interest in established theories of decision making, such as bounded rationality and prospect theory, as well as a wealth of new and modified theories from the rational choice tradition.

RATIONAL RESPONSE: THIN RATIONALITY AND NEW INSTITUTIONALISM

Although a few dogmatic economists and political scientists have continued to conduct research through assumptions of comprehensive rationality, most recent rational choice studies in political science have attempted to integrate empirical findings of individual preferences into renewed models of individual rationality. It is clear that the traditional rational choice assumption of self-interest has failed—that the self-interest axiom cannot explain instances of altruism, ethical restraint, or fairness—has lead an increasing number of theorists to allow maximization for any goal. The analyst refuses to make assumptions about motives for maximization. These steps may complicate the model or limit broad theoretical generalizability; however, the growing evidence documenting inconsistencies between rational choice assumptions and actual human behavior warrants a revision of the theory to include the potential rewards outside of material or status gain. Margaret Levi points out:

The addition of non-egoistic considerations or motivational norms [as community standards or fairness principles] does increase the complexity and difficulty of analysis…. The advantage of an assumption in which actors consider net payoffs that include both material and ethical factors is that it may better approximate reality albeit at the sacrifice of a neater and more parsimonious model. Nonetheless, as the maximand more closely approximates reality, the more reality it should capture’ (Levi 1997, 24–25).

Individual preferences may be modified to capture not only an individual’s desire to maximize gains, but an equally strong ‘ethical’ belief that his behavior harms no other individual. These rational choice models integrate assumptions of fairness, altruism, reciprocity, justice, etc., into the ordered preferences of their rational actor. The individual continues to hold ordered preferences, and continue to make decisions based on whatever will yield their most preferred outcome. The problem, of course, is that trade-offs among these various goals must be explicitly built into the model, and one of the primary findings of behavioral decision theory centers on the difficulties people have with such trade-offs.

Another hallmark of recent research in rational choice theory has been to give greater attention to the institutions in which political actors make choices. Institutional Rational Choice Theory examines how institutions provide specific information about the gains and losses available to actors in a given sphere. One variant institutional rational choice theory is especially concerned with understanding how organizational cultures permit cooperation in places where selfish maximization might prove problematic (Miller 1992; Ostrom 1998). In many situations, rational maximization at the individual level leads to collective disaster—as in Garret Hardin’s (1968) tragedy of the commons, where unlimited common grazing rights lead to destruction of the common pasture illustrates. Through learning or experimentation, organizational rules may develop that encourage collective outcomes that overcome the “tragedy” of individual self-maximization (Ostrom 1999a). Once in place, these rules are enforced through a system of formal and informal incentives that control selfish behavior. This framework provides ‘a general language about how rules, physical and material conditions, and attributes of the community affect the structure of action arenas, the incentives individuals face, and the resulting outcomes’ (Ostrom 1999b, 59).
These institutional and cultural extensions are the most important additions to rational choice theory to be put forward in recent years. Yet, by incorporating so many “soft” variables they move quite briskly away from the traditional approach. How can one predict whether an organizational culture will emerge to overcome rational self-interest? Clearly not always, and hence the institutional structure is dependent on some other context that is left specified in the model.

PART III: PRINCIPAL AGENCY AND ORGANIZATIONAL BEHAVIOR

In spite of improvements, rational choice models simply are insufficient to capture critical aspects of organizational behavior. Because this research remains focused on outcomes rather than processes, rational choice theories are overwhelmingly focused on problems of collective action and individual utility maximization, which provide only limited insights into organizational behavior. Bounded rationality can both examine the roots of collective action problems, but it also can shift research attention toward the more interesting questions of attention allocation and organizational information processing.

Recent studies in public administration have focused a good deal of attention on problems of delegation and control in organizational decision-making. Indeed, virtually all theoretical discussions of policy implementation have centered on these questions to the exclusion of such important topics as adaptive problem-solving and information-processing. While principal-agent dilemmas illuminate some important aspects of bureaucratic behavior, its prominence in studies of public policy is partially an artifact of the rational choice model of behavior. The fundamental problems of principal-agent dilemmas—those of information asymmetries, moral hazards, and incentive structures—are those that map neatly onto the most basic assumptions of individual utility maximization. Pioneering work in principal-agent models have spurred a rich discussion in public management; however we fear the privileged place such research now holds in the study of organizations provides students with a simplistic and misleading view of the dynamics of bureaucratic policy-making.

A more comprehensive understanding of bureaucratic decision-making can be gained through the lens of bounded rationality, which offers a scientifically sound alternative model of behavior while sacrificing little parsimony. Rather than focusing exclusively on control, bounded rationality focuses on how changes in the external environment shape information processing both within and across organizations. This approach to the study of organizations will invariably yield a more sophisticated understanding of bureaucracy, as it is better suited to link the procedures of human choice with broader policy processes.

POLICY IMPLEMENTATION AND AGENCY THEORY

Rational choice decision making as applied to the implementation stage of the policy process generally takes the form of principal-agent models of the bureaucracy’s interactions with the legislative, executive, and judicial branches of government. In this section we argue that the underlying assumptions of the principal-agent model of politico-bureaucratic relations render it inaccurate, not only in terms of adequately explaining the relationship, but also as a description of the relationship and policy processes more generally at the implementation stage.

Principal-agent models of behavior were originally developed in the economic analysis of firms (Alchian and Demsetz 1972); although an emphasis on efficient production and incentive structures in organizations dates back much earlier (Taylor 1911). The principal-agent relationship is essentially a contract. In order to increase production, the principal enters into an agreement with the agent for the production of the good. The principal’s goal is to ensure the efficient production of the good; meanwhile, the agent’s goal is to avoid time,
effort, and resources necessary for producing the good.

For example, if we were examining a business firm, we might consider the employer the principal and the employees as agents. In the model, employees have incentive to "shirk", that is, loaf. This means that managers must closely monitor them to insure efficient production of a given good. No system of monitoring can completely eliminate all free-riding, because the resources for such monitoring face the law of diminishing marginal returns. This is the essence of the principal-agent problem.

Why do principals face the problem of free riding? The answer lies in information asymmetries found in the relationship between the principal and the agent. In the principal-agent model, the agent holds two key informational advantages over the principal. The first informational advantage lies in the selection process. Principals cannot know a priori the exact qualifications of a given agent, while the agent has a better idea of his or her competence. The agent may exaggerate her ability in order to attain the position, and even if most exaggerations are caught by the employer's personnel policies, some incompetents will slip through. This first informational asymmetry is usually termed adverse selection. The second informational advantage held by the agent involves the monitoring of the agent after the contract has been established. Moral hazard is the result when the principal lacks the resources or ability to engage in constant monitoring of the agent. The agent is then able to shirk, reducing the efficiency of the firm. The task of the principal is to induce compliance on the part of the agent through a system of incentives.

In order to invoke the principal-agent framework, the analyst must assume goal incongruence. The goal of the principal is the production of some good, while the goal of the agent is to reap the rewards of the contract with the principal without expending the effort to produce the good. To put it simply, the goal of the agent is to shirk. If goal incongruence does not exist, then principal-agent models are inappropriate for examining the relationship. Underlying the assumption of goal incongruence is a further assumption about the relationship of the actors to their goals. The principal-agent model assumes that a given actor's goals are known. Formal principal-agent modeling has rightly recognized that there exists uncertainty about the goals of the other actors in the relationship (Bendor and Moe 1985; Bendor, Taylor, and Van Gaalen 1985; Bendor, Taylor, and Van Gaalen 1987a; Bendor, Taylor, and Van Gaalen 1987b; Moe 1984), but this does not take account of the uncertainty that shrouds an actor's knowledge of her own goals. Actors may not be able to define their own goals, especially not in contexts in which trade-offs must be made between competing goals.

The model also assumes that strategies for attaining given goals are easily attached to the goals. The principal's goal of efficient production is attained by developing the appropriate incentive structure such that it is to the agent's material disadvantage to shirk rather than work. Likewise, the agent's goal is to maximize the benefit to be gained from the contractual relationship with the principal through manipulation. The final assumption concerning goal oriented behavior is complete information. The model assumes that if the principal is able to observe agent outputs, the principal is able to judge whether or not these outputs conform to the goal of efficient production (Worsham, Eisner, and Ringquist 1997).

Finally, the principal-agent approach makes certain assumptions about the nature of the principal-agent relationship over time. The principal-agent approach is based in comparative statics (Bendor and Moe 1985). Relationships change only when exogenous 'shocks' to a system disturb a system otherwise in equilibrium. It should be noted that this is not entirely a problem of principal-agent theory per se. Comparative statics and equilibrium analysis also has roots in systems theories in both political science and sociology (Eisner, Worsham, and Ringquist 1996; Worsham, Eisner, and Ringquist 1997; Worsham, Rinquist, and Eisner 1998).

Given these assumptions, it is imperative to the success of the contractual relationship that the principal control the behavior of the agent. There is no room for shared goals or learning
and adaptation of behavior over time, or the potential counterproductive results on productivity of an environment of continual distrust and suspicion fostered by monitoring. The theme of control has been central in subsequent scholarship on the implementation stage of the policy process.

Rational choice approaches were used in the study of public policy long before the advent of the now-dominant principal-agent approach, most notably in Niskanen’s (1968; 1971; 1975) model of the budget-maximizing bureaucrat and Anthony Downs’ (1967) study of information limitations and bureaucracies. The principal-agent model really burst onto the scene in political science with works by Mitnick (1980; see also Mitnick 1973) and Miller and Moe (1983a; 1983b). This culminated in an article by Moe (1984), *The New Economics of Organization*, which formally introduced agency approaches to the study of policy implementation to political scientists. Moe insightfully unified principal-agency from the economic analyses of business firms to the issue of how democratically elected officials can control unelected public bureaucrats; hence principal-agency could be used to illuminate a critical problem in democratic theory. Unfortunately, it altered the perspective on bureaucracy from one of balancing problem-solving, information-acquisition, and the advantages and disadvantages of delegation, to one stressing a single-minded emphasis on formal procedures of monitoring and control.

It also reversed the old normative emphasis in public administration on keeping politics out of administration. At least since Woodrow Wilson (1887), political scientists have been concerned with the distinction between politics and administration, mostly from the perspective of trying to keep politicians from interfering with professional implementation by demanding special favors from supposedly neutral bureaucrats. Principal-agency suggests that all independent behavior of bureaucrats is not motivated by professional problem-solving, but is based in the desire to shirk. Principal-agency assumes that all attempts by politicians at control are motivated by “efficient production” rather than crass political gains that add inefficiencies to public production.

Moe (1984) argues that for principal-agent theory to be useful in the study of implementation, we must take account of the fact that politicians are not necessarily motivated by the efficient production of public service; they may be more concerned with political efficiency rather than production efficiency. Moreover, the major problem of control might not be shirking, but could involve several different possibilities, including material benefit of some sort, ranging from budgetary slack to promotion, but might also be policy related. At a minimum, scholars utilizing the principal-agent frameworks would need to grapple with the issues of political efficiency and diversity of goals—not trivial alterations.

**EX ANTE CONTROL OF IMPLEMENTATION**

With the formal introduction of principal-agent theory, work on policy implementation collapsed to analyses of how a political principal is to control a shirking bureaucracy. One group of scholars argues that political control of the bureaucracy is achieved ex ante (McCubbins 1985; McCubbins, Noll, and Weingast 1989). Democratically elected officials gain control of the bureaucracy through setting the ‘structure and process’ of organizations before their actual creation through detailed legislation specifying administrative procedure, personnel, and organizational structure (McCubbins, Noll, and Weingast 1989; Banks and Weingast 1992). In this view, Congress ‘stacks the deck’ against the agency during its creation in order to ensure subservience to Congress at later periods. The deck-stacking thesis focuses on political efficiency—the political coalition in control wishes to continue its advantage through later implementation. Other scholars argued that the political branches of government could also achieve ex post control of the bureaucracy through budgeting, personnel, staffing, political appointments, ongoing interactions with the bureau, and congressional oversight hearings (Bendor and Moe 1985; Bendor, Taylor, and Van Gaalen 1987a; Miller and Moe 1983b).
But is congress the principal in the American system of divided powers? Both congress and the president attempt to influence the bureaucracy (not to mention the judiciary). Some scholars argue that congress is the institution wielding the most influence or control on the bureaucracy (McCubbins 1985; McCubbins, Noll, and Weingast 1989; Weingast and Moran 1983). Other scholars see the president, with his ability to influence the bureaucracy through budgets, OMB, and political appointments, as the abler of the two branches to achieve control and focus their theoretical and empirical attention on him (Golden 2000; Wood 1988; Wood and Waterman 1991).

McCubbins, Noll, and Weingast (1989) put forth three hypotheses concerning agency structure and process and its effects on the subsequent regulatory environment. First, the agency’s design should, ‘create a political environment that mirrors the politics at the time of enactment’ (McCubbins, Noll, and Weingast 1988, 444). That is to say, winners and losers at the formulation stage of the policy process should benefit or lose in the same relative proportions at the implementation stage due to institutional design. Second, agency structure and process will favor those constituencies of the winning coalition at the formulation stage. Finally, agencies will exhibit an ‘autopilot’ characteristic (McCubbins, Noll, and Weingast 1988, 444). As the preferences of the favored constituency change, so will the preferences and thus behavior of the agency.

More recent scholarship is divided on the efficacy of structure and process as a means for controlling the bureaucracy. Balla (1998) finds that the open comment procedures for the new pay schedule and rules regarding Medicare did not favor the groups presumably favored by the dominant coalition in congress. Balla and Wright (2001) did find that advisory commissions correctly represented the political factions involved in the legislation of drinking water, but these authors did not examine the actions of these commissions.

Note that in the development of the principal-agent literature, the descriptive fit of the principal-agent model is somewhat strained. For example, in the deck-stacking thesis there is no room for politicians to prefer neutral competence in administration—the major aim of the Progressive Movement so important in American political life during the first quarter of the 20th Century. To the extent that Progressives were successful (and in many respects they were very successful), they left a legacy of removing political control and the politicized administration that the deck-stacking thesis suggests.

DESCRIPTIVE CRITIQUES

Beginning in the 1990s, scholars began to argue that descriptive flaws led agency theorists to explanations of bureaucratic behavior that were less than adequate (Eisner, Worsham, and Ringquist 1996; Hindera and Young 1998; Krause 1996; Meier, Pennington, and Eller 2001; Potoski and Nemacheck 2001; Worsham, Eisner, and Ringquist 1997). Principal-agent analysis generally has difficulty accounting for the fragmented nature of policy processes in the American system of government. That is, there are multiple, competing principals in the American system of separated powers (Choi 2001). Moreover bureaucracies themselves have constituencies and engage in advocacy policy-wise for these constituencies (Hindera and Young 1998; Meier, Pennington, and Eller 2001; Potoski and Nemacheck 2001). These bureaucratic constituencies could also be considered the principals.

All of these potential principals compete to influence or control the bureaucracy. These competing principals often have multiple, conflicting goals, especially under divided government. Once the political context of multiple, competing principals is taken into account, bureaucratic intransigence becomes much harder to pin down conceptually. To whom should the agency respond? Why should the agency respond to one principal to the exclusion of others? We argue that the more appropriate question to ask is, how do bureaucracies make trade offs among the competing and contradictory demands of their political environments? Answering this question requires that we take
an information processing approach to the study of bureaucratic behavior.

A second descriptive criticism of agency theory involves the goals of the agent. The goals of the agency are assumed to be adverse to those of the principal. Principal-agent modeling has difficulty addressing agents whose goals are not consistently adverse to those of the principal. Recent research at the street level calls into question the assumption that bureaucrats are primarily motivated by the goal of shirking (Brehm and Gates 1997). Krause (1996) argues that the development of goals over time is a two-way process. Agencies influence and even help to refine the goals of political principals. Second, using principal-agent models requires the researcher to acknowledge that both the principals’ and agents’ goals are known and easily prioritized. Bender and Moe (1985) note that the relationship between the bureaucracy and its political principals is likely dynamic, and the goals of the actors are likely to evolve and are clarified over time, both by the interactions of the bureaucracy with the political principals as well as through organizational factors. Worsham, Eisner, and Ringquist (1997) have argued that we cannot assume that bureaucratic agents are motivated to maximize material return to themselves; they may have other goals—some of which may correspond to those of the policymaking branches. This raises several questions concerning the type of behavior we should expect from the bureaucracy.

When confronted with a choice between attaining one of many possible goals, how does an actor choose? We are back to a serious problem in all rational theories of choice: to make any progress, some very strong assumptions must be made about what goals are used in the process of maximization. A bounded rational view of individual decision-making takes account of the difficulties of making trade-offs among multiple and competing goals. Principal-agent theories of policy processes also assume that the strategies, or means for attaining goals, are easily attached to the goals themselves. Although the studies outlined above all note that goals are not easily achieved, the uncertainty that surrounds the means to achieve the goals is rarely modeled.

How do principals know which strategies will lead to the desired outcomes?

Furthermore, it was argued long before principal-agent theory came to dominate implementation studies that success in terms of goals is difficult to measure (Lipsky 1980; Worsham, Eisner, and Ringquist 1997). How do principals know desirable outputs when they see them? Monitoring outputs can be counterproductive as agents shift away from productive problem-solving (as well as shirking) toward the goal that is monitored. This is a classic critique of standardized testing in schools as teachers “teach to the test.” The linkage between strategies and goals cannot be understood outside a dynamic relationship. This suggests that bureaucratic outcomes, rather than outputs, are the more important indicators of goal success, which, in turn, suggests the efficacy of a problem solving approach to implementation.

If bureaucrats’ interests are not necessarily adverse to those of the principal, an emphasis on control as a description of the process of implementation may be misplaced. Golden (2000) finds that bureaucrats in various regulatory agencies in the Reagan administration for the most part were responsive to the policy directions of the administration, even in the absence of formal control procedures. Brehm and Gates (1997) find that bureaucrats are not fundamentally interested in shirking. They argue that functional motivations and peers are the prime influences on intra-organizational behavior to the exclusion of influence by superiors within the organization. Bureaucrats, for the most part, work. Their research adds a strong empirical foundation to arguments advocating taking organizational context seriously (Lipsky 1980; March and Simon 1958; Simon 1947; Wilson 1989).

PART IV: INFORMATION-PROCESSING AND PUBLIC POLICY

How are we to come to grips with the fact that bureaucrats are often predisposed to be responsive, but that policy outcomes are not always exactly as political principals would
have them? If organizational decision-making is our focus, then we should approach the problem by asking how organizations process information from both their political and task environments (see Wilson 1989), and how this information is used in making trade-offs among competing demands and functional requirements (Jones 2001; Jones 2003).

An information-processing approach would analyze the ways in which administrative structure and processes are developed in ways that aid in focusing attention on particular problems and types of information. It would need to rely on bounded rationality as a decisional underpinning, because the cognitive limits of individuals and organizations color how they process information. This approach would view hierarchies or organizational structures as mechanisms for focusing attention (Jones 2001). Normatively, an information-processing approach sees structure and process aimed at enhancing the problem-solving capacity of both agencies and congress more generally, not just enhancing one partner at the expense of the other. Constraining agency decision-making renders the agency less able to adjust both to its task environment and its political environment (Jones 2001; Wilson 1989). Keeping politicians completely out of implementation can cause agencies to cut off an important source of feedback.

It is reasonable to argue that members of congress might well prefer a flexible agency that is adaptable both to its political and task environment; delay may be desirable in order to maintain policy stability (Carpenter 1996). Agencies are rightly cautious in responding too quickly to demands from their environments. Agency delay is a result of the agency reducing the uncertainty in their environment. Once the agency is sure that a given signal is meant to influence policy in a given direction, the agency adjusts accordingly.

This approach calls into question a theory of bureaucratic behavior that fails to consider the uncertainties inherent in the environments of agencies and questions the amount of resources that are devoted to questions of control without actually specifying how agencies go about making decisions—decisions made in environments characterized by ambiguity and uncertainty (Jones 2001; Jones 2003; Krause 2003). An emphasis on discovering the ways in which agencies balance competing, contradictory demands in their environments mandates that scholarly attention be turned toward specifying agency decision-making processes. It also mandates that scholarly attention be turned toward explaining bureaucratic behavior more broadly, rather than one facet of the relationship between bureaucracy and the elected branches of government.

**BOUNDED RATIONALITY AND POLICY IMPLEMENTATION**

If we adopt a problem-solving, information-processing perspective based in bounded rationality, we are equipped to examine both scientific and normative questions in a more productive light than if we begin with a rigid principal-agent model based in rational choice. Because bounded rationality begins with the assumption that both “principals” and “agents” are goal-seeking entities with multiple potential objectives, and assumes that both are fallible in the pursuit of these goals, it leads to normative analyses that neither treat the “professional bureaucrats” as invariably correct, as did the Progressives, nor the elected politicians, as do the proponents of principal-agency. Scientifically, it is much less likely to “get stuck” on one partial and incomplete understanding of legislative-bureaucratic relationships, because it has a much stronger inductive component than does principal-agency or any other models derived from rational choice.

Because bounded rationality begins by looking to the procedures of individual choices, it is well equipped to explore endogenous dynamics of interpersonal and hierarchical relationships at the heart of principal-agent dilemmas. But the approach isn’t limited to the internal dynamics of public administration. Because bounded rationality looks to how decision-makers interact with exogenous environmental changes in time, information flows, and attention—it is also able to understand
dynamic change in the policy outputs of organizations. In this regard, we find it curious that researchers in public administration are calling to expand principal-agent models of bureaucratic behavior (Waterman and Meier 1998; Bendor, Glazer and Hammond 2001).

A scientifically sound and parsimonious alternative approach connecting individual and organizational behavior already exists. To demonstrate the value of bounded rationality in the study of organizations, we describe first how the model might be applied to principal-agent dilemmas and to broader and more interesting questions about bureaucratic information processing. We contend that bounded rationality actually provides a more robust understanding of principal agency. More importantly, we argue that those researchers who study organizational behavior through the lens of bounded rationality look to a greater and more appealing range of research questions than the problems of delegation and control.

**BOUNDED RATIONALITY AND PRINCIPAL-AGENT MODELS**

Recent empirical challenges to the central assumptions of bounded rationality have spurred calls for reform in either the method or the behavioral model employed in studies of interpersonal relationships. In their elegant review *Theories of Delegation*, Bendor, Glazer, and Hamond (2001) conclude with a challenge to improve the study of principal-agent dilemmas, complaining in part that too little attention is given to a) the ‘real world’ institutional context of principal-agent dilemmas (266–267); b) pre-existing information and policy programs that might constrain an actor’s behavior; and c) the complexity of the task environment facing actors in a given environment (267). Waterman and Meier (1998) attack the limitations of standard principle agent dilemmas from another direction, deriving an interesting set of hypotheses by relaxing assumptions of information asymmetries and goal incongruence. They argue ‘all political-bureaucratic relationships are not a caldron of conflict. The environment is dynamic and conflict varies according to both the level of information that principals and agents posses and the level of goal conflict’ (197).

Perhaps the most troubling set of problems identified by critics of principal-agent models have been that they are founded in artificial and even incorrect assumptions of human behavior. Under empirical scrutiny, the assumption of goal incongruence generally, and agent shirking specifically, is less pronounced than principal-agent theory would have us believe. In many organizations, agent behavior is determined more by the organizational culture and feelings of agent solidarity than through incentive structures and principal monitoring (Brehm and Gates 1993). The importance of institutional norms and cultures inspired Gary Miller’s *Managerial Dilemmas*, one of the most exciting innovations in modern organization theory (Miller 1992). Miller’s work challenged the Skinnerian notion that workers in agency were most effectively induced to work through incentives and punishment, arguing instead that an organizational culture is far more effective in establishing worker compliance or non-compliance with principal goals (Miller 1992). Nothing in bounded rationality rejects that incentives and monitoring within organizations can shape individual behavior—and Miller’s interest in organizational culture is entirely consistent with the procedural rationality’s interest in the influence of external factors in individual choice.

Yet bounded rationality isn’t merely as good as new institutionalism in explaining principal-agent relationship—it performs better. This superiority stems from the considerable attention bounded rationality places in the cognitive constraints on decision-making. New institutionalism is an enormous improvement over traditional principal-agent studies because it built such exogenous factors as culture into the decision-making model. Unfortunately it adds, in an ad hoc way, whatever the analyst thinks actors are “maximizing”, and generally ignores the trade-offs implied between the standard assumption of self-interest and the other imputed goals that might get
maximized. Find a deficit in the theory, and a new goal is added.

Bounded rationality takes but one additional step by showing how biological hard-wiring and psychological traits shape how individuals act within organizations, including the complex relationship between incentives and response. Having moved so rapidly away from simple models of self-interest maximization, and having recognized that “informal norms” such as a sense of fairness can be maximized, why do rational choice theorists not simply take the next small step and introduce the individual and organizational cognitive architectures that bounded rationality emphasizes?

Bruno Frey’s research on ‘the crowding out effect’ (Frey 1993; Osterloh and Frey 2000) is an excellent example of how bounded rationality might improve models of principal-agent relationships, and it shows the practical danger of believing that agents simply respond to incentives provided by the principal. Drawing from psychological studies that show individuals respond negatively to control, Frey argues that because ‘under readily identifiable conditions, increased monitoring reduces agents overall work effort’ (1993, 663). Increased monitoring breaks the ‘psychological contract’ with the agent, lowering his or her incentive to respond to incentives or punishment. Not only is increased monitoring not terribly effective, but it might prove harmful to organizational productivity—a blunder based on theory rather than analysis. Frey then compares behavior across organizations to identify which types of organizations will produce a positive (increased productivity) or a negative (decreased productivity) by increasing monitoring—a much more scientific and inductive approach than the standard deductive stance taken by principal-agent theorists.

By looking to the psychological and environmental constraints on individual choice, bounded rationality allows researchers to capture realistic and counter-intuitive behavior that escapes rational choice analysis. New institutionalism and bounded rationality have both discovered the importance of organizational rules and norms in delineating individual preferences; however, bounded rationality distinguishes itself by capturing psychological as well as environmental constraints on choice. Although inductive research is more time consuming, such an approach will ultimately yield a more compelling portrayal of principal-agent dilemmas. In bounded rationality, individual responses to monitoring, incentive structures, routines, and organizational mission are no longer abstracted from artificial models, they are grounded in the science of observation.

**BOUNDED RATIONALITY AND POLICY OUTPUTS**

While bounded rationality may well improve research in principal-agent dilemmas, the thrust of this argument is not to rescue principal-agent theory. Instead, we hold that starting with the assumption of bounded rationality will stimulate a wide range of interesting research questions in organizational information processing, which will ultimately provide us with a far more robust understanding of bureaucratic behavior. This final section looks to the unique emphasis of bounded rationality on how individuals and organizations receive, process, and act upon a variety of information cues.

The promise of bounded rationality is that it allows researchers to connect individual behavior to organizational output without the rigid and misleading constraints of rational choice. For the study of public policy processes, such a model is enormously appealing. It allows researchers to explore political behavior in the aggregate by making assumptions about the behavior of the individual.

The earliest advances in organizational processing models emerged from the study of public budgeting (Simon 1947; Wildavsky 1964; Davis, Dempster, Widalsky 1966); see Thompson and Green 2001 for a critical review). Bounded rationality was used to explain incremental changes in public budgets. The high costs of generating alternatives and making fully optimal decisions, coupled with the uncertainty of outcomes, led to what might
be best characterized as risk-adverse policy incrementalism. At the core of many explanations of budget incrementalism was Simon's notion of satisficing. To limit the costs of making decisions, both individuals and organizations looked to alternatives that were good enough. Budget incrementalism emerged from the risk adverse convention of referring to the prior-year's budget when creating a new one. 'Because the reference point for decision-making is always some point decision made in the past, the outcomes of local search, combined with a propensity to limit bargaining and enforcement costs, are small or incremental changes in policy' (Thompson and Green 2001: 3).

While early models of organizational information processing focused on the politics of incrementalism, more recent studies of bounded rationality have focused on the dynamics of policy change. Behavioral models here have been driven by an intriguing empirical reality—neither budgets specifically (Padgett 1980; Carpenter 1996) nor policy agendas more generally (Baumgartner and Jones 1993) adhere to the strict pattern of incrementalism. Instead, budgets and agendas follow the trajectory of punctuated equilibriums—periods of incrementalism followed by a sudden flurry of policy activity (Baumgartner and Jones 1993). These recent studies have improved our understanding of the policy processes by integrating the role of shifting attention into models of organizational behavior (Jones 2003). Students of how governments set policy agendas emphasize the role of the allocation of attention in determining the behavior of policy makers (Cohen, March, and Olsen 1972; Kingdon 1995; Baumgartner and Jones 1993). Shifts in the external environment produce shifts in the preferences and goals of policymakers. When an institution or organization's attention is squarely focused on a specific policy domain, we might anticipate increased activity, legislation, and spending. However, because the allocation of attention is limited to a select few issues, many policy programs continue incrementally—and follow the conservative, risk-adverse path outlined by prior research in public budgeting.

Low salience issues will generally follow an incremental pattern—little new information or alternatives will be integrated into the solution searchers. When attention is focused on policy problem, we anticipate a broader solution search, greater generation of alternatives, and increased legislative activity. However, even when attention is squarely focused on a problem, decision-making will be bounded by cognitive and emotional constraints.

TOWARD AN INFORMATION-PROCESSING MODEL OF POLICYMAKING

In an upcoming manuscript on how government prioritizes problems, Bryan Jones and Frank Baumgartner (In press) outline the components of a comprehensive model connecting individual and organizational behavior. This model is drawn from bounded rationality, and attempts to capture both the dynamics of individual and collective choice at each stage of the policy process.

Figure 3.1 depicts their information-processing model. Notice at each stage the authors have characterized the decision-making process at the individual and the systems level, because organizations fall prey to the same kinds of cognitive limits as do individuals (Jones 2001).

At the core of the model is the problem of issue attention. Organizational agendas reflect what actors believe to be the most salient or pressing concerns. Because both organizational and individual attention is limited, agenda setting necessitates that organizations prioritize political problems, focusing first and foremost on the most urgent concerns. These concerns may be dictated less by informed reasoning and more by emotional responses to political problems, 'as emotions are critical in determining priorities' (Jones 2001: 73–74). In organizations, the high priority issues will receive the brunt of attention, while less salient concerns will fall to the wayside. If these less salient issues demand political action, they may follow the path of incrementalism and emphasis on 'pre-packaged' solutions.
The next step in the information-processing model takes the form of problem characterization and organizational problem definition. Received information is rarely neutral—individuals and organizations must decide how to characterize and approach a particular problem. Understanding how individuals characterize and organizations define the nature of a political task is essential in the policy process. For example, failing students in public education might be seen primarily as a problem of waning parental involvement, or it might be seen as a systemic failure of the school system. The received information may be characterized in any number of ways, depending on the preconceived attitudes of the decision-maker, or the larger organizational culture.

The final two stages of the model describe how individuals and institutions reduce a problem to a manageable set of alternatives, and then select a policy solution. For example—policy proscriptions addressing failing schools might take any number of forms—smaller class sizes, increased funding, increased oversight, school and student accountability, curriculum overhaul, etc. Because decision-makers struggle at evaluating and making tradeoffs (Sniderman et al., 1991; Tetlock 2000; Jones 2001), choices must be reduced to a few actionable alternatives. In organizations, this is often done through debate. Failing schools can be solved through increasing funding, or through increasing accountability and oversight. The choice is then made according to institutional procedures.
It is important to note that political decisions may not represent individual responses to politically neutral information. People often become attached to preferred solutions beyond their direct utility—a process Herbert Simon referred to as ‘identification with the means’ (Jones 2001). The emotional orientation toward a solution or particular set of solutions bounds the alternatives a decision-maker is willing to consider. In politics, political ideology affects how people prioritize problems, construct problem spaces, and organize solutions. Because ideologies help define “the self” the solution implied by the ideology will be very resistant to information.

The entire process of attention allocation, problem definition, and attaching solutions to problems must of necessity be a disjoint and episodic process. Governmental response will not be proportionate to the severity of the problem—the “exogenous shocks” of rational choice’s favored comparative statics. One of the major empirical predictions of bounded rationality is that, taken in dynamic terms, policy responses will be far more punctuated than would be predicted from a fully rational response alone (Jones and Baumgartner 2005).

CONCLUSION

This essay has distinguished between rational choice and bounded rationality in studies of public policy. We began with a comparison between the two approaches, then detailed the objections to rational choice stemming from the laboratory experiments in behavioral decision theory. Taking principal-agent theory as the major perspective used today to analyze policy implementation, we showed how it has mis-directed scholarly attention to one issue in bureaucratic behavior to the exclusion of many others, and produced a questionable approach to understanding how to improve bureaucratic performance. We briefly examined the information-processing approach, based in bounded rationality, which insists that we focus both on the complex and uncertain environment of policy-makers and on their cognitive capacities that channel how they respond to the uncertain and ambiguous information pressing down on them.

A growing body of evidence criticizing the empirical and theoretic inadequacies of comprehensive rationality has led an increasing number of researchers to call for a renewed model of choice in public decision-making. In political science, rational theorists have admitted to a variety of organizational and cultural constraints on the maximization of self-interest, to the extent that even altruism can be “rational”. This destroys parsimony and the deductive method, for one never knows what is being maximized without empirical examination. Yet rational choice theorists still refuse to take the final step, the admission that cognitive limits of individuals also affects the decision-making process.

A full return to bounded rationality is necessary to carry forward the new demands for realistic yet theoretically tractable models of individual choice. Only by dropping the increasingly weighty baggage of rational choice and its misleading theories represented by principal-agency can we proceed productively, with an increased appreciation for observation as well as theory. After all, the approach was born in the disciplines of political science, organization theory, and public administration. Instead of borrowing from economists, poorly equipped to study politics, we are simply re-acquainting ourselves with our past.

NOTE


REFERENCES


Hoffman, Elizabeth, and Mathew L. Spitzer. 1985. 'Entitlements, Rights and Fairness: An Experimental


True James L., 2000. ‘Avalanches and Incrementalism: Making Policy and Budgets in the


