INTRODUCTION

BY

IRWIN MORRIS AND JOE OPPENHEIMER

FROM

POLITICS FROM ANARCHY TO DEMOCRACY: RATIONAL CHOICE IN POLITICAL SCIENCE.
IRWIN MORRIS, JOE OPPENHEIMER, KAROL SOLTAN, EDS. STANFORD UNIVERSITY PRESS,
2004.

POLITICS FROM ANARCHY TO DEMOCRACY: AN INTRODUCTION

People have been analyzing politics and political institutions for a long time: at least since the time of the ancient Greeks. The prominence of politics in human history is certainly partial cause of its fascination. Aristotle tried to understand the implications of governmental type and form for the lives of citizens. Other questions have similar classic roots because our political institutions have set the parameters for our social lives, and they have done so since ancient times. They determine much of our freedoms, our prohibitions, our obligations, and our privileges. In sum, politics plays and has always played a large part in the determination of who we are or who we can be. It should come as no surprise then that politics has occupied our intellects for well over 2000 years.

And many of the central questions that interested the earliest political scientists still remain on our agendas. We still wonder about the origins of governments, their proper scope and authority, and their conditions for stability. Although we continue to grapple with some of these age-old questions, we are fortunate to live and work in a time when new and powerful tools are being developed to explore these fundamental questions. These intellectual tools, include a set which has come to be broadly characterized as 'rational choice theory.' This body of theory provides new leverage[1] for identifying answers to age-old questions: answers that can be tested and corrected. Although rational choice theory has a long lineage — going back at least as far as Thomas Hobbes — it has undergone a transformation in the past few decades, and its promise and potential as a tool for understanding politics has become clearer to many in the discipline.

With this volume, we illustrate this promise and potential, and, provide examples of some of the very best and most interesting current work in the rational choice research program. The essays have been written in a fashion to require little prior knowledge of the theory or its methods: as such the volume is self-contained. The set of essays represent several research frontiers, but may also serve as an introduction that may motivate the reader to further investigate rational choice theory by fostering an appreciation of the power of the theory among those interested in a rigorous, analytical study of politics.

But the volume also has a substantive bent: it is a book concerning foundational political issues: What is the state and what should it do? Where did it come from? How should the state be structured? The essays provide an intellectual bridge between the basic principles of rational choice theory and their current applications to these central questions of politics.

In order to highlight the way in which rational choice theory addresses traditionally important topics, we have divided the volume into an introduction and three substantive sections. The chapters in each section focus on particular aspects of a few specific fundamental questions:

1. Why do humans establish governments?
2. How can we design better political institutions?
3. What is needed to establish democracy?

And the essays do not stop at asking the questions: they formulate testable answers using rational choice theory. For example, the essays by Swistak and Bendor as well as the one by Miller and Falaschetti grapple with "how can social trust be established to stabilize valued institutions." Bates and Hardin try to answer "what is the proper role of the state and what are its alternatives."

[1] We clarify what is meant by 'leverage' when the discussion turns to knowledge claims and methods (see p. ??17).

Weingast analyzes "what it takes to establish a democracy" as well as what can be expected from it once it is established. Another chapter by Lupia identifies how to design institutions so that the information they give out will be trusted while Knight considers the properties that can lead to an independent judiciary. These are the questions that occupy the chapters to come. But we have gotten ahead of ourselves. We have already referred to rational choice theory several times, but we have yet to properly introduce it.

Although there are numerous formulations of the concept of "rational choice," for the purposes at hand, individuals are said to make "rational choices" when they chose a decision or strategy instrumentally so as to achieve some value. Traditionally the definition has been that rational choice is choice that maximizes the attainment of the self-determined value. (This has also been known as "optimizing" behavior.) However, certain more modern theoretical formulations (especially those related to evolutionary game theory as in the chapter by Bendor and Swistak in this volume, but also see Simon, 1986; Bendor, 1995) do not restrict rational choice to optimizing behavior. In these cases, making a decision that generates a pre-determined level of a certain value — though not necessarily the maximum level of that value — also fits under the rubric of rational choice. In contrast to optimizing, this type of behavior is often referred to as "satisficing" (Simon 1986).

Regardless of the specific definition chosen for rational choice, rational choice theories are designed to explain individuals' choices using their values and the environment in which they must make their choices. The environment consists mainly of the other individuals in the situation and the social institutions that determine each of their rights, endowments, and powers. In the jargon of these theories, we say that the outcome resulting from the situation depends upon the set of feasible outcomes, individuals' preferences, and the rules of the institutions. The institutions determine both the acts available to the individuals and the consequences that result from any pattern of acts taken by them. In sum, the acts of the individuals, and hence the outcomes which arise from these acts, depend upon both the choices of the individuals and the institutions which define the processes.

Analyzing what to expect from individuals in different contexts tells us neither which change is desirable, nor which is not. To do that one needs values. For example, we need content for the evaluative word 'better' or 'optimal'. In any case, these types of problems, relating individual decisions to collective decisions and outcomes for a group, are at the heart of both politics and institutional design. How our institutions relate these individual decisions to group choices determines the extent to which we are able to achieve both our own individual and collective goals. Given the centrality of the relation between the individual and the group to politics, political scientists need to understand the effect of the institutions in the translation of individual behaviors (say, votes) into social outcomes (say, choice of political leaders). As will be clear, rational choice theory is well-suited for analyzing these interrelated phenomena.

In this essay, we our main objective is to present applications and expansions of the central findings of the theory in accessible, and non-technical ways. In doing this we hope to introduce enough of the basic tools of the theory for the unschooled reader to follow the arguments that come later. Hence, although we do not attempt to provide a textbook-like presentation of the "basics" of rational choice theory, we do present a somewhat more intuitive pedagogical introduction. Our intended audience is students and professional scholars with an interest in politics and public policy and a curiosity regarding the applicability of rational choice to their own interests. We do not

presume prior knowledge of rational choice theory. Of course, in such an introduction we must also introduce and explain a variety of rational choice concepts and terms., but that is not our main objective.

Classic Illustration 1: Two-person and n-person Prisoner Dilemmas

Suppose you and a partner have just robbed a bank. While you had time to hide the money stolen from the vault, you and your partner were unsuccessful in your efforts to evade the police. Once caught, you were immediately separated. You are now in the interrogation room. You assume that your friend is in some other interrogation room somewhere in the police station, but you have no way of knowing for sure. The detective who has been questioning you has just returned with a mug of coffee, and as he gets comfortable, he details your situation. You can either admit to the crime, or not, but the penalty you receive if you confess (or not) depends upon the choice made by your partner (who is given exactly the same choice). If you both confess, you both receive ten years in jail. If neither you, nor your partner, confess, you both receive three years in jail (some circumstantial evidence links both you and your partner to the crime). If either you or your partner confesses and the other keeps silent, the silent partner receives twenty years in jail (you stole a lot of money) and the confessor goes free. What do you do?

Thus begins the normal motivation of the prisoners' dilemma game. The game is but one simple model in a complex theory of how an individual goes about calculating what to do when achieving her goals requires that she take into account the behavior of others. Here is another depiction of the same sort of model: a prisoner's dilemma game, but this time with a larger number of actors.

You are a member of Congress and must decide whether or not to request federal funding for a local public works project (a levee for a river in your district, for example). You know that the project is expensive and that the local benefit is significantly less than its overall cost (the river rarely floods), but it would be popular with your constituents (who would only bear a tiny (1/435th) portion of the cost but would enjoy the entire benefit). You worry that many of your colleagues want to make similar requests. And your next electoral opponent certainly won't let your constituents forget about the projects their tax dollars funded in other places and the opportunity you wasted if they don't get their levee. What do you do? Are you fiscally responsible or politically savvy?

In both of these scenarios, you face a choice, and that choice has serious consequences that are a function of the choices made by one or more other people. In the first scenario, the classical prisoners' dilemma, your freedom is at stake. In the second, an explicitly political scenario that occurs all too frequently, your career is at stake. Given the circumstances, you want to make the best choice. The question, of course, is *what is the best choice*? And the answer is: it depends on how your choice relates to those of others.

[Table A about here]

In the first situation, see Table A,[2] regardless of your partner's choice, you are better off if you confess. That is, confessing *always* yields a better outcome: one says that confessing *dominates* the alternative choices or strategies.[3] This is so even though the most obvious best response for both of you *collectively* is to keep silent. Collectively, total jail time is minimized if both players keep their mouths shut. Unfortunately, *individually* the best response is to confess for *both* parties. Collectively, of course, this behavior on both your parts produces a suboptimal outcome (you would both get ten-year sentences) and you could have both done better (three-year sentences).

The same dynamic plays out in the legislative pork barrel scenario. The collectively optimal choice is for no one to receive funding for a project in which the overall costs exceed the overall benefits (see Box B and Table B). But in the current scenario, if any individual legislator chooses not to request a pork barrel project for her district, that legislator receives nothing for her district *and* her districts pays its share of the costs of *all* of the other pork barrel projects received by other districts--clearly, a politically inferior outcome. In both cases, the individual has a dominant strategy: Further, individually-optimal behavior generates a collectively suboptimal outcome (see the extended discussion on page ??8 for a more careful depiction of optimality and suboptimality).

[Text Box B and Table B about here]

Due to the structure of both of these situations, optimizing behavior induces worse outcomes than the participants would want. And better outcomes are clearly obtainable. Naturally, this suggests (policy) question, "How can we obtain the better outcomes?" For example, what if you and your partner were both members of a criminal syndicate, a syndicate that had a known policy of severely punishing confessors (in or out of prison). Would knowing this lead you to make a different choice? If you faced a severe punishment for "ratting" on your accomplice—death, let's say—then that might increase the costs associated with ratting to such an extent that you would keep your mouth shut regardless of the deal offered by the authorities. If your partner had a similar understanding of the situation, neither of you would rat out the other, and both of you would receive light prison sentences.

In the pork barrel case, would you make a different choice about your district's project if you knew it would be difficult (if not impossible) for other legislators to win projects for their districts (because of some formal institutional constraint such as *germaneness* or a rule prohibiting amendments in general)? In those altered situations the incentives shift and so you might just change your mind. If others find it difficult to win pork barrel projects for their districts, you might find efforts to win a pork barrel project for your own district equally unprofitable. Following the logic of the same theory permits one both to analyze behavior within the context of alternative hypothetical institutions, and to think about changing the institutions to alter the outcomes.

Classic Illustration 2: The role of the state

[2] In the table, 'you' get to choose the *row* (**not** the outcome) and your partner chooses the column. Together it is the pair of choices that determines the outcome (or cell or box) you end up in. The first number in each box represents the payoff to you, and the second to your partner were you to end up in that box. Note that you are best off in Row 2, regardless of which column you are caught in by the choice of your partner.

[3] The way to see this is that the payoffs for confessing are bigger for the row player regardless of which column the player finds himself in.

What ought to be the role of the state in the lives of the citizen? Many an argument has hinged on this. In the seventeenth century, Thomas Hobbes wrestled with the question of the necessity of the state. He saw coercive institutions as requisite for the development of any of the creature comforts of civilization; indeed, he argued that coercive institutions were a necessary precursor to civilization itself. With the advent of game theory and the extension of economic reasoning to the problems of collective action, the picture has clarified. Rather than a single monochromatic story as Hobbes told it, we now have a multiplicity of possibilities of cooperation based on incentive structures other than mere coercion.[4] In a simple one-shot prisoner's dilemma game, as in Table A or B, coercive solutions seem quite natural. But, as we point out below (see page xx) there are numerous other possibilities to support human cooperation.

And beyond the basic structure of the state lie other questions. What should be its responsibilities and functions? Socialists and capitalists have argued the point for centuries. Are there any ways of judging the answers? Can we even discern what are the implications and consequences of the different conceptions of the state's proper role? Certainly, much has been learned in economics of the power of markets: how, under a certain set of assumptions, they produce efficient collective outcomes by giving individuals with diverse values a set of incentives to guide their own highly decentralized choices.[5]

Unfortunately, the set of assumptions required to generate efficient outcomes in market settings are quite restrictive. In many real-world settings, markets fail to satisfy the admittedly restrictive assumptions that the efficient market outcomes demand. In some cases competition is imperfect. In others, goods are not private or there is no informational symmetry between buyer and seller. The traditional prescription for market failures was state intervention. But since at least the early 1960s, a school of thought within the rational choice tradition commonly referred to as public choice theory has critiqued state-oriented responses to market failures. For example, Coase (1960) argued that the solution to market failures did not require political intervention. Other students of public choice—Buchanan and Tullock in particular—argued that state intervention designed to prevent market failures might actually generate worse outcomes than those associated with the original market failure. In response to this literature, game theorists Aivazian and Callen (1981 and 1987) showed that under a set of conditions it is impossible to arrive at an adequate solution without governmental intervention. More specifically, they demonstrated that there were conditions under which the parties to 'market failure' could not be expected to negotiate a stable, socially-optimal agreement among themselves. At about the same time, Schwartz (1981) made similar discoveries regarding the role of state_assured property rights to stabilize market transactions.[6] As our understanding of organizations other than markets developed, the evaluation of the relative costs and benefits of the two sorts of institutional (market and governmental) failures began to be taken seriously, and rational choice theory has provided a set of conceptual tools to make just these sort of crucial evaluations.

---

[4]Bates' chapter clarifies precisely what trade offs are involved in the granting of coercive power to the governing authorities of the society.

[5]Hardin's chapter tackles this question by considering how the distribution of information affects the degree to which governmental institutions can satisfy the basic goals people have.

[6]Weingast's view of the stabilization of inherently unstable political agreements parallels the problem of markets in politics.

All this is to say that rational choice theory has improved our understanding of the relationship between institutional designs and the implications they generate. Perhaps we can now hope to improve further our design of institutions to reconcile individual behaviors with group needs without requiring oppressively strong governments or massive and continual intervention. Thus, we may be on the cusp of being able to answer questions dealing with the overall structure of government, the design of the specific institutional structure of government, and the extent to which certain types of governments and institutions can be expected to deliver the social benefits we would like.

We offer these classic illustrations to suggest the theoretical power of the rational choice perspective. In short, rational choice theory has important things to say about the fundamental questions that fascinate political scientists. Let us now consider, in somewhat more detail, the foundations—or building blocks, if you will—of rational choice theory itself.

Rational Choice Theory: Its Premises and Methods

There are at least three major aspects for the study and understanding of the theory illustrated above. First, we must understand the underbelly of choice: its structure, motivation or *psychology*. Second, we must come to understand the aggregation problem: how the choices of many individuals relate to what becomes understood as a 'group' choice. Both the structure of choice and the aggregation problem are mediated by *social institutions*. Finally, there is a need for *performance criteria*: measuring rods by which we can decide whether one change or another might be more advantageous.

Much of the leverage in social science theory in understanding these aspects of politics, has come from the applications of particular assumptions about human behavior to particular social institutions. Rational choice theory is a conglomeration of "models of choice" made up of a family of psychological premises applied to behavior in a variety of institutions. The models are put together by a methodology of logical inference from a set of premises. The premises, stem from the 'rational choice theory' and a stylized approach to the description of institutions: one which synergistically weaves a deductive story from the psychological premises. The result has been a family of models with wide acceptance in economics and a broad influence and role in the other social sciences: including political science. They consist of three elements: a psychology, a style of institutional description, and a method of careful logic.

--The Psychological Premises of Rational Choice Theory

Rational choice is choice with purpose, goal seeking, even maximizing. As such, it presumes goals, and is not itself an argument about how goals are identified, nor how the underlying values are acquired. Rather, it allows the theorist to begin analysis assuming a set of goals. It follows, immediately, that rational choice theory is about choosers. This often (but not necessarily) gives it an *individualistic* focus and starting point.[7]

[7]This is referred to as "methodological individualism" in the philosophical literature surrounding questions of behaviorism in the social sciences. In principle, any agent 'with purposes' will do whether it be an individual or a bureau

Theorists conjecture that the behavior of groups, or institutions, may best be understood as the outcome of the aggregate of individual decisions. Hence, one must, generally, analyze the behavior of the constituting decision makers and their aggregation. This does not mean, however, that group memberships and group interactions are unimportant determinants of human behavior. It simply means that the making of decisions is a starting point.

The trick (or difficulty, depending upon your perspective) is the development of arguments which show how characteristics of group behavior can be understood to stem from individualistic premises. These arguments usually take the form of combining the psychology with the structure of the institutions within which the decisions are made. This structure is specified as that which is determining the rights and powers of individuals: more specifically, the acts available to players and the consequences which result from any pattern of acts taken by the set of players. Thus the social choice from among the feasible alternatives depends upon both the choices of the individuals and the institutions which define the process. In the jargon of game theory: the outcome results from the set of feasible outcomes, the individuals' preferences, and the rules of the situation (see Plott, 1978, p. 207).

--Psychological Elements

As indicated above, rational choice theory has to do with goal seeking, purposive behavior. The traditional starting point is the assumption that the actors have well defined, stable, and ordered preferences. The idea is that all decision makers can make sensible decisions: they can make the necessary comparisons between the alternatives they face, and can and will choose the best available alternative. In facing choices, this leads them to have identifiable *preferences* among the alternatives they face. Rational choice theory is built upon two substantial assumptions which permit the theorist to assume that the individuals' choices are consistent with these preferences. This permits the theory to be used to generate interesting results.

The first assumption is simply that the individual can always make sensible judgments between alternatives. Substantively this means that someone (call her Iris) can choose which of say two alternatives are preferable, or whether they are equally attractive to one another. Thus, in comparing x and y, Iris can say whether she prefers x to y (usually written as xP_iy), y to x (yP_ix) or whether they are equally good and she is indifferent (xI_iy). This assumption, that all items are comparable, is known as '*completeness*.'[8]

The second assumption is that the individuals' preferences make sense across many pairs: that is, they allow the decision maker to order alternatives from 'best to worst.' This permits a sensible choice of 'best' alternatives. To insure this, it is assumed that the preferences are '*transitive*.' Transitivity means that if x is preferred to (or as good as) y, and y is preferred to (or as good as) z, then x is preferred to (or as good as) z. This can be written: xP_iy and yP_iz, then xP_iz. Taken together with 'reflexivity' (see footnote ??21), this means that Iris's preferences generate an

---

or a state. However, one of the principle findings of the theoretical efforts has been that only rarely can we expect collective decision makers (e.g. governments) to behave in the same fashion as purposive individuals, Arrow (1951). On the other hand models of rational choice have been applied to other actors making choices: nation states included.

[8]It is technically coupled with the notion that something is equally good as itself, which is known as '*reflexivity*.'

'ordering' of her options such that she can say which is best or 'at the top,' which is worst or 'at the bottom' and which is to be placed at each position in between.

Combine this with the notion that Iris chooses on the basis of her preferences, and these assumptions imply that when given the chance, she is going to choose her best outcomes, or maximize. Mix a little substance with these formalities (that the individual has some degree of private interests or values which she wants to achieve and we have a sufficient brew to create the menu of models developed above, plus many others: including those in the chapters below.

--Some Criticisms

The perspective of rational choice has not been without serious detractors. Over the last 25 years or so, psychologists have developed a number of tests, and critiques of this model.[9] Most of these results have focused on two sorts of limitations of the rational choice theories. First, they have questioned the stability of the values or goals held by individuals in their making of choices. Goals of individuals were found to be a function of the framing of the decision, rather than independent of it as the theory required. Second, individuals were discovered to have severe limitations in their ability to handle probability calculations in a consistent manner.

These empirical questions certainly have come to define the deep research frontiers of rational choice modeling. Indeed, rational choice theorists themselves are the researchers who have been among the most involved in trying to understand the shifting foundations of their theories. But one can make too much of such threats. Similar research problems sit in every scientific discipline. As Popper (1959, p.111) has said:

> The empirical basis of objective science has thus nothing 'absolute' about it. Science does not rest upon rock-bottom. The bold structure of its theories rises, as it were, above a swamp. It is like a building erected on piles. The piles are driven down from above into the swamp, but not down to any natural or 'given' base; and when we cease our attempts to drive our piles into a deeper layer, it is not because we have reached firm ground. We simply stop when we are satisfied that they are firm enough to carry the structure, at least for the time being.

The questions raised of the foundations seem not to threaten the edifice itself. Rather, recent developments in pulling together evolutionary game theory and showing its relations to the more traditional modes of analysis has made the work of rational choice theorists more stable and less dependent upon highly stylized definitions. For example, the results in the evolutionary game literature (Bendor and Swistak) don't require what might be called full rationality, but rather only a psychology that generates choices that improve one's situation.[10] What then are some of the other characteristics of this edifice? And then, what is the mortar which holds the pieces together?

---

[9]Good summaries of this literature abound. The reader might look at Quattrone and Tversky (1988), Osherson (1995) and Tversky and Kahneman (1986).

[10]See Bendor, who has developed a useful catalogue of possible reformulations of the psychology. On the other hand, Kreps, 1990 and Dixit and Skeath, 1999 give interesting introductions to the theory of non-cooperative games and its less demanding foundations.

--Method

To this point, we have described three of the four elements which characterize the analysis which is referred to as rational choice theory: the psychology, the performance criteria, and the way in which institutions are sketched to generate 'analytic results.' We have implicitly also given examples of the 'method' of analysis, but here we will sketch out its elements in a bit more detail.

Knowledge Claims

Let us take a step back from our exercise for a moment. The theory, and its models extend our knowledge by making 'knowledge claims.' But how, precisely, do the models facilitate this? Indeed, what is a 'knowledge claim?' And for that matter, what is 'knowledge?' To understand knowledge one ought to juxtapose it against something weaker: say, 'belief.' One can believe that the world is flat, but it isn't. And how would we deal with someone who says "I know it is flat!" We might not say, "You are wrong," for we do not easily count as knowledge, something which is believed but which can not be justified. Knowledge is usually referred to as "justified true belief."[11]

But what in the world is justified true belief? In the fields of science and social science, truth of a statement (i.e. a theory, a conjecture, a belief) is usually understood best in terms of the correspondence of the statement to 'reality.'[12] This definition of truth permits one a relatively straight forward understanding of the standard testing of hypotheses and models as they occur in our discipline.

A knowledge claim is then merely a 'justified' claim or a 'justified' conjecture that the world is like a particular statement or set of statements. But what does it mean to be 'justified.' To justify a statement is to show why it is to be accepted. The strongest form of justification is what we have already alluded to: the careful exercise of logic to show that it can be implied from some other better known facts or conjectures. This exercise of logic is to generate a relation between the premises of an argument and its conclusion: a relation of deduction.[13]

What is deduction? Deduction is a method of generating conclusions from premises such that the conclusions are true if the premises are true. Conversely, deduction requires that if the conclusions are false, the premises can't be true. But this is insufficient to fully describe what is going on in a rational choice argument.

--Rational Choice Theory and Deductive Modeling

Rational choice theory is equally known for its methods as it is for its substantive conclusions. The methods are carefully deduced from the premises of institutional and psychological elements as described above. The bold step is to see that a political phenomenon, such as a legislative struggle,

[11]A useful introductory treatment of this can be found in Giere (1984), but almost any standard treatment of epistemology in philosophy will give the same perspective and definition.
[12]Again, almost any introductory work on truth will do as a serious starting point. But we recommend the very fine short book by White (1970) or his shorter essay (1967).
[13]It is surprising how many persons do not know that logic is a set of rules: indeed, that logic is the set of rules which preserves truth in argument. Any introductory text is useful. But an especially useful one is Jeffrey (1981).

can be covered by, or understood as, a 'rational choice system.' (See Giere, 1984, for a careful description of this step in the role of scientific understanding.)

This is the radical, inductive leap: one which says, 'I think this thing over here actually works like this system which we already know.' To show this, one must develop the deductive steps which show that the psychological and institutional premises actually lead to the conclusions one wants to explain (as in the median voter model above on page ??). And this takes the careful steps of logic.

So the models exhibit deductive reasoning. And this is often formulated as mathematical reasoning. Why is this? The core assumptions of purposive behavior are often consistent with the properties of maximization (see above, p. ??). Maximization of a function is a particular mathematical operation and this has permitted the arguments of rational choice to be developed with mathematical rigor.[14] It has made it easier for rational choice theory to be developed in a fundamentally deductive fashion. The use of logical and mathematical methods of argument is precisely what makes the models generate easily testable 'knowledge claims.'

On the other hand, just because the premises *lend themselves to mathematical expression* doesn't mean they require it. Though often presented symbolically, rational choice theory is not *inherently* symbolic or mathematical in character. While rational choice theories are often presented in non-verbal form (either through a series of equations or geometric diagrams), the nature of deduction does not absolutely require the use of such symbols. Rational choice theory can be, and often is, presented verbally. Examples abound: from the realm of normative theory, Hobbes' *Leviathan* is clearly a rational choice theory presented completely in prose. Much more recently, David Mayhew's path breaking *The Electoral Connection* is an argument presented verbally (or in prose) based on rational choice.

Whether or not deductive models are highly *mathematized,* their value depends upon the extent to which they illuminate real-world political dynamics. But how are these deductive models to be applied to the world of politics? After all, for most of us, the aim of our study is not to test a theory, but rather to explain real events, and the functioning of real institutions. The rational choice models are used to generate results that can be redefined as bold hypotheses which apply the core premises to specific sets of social events or political phenomena. This is done by identifying the relevant actors, the institutional structure and rules of the situation, and the characterized plausible set of individuals' preferences to see if one can develop explanations for the sorts of outcomes as a result of these elements (as in the case of the committee in the legislature above, see page ??).

As such, the empirical application of rational choice theory is a difficult combination of deductive and inductive endeavor. One must take general models and apply them to specific situations which have been understood (inductively) to develop a deductive argument of the patterns which one wishes to explain. Each study not only investigates the particular facts using the model at hand (e.g. is the model of the legislature we sketched above a useful description of the committee system in our state's legislature?) but it also lends itself as a small element in the discussion of the utility and accuracy of the overall theory. To demonstrate how rational choice models are built and

[14]The relation between mathematics and logic is a bit complicated but let it be said that much (but not all) of mathematics can be shown to be a subset of logic. Again, many texts suffice to show this, but Jeffrey (????) is particular clear.

then used to develop testable implications, we describe a prominent strain of rational choice theory in the next section—the spatial model perspective.

Seeing Democracy and its Institutions from a Spatial Model Perspective

How does an institution structure outcomes? To answer this question we begin with models generating a rational choice analysis of some aspects of democracy: starting with the simplest, and moving to more complex ones so that the reader can understand the importance of the details.

--Simple Majority Voting Institutions:

As an example, take a simple case of voting in a one-street village. After a damaging fire in a nearby neighborhood, everyone has agreed to install a fire hydrant on the street. Indeed, each person, fearful of the consequences of a bad blaze, would like the fire plug as near as possible to their own address. This permits us to develop what we might call a traditional, one - dimensional '*spatial*' voting model in which it can be shown there is a simple way to predict the outcome of the vote.'[15]

Predictions:

As in most of these models there will be assumptions regarding preferences, the choice environment (i.e. the rules of decision making and the set of alternatives. The preference assumption which generates the analysis is that each voter has an ideal point "along a line" such that the further the collective choice is from the voter's ideal point, the worse off the voter is.[16] In this case, that implies that the further the fire plug is from the house the less happy the voter is about the outcome. The major result of the analysis is that the median voter's position which will win under most implementations of majority rule.

A typical sketch of a proof of what to expect concludes that there is a predicted equilibrium which we can identify as follows:

[Figure 1 about here]

Begin by noting that x_i is the i^{th} individual's ideal point and x_{med} is the median individual's ideal point. The median position, x_{med}, with respect to the set of voter ideal points, $\{x_i\}$, means that the number of points at or to the left, N_L and at or to the right N_R of x_{med} are equal and both are less than $N/2$.[17] (See Figure 1.) The proof then proceeds as follows:

[15]Good introductory descriptions of these models can be had in Mueller (1989) and Enelow and Hinich (1984). The major elements of the theory were worked out by Duncan Black in the 1940's, and are reported in Black (1958).
[16]Many variations are possible. Specifically, for example, the analysis can be done even if the preferences fall off on a non symmetric basis to the left and right.
[17]Here note that more generally we need to deal in a more detailed fashion with cases where more than 1 voter share having the median ideal point, etc.

Consider x_{med}, and j's (some other voter's) ideal point, x_j to the left of x_{med}. Now all ideal points at or to the right of x_{med} are closer to x_{med} than they are to x_j. By definition of x_{med} there are more point at and to the right of x_{med} than to the left of x_{med}. Therefore more voters would prefer x_{med} than would prefer x_j. Hence, if voter's 'vote their preferences,' they would defeat any x_j ... x_{med}. Note that no majority could form which could *guarantee* itself outcomes better than that of the ideal point of the median voter.

Performance:

How well does such a system perform?

One answer would be to see if we can 'characterize' the predicted outcome in a satisfactory way. And here the geometry is helpful. Note that the result in the above example is not in some 'outlying' area of the street, but rather in the neighborhood of the people. Thus, in some very rough sense, the system at least hits the target. Further, it also delivers the hydrant not to the edge of the neighborhood, but in front of the median location: toward a central location. Of course, this is not the same as conforming with some other desirable characteristic - such as minimizing the average distance from the hydrant to the homes (that would require placement at the mean of the ideal points), nevertheless by choosing a median outcome, it does move in the direction of moderation.

Another answer would be to see if the system satisfies a performance criteria. In this case we will consider whether it satisfies the notion of optimality. The traditional notion of optimality was given to us by Vilfredo Pareto, an Italian Economist of the late 19th and early 20th centuries. It is called Pareto optimality. What is Pareto optimality? It is a condition of a situation that is desirable and is best first described by its failure. When a situation is *not* optimal, or is suboptimal, at least some of the individuals could be made better off without hurting anyone. That means, if a situation is optimal, to make someone still better off will require redistribution: some in the group must be hurt. The Pareto set would be the set of situations which are Pareto optimal. Again, look at Figure 1. Consider two points, X and Y which are not part of the Pareto set. To show this, we need to show that everyone can agree to move from either of them to some point between *i* & *m*. Then we would need to show that all the points between *i* and *m* are in the Pareto set.

The proof of this is straight forward: Consider, first the argument that Y is not Paretian. Note, Y is to the left of the left most house on the street: that of individual *i*. Imagine that someone suggests another point for the hydrant, call it Z, between Y and *i*: This would meant that everyone in the group would find Z closer to their house, and hence better, than Y. They would therefore prefer Z to Y, and thus, Y is not in the Pareto Set.

We must still show that the points between *i* and *m* are in the Pareto set. Consider a *k* somewhere between *i* and *m*. Does there exist some other point, call it *q*, which everyone would prefer? Consider a *q* slightly to the left of *k*. Then those at *k* or to the right of *k* prefer *k* to *q*. And what if *q* is to the right of *k*? Then those at or the left of *k* prefer *k*.

So majority rule, when all voters have the preferences of a maximum along a single line, and all prefer being closer to their maximum than further, delivers Paretian, or optimal, results.

Extensions:

There are numerous complex models constructed much like tinker toy structures by putting together the elements of the above simple institutional structure. To illustrate, imagine a legislature using a simple majority rule and considering a change of a law. The relevant parameters are the placement of the status quo, the distribution of the preferences of the legislators, and some other structural issues. What are the mechanisms of bill writing in the legislature? Does it work by specialized committees taking up the initial task of writing the proposed bill? Will such a committee have to reach its decisions by majority rule? Assume so. Further, assume that the voters' loss of value is symmetric around their ideal preference points.

[Figure 2 about here]

We can now use the above spatial voting model to predict the outcome. As an example, consider the committee's strategy. Note, in Figure 2 the committee members' preferences (shown by the short line near the top of the diagram, are distributed to the right of both the preference of the legislature's median member (marked as L_{med} and the status quo (see the bottom line). Note that normally we might predict that the median voter on the committee (C_{med}) would win the day, and that would be the bill reported out of the committee. But C_{med} is further from the median in the legislature than is the status quo. And the committee's members obviously would like to move to the right of the status quo as far as possible: certainly at least to the median voter of the legislature. But an outcome as far to the right as C_{med} will not beat the status quo. For all at and to the left of the L_{med} would find themselves closer to the status quo than to C_{med} and they will vote for the status quo. So the committee will put forward a bill that appeals to the majority without mirroring its own median. To do this, they will select a proposal that is almost as far to the right from the median as the status quo is to the left of it. This would mean that all those at the median and those to the right of the median will find themselves closer to the proposal than they are to the status quo. Hence, they will prefer it. And they constitute a majority. So that is the best that the median voter on the committee can achieve.

Consider the performance of such a system. Obviously, the status quo can end up being altered, but not in a manner so simple, and intuitively 'benign,' as we found above. Here the result is not necessarily a significantly better outcome than the status quo, but rather a move which could be considered a manipulation of the system for the special interests represented on the committee.

Similar models are used to explain why governmental bureaus may be over-funded, how school boards negotiate for bond issues, and the like (see Eavey and Miller, 1984; as well as Romer and Rosenthal, 1978). Many other models of similar multi-stage negotiations and votes (often discussed as 'structure induced equilibria,' a phrase introduced by Shepsle and Weingast (1981)) have been developed to explain other phenomena which reflect the above reasoning. Multiple institutions can generate equilibria, much as the legislature and its committee structure (see page ??) can. Numerous models have been developed to analyze the nature of American national legislative / Presidential structures (see Miller and Hammond, 1987 and 1990 for a particularly good example), as well as that of other institutions such as the Federal Reserve Bank (see Morris, 2000).

And what happens when there is more than one dimension along which all proposals can be located. We go into the extensions of the analysis of majority rule beyond a single dimension in the next section. But let it be noted that neither of these nice properties (equilibrium, Paretian outcomes) carry over to this more general case, once we move from the single dimension to a more generalized, or less restricted conception of allowable preferences. Rather, what we are faced with then are very difficult normative problems and the analytic literature of these cases is extraordinarily rich. The interested reader might well seek a more elaborate discussion in a good text book.[18]

More Complex Political Situations & Institutions:

Models of problems in real democracies, of course, often cannot be restricted to 'single issues' which have the characteristic that each voter has an ideal preference point along a line and preferences which fall off as the distances from the ideal increase. We have many options to describe what happens when we depart from the simple model above. Easiest in exposition, probably, is the expansion of the model to multiple dimensions. Institutions must still aggregate the decisions of many voters across many issues. The few problems which we identify here, looking at a simple expansion from one to two dimensions, will help us understand many of the other difficulties which are endemic to democratic decision making.

[Figure 3 about here]

To help in the analysis, consider Figure 3. There we show the decision problem facing 3 individuals (A, B, and C) who are to decide a two dimensional issue. Such an issue, to draw the parallel with the simpler case, might be where to put commonly held fire fighting equipment on a prairie to protect each of the voters' homes. We might consider an outcome inside the triangle (say at Y) as a useful benchmark: a good compromise. As before, assume that each wishes the equipment to be as close as possible to their own house. That would mean, that each individual would prefer (and vote for) points closer to their corner of the triangle than Y.

The analysis follows directly from the observation that, in this case, each individual would prefer (and presumably vote against Y for) any point inside that circle which is centered at their house, and which has Y as a point on its edge (these are shown with the dotted lines). What can we predict to be the outcome of a simple majority rule decision? We can see in Figure 3 that there are 3 "lenses" (made from where the circles intersect) which stretch from Y to a point outside of each of the 3 sides of the triangle (X is an example of such a point). Take, as an example, any point inside the lens XY. Both A and C would vote for such an alternative as it is closer to both of their houses. In other words, movement from Y to a point in any one of those lenses permits *a majority* (two) of the voters to be better off. Thus, Y is not a stable outcome. For at Y any of three 2-person coalitions can form to improve the position of their coalition members. And an analysis shows that there is NO point that can't be improved upon for a majority. Indeed, majorities can even swing the choice outside the triangle![19]

Note that we can analyze this using our previously discussed notion of optimality. Examination of the problem will lead you to see that if we begin outside the triangle we can find points which all

[18]See for example, the discussion of Arrow and multi dimensional voting in a 'standard text' such as Mueller (1989).
[19]Recall any point in the lens will beat the mid point Y. And all three lenses have a portion outside the triangle.

can agree to as an improvement. Thus, from X, A and C would prefer movement inside the lens going from X back toward Y, and B would also.[20] Once in, or on, the triangle, no further movement improves all three individuals. So the triangle is the Pareto set, and the set of points outside are suboptimal.

Similar discouragement is found if we seek stability of majoritarian decisions on such questions as redistributive policies, vote trading, and more generally for virtually all democratic procedures when analyzed from this perspective (see footnote for some basic references). But we can still say a number of things beyond the discussion of structure induced equilibria.

In much of the analysis, a specific form of equilibrium will be considered. It is a notion of equilibrium which arises 'endogenously' from the situation itself. The idea is that the actors' set of choices is in equilibrium with each other when no single party has an incentive to change their choice, given the choices of the others. In a simple two person interaction, or game, as in the simple case of the prisoner's dilemma sketched in this introduction, the simplest solution concept - the **Nash equilibrium** - is reached when neither player has an incentive to unilaterally deviate from the strategy she has played. Hence, when a Nash equilibrium is reached, neither player changes her strategy, and the outcome is stable. Reexamining Table A should satisfy the reader that this is a quality implied by all players having dominant strategies as in the prisoner's dilemma. But it turns out that all games have Nash equilibria. So, for example, consider the simple 'chicken' game depicted in Table D. In this game neither player has a dominant strategy and yet there are two Nash equilibria. (Can you identify them?)

[Table D about here]

Of course, the fact that no one actor has a unilateral incentive to change their choice, does not mean that there may not be other ways the equilibrium might be upset. So for example, the players could come together in groups, and reach an understanding as to what might be a better 'joint' or 'cooperative' strategy. In the simple case of the prisoner's dilemma sketched at the top of the introduction, this was illustrated by the notion of a criminal syndicate to coerce the preferred outcome. But this requires an ability to make a binding agreement, often requiring some mechanism or institution which may not be an inherent part of the situation.

The Nash equilibrium is a very useful solution concept, especially when all the elements which may be needed to support a cooperative outcome are not available. But Nash has some limitations. The most severe limitation, for our purposes, is its application to some situations which are repeated, or continuous and structured over time. When modeled as games, these situations are usually referred to as 'multi_stage games.' In such games, the problem is that some choices that are Nash equilibria are "unreasonable." This is relatively easy to illustrate. For example, consider the chicken game in Table D. There we showed the game as it is normally considered: a simultaneous move game, where the actor doesn't know the choice of her opponent prior to making her own choice.

But what happens if we 'stretch' the game over a bit of history? Now let one player (say row) move first, and let us assume that the second player observes the play and then chooses her response. We can analyze this new situation quite directly: row still has the same two possibilities,

[20]The solid arc going through X is B's indifference curve and B would prefer to be inside that arc.

but column now has 4 ways of responding (see Table E). She can choose 1) always to refuse to swerve (clearly not a good idea if her opponent has done likewise), or 2) always swerve (not the best again because it is too inflexible: what if the row player has swerved), or 3) to match her opponents move (swerve for swerve, head on for head on: clearly the worst choice), or 4) to play the opposite of her opponent (her dominant strategy).

[Table E about here]

Examining the depiction of this in table form, we find that there are an embarrassing number of Nash equilibria: 3, rather than the one associated with the dominant strategy. The two which are associated with the dominated strategies wouldn't normally be reached: the column player would play her rational (dominant) strategy. We can 'explain away these Nash points' but they are misleading. A better way has been found to analyze the problem of choice structured by time: the use of a 'game tree' which depicts the moves of each player, given the choices which have been made to that point in the extensive form game. To show how to analyze this, we show the game tree for the Chicken game in Figure 3b.

[Figure 3b about here]

In any game tree the following is depicted:

1. decision points (or moves) of the players (called nodes). The nodes are usually depicted as circles with the label of the player in the circle. One of these is the (single) starting node; others are the end nodes one of which would be reached by some pattern of play. Other end points would require other moves.
2. At each node, the options from which the choice will be selected is depicted as 'branches' from the node to successor nodes.
3. Now the whole two move chicken game is depicted by the tree. But after the row player moves, to begin the game, the column player finds herself in the tree with part of the game left. Note that any section of the tree can be thought of as a game beginning from a node and continuing with all the branches from that point down the tree. When the players can identify where they are in the tree such subsets of the game are referred to as 'subgames.' Thus, the simple tree has 3 subgames, one starting from the top (labeled 'I', and the other two starting from each of the column player's decision points, labeled 'II', and'III').

Putting the game in this form gives us a simple and intuitive form of 'solving' the game called *backward induction*.[21] For example, the first player (*R*) looks to the *end* of the game: the last choices being made by the last player to choose. Consider Figure 3b. The first player can ask, what would happen, at the last move, if the last player were on the right hand branch of the tree: subgame III (i.e. player one played 'head on,' or H). At this point *C* could choose H or S. H would lead to a payoff of -100, and S would lead to -5. Clearly, *C* is better off choosing S, and hence can be expected to choose S. Similar analysis of subgame II would lead *R* to conclude that *C* would be better off choosing H (gaining 10, rather than nothing), and that she can therefore be expected to choose H. This being the case, *R* can now readily predict the outcomes of his 2 strategies: choosing H will lead to *C* swerving, and *R* will get 10; his choice of S would lead to *C* going head on and *R* getting _5. So *R* chooses H.

[21]This technique of solving games works on all games with 'perfect and complete' information, and only some of the others. For a fuller discussion of this, Dixit and Skeath, p. 48_52.

We are now ready to discuss those three Nash equilibria in the chicken game 'stretched' over time. They show up as each understandable in one of the 3 subgames of the game: To see this, let us consider each of the three subgames separately. Were Column considering how to play the game in subgame II, after Row choose to swerve, she would find that a strategy of always going head on would be in equilibrium (and equivalent to "Opposite" for the subgame she was in). Similarly, were Column considering how to play in subgame III, after Row choose to move head on, she would find that a strategy of always swerving on would be in equilibrium (and equivalent to "Opposite" for subgame III). Of course, thinking about the game as a whole, i.e. from the point of view of subgame I, the first player might not always swerve (indeed, given the dominant fourth strategy, she won't), and the rational response is to play 'Opposite.' Thus only the Nash Equilibrium associated with the dominant strategy is sensible for the entire game (subgame I). If a Nash equilibrium holds for the entire game, and its implied rules are a Nash equilibrium in each of the subgames, then it is called a subgame perfect Nash equilibrium.

It is this last sense of equilibrium which will be important in a number of the papers in this volume and the reason is relatively easy to understand. When a game is played over time, players may adopt 'threats' to convince their opponents to adopt specific strategies in the future. Consider the above chicken game. Perhaps row can't be fully committed to the 'head on' play which she seems to open with. Perhaps she can be dissuaded by a threat to play "Always Head On" by Column. Were that threat communicated, and believed, after all, the equilibrium first move would be 'swerve.' Of course, a careful analysis of the situation by Row could lead her to the belief that such a threat by Column would not be credible. The subgame perfect equilibrium (Nash is often dropped) would then tell us what to expect upon the rejection of the non_credible threats. Hence, in games of this type an outcome based solely on one player's threat to behave irrationally may satisfy the Nash criteria, but we would consider this an unlikely / unreasonable solution since it could sometimes depend upon blatantly irrational behavior. In games of this type, the Nash solution concept will also fail to identify a unique outcome, which means that the concept will not let us predict a determinate outcome to the game.

To avoid these problems with the Nash equilibrium, anlaysts often use the more restrictive solution concept _ subgame perfect equilibrium. Obviously, although all subgame perfect equilibria are Nash equilibria, the converse is not true, and it is those Nash equilibria that are based on threats of irrational behavior that are eliminated. Because of the nature of the situations being analyzed by the authors of the papers in this volume, subgame perfect equilibria play a prominent role in some of them. Especially in those situations where there is no institution or outside mechanism to help enable cooperation, a recent idea has been to predict what strategies can stabilize an agreement. Here we would explore the ways behaviors can 'play out' so as to induce the parties to honor an agreement: a notion strongly related to the subgame perfect equilibrium concept. The chapter by Miller gives a good exposition of the problem and the chapter by Weingast gives a further illustration of this.

Collective Action Illustration

Perhaps surprisingly, the simple set of assumptions behind rational choice can also generate quite powerful conclusions without any assumptions about institutional contexts. Individuals each maximizing their payoffs create problems for the achievement of shared goals among members of

groups. This is the famous insight behind what is often called the 'logic of collective action' (Olson, 1965). To see this, consider the case of Irisin a group of some size (say *10* persons) like-minded individuals. Let them share a costly goal and assume the 'institutional' structure of the group is rudimentary: individuals must voluntarily contribute to the obtaining of the goal. Consider then the following simple case.

Let each individual, have some endowment of 'private' resources. So Iris might start with $10. Consider a case where the goal can be partially achieved and to the degree that it is, it rebounds in value symmetrically to each member. So, we might say any $1 spent on the attainment of the goal generates $.25 of value to each individual. Then each individual, say Iris, might note that for her, it is not worth her while to invest in the group goal. For every $1 she spends out of her endowment, she loses $.75 (she, like each of the others gets $.25 worth of value from the $1's worth of partial attainment of the group's objective). We can put her situation in perspective by specifying her options more carefully as depicted in Table 1.[22]

[Table 1 about here]

In Table 1, Iris, as the typical member of the 10 person group, contemplates giving $1 or nothing toward the achievement of the goal. That is, Iris chooses between the two rows of the table. In keeping with our substantive understanding of the premises of rationality theory Iris will examine her options and choose the one which she prefers. Assuming that the financial calculations properly capture her values, how ought she to choose what to do? Iris can't select the column with which she "is faced." Rather, this is determined by the choices of others. But she can note that regardless of which situation she is faced with (i.e. in which 'column she is') the second row yields preferred outcomes to the first row (it is 75 cents better).

But, recall that though each dollar's partial achievement of the group's goal generates only a return of a quarter of a dollar, it does so for every member of the group as they all share the partial achievement of the goal of the group. Thus, each dollar given generates $2.50 worth of benefit in total if we add up the difference it makes for each person (10 persons at $.25 each). In other words, each individual following their own incentives, leads to a situation which is not very good: certainly not '*optimal*.' Specifically, if all follow their incentives, the group could agree *unanimously* that it would have been better had they all contributed.

Consider the implications of this: Iris would be substantially better off were she and everyone else be forced to contribute. Then each of them would have an outcome worth $11.50, rather than $10.

1. Iris, and all others find that it pays them not to contribute.
2. Iris, and all others in the group would prefer that they did contribute.
3. The group is saddled with a suboptimal outcome.

[22]This model reflects the perspective taken in Hardin (1971). Other ways of modeling the problem involve either no game theory (e.g. Olson, 1965), or other forms of games (see Frohlich and Oppenheimer, 1970; and below, page ?? for

[Figure 4 about here]

Some would argue the situation sketched here, properly generalized of course, underlies many of the dilemmas of politics. To understand the links between individual choice and more general notions of collective action and group patterns of behavior, we might begin with a simple observation: individuals will usually not find it worth while to help achieve shared interests: they will find it useful to shirk their responsibilities. Why does this happen? One of the major contributions of the theory of rational choice is that it gives a relatively simple account of this:[23] each contribution to a group or collective project yields benefits for all members of the group, but costs accrue only to the individual who makes the effort. For the group and all its members, it is not difficult to see that it would be best if all contributed (see Figure 4, which is based on work by Schelling, 1975). In the diagram giving is worth less (the lower line) than not giving (the upper line) regardless of how many others give (depicted on the horizontal axis). But if all give, one gets more by giving than by not giving when all don't give: the right hand end of the bottom line is above the left hand intercept of the upper line). If the individually accruing benefits don't fully compensate for the individual's costs the incentive is not to cooperate. Thus, the group goal would not be easily achieved until the members of the group worked out a more complex institution. With this depiction, the general situation can often be analyzed as an *N persons prisoner's dilemma game*. In such a game, the group optimum is not achieved because each of the group members is in *equilibrium* (i.e. can't do better by unilaterally changing their choice) when they decide to not contribute or to shirk their responsibilities.

The theoretical predictions of such simple models don't generalize or often don't predict well. People *do* contribute to collective efforts, so the theory appears false. But the modeling of collective action doesn't stop with the single shot prisoner's dilemma analysis. Other models of the problem have been developed.[24]

Some Elaborations

Of course, not all situations lead to such bad outcomes, and we might wonder why. There are a number of possibilities. First, not all games need be prisoner's dilemmas. Much has been made of the possibility that some collective action games are actually assurance games, and not prisoner's dilemmas. In assurance games the lines showing the value of contributing and not cross each other at some point (see Figure 5). If enough persons come together to solve a problem, it becomes worth the individual's while to join the project, rather than to 'shirk.' A number of studies have

[23]This is not the place to give a comprehensive account of the theory of collective action. The reader is referred to numerous volumes and articles on this. Any list would have to begin with the work of the late Mancur Olson (1967), but Hardin (1971) showed that the relationship could sometimes be considered a prisoner's dilemma game. Ledyard (1995) reviews the experimental studies of the subject.

[24]Green and Shapiro also criticize the plethora of models by saying that such a surfeit means that the theory is not falsifiable. But this is too simple. The different models are for different conditions, and the issue is which conditions, or institutional designs, lead to group goals being achieved, and which do not. For example, Olson's perspective was *not* game theoretic, and certainly some contexts don't lend themselves to Hardin's game theoretic interpretation (see Dixit and Skeath, Chapter 2). But beyond this, if the collective goals are not 'continuously' differentiable, or, in other words, are 'step functions' or 'lumpy,' then other forms of analysis are required (see Frohlich and Oppenheimer, 1970, 1978). Still other forms of analysis are required if there is repetition of the interactions, etc. (See Dixit and Skeath, Chapters 8, 10 and 11).

shown that the leadership of the American civil rights movement transformed the situation into an assurance game, and thereby greatly improved the chances for success (see for example Chong, 1991, and Dixit and Skeath, 1999).

[Figure 5 about here]

Another complicating factor is the possibility that the games are not 'one-shot' but rather are repeated. When interactions similar to the one developed above are repeated, the incentives are altered. This is especially clear if we restrict the discussion to two-person situations. Then things improve if each person can make their choices contingent upon the choices of the other person. A repeated event, often referred to as a *repeated game*, requires the careful depiction of possible ways of choosing throughout the repetitions: called a *strategy*. An example might be Iris will contribute and continue to do so if Jerry also contributes. If she doesn't, then Iris will stop contributing until such time as Jerry contributes again. Such a strategy is referred to as 'tit for tat.'

All these models have a number of elements in common: a set of decision makers with specifiable alternatives and goals, an institutional context, a relationship (or set of rules) which maps the choices into outcomes, and some criteria (e.g. Paretian optimality as described above) by which we can evaluate performance of the institution. Not listed above is the 'maximizing' behavior at the heart of our description of 'rationality.' This is because some recent works have inquired whether we could change our assumptions. For example, rather than assume maximizing, what if we instead assume that people tend to change the way they make their choices so as to improve their outcomes (some form of learning). This can take various forms (see Bendor, 1995) and all lead in a similar direction in one's arguments.

One such form has been especially useful in understanding repeated events: the idea that people observe the payoffs of others and the 'norms' they follow. Individuals then change their strategies or norms when they find others with better payoffs. Such a perspective is called 'evolutionary' because it mimics quite accurately notions of evolution in biology (see the essay in this volume by Bendor and Swistak). With an evolutionary perspective, one can radically weaken the psychological assumptions and still come up with interesting and testable results. Indeed, one of the surprising discoveries is that the evolutionary analysis of a situation yields, in equilibrium, what can be seen as an equilibrium in the more traditional mode of analysis (see Dixit and Skeath, 1999, p. 347).

--Citizens

Rational choice theory is also useful in understanding the characteristics of the behavior of the citizens of democratic systems. Normative theorists have often argued that the democratic citizen ought to be vigilant, attentive to the public and the government which she helped elect. But an extension of the logic of collective action leads one to believe that the average citizen is highly unlikely to be well informed. Rather, the individual's rational choice is to take into account the great likelihood that she is not going to affect the outcome, and then calibrate her effort at information acquisition accordingly.

The conclusion is that we can expect citizens to remain 'rationally ignorant' (a conclusion first developed by Downs, 1957). The implications of this, normatively, are severe. For just as one can

argue that the voters ought to be informed, to note that the rational voter won't be, is also, in part, to excuse the voter from the responsibility of being informed (see Oppenheimer, 1985). Certainly all that the law requires, for excuse from responsibility is reasonable care. Since this excuses the citizenry from the responsibility of exercising control over the government (the defense of the 'good German': "I didn't know what was going on," comes to mind) concern has been developed about two aspects of expectable ignorance.

But, of course, the first concern is that of governmental performance. All the discussion above regarding democratic governmental performance assumed that the government *knew* the preferences of the citizenry. But if the citizens don't know the activities of the government, and do not have an incentive to communicate their preferences to the government what can we expect of governmental performance? A number of models and experimental studies have shown that we can expect substantially better governmental performance than we might intuitively believe (see Collier, et. al., 1989; Lupia and McCubbins, 1997; and Popkin, 1991). More recently, analysts have begun to consider how institutions might be better designed to insure adequate information flows and absorption by citizens (see Lupia's chapter in this volume).

Rational Choice's Attractiveness

So we can see that rational choice theory is a useful tool for answering many fundamental and other questions. It is useful because it provides a means for predicting behavior, and identifying the gap between predicted behavior and optimal performance (depending upon one's standard of optimality). Thus, rational choice theory can tell us what we can reasonably expect in the way of performance and how we ought to change institutions to improve performance. In any particular institutional setting and social environment one can also usually measure the gap between the predicted and the actual behaviors. Thus it can be tested and improved. For these reasons, rational choice theory is attractive to individuals interested in solving real world problems.

Description of the Volume

The volume is divided into three substantive sections within which the chapters are organized. The first section, "From Anarchy to Society," deals with the formation and underlying social conditions for the development of political societies and states. The second section, "Institutional Design," deals with the workings of the modern state. And finally, the section "To Democracy," deals with issues of political development. An introductory essay precedes each of the three sections in the volume, and Karol Soltan concludes with some thoughts on rational choice and theoretical history. A technical appendix, bibliography, and index follow the conclusion.

The two chapters in the first section, "From Anarchy to Society," focus on the conditions in society needed to support any sort of cooperation, and the original organization of the state. **Jonathan Bendor and Piotr Swistak** develop a rational choice model of the evolution of the social norms and informal institutions which are needed as a foundation for cooperation. In doing so, they provide an individualistic foundation for the types of social characteristics that are often presumed to precede the formation of states. One of the most interesting aspects of their analysis is the manner in which they "measure" the extent to which a particular norm is manifest in social behavior. Once they develop a technology for evaluating the ubiquity of a norm, they analyze the

preconditions for lead to the widespread diffusion of these norms. And they demonstrate that certain important norms can gain a foothold in a society without being imposed authoritatively from the top. These results have implications for our understanding of the role of government (or the need for government) in generating certain social outcomes.

Robert Bates (and coauthors) joins economics and politics to explain what economic circumstances lead to the development of states within traditional societies. This work has implications for our understanding of both the potential for economic development in stateless societies and the relationship between economic development and political development. Bates' model poses the alternative starkly as either poverty and no coercive power of the state, or coercion and a possible escape from want. Bates finds that economic and political development are necessarily linked, though not in the manner often thought. Bates' model implies that economic development precedes (is, in fact, a precondition for) the development of the state. As is often true, the logic of the model is intuitive after the fact, but it does suggest a rethinking of the more conventional perspective towards the relationship between political and economic development (economic development depends on a set of pre-existing political institutions—or political development must, at least to a degree, precede economic development). **[Needs to be revised to correspond to current paper.]**

The second section, "Institutional Design," contains three essays. In this case the authors deal with two normative elements in politics: trust (Miller and Falaschetti) and responsiveness to citizen welfare (Hardin) and a property which is in part normative: the dispersal of credible information (Lupia). **Russell Hardin** discusses the prevalence of rational choice principles in political philosophy—at least since Hobbes, and argues against the characterization of rational choice as an *alternative* rather than as a *foundation* for moral behavior. He then goes on to argue that the necessary lack of available and collectively held information and knowledge at the center leads to a necessary prescription for a limited governmental structure. **Miller and Falaschetti**'s chapter on institutions deals with the central problem of organizational design: how does one structure an organization to achieve collective goals with individually self-interested members. They demonstrate that trust is the central element which must be addressed and argue that no final, general solution to this problem exists. They do identify several potential responses to dealing with this problem in particular contexts. Finally, **Skip Lupia** develops the conditions under which information is likely to believed. He shows that there needs to be knowledge of the incentives that the dispenser of the information faces for the dispensed information to be accepted. He then presents an experimental study of individual decision-making and information acquisition. Lupia shows that institutional structures may have a significant impact on the substantive transmission of information, and he demonstrates how a sophisticated understanding of the psychology of choice can be used to design better democratic institutions which could generate better informed citizens.

The third and final section, entitled "To Democracy," consists of two chapters investigating the development and internal political dynamics of institutions in democracies. **Barry Weingast** is concerned with the establishment of democracy: how is it possible. He notes that the problem is one of agreement among the elite, or the power brokers. His concern is the conditions under which any such agreement is forthcoming and stable. Finally, **Lee Epstein** and **Jack Knight** discuss the use of rational choice theory to understand the politics and the development of an independent

judiciary, especially one empowered to review the practice of government to preserve constitutional rules and agreements.

These chapters, individually and in their collective methodological and substantive overlap, highlight the characteristics of the best current work in the rational choice paradigm and demonstrate why rational choice theory is such a powerful tool for understanding politics. Rational choice theory often generates counterintuitive implications: results which can suggest that our unsophisticated understandings may be incomplete or misguided. For example, Lupia shows how identical communications can have very different informational content depending upon the institutional structure within which the communications are transmitted. Similarly, rational choice theory frequently highlights the importance of variables that had been under-examined or completely ignored. In several chapters (Bates, Bendor and Swistak, Knight, Miller and Falaschetti, and Weingast), time – or more specifically *time horizons* – play a particularly important role in determining the extent to which desirable outcomes can be achieved.

The practitioners of rational choice theory have spent considerable efforts at testing, and at trying to improve the theory to take into account the empirical problems these tests have uncovered. Further, the theory itself often can be folded back on itself to show potential difficulties with models which have been forward. This effort to improve the analysis is certainly part of what is going on in the chapters by Miller and Falaschetti as well as Bendor and Swistak.

These chapters also highlight two important aspects of much of current rational choice theory that have not always been true of work in this paradigm. First, rational choice theory and sophisticated empirical analysis are complementary endeavors that are extremely productive when melded effectively. Several of the chapters in this volume explicitly connect the implications of formal models with specific empirical phenomena (see Lupia and Weingast's chapters for two very different enterprises ranging from experimental results to case studies).

On the other hand, the essays in this chapter also suggest that rational choice theory – as it is currently understood – is inadequate to the task of providing a general theory of socio - political relations. Some authors (such as Knight, Lupia, and Miller and Falaschetti) deal directly with the limitations of conventional rational choice theorizing. Others emphasize the importance of factors or variables (such as leadership and *fortuna* in Weingast) for which rational choice has yet to provide a compelling description and explanation. For all of these reasons, we expect rational choice theory to play a major supportive role in the future development of our understanding of political phenomena. We hope that these essays (and our comments about them) will foster a similar appreciation of the theory among our readers.

References

Aivazian, V.A. and Jeffrey L. Callen, (1981) "The Coase Theorem and the Empty Core." *Journal of Law and Economics*. v.24: 175 _181.
Aivazian, V.A., J.L. Callen and I. Lipnowski, (1987) "The Coase Theorem and Coalitional Stability." *Economica*, 54: 517_520.
Arrow, Kenneth J. (1963). *Social Choice and Individual Values*, 2nd ed. Yale: New Haven.
Bendor, Jonathan (1995). "A Model of Muddling Through." *American Political Science Review*, vol. 89 (#4, December): 819 _ 840.
Black, Duncan (1958) *The Theory Of Committees And Elections*, Cambridge: Cambridge University Press.
Brandl, John E. (1998a) *Money and Good Intentions Are Not Enough Or, Why a Liberal Democrat Thinks States Need Both Competition and Community*. Brookings.
Brandl, John E. (1998b) "Governance and Educational Quality."(pp. 55_82) in Paul Peterson and Bryan C. Hassel (editors), *Learning from School Choice*. Washington, DC: The Brookings Institution Press.
Chong, Dennis (1991), *Collective Action and the Civil Rights Movement*. Chicago: Chicago University Press.
Coase, R. "The Problem of Social Cost," *Journal of Law and Economics*. v. 3, 1960: 1 _ 44.
Collier, Kenneth, Peter C. Ordeshook, and Kenneth C. Williams. (1989) The Rationally Uninformed Electorate: Some Experimental Evidence." *Public Choice*. vol. 60: 3_29.
Dixit, Avinash & Susuan Skeath (1999). *Games of Strategy*. New York: Norton.
Downs, Anthony (1957) *An Economic Theory Of Democracy*. New York: Harper and Row.
Eavey, C. & G. Miller, "Bureaucratic Agenda Control: Imposition or Bargaining?" *American Political Science Review*, 78, September, 1984: 719_733.
Enelow, J. and M. Hinich. (1984) *The Spatial Theory of Voting*. Cambridge Univ. Press: Cambridge, UK.
Friedman, Jeffrey (ed.), (1995) *The Rational Choice Controversy: Economic Models of Politics Reconsidered*. New Haven, Conn.: Yale University Press.
Frohlich, Norman and Joe A. Oppenheimer (1970). "I Get by with a Little Help from My Friends," *World Politics,* XI (October, 1970), pp. 104_121.
Frohlich, Norman and Joe A. Oppenheimer (1978). *Modern Political Economy*, Englewood Cliffs, New Jersey: Prentice Hall.
Giere, Ronald N. (1984) *Understanding Scientific Reasoning,* 2nd ed. Holt, Rinehart and Winston: Chicago, Ill.
Green, Donald P. and Ian Shapiro (1994) *Pathologies of Rational Choice Theory: A Critique of Applications in Political Science*. New Haven, Conn: Yale University Press.
Grossman, Philip and Catherine C. Eckel (1996), "Anonymity and Altruism in Dictator Games" *Games and Economic Behavior*:
Hardin, R. (1971) "Collective Action as an Agreeable N-Prisoners' Dilemma" *Behavioral Science*. 16 (no. 5): 472 - 479.
Hempel, Carl G. (1966) *The Philosophy of the Natural Sciences*. Englewood Cliffs: NJ. Prentice Hall.
Hoffman, Elizabeth, Kevin McCabe and Vernon L. Smith (1996). "Social Distance and Other_Regarding Behavior in Dictator Games." *American Economic Review*. V. 86, No. 3 (June): 653 _ 660.
Jeffrey, Richard. (1981) *Formal Logic: Its Scope and Limits*. (2nd ed.). McGraw Hill: NY.

Ledyard, John O. (1995). "Public Goods: A Survey of Experimental Research," in *The Handbook of Experimental Economics*, John H. Kagel and Alvin E. Roth, eds. Princeton University Press: Princeton. pp. 111 _ 194.
Kreps, David M. (1990) *Game Theory and Economic Modelling*. Clarendon Lectures in Economics. Oxford University Press: Oxford.
Landa, Janet. 1986. "The Political Economy of swarming in honeybees: Voting with the wings, decision_making costs, and the unanimity rule." *Public Choice* 51 (1): 25 _ 38.
Lupia, Arthur and Mathew D. McCubbins (1995). *The Triumph of Reason: Knowledge and the Foundation of Democracy.* (draft)
Miller, Gary J. and T. H. Hammond, (1987) "The Core of the Constitution" *American Political Science Review*. Vol 81, 1156 _ 1174.
Miller, Gary J. and T. H. Hammond, (1990) "Committees and the Core of the Constitution," *Public Choice*. Vol 66, No. 3, 201 _ 228.
Morris, Irwin (2000). *Congress, the President, and the Federal Reserve: The Politics of American Monetary Policymaking.* Ann Arbor: Univ. of Michigan Press.
Moe, Terry (1984), 'The New Economics of Organization,' *American Journal of Political Science*, vol. 28 (November): 739_777.
Mueller, Dennis C. (1989). *Public Choice II.* Cambridge Univ Press, Cambridge, UK.
Olson, Mancur, (1965) *The Logic of Collective Action*. Harvard University Press: Cambridge.
Oppenheimer, Joe A. (1985) "Public Choice and Three Ethical Properties of Politics." *Public Choice.* 45: 241 _ 255.
Osherson, Daniel N. (1995). "Probability Judgement," in Edward E. Smith and Daniel N. Osherson, eds. *An Invitation to Cognitive Science: Thinking* Volume 3, Second Edition, Cambridge, Mass.: MIT Press. 35 _76.
Charles R. Plott, (1978) "Rawls's Theory of Justice: An Impossibility Result." in Hans W Gottinger and Werner Leinfellner (eds.), *Decision Theory and Social Ethics, Issues in Social Choice.* Reidel Publ.: Dordrecht, Holland.
Popkin, Samuel L. (1991). *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns.* Chicago University Press: Chicago, Ill.
Popper, Karl (1959) *The Logic of Scientific Discovery*. New York: Harper and Row.
Quattrone, George A. and Amos Tversky, (1988) "Contrasting Rational and Psychological Analyses of Political Choice." *American Political Science Review*. (82, No. 3 Sept.) 719_736.
Romer, Thomas and Howard Rosenthal (1978) "Political Resource Allocation, Controlled Agendas and the Status Quo," *Public Choice.* 33 (4): 27 _ 43.
Schelling, Thomas C. (1973) "Hockey Helmets, Concealed Weapons, and Daylight Savings: A Study of Binary Choices with Externalities." *Journal of Conflict Resolution*, v. 17, No. 3 (September), pp. 381 _428.
Schwartz, T. (1981) "The Universal Instability Theorem," *Public Choice*, 37, no. 3, 487 _ 502.
Sen, A.K. (1977) "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory," Philosophy and Public Affairs v. 6 (no. 4, Summer): 317_344. Reprinted in Jane J. Mansbridge, ed. BEYOND SELF_INTEREST. Chicago U. Press: Chicago (1990) (p. 25_43).
Shepsle K. & Barry Weingast (1981), "Structure Induced Equilibrium and Legislative Choice," *Public Choice*, 37, no. 3, 503_520.
Simon (1986) *Journal of Business,* v. 59, no. 4 pt. 2, pp.

Tversky, Amos and Daniel Kahneman (1986), "Rational Choice and the Framing of Decisions," *Journal of Business*, v. 59, no. 4 pt. 2, pp. s251_s278. Reprinted in Karen Schweers Cook and Margaret Levi, eds. The Limits of Rationality. Chicago: U of Chicago Press, 1990: 90_131
White, Alan R. (1967) "Coherence Theory of Truth," *Encyclopedia of Philosophy*. Vol. 2, Macmillan: New York, pp. 130_132.
White, Alan R. (1970) *Truth*. Doubleday Anchor: Garden City, New York.

Table of Contents: POLITICS FROM ANARCHY TO DEMOCRACY: AN INTRODUCTION

The Structure of Institutions

Although popular impressions are that rational choice models are radically simplified, the models' conclusions are always contingent on the details of institutional and political settings. Hence, explaining actual individual choices and how they aggregate into group outcomes, always depends on a rich understanding of fundamental aspects of the choice environment: the characteristics of the institutions, the nature of the options or alternatives, the constraints and values held by the individuals, and the information the individual has (see, as an example, the chapter by Weingast).

But not *all* institutional details are needed to explain the functioning of institutions. Rather, each model specifies those aspects which are relevant to the general patterns of individual choice induced by the institutional structures themselves.[25] To analyze the more general problem of institutions, and institutional design, one must examine the manner (i.e. rules) by which individual behaviors are expressed and aggregated and how incentives are structured in those choice situations. Often one can infer how the structures are likely to work, without details of the values of the particular individuals underlying each individual's decision. This is because these very specific aspects of institutions and their social context can give us enough of an answer to the above aspects that much of the general pattern of behavior can be inferred.[26] (This is certainly much of the thrust in the chapters by Miller and Falaschetti, and Bates.)

The application of rational choice to the field of 'institutions' has led to what some have come to call the 'new economics of organization' (Moe, 1984). Its hallmarks are captured in Brandl (1998a & b) who describes the major premises as:

1. We can best understand the behavior of an organization as the aggregated behaviors of its members.
2. The members of an organization are ordinarily best understood as behaving to further their own interests. So we should begin with their preferences.
3. The interests of the members will diverge from one another as well as from the supposed publicly stated purposes of the organization and we should therefore not presuppose the outcome by the statement of the function of the institution.

Two other premises are used to explain why the performance of institutions often are not what we would want or expect. These two premises can be stated as.

1. It is costly, if not impossible, to hold the individuals responsible for their behaviors.
2. Organizational 'failure' stems from behavior motivated by these divergences of individual interest from one another and the goals of the organization because of a lack of constraint on their behaviors.

We can often reuse the same rational choice models of behavior to understand different organizations and situations. We now briefly consider some of the basic models that give us a foundation to understand the arguments in the chapters below. The theoretically informed reader

[25]This implies that behind one's analysis of a particular problem lies a proper choice of model (see Giere, 1984). This is similar to medical practice: proper diagnosis is critical for a good application of the biological disease theory.

[26]Many of the essays in this volume bring a rational choice model to bear on a more detailed, more contextualized understanding of a single case, so as to shed light on how one goes about analyzing the particulars of politics which make up the subject matter of our discipline.

may wish to skip ahead to page ??, or if uninterested in some of the epistemological questions, to the chapters themselves.

Table 1: 2 Person Prisoners? Dilemma Game		
Your Strategy	**Your Partner?s Strategy**	
	Don?t Cooperate w Police	Cooperate w Police
Don?t Cooperate w Police	-3,-3	-20,0
Cooperate w Police	0,-20	-10, -10

Table 2:[27] **Voting for a Public Works Bill ? an n-Person Prisoner Dilemma Game**

NOTE: The Benefit to District i of a project in the district = B_I
The Benefit to District i of a project in any other district = 0
The Cost of each project = C
Cost of each project to any one district, I: $C_I = C / 435$
$C > B > C/435$

Your strategy	**434 others request n projects where n equals**	
	434	**0**
Don?t request a project	$0 - 434C_I / 435 << 0$	0
Request a project	$B_I ? 435C_I / 435 < 0$	$B_I ? (C_I / 435) > 0$

[27] **In each cell in this table, the entry in the cell is the payoff to your district given the behavior of all others, as summarized by the heading of the column you find yourself in.**

Table 3: A Chicken Game

Row Strategies	**Column Strategies**	
	Swerve	**Head On**
Swerve	**0,0**	**<u>-5,10</u>**
Head On	**<u>10,-5</u>**	-100,-100

Note that the Nash equilibria are indicated in bold and underlined.

Table 4: A Chicken Game Over Time

Row Strategies	**Column Strategies**			
	Always Swerve	**Always Head On**	**Match**	**Opposite (dominant strat)**
Swerve	**0,0**	**-5,10**	0,0	-5,10
Head On	**10,-5**	-100,-100	-100,-100	**10,-5**

Note that the Nash equilibria are indicated in bold and underlined.

Table 5: Value of Donating, Given Behaviors of Others:
Illustrating the Problem of Collective Action

Iris? (Typical Person's) Strategy	Number of Others Who Donate					
	9	8	...	2	1	0
Contribute	$11.50	$11.25	...	$9.75	$9.50	$9.25
Don't	$12.25	$12.00	...	$10.50	$10.25	$10.00

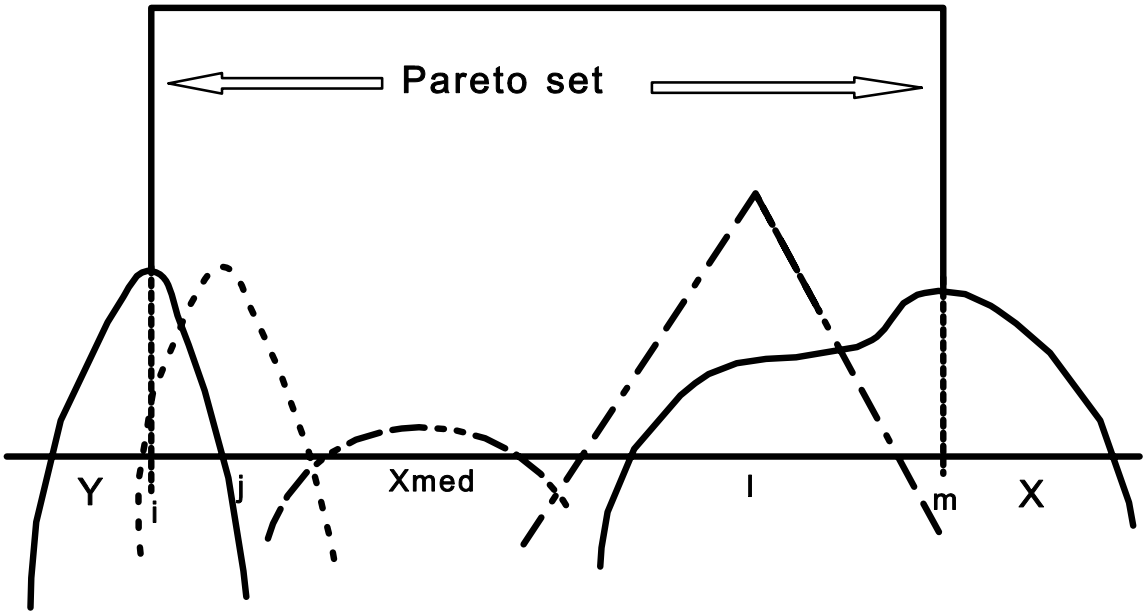
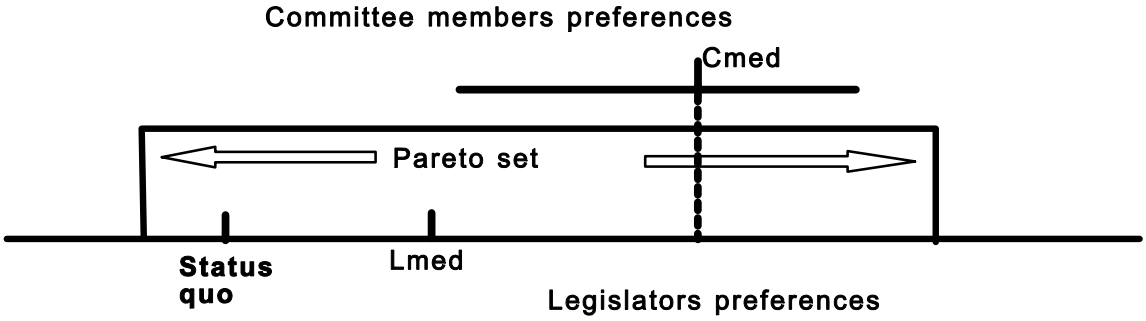

Pareto set
Y
i
j
Xmed
l
m
X

Figure 1-2: Modeling a Legislative Committee's Strategy for Writing a Bill



Apostrophe needed after “members” and “Legislators.”

Figure 1-3: Majority Rule when Preferences are Single-Peaked in 2 Dimensions

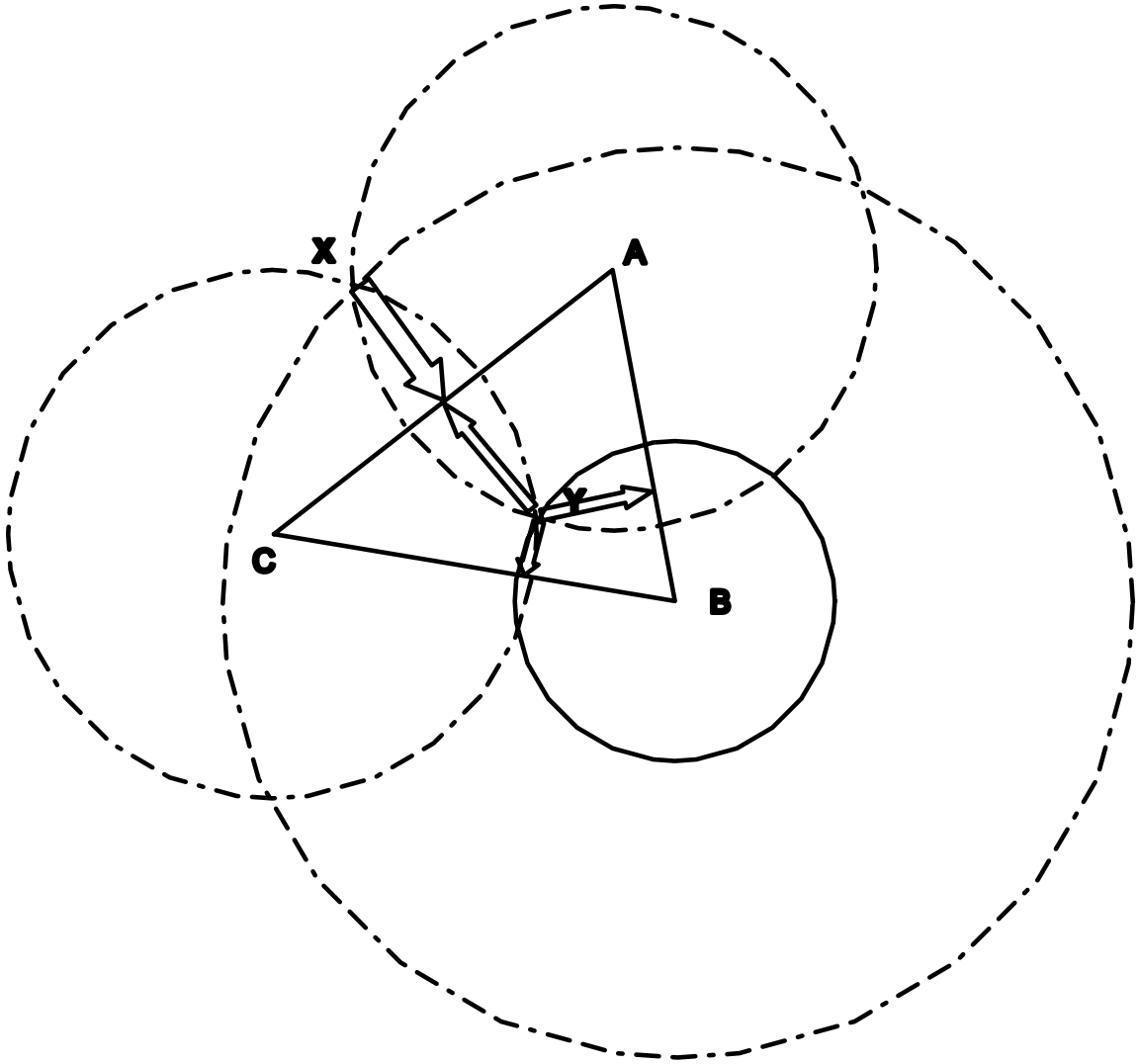


Figure 1-4 Game Tree for the Game of Chicken

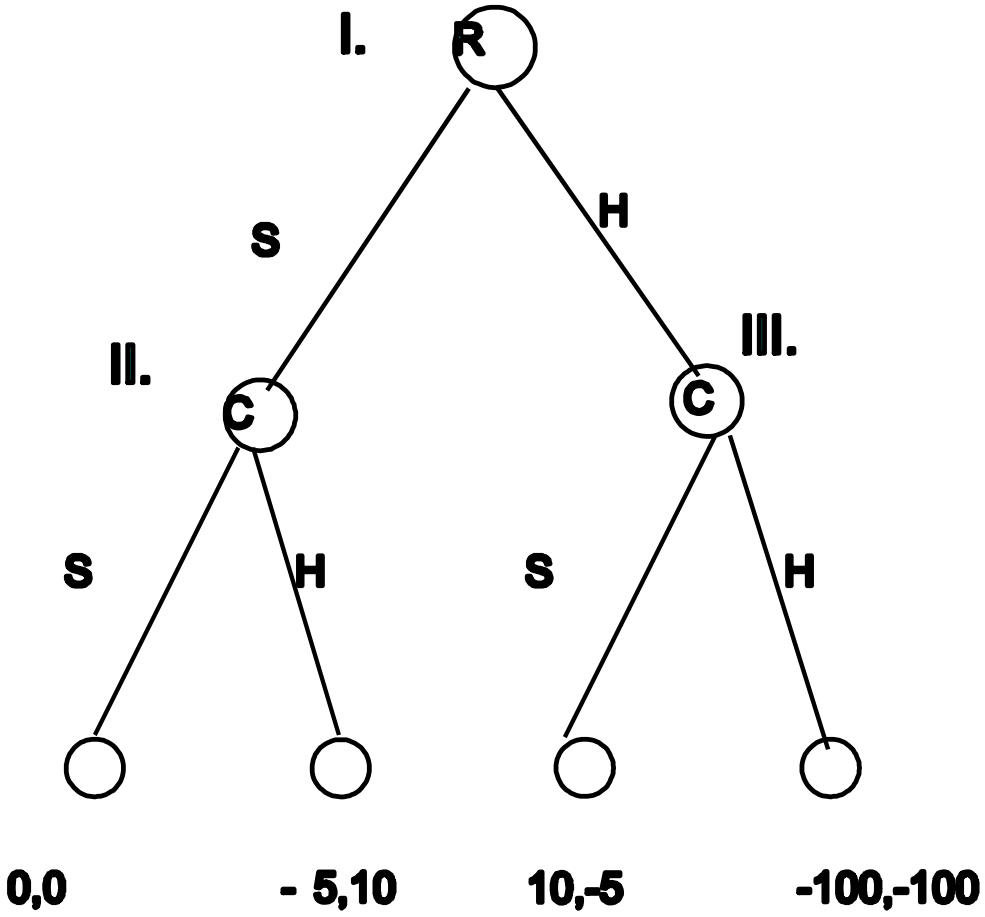


Figure 1-5: Graph Showing Value of Contributing or Not Given Others' Behaviors in a PD

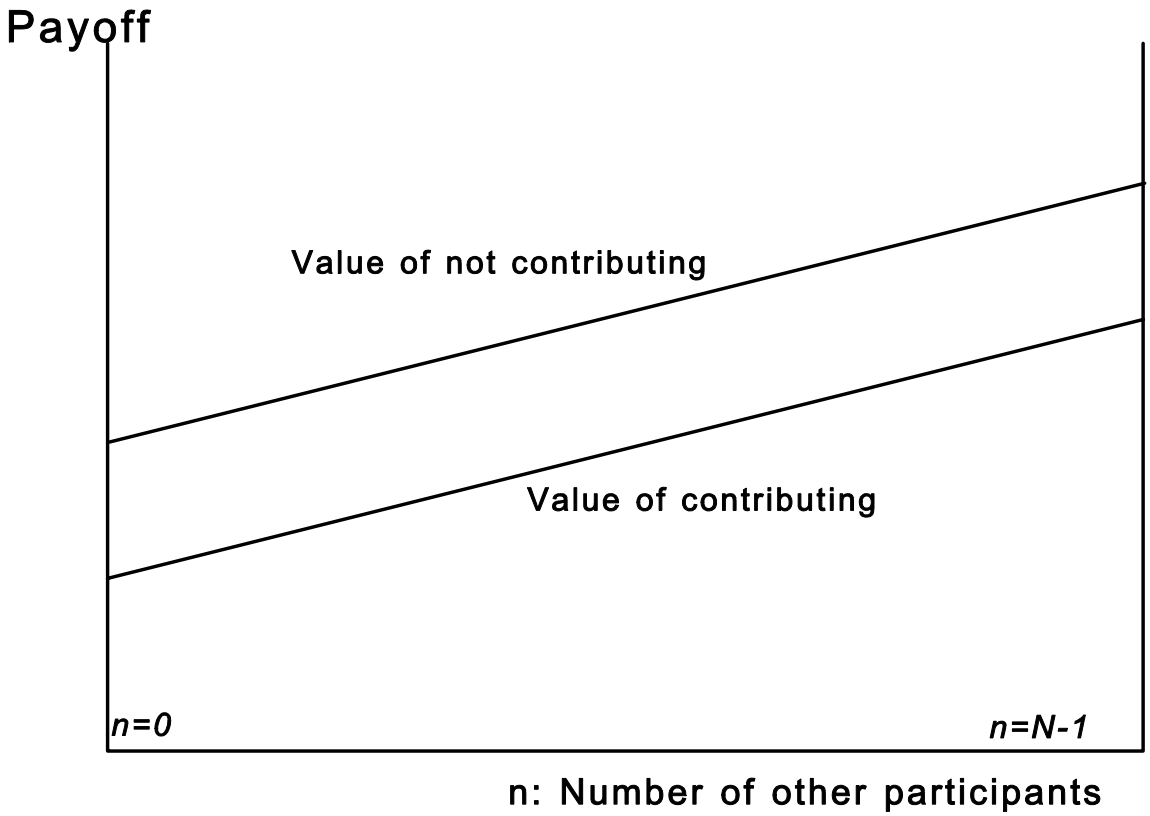
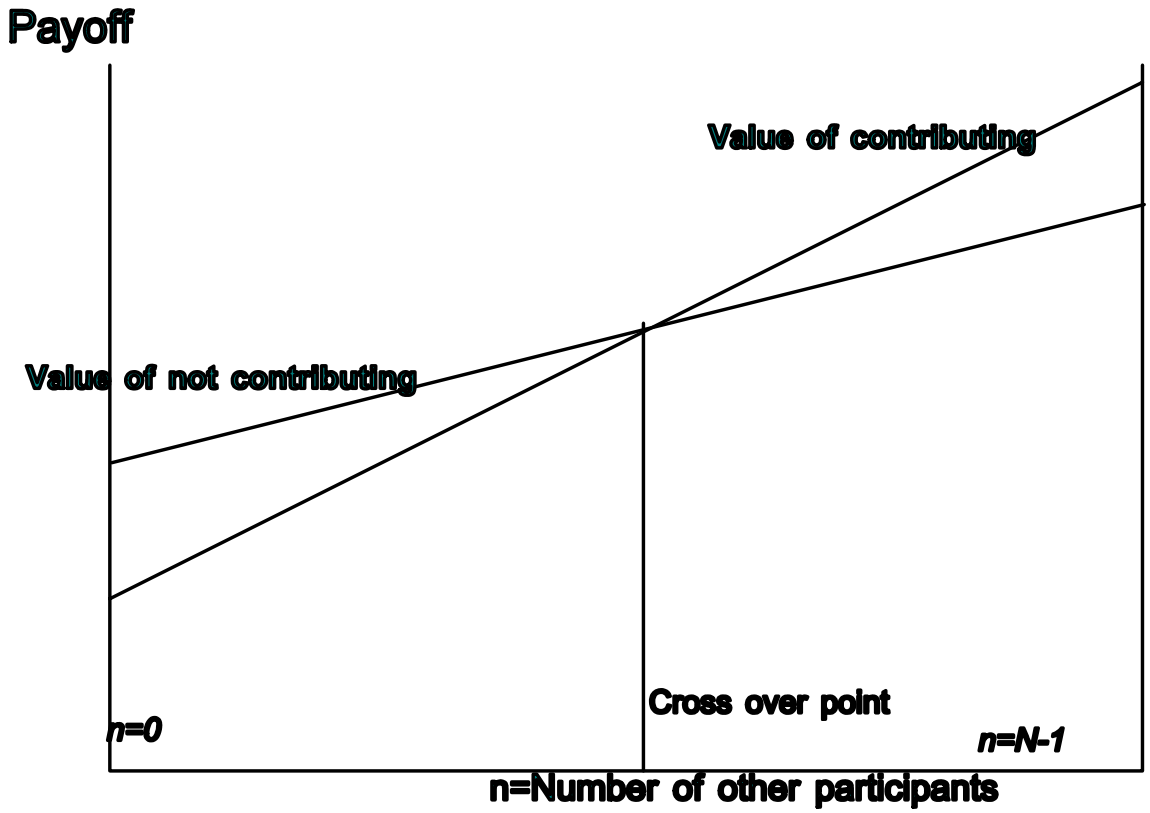


Figure 1-6: Graph Showing Value of Contributing or Not Given Others' Behaviors in an Assurance Game



"Crossover" should be all one word.