1 Introduction

In a rather short period of time, game theory has become one of the most powerful analytical tools in the study of politics. From their earliest applications in electoral and legislative behavior, game theoretic models have proliferated in such diverse areas as international security, ethnic cooperation, and democratization. Indeed all fields of political science have benefited from important contributions originating in game theoretic models. Rarely does an issue of the *American Political Science Review*, the *American Journal of Political Science*, or *International Organization* appear without at least one article formulating a new game theoretic model of politics or one providing an empirical test of existing models.

Nevertheless, applications of game theory have not developed as fast in political science as they have in economics. One of the consequences of this uneven development is that most political scientists who wish to learn game theory are forced to rely on textbooks written by and for economists. Although there are many excellent economic game theory texts, their treatments of the subject are often not well suited to the needs of political scientists. First and perhaps most important, the applications and topics are generally those of interest to economists. For example, it is not always obvious to novice political scientists what duopoly or auction theory tells us about political phenomena. Second, there are topics such as voting theory that are indispensable to political game theorists but receive scant coverage in economics texts. Third, many economics treatments presume some level of exposure to ideas in classical price theory. Consequently, the entry barriers to political scientists include not only mathematics but also knowledge of demand curves, marginal rates of substitution, and the like.

Certainly, there have been a few texts by and for political scientists such as those by Ordeshook (1986) and Morrow (1994). We feel, however, that each is dated both in terms of the applications and in terms of the needs of modern political science. Ordeshook remains an outstanding treatment of social choice and spatial theory, yet it was written well before the emergence of noncooperative theory as the dominant paradigm in political game theory. Morrow provides an accessible introduction to the tools of noncooperative game theory, but the analytical level falls short of the contemporary needs of students. Further, it has been a decade since its publication – a decade in which there have been hundreds of important articles and books deploying the tools of game theory. In a more recent series of books, Austen-Smith and Banks (1999 and 2005) address part of this need. The first book, *Positive Political Theory I*, provides a thorough treatment of social choice theory, a topic to which we devote only one chapter. The second book, *Positive Political Theory II*, deals with strategy and institutions, but presumes a knowledge of game theory atypical of first-year students in political science. It is also organized by substantive topics rather than game theoretic ones.

So we have several goals in writing this book. First, we want to write a textbook on political game theory instead of a book on abstract or economic game theory. Consequently, we focus on applications of interest to political scientists and present topics unique to political analysis. Second, in writing a book for political scientists, we want to be cognizant of the diversity of backgrounds and interests of young political scientists. We recognize that most doctoral students in political science enter graduate school with limited mathematical and modeling backgrounds. We feel, however, that it does not serve even those students to ignore the mathematical rigor and key theoretical concepts on which contemporary political models are based. For students needing more remediation, we include a detailed mathematical appendix covering some necessary tools ranging from set theory and analysis to basic optimization and probability theory. Some students enter graduate study in political science with stronger backgrounds in mathematics and economics. We want our book to be useful to this audience as well. Thus, we provide in-depth coverage of some of the more difficult and subtle concepts. We include a number of advanced sections (denoted by * or **) that provide more detail about the analytical and mathematical structure of the models we encounter. These sections can be safely skipped upon first readings by those not quite ready for the more technical material.

1. Organization of the Book

Organizationally, our book departs from standard treatments, because it includes a number of topics that are either directly relevant for political science or designed for remediation in areas in which students of political science have limited backgrounds. Chapter 2 is a self-contained exposition of classical choice theory under conditions of certainty. In this chapter, we introduce the basic ideas of preferences and utility theory. We prove a few key results. Some of these proofs are quite simple, and others appeal to more advanced mathematics and appear in starred sections. The focus of this chapter, however, is on providing the intuition and language of rational choice theory. We also include a section on spatial or Euclidean preferences. This class of preferences plays a central role in voting theory and its application to electoral and legislative politics.

In Chapter 3, we describe how game theorists model choices under uncertainty. The focus is the standard von Neumann-Morgenstern expected utility model, but we also consider some of the most serious criticisms leveled against it. In addition to a standard treatment of risk preferences, we discuss the special implications for risk when actors have spatial preferences.

Chapter 4 provides a cursory review of social choice theory. The chapter is not intended to be a replacement for full-length texts such as those by Peter Ordeshook (1986) and David Austen-Smith and Jeff Banks (1999). Instead it is primarily a reference for those ideas and concepts that have become integral parts of formal political science. These include Arrow's impossibility theorem, the emptiness of the majority core, and the role of single-peaked preferences. This chapter also presents Gibbard-Sattherwaite's theorem about the ubiquity of strategic behavior in social decisions.

Chapter 5 begins our treatment of the heart of contemporary formal political theory: noncooperative game theory. We examine normal form games with complete information and present the most fundamental solution concepts, dominance and Nash equilibrium. Our theoretical development is fairly standard, but we include a number of important political applications. We review the standard Downsian model of electoral competition as well as the extensions developed by Donald Wittman and Randy Calvert. We also present several models of private contributions to public goods based on the work of Thomas Palfrey and Howard Rosenthal. In Chapter 6, we extend the normal form model to cases where agents are uncertain about the payoffs associated with different strategy combinations. After presenting solution concepts for such games, Bayesian Nash equilibria, we consider incomplete information versions of many of the models reviewed in Chapter 5. These comparisons aid understanding of the strategic implications of uncertainty.

Chapter 7 considers dynamic, multistage games of complete information and develops the notion of subgame perfection. Here we focus on a number of applications from legislative politics, democratic transitions, coalition formation, and international crisis bargaining. In Chapter 8, we consider dynamic games in which some players are imperfectly informed about the payoffs of different strategic choices. After explaining how these models are solved, we explore applications drawn from legislative politics, campaign finance, and international bargaining. Signaling games, which have increasingly important applications in political science, are the focus of much of this chapter.

Chapter 9 reviews the theory of repeated games and its application to political science. The role of time discounting and the structure of folk theorems in repeated games are the primary focus of the chapter. Applications include interethnic cooperation and trade wars.

In Chapter 10 we consider applications of bargaining theory. Beginning with the canonical models of Nash and Rubinstein, we focus on the majority-rule bargaining game developed by Baron and Ferejohn. We then consider several examples of bargaining with incomplete information.

In Chapter 11, we illustrate the mechanism design approach to modeling institutions. Our focus is the selection of games that induce equilibrium behavior that meets certain prespecified goals. After presenting the revelation principle and incentive compatibility conditions, we trace out a number of recent applications to electoral politics and organizational design. Building on Chapter 8, we then draw connections between signaling games and mechanism design.

Finally, to keep the book as self-contained as possible, Chapter 12 provides a review of all of the mathematics used. Topics that are integral to the development of key theoretical results or tools for analyzing applications are drawn from the fields of set theory, real analysis, linear algebra, calculus, optimization, and probability theory. Indeed this chapter may serve as a basis for review or self-study. Students interested in working at the frontier of political game theory are encouraged to seek additional course work in order to gain comfort with the mathematical concepts summarized in this appendix.

2 The Theory of Choice

Much of political game theory is predicated on the idea that people rationally pursue goals subject to constraints imposed by physical resources and the expected behavior of other actors. The assumption of rationality is often controversial. Indeed one of the most lively debates in the social sciences is the role of rationality and intentionality as a predictor of behavior. Nevertheless, we omit the debate between *Homo economicus* and *Homo sociologicus* and jump immediately into the classical model of rational choice.

For almost all of our purposes, it is sufficient to define rationality on a basis of two simple ideas:

(1) Confronted with any two options, denoted x and y, a person can determine whether he does not prefer option x to option y, does not prefer y to x, or does not prefer either. When preferences satisfy this property, they are *complete*.
(2) Confronted with three options x, y, and z, if a person does not prefer y to x and does not prefer z to y, then she must not prefer z to x. Preferences satisfying this property are *transitive*.

Roughly speaking, our working definition of rational behavior is behavior consistent with complete and transitive preferences. Sometimes we call such behavior *thinly* rational, as properties 1 and 2 contain little or no substantive content about human desires. Thin rationality contrasts with *thick* rationality whereby analysts specify concrete goals such as wealth, status, or fame. The thin characterization of rationality is consistent with a very large number of these substantive goals. In principal, thinly rational agents could be motivated by any number of factors including ideology, normative values, or even religion. As long as these belief systems produce complete and transitive orderings over personal and social outcomes, we can use the classical theory of choice to model behavior.

Although it is appealing to avoid explicit assumptions about substantive goals, it is often necessary to make stronger assumptions about preferences. For example, a model might assume that an interest group wishes to maximize the wealth of its members or that a politician wishes to maximize her reelection chances. In subsequent chapters, we explore models that make such assumptions about agent preferences. But rational models may be just as useful in developing models of activists who wish to minimize environmental degradation or the number of abortions for principled, nonmaterial reasons.

In the following sections, we develop the classical theory of choice under *certainty*. By certainty, we mean simply that each agent has sufficient information about her available set of actions that she can perfectly predict the consequences of each. Later we examine choice under uncertainty – where the actor's lack of information forces her to choose among actions with uncertain consequences.

1. Finite Sets of Actions and Outcomes

We begin with the simplest description of a choice problem: an agent chooses an action from a finite list. We denote these alternatives as a set $A = \{a_1, \ldots, a_k\}$. A leader involved in an international crisis might face the following set of alternatives: $A = \{$*send troops*, *negotiate*, *do nothing*$\}$. An American voter might choose among $A = \{$*vote Democrat*, *vote Republican*, *abstain*$\}$.

As mentioned, we assume, for now, that agents have *complete* information – they are sufficiently knowledgeable that they perfectly predict the consequences of each action. To formalize this idea, we define outcome sets as $X = \{x_1, \ldots, x_n\}$. In our crisis example, let $X = \{$*win major concessions and lose troops*, *win minor concessions, status quo*$\}$. The assumption of certainty implies that each action $a \in A$ maps directly onto one and only one $x \in X$. Formally, certainty implies that there exists a function $x : A \to X$ that maps each action into a specific outcome. For convenience, we also assume that all of the outcomes listed in X are feasible – each outcome is the consequence of at least one action. Thus, x_i is feasible if there exists an $a \in A$ such that $x(a) = x_i$. With certainty and feasibility, it makes no difference whether we speak of an agent's preferences over actions or his preferences over outcomes. Consequently, we concentrate on the agent's preferences over outcomes. In Chapter 3, the assumption of uncertainty or *incomplete* information makes the distinction between actions and outcomes relevant.

To generate predictions about choice behavior, we require a more formal notion of preferences. *Weak preference* is captured by a binary relation R where the notation $x_i R x_j$ means that outcome x_j is not preferred to policy x_i. If $x_i R x_j$, x_i is "weakly" preferred to x_j.[1] By way of analogy, note that R is similar to the binary relation $\geq$ (greater than or equal) that operates on real numbers.

Beyond the weak preference relation R, we define two other important binary relations: strict preference and indifference.

DEFINITION 2.1 *For any $x, y \in X$, xPy (x is strictly preferred to y) if and only if xRy and not yRx. Alternatively, xIy (x is indifferent to y) if and only if xRy and yRx.*

Accordingly, P denotes strict preference and I denotes indifference. Returning to the analogy of $\geq$, the strict relation derived from $\geq$ is equivalent to the relation $>$ and the indifference relation derived from $\geq$ is equivalent to the relation $=$.

Although preferences expressed in the form of binary relations are useful concepts, we are ultimately interested in behavior. Given a set of preferences, an agent's behavior is rational so long as she selects an outcome that she values at least as much as any other. Consequently, a rational agent chooses an $x^* \in X$ (read x^* in X) such that $x^* Ry$ for every $y \in X$. Without adding more structure to preferences, however, there is no guarantee that such an optimal outcome exists. We now turn to the conditions on X and R to ensure that such a best choice is meaningful and well defined. We begin with the following formal definition.

DEFINITION 2.2 *For a weak preference relation R on a choice set X, the maximal set $M(R, X) \subset X$ is defined as $M(R, X) = \{x \in X : xRy\ \forall\ y \in X\}$ (read as $M(R, X)$ is the set of x's in X such that xRy for all y in X).*

The fundamental tenet of rationality is that *agents choose outcomes from the maximal set.* Of course, this requirement is meaningful only if

[1] Formally, a binary relation R is a subset of $X \times X$ such that if $(x, y) \in R$ then xRy.

the maximal set contains at least one outcome. Consequently, we are interested in the properties of preferences that guarantee that $M(R, X)$ is nonempty.

The easiest way for the maximal set to be empty is for R to be silent between a pair of outcomes. If neither xRy or yRx, it is not clear what a rational choice is. Two conditions ensure that all elements of X are ordered.

DEFINITION 2.3 *A binary relation R on X is*

(1) complete if for all $x, y \in X$ with $x \neq y$, either xRy or yRx or both.
(2) reflexive if for all $x \in X$, xRx.

Completeness means simply that the agent can compare any two outcomes. This may not be a terribly controversial assumption, but we all know people who cannot seem to make up their minds.[2] Reflexivity is a more technical condition. Some authors choose to define completeness in a slightly different manner that also captures reflexivity.[3]

Although these properties rule out the noncomparability problem, completeness and reflexivity do not ensure that rational choices exist. We also must rule out the following problem: xPy, yPz, and zPx. The problem is that there is no reasonable choice – why choose y when you can choose x, why choose x when you can choose z, and why choose z when you can choose y? Each of the following restrictions on preferences resolves this problem.

DEFINITION 2.4 *A binary relation R on X is*

(1) transitive if xRy and yRz implies xRz for all $x, y, z \in X$.
(2) quasi-transitive if xPy and yPz implies xPz for all $x, y, z \in X$.
(3) acyclic if on any finite set $\{x_1, x_2 \ldots, x_n\} \in X$ x_iPx_{i+1} for all $i < n$ implies x_1Rx_n.

Note the subtle differences among these definitions. Transitivity and quasi transitivity may seem innocuous, but they are strong assumptions that might be violated even by very reasonable behavior. For example,

[2] Many economists and psychologists, however, have been concerned about the assumption of completeness. A theory of choice without this condition has been derived.
[3] For all $x, y \in X$, either xRy or yRx or both.

suppose X is a set of 1,000 different bottles of beer. Beer b_1 has had one drop of beer replaced with one drop of plain water, b_2 has had two drops replaced, and so on, to $b_{1,000}$. Unless one is a master brewer, $b_1 I b_2$, and $b_2 I b_3, \ldots,$ and $b_{999} I b_{1,000}$. Because $x I y$ implies $x R y$ (by the definition of I), then $b_{1,000} R b_{999}, \ldots, b_2 R b_1$. If the relation is transitive, we derive $b_{1,000} R b_1$. But clearly, $b_1 P b_{1,000}$.[4] The assumption of acyclicity does not suffer from this problem, however, and is typically sufficient for our purposes. Despite the problems associated with transitivity, we maintain it as an assumption (rather than acyclicity) to simplify many of the results that follow.

The properties of completeness, reflexivity, and transitivity together form the basis of a *weak ordering*.

DEFINITION 2.5 *Given a set X, a weak ordering is a binary relation that is complete, reflexive, and transitive.*

Our recurring analogy of $\geq$ satisfies all of the conditions for a weak ordering. We now state our first result.

THEOREM 2.1 *If X is finite and R is a weak ordering then $M(R, X) \neq \emptyset$.*

Theorem 2.1 guarantees that there is a best choice so long as the choice set is finite and that R is complete, reflexive, and transitive. Its proof follows.

Proof Let X be finite and R be complete, reflexive, and transitive. We establish the result by induction (see Mathematical Induction in the Mathematical Appendix) on the number of elements in X.

Step 1: If X has one element, $X = \{x\}$. From reflexivity $x R x$, $M(R, X) = \{x\}$.

Step 2: We show that if the statement of the theorem is true that for any set X' with n elements and weak ordering R' on X' then it must be true for any X with $n+1$ elements and weak ordering R on X.

Proof of Step 2: Assume that $M(R', X') \neq \emptyset$ for any X' with n elements and weak ordering R'. Now consider a set X with $n+1$ elements and any weak ordering R. For an arbitrary $x \in X$, $X = X' \cup \{x\}$ with X' a set having n elements. Let R' denote the restriction of R to X' (i.e.,

[4] This is approximately the difference between Guinness and Coors Light.