1 Introduction

1.1 What is game theory?

GAME THEORY aims to help us understand situations in which decision-makers interact. A game in the everyday sense—"a competitive activity ... in which players contend with each other according to a set of rules", in the words of a dictionary—is an example of such a situation, but the scope of game theory is very much larger. Indeed, I devote very little space to games in the everyday sense; my main focus is the use of game theory to illuminate economic, political, and biological phenomena.

A list of some of the applications I discuss will give you an idea of the range of situations to which game theory can be applied: firms competing for business, political candidates competing for votes, jury members deciding on a verdict, animals fighting over prey, bidders competing in an auction, the evolution of siblings' behavior towards each other, competing experts' incentives to correctly diagnose a problem, legislators' voting behavior under pressure from interest groups, and the role of threats and punishment in long-term relationships.

Like other sciences, game theory consists of a collection of models. A model is an abstraction we use to understand our observations and experiences. What "understanding" entails is not clear-cut. Partly, at least, it entails our perceiving relationships between situations, isolating principles that apply to a range of problems, so that we can fit into our thinking new situations that we encounter. For example, we may fit our observation of the path taken by a lobbed tennis ball into a model that assumes the ball moves forward at a constant velocity and is pulled towards the ground by the constant force of "gravity". This model enhances our understanding because it fits well, no matter how hard or in which direction the ball is hit, and applies also to the paths taken by baseballs, cricket balls, and a wide variety of other missiles, launched in any direction.

A model is unlikely to help us understand a phenomenon if its assumptions are wildly at odds with our observations. At the same time, a model derives power from its simplicity; the assumptions upon which it rests should capture the essence of the situation, not irrelevant details. For example, when considering the path

AN OUTLINE OF THE HISTORY OF GAME THEORY

Some game-theoretic ideas can be traced to the 18th century, but the major development of the theory began in the 1920s with the work of the mathematician Émile Borel (1871–1956) and the polymath John von Neumann (1903–57). A decisive event in the development of the theory was the publication in 1944 of the book *Theory of games and economic behavior* by von Neumann and Oskar Morgenstern, which established the foundations of the field. In the early 1950s, John F. Nash (see the box on page 23) developed a key concept (Nash equilibrium) and initiated the game-theoretic study of bargaining. Soon after Nash's work, game-theoretic models began to be used in economic theory and political science, and psychologists began studying how human subjects behave in experimental games. In the 1970s game theory was first used as a tool in evolutionary biology. Subsequently, game-theoretic methods have come to dominate microeconomic theory and are used also in many other fields of economics and a wide range of other social and behavioral sciences. The 1994 Nobel Prize in Economic Sciences was awarded to the game theorists John C. Harsanyi (1920–2000), John F. Nash (1928–), and Reinhard Selten (1930–).

taken by a lobbed tennis ball we should ignore the dependence of the force of gravity on the distance of the ball from the surface of the earth.

Models cannot be judged by an absolute criterion: they are neither "right" nor "wrong". Whether a model is useful or not depends, in part, on the purpose for which it is used. For example, when I determine the shortest route from Florence to Venice, I do not worry about the projection of the map I am using; I work under the assumption that the earth is flat. When I determine the shortest route from Beijing to Havana, however, I pay close attention to the projection—I assume that the earth is spherical. And were I to climb the Matterhorn I would assume that the earth is neither flat nor spherical!

One reason for improving our understanding of the world is to enhance our ability to mold it to our desires. The understanding that game-theoretic models give is particularly relevant in the social, political, and economic arenas. Studying game-theoretic models (or other models that apply to human interaction) may also suggest ways in which an individual's behavior may be modified to improve her own welfare. By analyzing the incentives faced by negotiators locked in battle, for example, we may see the advantages and disadvantages of various strategies.

The models of game theory are precise expressions of ideas that can be presented verbally. Verbal descriptions tend to be long and imprecise; in the interest of conciseness and precision, I frequently employ mathematical symbols. Although I use the language of mathematics, I use few of its concepts; the ones I use are described in Chapter 17. I aim to take advantage of the precision and conciseness of a mathematical formulation without losing sight of the underlying ideas.

JOHN VON NEUMANN

John von Neumann, the most important figure in the early development of game theory, was born in Budapest, Hungary, in 1903. He displayed exceptional mathematical ability as a child (he had mastered calculus by the age of 8), but his father, concerned about his son's financial prospects, did not want him to become a mathematician. As a compromise he enrolled in mathematics at the University of Budapest in 1921, but immediately left to study chemistry, first at the University of Berlin and subsequently at the Swiss Federal Institute of Technology in Zurich, from which he earned a degree in chemical engineering in 1925. During his time in Germany and Switzerland he returned to Budapest to write examinations, and in 1926 obtained a Ph.D. in mathematics from the University of Budapest. He taught in Berlin and Hamburg, and, from 1930 to 1933, at Princeton University. In 1933 he became the youngest of the first six professors of the School of Mathematics at the Institute for Advanced Study in Princeton (Einstein was another).

Von Neumann's first published scientific paper appeared in 1922, when he was 19 years old. In 1928 he published a paper that establishes a key result on strictly competitive games, a result that had eluded Borel. He made many major contributions in pure and applied mathematics and in physics—enough, according to Halmos (1973), "for about three ordinary careers, in pure mathematics alone". While at the Institute for Advanced Study he collaborated with the Princeton economist Oskar Morgenstern in writing *Theory of games and economic behavior*, the book that established game theory as a field. In the 1940s he became increasingly involved in applied work. In 1943 he became a consultant to the Manhattan Project, which was developing an atomic bomb, and in 1944 he became involved with the development of the first electronic computer, to which he made major contributions. He stayed at Princeton until 1954, when was appointed to the U.S. Atomic Energy Commission. He died in 1957.

Game-theoretic modeling starts with an idea related to some aspect of the interaction of decision-makers. We express this idea precisely in a model, incorporating features of the situation that appear to be relevant. This step is an art. We wish to put enough ingredients into the model to obtain nontrivial insights, but not so many that we are led into irrelevant complications; we wish to lay bare the underlying structure of the situation as opposed to describing its every detail. The next step is to analyze the model—to discover its implications. At this stage we need to adhere to the rigors of logic; we must not introduce extraneous considerations absent from the model. Our analysis may confirm our idea, or suggest it is wrong. If it is wrong, the analysis should help us to understand why it is wrong. We may see that an assumption is inappropriate, or that an important element is missing from the model; we may conclude that our idea is invalid, or that we need to investigate it further by studying a different model. Thus, the interaction between our ideas and models designed to shed light on them runs in two directions: the implications of models help us determine whether our ideas make sense, and these ideas, in the light of the implications of the models, may show us how the assumptions of our models are inappropriate. In either case, the process of formulating and analyzing a model should improve our understanding of the situation we are considering.

1.2 The theory of rational choice

The theory of rational choice is a component of many models in game theory. Briefly, this theory is that a decision-maker chooses the best action according to her preferences, among all the actions available to her. No qualitative restriction is placed on the decision-maker's preferences; her "rationality" lies in the consistency of her decisions when faced with different sets of available actions, not in the nature of her likes and dislikes.

1.2.1 Actions

The theory is based on a model with two components: a set A consisting of all the actions that, under some circumstances, are available to the decision-maker, and a specification of the decision-maker's preferences. In any given situation, the decision-maker is faced with a subset[1] of A, from which she must choose a single element. The decision-maker knows this subset of available choices, and takes it as given; in particular, the subset is not influenced by the decision-maker's preferences. The set A could, for example, be the set of bundles of goods that the decision-maker can possibly consume; given her income at any time, she is restricted to choose from the subset of A containing the bundles she can afford.

1.2.2 Preferences and payoff functions

As to preferences, we assume that the decision-maker, when presented with any pair of actions, knows which of the pair she prefers, or knows that she regards both actions as equally desirable (in which case she is "indifferent between the actions"). We assume further that these preferences are consistent in the sense that if the decision-maker prefers the action a to the action b, and the action b to the action c, then she prefers the action a to the action c. No other restriction is imposed on preferences. In particular, we allow a person's preferences to be altruistic in the sense that how much she likes an outcome depends on some other person's welfare. Theories that use the model of rational choice aim to derive implications that do not depend on any qualitative characteristic of preferences.

[1] See Chapter 17 for a description of mathematical terminology.

How can we describe a decision-maker's preferences? One way is to specify, for each possible pair of actions, the action the decision-maker prefers, or to note that the decision-maker is indifferent between the actions. Alternatively we can "represent" the preferences by a *payoff function*, which associates a number with each action in such a way that actions with higher numbers are preferred. More precisely, the payoff function u represents a decision-maker's preferences if, for any actions a in A and b in A,

$$u(a) > u(b) \text{ if and only if the decision-maker prefers } a \text{ to } b. \qquad (5.1)$$

(A better name than payoff function might be "preference indicator function". In economic theory a payoff function that represents a consumer's preferences is often called a "utility function".)

◆ EXAMPLE 5.2 (Payoff function representing preferences) A person is faced with the choice of three vacation packages, to Havana, Paris, and Venice. She prefers the package to Havana to the other two, which she regards as equivalent. Her preferences between the three packages are represented by any payoff function that assigns the same number to Paris and Venice and a higher number to Havana. For example, we can set $u(\text{Havana}) = 1$ and $u(\text{Paris}) = u(\text{Venice}) = 0$, or $u(\text{Havana}) = 10$ and $u(\text{Paris}) = u(\text{Venice}) = 1$, or $u(\text{Havana}) = 0$ and $u(\text{Paris}) = u(\text{Venice}) = -2$.

? EXERCISE 5.3 (Altruistic preferences) Person 1 cares about both her income and person 2's income. Precisely, the value she attaches to each unit of her own income is the same as the value she attaches to any two units of person 2's income. For example, she is indifferent between a situation in which her income is 1 and person 2's is 0, and one in which her income is 0 and person 2's is 2. How do her preferences order the outcomes $(1,4)$, $(2,1)$, and $(3,0)$, where the first component in each case is her income and the second component is person 2's income? Give a payoff function consistent with these preferences.

A decision-maker's preferences, in the sense used here, convey only *ordinal* information. They may tell us that the decision-maker prefers the action a to the action b to the action c, for example, but they do not tell us "how much" she prefers a to b, or whether she prefers a to b "more" than she prefers b to c. Consequently a payoff function that represents a decision-maker's preferences also conveys only ordinal information. It may be tempting to think that the payoff numbers attached to actions by a payoff function convey intensity of preference—that if, for example, a decision-maker's preferences are represented by a payoff function u for which $u(a) = 0$, $u(b) = 1$, and $u(c) = 100$, then the decision-maker likes c a lot more than b but finds little difference between a and b. *A payoff function contains no such information!* The *only* conclusion we can draw from the fact that $u(a) = 0$, $u(b) = 1$, and $u(c) = 100$ is that the decision-maker prefers c to b to a. Her preferences are represented equally well by the payoff function v for which $v(a) = 0$, $v(b) = 100$, and $v(c) = 101$, for example, or any other function w for which $w(a) < w(b) < w(c)$.

From this discussion we see that a decision-maker's preferences are represented by many different payoff functions. Looking at (5.1), we see that if u represents a decision-maker's preferences and the payoff function v assigns a higher number to the action a than to the action b if and only if the payoff function u does so, then v also represents these preferences. Stated more compactly, if u represents a decision-maker's preferences and v is another payoff function for which

$$v(a) > v(b) \text{ if and only if } u(a) > u(b),$$

then v also represents the decision-maker's preferences. Or, more succinctly, if u represents a decision-maker's preferences, then any increasing function of u also represents these preferences.

? EXERCISE 6.1 (Alternative representations of preferences) A decision-maker's preferences over the set $A = \{a, b, c\}$ are represented by the payoff function u for which $u(a) = 0$, $u(b) = 1$, and $u(c) = 4$. Are they also represented by the function v for which $v(a) = -1$, $v(b) = 0$, and $v(c) = 2$? How about the function w for which $w(a) = w(b) = 0$ and $w(c) = 8$?

Sometimes it is natural to formulate a model in terms of preferences and then find payoff functions that represent these preferences. In other cases it is natural to start with payoff functions, even if the analysis depends only on the underlying preferences, not on the specific representation we choose.

1.2.3 The theory of rational choice

The theory of rational choice may be stated simply: in any given situation the decision-maker chooses the member of the available subset of A that is best according to her preferences. Allowing for the possibility that there are several equally attractive best actions, **the theory of rational choice** is

> *the action chosen by a decision-maker is at least as good, according to her preferences, as every other available action.*

For any action, we can design preferences with the property that no other action is preferred. Thus if we have no information about a decision-maker's preferences, and make no assumptions about their character, any *single* action is consistent with the theory. However, if we assume that a decision-maker who is indifferent between two actions sometimes chooses one action and sometimes the other, not every *collection* of choices for different sets of available actions is consistent with the theory. Suppose, for example, we observe that a decision-maker chooses a whenever she faces the set $\{a, b\}$, but sometimes chooses b when facing the set $\{a, b, c\}$. The fact that she always chooses a when faced with $\{a, b\}$ means that she prefers a to b (if she were indifferent, she would sometimes choose b). But then when she faces the set $\{a, b, c\}$, she must choose either a or c, never b. Thus her choices are inconsistent with the theory. (More concretely, if you choose the same dish from the menu of your favorite bistro whenever there are no specials, then, regardless of your preferences, it is inconsistent for you to choose some other item *from the menu* on a day when there is an off-menu special.)

If you have studied the standard economic theories of the consumer and the firm, you have encountered the theory of rational choice before. In the economic theory of the consumer, for example, the set of available actions is the set of all bundles of goods that the consumer can afford. In the theory of the firm, the set of available actions is the set of all input–output vectors, and the action a is preferred to the action b if and only if a yields a higher profit than does b.

1.2.4 Discussion

The theory of rational choice is enormously successful; it is a component of countless models that enhance our understanding of social phenomena. It pervades economic theory to such an extent that arguments are classified as "economic" as much because they involve rational choices as because they involve particularly "economic" variables.

Nevertheless, under some circumstances its implications are at variance with observations of human decision-making. To take a small example, adding an undesirable action to a set of actions sometimes significantly changes the action chosen (see Rabin 1998, 38). The significance of such discordance with the theory depends upon the phenomenon being studied. If we are considering how the markup of price over cost in an industry depends on the number of firms, for example, this sort of weakness in the theory may be unimportant. But if we are studying how advertising, designed specifically to influence peoples' preferences, affects consumers' choices, then the inadequacies of the model of rational choice may be crucial.

No general theory currently challenges the supremacy of rational choice theory. But you should bear in mind as you read this book that the theory has its limits, and some of the phenomena that you may think of explaining by using a game-theoretic model may lie beyond these limits. As always, the proof of the pudding is in the eating: if a model enhances our understanding of the world, then it serves its purpose.

1.3 Coming attractions: interacting decision-makers

In the model of the previous section, the decision-maker chooses an action from a set A and cares only about this action. A decision-maker in the world often does not have the luxury of controlling all the variables that affect her. If some of the variables that affect her are the actions of *other* decision-makers, her decision-making problem is altogether more challenging than that of an isolated decision-maker. The study of such situations, which we may model as *games*, occupies the remainder of the book.

Consider, for example, firms competing for business. Each firm controls its price, but not the other firms' prices. Each firm cares, however, about all the firms' prices, because these prices affect its sales. How should a firm choose its price in these circumstances? Or consider a candidate for political office choosing a policy platform. She is likely to care not only about her own platform, but those of her rivals, which affect her chance of being elected. How should she choose her platform, knowing that every other candidate faces the same problem? Or suppose you are negotiating a purchase. You care about the price, which depends on both your behavior and that of the seller. How should you decide what price to offer?

Part I presents the main models of game theory: a strategic game, an extensive game, and a coalitional game. These models differ in two dimensions. A strategic game and an extensive game focus on the actions of individuals, whereas a coalitional game focuses on the outcomes that can be achieved by groups of individuals; a strategic game and a coalitional game consider situations in which actions are chosen once and for all, whereas an extensive game allows for the possibility that plans may be revised as they are carried out.

The model within which rational choice theory is cast is tailor-made for the theory. If we want to develop another theory of a single decision-maker, we need to add elements to the model in addition to actions and preferences. The same is not true of most models in game theory: strategic interaction is sufficiently complex that even a relatively simple model can admit more than one theory of the outcome. We refer to a theory that specifies a set of outcomes for a model as a "solution". Chapter 2 describes the model of a strategic game and the solution of Nash equilibrium for such games. The theory of Nash equilibrium in a strategic game has been applied to a vast variety of situations; a few of the most significant applications are discussed in Chapter 3.

Chapter 4 extends the notion of Nash equilibrium in a strategic game to allow for the possibility that a decision-maker, when indifferent between actions, may not always choose the same action.

The model of an extensive game, which adds a temporal dimension to the description of strategic interaction captured by a strategic game, is studied in Chapters 5, 6, and 7. Part I concludes with Chapter 8, which discusses the model of a coalitional game and a solution for such a game, the core.

Part II extends the models of a strategic game (Chapter 9) and an extensive game (Chapter 10) to situations in which the players do not know the other players' characteristics or past actions.

The chapters in Part III cover topics outside the basic theory. Chapters 11 and 12 examine two theories of the outcome in a strategic game that are alternatives to the theory of Nash equilibrium. Chapter 13 discusses how a variant of the notion of Nash equilibrium in a strategic game can be used to model behavior that is the outcome of evolutionary pressure rather than conscious choice. Chapters 14 and 15 use the model of an extensive game to study long-term relationships, in which the same group of players repeatedly interact. Finally, Chapter 16 uses extensive and coalitional models to gain an understanding of the outcome of bargaining.

Notes

Von Neumann and Morgenstern (1944) established game theory as a field. The information about John von Neumann in the box on page 3 is drawn from Ulam (1958), Halmos (1973), Thompson (1987), Poundstone (1992), and Leonard (1995). Aumann (1985), on which I draw in the opening section, contains a very readable discussion of the aims and achievements of game theory. Two papers that discuss the limitations of rational choice theory are Rabin (1998) and Elster (1998).