This article discusses the value and limitations of game theory's use in the social sciences. The role of game theory is discussed and contrasted with exaggerated expectations of the subject. The importance of a modeling dialogue between theorists and empiricists is reviewed. The basic limitations of game theory are discussed, including the rationality and intelligence assumptions and the problem of multiple equilibria. The appropriate interpretation of randomized equilibria is illustrated.

On the Value of Game Theory in Social Science

ROGER B. MYERSON
Northwestern University

THE ROLE OF GAME THEORY AND OTHER EXPECTATIONS

We begin with a recognition that behavior responds to incentives.[1] That is, economists and other rational choice theorists try to predict and understand people's behavior by assuming that they are motivated by the anticipated consequences of their actions. But what if each person anticipates that the mapping from his actions to consequences will also depend on other people's actions? Then each person's incentives depend on other people's behavior, which in turn should depend on their incentives. In such situations, the behavior and incentives of different people must be seen as an interdependent system. Such interdependent incentive systems can become quite complicated, and so we need a special conceptual structure for organizing our thinking about them and predicting their effects. Game theory is this conceptual structure.

Since von Neumann and Morgenstern (1944) published their seminal book, the promise of game theory has generated great excitement. Game theory provides a fundamental and systematic way of thinking about questions that concern all of the social sciences. If the methods of game theory seem esoteric to the uninitiated, that perception can lead to even higher

RATIONALITY AND SOCIETY, Vol. 4 No. 1, January 1992 62-73

estimates of its power. But high expectations may be followed by disappointment when there is a perception of promises unfulfilled. Such disappointment has been expressed in the articles by Gordon Tullock (1992) and Michael Hechter (1992) in this issue of *Rationality and Society*.

Some authors almost seem disappointed that game theory is not omnipotent. Tullock believes that game theorists could not have helped Nelson with strategic advice on the morning of the battle of Trafalgar, and he suggests that this is a failure of game theory. I also doubt that anyone who has spent his life studying applied mathematics could have recommended a better battle plan to an admiral with a lifetime of experience in naval warfare. I would also doubt that a theoretical physicist could have significantly helped Michelangelo on the day when he began work on his *David*. But I believe that students of sculpture can benefit from studying something about the physics of materials, and young naval officers can benefit from a study of game theory. In each case, the academic discipline provides a vocabulary and a conceptual structure for organizing and learning from subsequent experiences. I believe that the chances of a young naval cadet becoming a great tactician like Nelson would be enhanced by including some study of game theory in naval college curriculum.

Similarly, Hechter (1992) complains that game theorists have not yet been able to provide a robust basis for the solution of collective action problems. This would be a failure of game theory only if there were a robust basis for achieving Pareto-efficiency in all collective action problems, which game theorists have somehow overlooked. In this patently imperfect world, a model should not be considered invalid merely because it predicts some Pareto-inefficiency. Indeed, it may be a great success of game theory that it helps us to understand the failures of cooperative behavior.

Tullock (1992) says, "The end result is not that the effort in game theory was totally wasted because it clearly has improved our understanding of a lot of problems even if it has solved few" (p. 31). It is ungenerous to say only that the effort in an academic discipline is "not totally wasted" because it merely increases understanding without definitively solving major real-world problems. There are few branches of social science that could defend a long list of definitive solutions to significant practical problems.

THE IMPORTANCE OF A DIALOGUE ABOUT MODELING

Certainly, it is fair to ask scholars to try to go beyond academic understanding and actually contribute something to the solution of social problems. We must recognize, however, that no analytical discipline can solve real social problems unless it is complemented by a modeling dialogue or sensitivity analysis. Critics of game theory can underestimate its practical value if they look for solutions to real problems without a modeling dialogue.

By a *modeling dialogue*, I mean a process in which theorists and empiricists work together interactively on the difficult task of finding tractable models that capture and clarify the important aspects of real situations. An analytical discipline provides both a vocabulary for defining models and procedures for generating predictions from these models. The predictions from these models are the theoretical output that empiricists need to guide their search for patterns in empirical data. However, there is never a clear answer to the question of which model best represents a given real situation, and so empiricists must help theorists to refine and extend their models. No conceptual framework can give us tractably simple models that fully represent every relevant fact that is known about complex real-life situations. To do any kind of analysis, be it game-theoretic or sociological or everyday verbal analysis, we must make some simplifying assumptions at the modeling stage. These simplifying assumptions must be tested and challenged in a modeling dialogue. We must constantly compare the predictions of our simple models with what we know about the real world and ask whether the appropriate simplifications have been made. We should challenge each other with different models and compare their predictions.

As we debate the validity of a model, however, we should also remember that a useful model does not have to be a picture of an entire situation. In scientific practice, theoretical models are usually not intended to represent entire situations but are only intended to describe certain relationships within these situations. For example, suppose that we have a conjecture that the price in a market will decrease as the number of suppliers increases. To analytically evaluate this conjecture, we construct some models, each of which must include variables for "the number of suppliers" and "the price." Our conjecture might be true (price does decrease as the number of suppliers increases) in some of our models, while it might be false in other models. In this case, when we ask which of these models are most accurate, we only need to ask about the qualitative validity of the factors in the model that actually determine the relationship between the number of suppliers and the price. In constructing a good model, we can simplify, distort, or even ignore factors that are not relevant to the relationship that interests us. When an empiricist objects to our distorting or ignoring these factors, we can defend the validity of our model for the purpose of probing this relationship if we can show that taking these factors correctly into account would not change our conclusions about this relationship.

An analytical discipline effectively allows us to redefine the process of thinking about the world as a process of building and comparing models. This point is implicit in the article by Aaron Wildavsky (1992 [this issue]), who sees a tension between individualists and collectivists expressed in a contrast between the models that they study. Of course, the fact that different people prefer different models does not imply that the different models are being used only for propagandistic purposes. When people advocate different policies but share an analytical methodology, comparing the models that lead them to their respective policy conclusions may help to clarify the essential basis of their disagreement. That is, an argument over policies may become clearer by converting it into an argument over the relative accuracy of different models.

So, a critic should not dismiss game theory's practical utility merely by looking at what it does with one model. Criticism of one model should be only the first step in a modeling dialogue. To illustrate, let us now consider Tullock's (1992) discussion of the model of Hillman and Samet (1987). This model is an auction game, in which two bidders simultaneously submit bids for a fixed prize of $100. The high bidder gets the prize, but both bidders must pay their bids in this auction. Assuming that bidders want to maximize expected monetary profit, Hillman and Samet find a randomized equilibrium in which each player independently chooses his bid randomly according to a uniform distribution over the interval from $0 to $100.

Tullock criticizes this model by saying that it neglects the fact that in real life, bidders may be concerned about competing with each other in more than one such auction and each bidder may hope that the other could be forced to withdraw from subsequent auctions if his capital is exhausted by early losses. Tullock is wrong to suggest that the limitations of one model are proof of the limitations of game theory because we can easily construct more sophisticated game models that take these factors (limited funds and subsequent auctions) into account. For a very simple such model, suppose that there will be a sequence of two such auctions, each for a prize of $100, and each bidder starts with a limited initial endowment of $100. Suppose also that both bids in the first auction will be publicly known when the second auction is held. Then there is a subgame-perfect equilibrium in which both bidders bid $100 in the first auction and the winner (randomly chosen from the two tied bidders) takes the second prize for an infinitesimal bid. The analysis of this model tends to confirm Tullock's intuition that these factors (subsequent auctions and limited funds) may tend to increase bids in the initial auction.

I am sure that Tullock could find other factors that are omitted by my new model and are significant in the real-world situations that concern him. There may be more than two auctions; the number of auctions may be uncertain; each bidder may be uncertain about the initial capital available to his opponent; bids might not be made public; bidders may be risk averse; and so on. Game-theoretic models can be constructed to take account of any or all of these factors, if they are judged to be significant elements of the real situation. Of course, the analysis of our model will become more difficult as the model becomes complicated by the need to take account of more factors. Rather than try to analyze one complicated model that takes account of all of these factors, it might be better to consider a sequence of simpler models, each of which serves to illustrate the potential impact of one or two of these factors. The power of game theory to improve our understanding of the given real situation is revealed only after we have worked through such a modeling dialogue.

THE LIMITATIONS OF GAME THEORY

In a modeling dialogue, the scope of game theory itself is challenged only when a critic calls for consideration of a factor that is intrinsically impossible to represent in a game-theoretic model. Like any other discipline, game theory is intrinsically restricted by its fundamental assumptions. The most important of these assumptions are *rationality* of agents and *common knowledge* of the model. Game theory is also limited in its predictive power when there are multiple equilibria. Each of these limitations merits some discussion here.

Rational utility maximization has been compellingly derived from axioms of consistency in decision making. Unfortunately, there is much experimental evidence that people often violate these axioms in real life. So rational choice models cannot be taken as the last definitive word in describing human behavior. However, if we want to do social analysis, it is not enough to reject the rationality assumption. We would need to replace it with some other disciplined way of characterizing human behavior, and other known behavioral disciplines have their own shortcomings.

For example, we might contrast rational choice models, which assume that behavior is determined by the anticipated consequences of actions, with perceptual stimulus-response models, which assume that behavior is determined by the perceptual antecedents of the action and comparable prior experiences. Rational choice and stimulus-response models offer different and complementary insights into human behavior, but each has significant limitations. For example, the stimulus-response approach would not be very useful for predicting how a change in the rules of a sealed-bid auction (say, changing from a first-price auction to a second-price auction) would affect the bids submitted, because each bidder's perceptual environment is almost unaffected by the change. (In either case, he will sit at the same typewriter to compose a bid on a sheet of paper that looks essentially the same.) Thus, when we are trying to understand how human behavior responds to changes in the way that material consequences depend on actions, we need something more like rational choice modeling. In view of the normative appeal of the weak rationality axioms, it seems quite reasonable to take rational utility maximization as a good first approximation to human behavior in such situations.

The second crucial assumption of game theory is that the model is common knowledge among the players. This assumption can itself be derived from a more fundamental assumption that the players are intelligent. When we say that players are *intelligent*, we mean that each player knows at least as much about the game as we, the social scientists, who are studying the game (see Myerson 1991). So whatever model we study, if it is supposed to be an appropriate model of the situation, then we must assume that each intelligent player knows that this is an appropriate model of their situation. Then, because we know that they all know our model, each player must also know that they all know our model. However, to avoid the assumption that we know anything more about the game than do the players themselves, we must assume that every player knows that every player knows that . . . every player knows our model of the situation. That is, a respect for the players' intelligence leads us to assume that our model is common knowledge among the players.

The requirement that the players know all the information that we specify in the game model does not imply that players have no uncertainty about each other. A game model may specify that each player privately observes a random variable for which the model does not specify the actual value, only its range of possible values and its probability distribution. In this way, our game model can take account of each player's uncertainty about others players' information. Following Harsanyi (1967-68), the random variable that represents a player's private information may be called his *type*. Mertens and Zamir (1985) showed that, in principle, the players' sets of possible types can always be made large enough for the model to represent all relevant uncertainty that players may have about each other. However, modeling greater uncertainty requires larger sets of possible types, which greatly increases the complexity of the model.

The third factor that limits the power of game theory is the fact that the set of equilibria (or other solutions), which forms the predictive output of game-theoretic analysis, may be quite large for some games. When a game has multiple equilibria, anything that focuses the players' attention on one equilibrium may make them expect it and hence fulfill it as a self-fulfilling prophecy (see Schelling 1960). But the question of what factors will most attract and focus people's attention is a cultural or psychological question. Thus, within its own terms, game theory calls for complementary analysis from psychological and sociological disciplines, in analysis of this *focal point* effect for games with multiple equilibria. That is, when there are multiple equilibria, predicting the focal equilibrium that will be played may take us beyond the economic analysis of incentives and into the domain of social psychology and cultural anthropology. Such a conclusion should not be surprising, however, other than to people who believe in pure economic determinism. Indeed, the focal-point effect may be seen as a way that game theory can offer perspectives and new research agendas for psychological and anthropological research.

The focal point effect leads me to disagree with Wildavsky (1992) when he says, "As a very crude first approximation, one might think of culture as being implicit in a model's payoffs" (p. 12). I think of culture as being expressed in the way that equilibria are selected in games that have multiple equilibria. At the very least, we should distinguish two very different roles for culture: (1) influencing how people's utility payoffs (which are expressed in the game model) are derived from material outcomes and (2) influencing how focal equilibrium expectations are determined (after the preference structure of the game has been specified and a large set of multiple equilibria has been found).

For example, let us follow Wildavsky and imagine a repeated Prisoner's Dilemma game in which the police regularly arrest the same pair of suspects and accuse them of joint crimes that are punishable by up to six days in jail. After an arrest, each suspect has an opportunity to confess, which will reduce his expected jail term by one day but will add five days to the other suspect's expected jail term. There are many equilibria of this repeated Prisoner's Dilemma game: the cooperative tit-for-tat equilibrium, the always-confess equilibrium, and so on. We should expect that the equilibrium that the suspects actually play will depend on their cultural perceptions, for example, on whether they view themselves as comrades or rivals. If the suspects are frequently innocent and their guilt or innocence is always common knowledge among them, then there is even an equilibrium in which they play cooperative tit-for-tat in the set of arrests where they are innocent, but they always confess in the arrests where they are guilty. Much faith in our legal system may be based on a cultural ideology that the players should focus on equilibria like this one.

In summary, the only assumptions necessary to bring us into the domain of game theory are that individuals' behavior is consistently determined by their goals and that we social scientists are not more intelligent than the individuals whom we are studying. Of course, these assumptions are never perfectly accurate: Inconsistency and ignorance are common human attributes. But a social scientist could explain almost anything by appealing to the inconsistency and ignorance of the agents involved. So we may naturally ask that a first analysis should try to follow the discipline of game theory. Beginning with game-theoretic models does not preclude subsequent consideration of other models or other methods of analysis after specific factors of inconsistency or ignorance have been raised by empiricists in a modeling dialogue or after multiple equilibria have been found. The game theorist's doctrine is only that we should ask how much can be explained by rationality and intelligence before we attribute people's behavior to their inconsistency and ignorance.

AN EXAMPLE WITH A RANDOMIZED EQUILIBRIUM

To illustrate these ideas, consider the first example discussed in Tullock's article. Following Tullock (1992), let us suppose that the monetary payoffs for players 1 and 2 respectively depend on their actions as follows:

		Player 2's action: A	Player 2's action: B
Player 1's action:	C	\$5, –\$5	–\$5, \$5
	D	–\$1, \$1	\$1, –\$1

If these monetary payoffs can be identified with utility payoffs (that is, if the players are *risk neutral*), then this game has a unique randomized equilibrium in which player 1 chooses C with probability 1/6 and player 2 independently chooses A with probability 1/2.

However, Tullock objects that the players are probably risk averse. In this case, player 1 would prefer a 50-50 gamble between –\$1 and +\$1 to a 50-50 gamble between –\$5 and +\$5, and so player 1's best response to 2's randomized strategy would be to choose D for sure. Of course, this argument is not an objection to game theory per se but is only a call for a second model, in which risk aversion is taken into account.

To keep things simple, let us suppose that the each player has essentially linear utility for monetary payoffs between –\$1 and \$5, but risk aversion may substantially decrease his utility for –$5. So, when we translate payoffs into a proper von Neumann-Morgenstern (1944) utility scale, we may get the following representation:

		Player 2's action:	
		A	B
Player 1's action:	C	5, y	x, 5
	D	–1, 1	1, –1

where x is player 1's utility for –$5 and y is player 2's utility for –$5. Risk aversion implies that $x < -5$ and $y < -5$. Of course, this game is no longer zero sum, but modern game theory does not require any restriction to zero-sum games.

If x and y are common knowledge, then this game has a unique randomized equilibrium in which player 1 chooses C with probability 2/(7-y) and player 2 independently chooses A with probability (1-x)/(7-x). Notice that these probabilities are 1/6 and 1/2, respectively, when $x = -5 = y$, and they approach 0 and 1, respectively, as x and y go to $-\infty$. For any value of the pair (x, y), the equilibrium leaves each player exactly indifferent between his two options, so that he is willing to leave his decision to chance. For example, when $x = -7.5$,

$$((1-x)/(7-x))(5) + (6/(7-x))(x) = .5862(5) + .4138(-7.5) = -.1724,$$
$$((1-x)/(7-x))(-1) + (6/(7-x))(1) = .5862(-1) + .4138(1) = -.1724,$$

so player 1 gets the same expected utility from C as from D in equilibrium. Notice that his expected monetary payoff is greater from choosing C (because player 2 has a probability of choosing A that is more than 1/2), but the risk is also greater from choosing C. Player 1's risk aversion, as expressed by his utility function, is such that these differences in riskiness and expected monetary value exactly cancel out and leave him indifferent between C and D. So, player 1 is willing to randomize as the equilibrium requires.

However, risk aversion is a matter of personal preferences, and so it is very likely that the players have some uncertainty about each other's risk aversion. As Tullock (1992) says, "Each party must now make an estimate of the utility value of the other, and there is not the slightest reason to believe that these estimates would be highly accurate" (p. 27). Again, this observation is not an objection to game theory but is only a call for a third model in which this uncertainty is taken into account.

So let us suppose that x and y are random variables and, when the game is played, player 1 knows x but not y, and player 2 knows y but not x. That is, each player knows his own utility for –$5 but is uncertain about the other player's utility for –$5. To characterize each player's beliefs about the other player's type, let us suppose that x and y are independently drawn from a uniform distribution over the interval from –5 to –10. The resulting game has a unique equilibrium in which each player chooses his action depending on his private information as follows:

Player 1 chooses C if $x > -5.683$.
Player 1 chooses D if $x < -5.683$.
Player 2 chooses A if $y > -7.635$.
Player 2 chooses B if $y < -7.635$.

In this equilibrium, the probability of player 1 choosing C is $(-5 - -5.683)/(-5 - -10) = .1366$, and the probability of player 2 choosing A is $(-5 - -7.635)/(-5 - -10) = .5269$. For comparison, in our earlier (second) model where x and y are assumed to be common knowledge, if x and y are both equal to –7.5 (the expected value of the uniform distribution on the interval from –5 to –10), then the probabilities of C and A are respectively .1379 and .5862 in equilibrium. So, the transition to a model in which players are uncertain about each other's risk aversion does not greatly change the equilibrium probabilities of each action.

Notice that in the equilibrium of this third model, each player is almost sure to have a unique best action, given the anticipated equilibrium strategy of his opponent. When player 1 expects player 2 to use the equilibrium strategy described earlier, then player 1's expected utility from each action depends on x as follows: Player 1's expected utility from C is $.5269(5) + .4731(x) = 2.6345 + .4731x$, and player 1's expected utility from D is $.5269(-1) + .4731(1) = -.0538$. Notice that if $x > -5.683$, then $2.6345 + .4731x > -.0538$ and so player 1 would strictly prefer C, but if $x < -5.683$, then $2.6345 + .4731x < -.0538$ and so 1 would strictly prefer D. Thus, in equilibrium, player 1 is never indifferent between his two options (except in the zero-probability event that his type is exactly the cut point –5.683), and that player 1 always uses the option that he strictly prefers.

Laypersons are often confused about how game theorists interpret randomized strategies. Tullock (1992) wonders about an "apparent feeling that playing a mixed strategy and actually throwing the dice somehow or other reduces your risk" (p. 30). This feeling is not apparent to me in the game theory literature, and I am not aware of any game theorist who advocates randomized equilibria for risk-reduction reasons. Tullock also worries about the fact that players have no strict incentive to follow their randomized strategies. The response to this concern is a generalization of the last result that we saw in the third model.

There is nothing in the logic of randomized equilibria that requires a player to base his decision on the toss of a die or the spin of a roulette wheel. The only logical requirement is that each randomizing player should base his decision on factors that the other players cannot observe. If our simple game model explicitly included every relevant factor in a real situation, then implicit unobservable factors that determine a player's move would have to be irrelevant, like dice and roulette wheels. When we think about real situations, however, this conclusion is almost always wrong because the premise that the simple game model includes every relevant factor in the real situation is wrong. As Harsanyi (1973) observed, players always have at least some minor uncertainty about each other's preferences. If a player has some minor information about his own payoffs that no one else knows, he can generally expect to do better by basing his decision on this payoff-relevant information than on a totally payoff-irrelevant toss of a die.

So, when I find a randomized equilibrium in a simple game model, I should interpret the randomization as a prediction that each player will base his decision on factors that he observes privately and that influence his utility payoff but were omitted from our simplified model presumably because they had only minor influence on payoffs. The fact that many games have only randomized equilibria tells us that examples of such dependence on minor private information may be quite abundant. That is, there are many situations in which we should expect players to base their actions on their private information in such a way that (1) players have substantial uncertainty about each other's moves, (2) each player's uncertainty about the others makes him almost indifferent among some of his own possible moves, and so (3) his optimal decision will depend on minor factors that only he can observe. When these privately observed minor factors are made explicit in our game model, as in the transition from the second model to the third model given earlier, then the randomized equilibrium becomes a pure-strategy equilibrium in which each player chooses a uniquely optimal move that depends on his privately known type.

NOTE

1. The initial sentence of this article is based on Paul Milgrom's succinct summary of the essential insight that underlies most economic reasoning.

REFERENCES

Harsanyi, J. C. 1967-68. Games with incomplete information played by "Bayesian" Players. *Management Science* 14:159-82, 320-34, 486-502.

Harsanyi, J. C. 1973. Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. *International Journal of Game Theory* 2: 1-23.

Hechter, M. 1992. The insufficiency of game theory for the resolution of real-world collective action problems. *Rationality and Society* 4:33-40.

Hillman, A. L., and D. Samet. 1987. Dissipation of contestable rents by small numbers of contenders. *Public Choice* 54:63-82.

Mertens, J.-F., and S. Zamir. 1985. Formulation of Bayesian analysis for games with incomplete information. *International Journal of Game Theory* 14:1-29.

Myerson, R. B. 1991. *Game theory: Analysis of conflict*. Cambridge, MA: Harvard University Press.

Schelling, T. C. 1960. *The strategy of conflict*. Cambridge, MA: Harvard University Press.

Tullock, G. 1992. Games and preference. *Rationality and Society* 4:24-32.

Von Neumann, J., and O. Morgenstern. 1944. *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Wildavsky, A. 1992. Indispensable framework or just another ideology? The Prisoner's Dilemma as an antihierarchical game. *Rationality and Society* 4:8-23.