Economics and Philosophy, 9 (1993), 101–133. Printed in the United States of America.

GAME THEORY AND THE HISTORY OF IDEAS ABOUT RATIONALITY

An Introductory Survey

ANN E. CUDD
Occidental College

Although it may seem from its formalism that game theory must have sprung from the mind of John von Neumann as a corollary of his work on computers or theoretical physics, it should come as no real surprise to philosophers that game theory is the articulation of a historically developing philosophical conception of rationality in thought and action. The history of ideas about rationality is deeply contradictory at many turns. While there are theories of rationality that claim it is fundamentally social and aims at understanding and molding all facets of human psychological life, game theory takes rationality to be essentially located in individuals and to concern only the means to achieve predetermined ends. Thus, there are some thinkers who have made important contributions to this history who do not appear in the story of game theory at all, among them, Plato, Kant, and Hegel. There is, however, a clear trail to follow linking theories of instrumental rationality from Aristotle to the nineteenth-century marginalist economists and ultimately to von Neumann and Morgenstern and contemporary game theorists, that historically grounds game theory as a model of rational interaction.[1]

I thank Neal Becker, the University of Kansas Philosophy Department colloquium, Dan Hausman, and several anonymous referees for very helpful comments on earlier drafts of this essay. I gratefully acknowledge the University of Kansas for a New Faculty Research Grant for financial support of this project.

1. Game theory has also been used to model nonrational processes, especially evolution. I focus on the predominant use of game theory as a model of rational interaction here.

This trail is worth following for more than historical reasons. For game theory shares its origins with nearby current theories of morality and agency, suggesting that game theory is a part of a complete and coherent picture of agency. So, while its adequacy as a descriptive theory of human behavior may be doubtful, in its consistency with nearby theories of rationality in belief and individual decisions and the theory of utility, game theory can be judged an enormous success. Further, there is a therapeutic value for game theory in this enterprise, for the landmarks that we shall examine have garnered criticisms over the years, criticisms that will apply to game theory insofar as it remains consistent with those theories.

Game theory is that part of rational choice theory that is concerned with the interaction of rational agents. It is to be distinguished from individual decision theory, which applies to situations in which there is one effective rational agent, and social choice theory, which is concerned with the decisions of groups of agents who must ultimately act as one. There are many connections between individual decision theory and game theory. Some decision theorists argue that for practical purposes decisions ought to be modeled as individual decisions, and some argue that game theory is a subset of decision theory (cf. Skyrms, 1990, chap. 1). I shall attempt to show how game theory makes a unique contribution to the understanding of instrumental rationality, and that it cannot be subsumed by decision theory. The matter is rather the reverse in my view: Game theory incorporates the picture of rationality offered by decision theory and then goes one crucial step further.

Game theory is inspired by three main developments in the history of ideas about rationality: (1) the idea that rationality is utility maximization; (2) the idea that rational beliefs and rational expectations (that is, of utility) can be formalized using probability theory; and (3) the idea that rational interaction, or interaction among rational agents, is strategic. The idea that rationality is utility maximization is the idea that rational agents represent their desires in utility terms, and rank their options in order to find the best way to satisfy most of their desires.[2] This requires first that we see rationality as a kind of calculation,[3] in particular a maximization, and second that the maximization has to do with getting what we desire as expressed in utility terms. Additionally, it assumes that rationality is a capacity inhering primarily in individuals, not in groups (though it is sometimes assumed that a group has univocal belief and desire systems in order to model its decisions game theoret-

2. I am not making a commitment here to a realist interpretation of game theoretic models. One may equally well for my purposes read this sentence as saying that agents act as if they were representing their desires . . . .
3. This term *calculation* contains the so-called "ing/ed" ambiguity. It is a debatable topic whether a theory of rationality owes us merely a substantive final answer on any question or a model of the procedure by which it is reached, as well.

ically). Game theory models agents as using their beliefs about their situations and their desires to decide on optimal actions. And, as a model of rationality, it incorporates the idea that these subjective beliefs can be judged rational or irrational. Thus, the second necessary development for game theory is the idea that beliefs can be formalized using probability theory, as desires are in utility theory. These first two aspects of rationality come together in individual decision theory. The third main idea distinguishes game theory from individual decision theory, the idea that in order to act rationally in situations of interaction with other rational agents one must act strategically. That is, one must take others to be responding to one's own planned actions with the same theorizing ability. Each of these ideas may be traced back far before Borel's paper of 1924 or von Neumann's paper of 1928, which are acknowledged to be the first papers in game theory. I shall discuss each idea, in historical order.

1. RATIONALITY AS UTILITY MAXIMIZATION

Utility theory has its origins long before Bentham and the utilitarians. Beginning with Aristotle, it has been common in some philosophical traditions to understand rational action as choosing the best means to attain one's desires. Aristotle called the practical reasoning that results in choice *deliberation*. On this view rationality is a process, a calculation, at least metaphorically. "Excellence in deliberation involves reasoning. . . .the man who is deliberating, whether he does so well or ill, is searching for something and calculating" (Aristotle, 1941, Bk. VI, chap. 9). One deliberates for the means, but rational action involves not only finding the right means, but also seeking the proper ends. The ends of action are not the result of rational deliberation, but we can determine whether they conform to reason by exercising practical wisdom. In Aristotle's view, to say that one acts irrationally is to say that either one's actions are not based on sound deliberations, or that one's ends are not derived from practical wisdom. "Excellence in deliberation will be correctness with regard to what conduces to the end of which practical wisdom is true apprehension" (Aristotle, 1941, Bk. VI, chap. 9). Aristotle separates deliberation, the calculation of means, from practical wisdom, the apprehension of the correct ends of action, but both are required for rationality in action. This separation of the means from the ends becomes, in later thinkers, a distinction between rational and nonrational processes.

The seventeenth century saw the further development of the idea of reason as a calculation, in what has been called the mechanical model of reason, through the work of Thomas Hobbes. Hobbes holds that reasoning is nothing more than rapid calculations on thought "parcels." The picture of reason and deliberation is that it, too, is a kind of "adding and substracting" that works at a physical level on analogy with a clock.

In the rational person these calculations proceed according to arithmetic rules. In the *Leviathan* he writes: "For Reason . . . is nothing but Reckoning (that is, adding and substracting) of the Consequences of generall names agreed upon, for the marking and signifying of our thoughts" (Hobbes, 1982, p. 111). Hobbes's model of action is also a (rather crude) mechanical model. Humans act on the impulse of appetites and aversions, with greater or lesser deliberation in between. Deliberation consists in imagining the possible consequences of our actions, and ultimately we act or not according to whether the last consequence before the decision incites our appetites or aversions. Deliberation can go wrong if one does not reason well, but if reason properly determines the will by conjuring up the right consequences of action, then one's life will be felicitous. "He who hath by Experience, or Reason, the greatest and surest prospect of Consequences, Deliberates best himself" (Hobbes, 1982, p. 129). Hobbes does not clearly set out the process by which reason represents desire and belief, but in showing that rational action depends on the propriety of this representation, he grasps the rudiments of a modern computational theory of action.

Hobbes also contributes to the development of the modern subjective conception of value on which the theory of utility is based. Desires, which are the objects of what he calls "appetites," are for a person's own good, and the test that something is good is that it is desired. "Whatsoever is the object of any mans Appetite or Desire; that is it, which he for his part calleth Good. . . . These words of Good, Evill, and Contemptible, are ever used with relation to the person that useth them: There being nothing simply and absolutely so" (Hobbes, 1982, p. 120). An appetite that results from deliberation is a voluntary act. Thus, an action to satisfy a subjective desire for a good is a rational appetite if rational deliberations show that the good desired will result from the action proposed to obtain the good. The Hobbesian account of desire has an objective element as well. Hobbes held that our appetites and aversions are determined physiologically by the survival value of the object in question. We have an appetite for what conduces to our survival and an aversion to what does not. Thus, appetites are, properly, for what helps survival. There is, then, a tension between the two accounts of proper actions: A rational appetite may be for something subjectively desired that does not help survival.[4]

4. There are problems with resolving this tension in either direction. If we suppose that Hobbes's account of value is subjective, then the account seems to conflict with his psychological assumptions. If we suppose that values are objective in the sense consistent with his psychological assumptions, then Hobbes's moral system degenerates into a purely descriptive account, and one without much empirical adequacy. I want to maintain only that it is plausible to interpret at least some of what he says in *Leviathan* as positing subjective value, since subjective value makes possible the development of utility theory.

Hume agrees with Hobbes against Aristotle in holding that reason serves only to decide on the means for obtaining independently given ends and to clarify the effects of actions under consideration. Desires, or ends, are a mere matter of taste, so that one could neither persuade someone of another desire nor criticize her *as irrational* for her desires. Unlike Aristotle, Hume holds that there is a sharp distinction to be drawn between reason and passion (the seat of desire), and so concludes, contrary also to Hobbes, that the very idea of "rational appetite" is a category mistake. Rationality, according to Hume, applies to well-formed beliefs, not desires. Beliefs are rational so long as one has good evidential reasons for holding them, but desires can never be called rational or questioned by rationality. Hume accepts from Hobbes the mechanical model of reason and the notion that only subjectively held desires can motivate actions. Reason enters into rational calculation and action in an advisory role: "Reasoning takes place to discover this relation; and according as our reasoning varies, our actions receive a subsequent variation. But 'tis evident in this case that the impulse arises not from reason, but is only directed by it. 'Tis from the prospect of pain or pleasure that the aversion or propensity arises towards any object" (Hume, 1981, p. 414). Reason affects action by changing the beliefs about what consequences follow from the action. Only the consequences for pleasure and pain motivate action. This reduction of motivation to just two independent sources leads to the theory of utility in which motivation is homogeneous.

Though the word *utility* with some connection to the ends of action had been around since Horace,[5] the utilitarians pursued the notion that diverse desires might be measured on a *single scale* and thus compared with each other and with desires of others (or so Bentham thought). Bentham's idea was that to determine what ought to be done one could compare different courses of actions according to the pleasure or pain persons derived from them, and he combined the measures of pleasures and pains to form a single sum of happiness. Although he considered the measurement of happiness, the "hedonic calculus," to be a "fiction" or a model, he argued that ultimately the sum of pleasure and pain is the motivation for all actions and that in principle the pleasures of different people for different goods could be compared along a single scale.

Later utilitarians abandoned the hedonistic hypothesis that pleasure and pain were the (only) ultimate motivations of actions, but still maintained that ends could be compared on a single scale by how strongly persons desired them. This severed utility theory's connection to happiness that Bentham had postulated. At this point, utility theory was linked to moral and political theory, but not by way of rationality as it

5. In his "Article on Utilitarianism" Bentham traces the history of the term *utility* and the greatest happiness principle, including the transformations in his own work, which he speaks of in the third person (cf. Bentham, 1983).

is now. The greatest happiness principle of the utilitarians holds that the right action in any given circumstance is the one that maximizes the happiness of all. This is not a matter of rationality, either individual or social, but of the Good.[6]

Economists first connected utility theory with the notion of individual rationality. The marginalist revolution in the 1870s in Austria with Carl Menger, in Switzerland with Leon Walras, and in England with W. S. Jevons, introduced a new model of microeconomic behavior in which agents are assumed to be rational and to be concerned with subjective desires.[7] By "rational" they mean that these agents act to maximize their subjectively given (from the point of view of the model) utility, subject to their budget constraints, that is, to the limits on their choices imposed by the goods they own that they could trade with others for what they want. Mathematically, this is a very neat theory. The constrained maximization is a simple matter of solving a Lagrange equation constructed from a functional representation of the utility function and the budget constraint. The result is that rational individuals will make trades until the ratios of the marginal utilities they receive from the last item of each good is equal to the ratios of the prices of the goods.

This microeconomic theory has had enormous implications for contemporary rational choice theory. First, it formalized the notion of utility into a functional representation.[8] Second, it linked the notion of rationality with the notion of utility maximization. Third, the mathematics by which utilities are represented suggest a dynamic process and a notion of stability: Rational individuals will continue to trade until the ratios of their marginal utilities equal the ratios of the prices, at which point they will be at a stable equilibrium. Although the marginalist economists originally intended their theory to account only for choices in the economic sphere, that is, consumption and production decisions, it has been expanded by their intellectual descendants in Austria by Ludwig von Mises and in the Chicago school, especially by Gary Becker, to all contexts of human behavior, for example, crime, education decisions, and the family (Becker, 1976). James Buchanan and Gordon Tullock,

6. On one understanding of J. S. Mill's derivation of the greatest happiness principle in *Utilitarianism*, another kind of rationality is being appealed to. That is, Mill may be taken to argue from some sort of social rationality principle that says that it is rational to seek to attain what is socially desired. This is a principle that would often conflict with the individual utility maximization principle that game theory endorses. But see Harsanyi (1984) for an attempt to ground utilitarian moral theory in game theory.
7. Heinrich Gossen, who discovered marginal utility analysis already in 1854, may have been the first marginalist, but his work was not taken seriously by economists. See Stigler (1950, pp. 314–15).
8. The original marginalists' utility functions were for a single good. Edgeworth (1881) was the first to give it a generalized form.

James Coleman, and their followers have applied marginal utility analysis to collective political and social decisions (Buchanan and Tullock, 1962; Coleman, 1986). Recent experimental work has even extended microeconomic theory to animals. (See, for example, Battalio, Kagel, and MacDonald, 1985.) Thus the theory of rational behavior has reached broadest possible scope.

The final step in making this economic theory into an explicit theory of rational choice was to deduce utility functions from a set of axioms describing rationality in preferences and actions and to couple this with the principle that rationality is utility maximization. This step is taken when the axioms that characterize coherent preferences – for example, the requirement that they be complete and transitive – are shown to imply a utility representation of the preferences. Such a result is called a *representation theorem*. This theorem shows precisely what must be assumed about agents, and in particular their preferences, in order for a numerical utility function to apply to their preferences. There are two kinds of utility scales for which representation theorems exist: ordinal and cardinal utility scales. An ordinal scale is one that represents only the order of goods according to one's preferences; it is unique only up to positive monotone transformations.[9] A cardinal scale is unique up to positive linear transformations. It thus represents the relative strength of one's preferences as well as the order.[10] Since cardinal scales claim to represent more information about agents' preferences, the axiom system has to be stronger.

9. To say that a function, $u(x)$, is "unique up to positive monotone transformations" means that any positive monotonic transformation of $u(x)$ yields a function, $u'(x)$, that is essentially equivalent to $u(x)$, where a positive monotonic transformation is a transformation, f, such that (1) f preserves the sign of $u(x)$, i.e., $f(u(x)) > 0$ if and only if $u(x) > 0$ and $f(u(x)) < 0$ if and only if $u(x) < 0$; and (2) the first derivative of $f(u(x))$ is always positive.
10. There are several ways of constructing cardinal scales, though, that support different interpretations of the underlying preferences. The von Neumann-Morgenstern utility function is the most well known and the one used almost exclusively in game theory. von Neumann and Morgenstern claimed that this utility function does not support comparisons of differences in preferences. That is, one cannot say that since $u(a) - u(b) > u(c) - u(d)$, the change from b to a is preferred to that from d to c. The problem is not one of mathematics; clearly that inference could be drawn if ">" is identified with "is preferred to." The point is rather that if we derive a claim from the mathematical relationship of the utilities for lotteries to the underlying preferences for states of affair, then we are assuming that there is an underlying preference for the changes of a to b and c to d that we are simply *measuring* with the utility function, or perhaps one that we are *prescribing*. In either case this violates the notion that a utility function is merely a representation of the stated preferences, which were over lotteries. What it does support is comparisons of preferences about risky events. For further explanation and consideration of other cardinal utility scales, see Fishburn, 1989, pp. 129–36, and Luce and Raiffa, 1957, pp. 19–32.

In the early years of the marginalist revolution, economists assumed utility in a simple cardinal functional form, where the utility of a collection of goods is equal to the sum of the utility of each good taken separately, that is, $u(a_1, a_2, \ldots, a_n) = u(a_1) + u(a_2) + \ldots + u(a_n)$. As early as the late nineteenth century Vilfredo Pareto recognized that it was not necessary to assume cardinal utilities in order to derive the theorems of consumer behavior. He renamed utility, *ophelimity*, in order to avoid confusion with the other uses of the term *utility*, although the new term never caught on. Pareto noticed that there were problems with the standard conception of utility. First, he questioned whether utility exists in the cardinal form that was then standardly assumed.[11] At that time, economists assumed that a cardinal scale of utility was (at least theoretically) available on which everyone's subjective desires could be measured and compared. But there existed no method for measuring desires. Whereas heat can be measured by thermometers, so that one can say more than "this is hotter than that," with desires all that we do have is "I prefer *a* to *b*," or "I am indifferent between *a* and *b*," Pareto's objection to the measurability of desires has been considered a serious problem ever since, and economic theory abandons this assumption whenever it can. Edgeworth had previously shown how to derive indifference curves from cardinal utilities. Pareto turned this around, deriving ordinal "ophelimity indices" from the indifference curves alone and showing that if an agent orders sets of bundles of goods to which she is indifferent, then from that information an ordinal utility ranking can be derived. Pareto thus derived the ordinal representation theorem (though not by this name), which says that if an agent's preferences are such that for any bundles of goods x and y, either x is preferred to y or y to x or the agent is indifferent between x and y, and only one of these disjuncts holds. In addition, if the preferences are such that they form indifference curves that are convex to the origin, then the preferences can be represented by an ordinal utility function (Pareto, 1971, pp. 191–99).[12] From his ophelimity indices Pareto derived the competitive equilibrium for perfect competition. Once he had shown that the most important results of microeconomic theory of his time could be derived from ordinal utilities alone, Pareto argued that we ought to do without

11. Schumpeter (1954) traces this objection of Pareto to 1900 in some lectures he gave at the École des Hautes Études in Paris, and to Pareto's "Sunto di alcuni capitoli di un nuovo trattato di economia pura," in the March and June numbers of the *Giornale degli Economisti*, 1900. He also gives this objection in *Manual of Political Economy* (1971, pp. 191–92).

12. Clearly he is also assuming that the preferences are transitive.

cardinal utility functions because of the serious philosophical problems he associated with them.[13]

Cardinal utility functions can be constructed in several ways that may be neatly divided into two groups according to their basic orientation toward comparing preference differences. The older construction (first made explicit by Pareto) relies (implicitly or explicitly) on a comparison of preference differences. That is, it constructs the utility scale by brute force, asking agents for an ordering of (combinations of) goods and then for the intervals between the (combinations of) goods. In the Appendix to his *Manual of Political Economy*, Pareto shows that cardinal scales constructed in this way require strict, and rather implausible, assumptions. Either we have to assume that our preferences for consuming good x depend only on the amount of good x (and not on what else we have), or we have to be able to perform experiments that directly reveal the intervals between combinations of goods. Thus, he formulated his economic theory without cardinal scales. However, Pareto's ordinal analysis assumed that outcomes of events are determined with certainty by the actions of agents.

In 1926, Frank Ramsey produced the first representation theorem for preferences by cardinal utilities that dispensed with these implausible assumptions. Ramsey constructed cardinal utilities by assuming the existence of an "ethically neutral" proposition p, which has the property that if an agent is indifferent between two states of the world, A and B, then he is indifferent between the following options: (1) A if p is true and B if p is false; and (2) B if p is true and A if p is false. Now he can define the difference in value between two states A,B of the world as equal to that between two other states C,D just in case the agent is indifferent between: (1) A if p is true and D if p is false; (2) B if p is true and C if p is false. This essentially gives him a way to compare preference differences by comparing lotteries over states of the world without assuming that agents can compare the preference intervals directly. Several other axioms concerning preferences are also required for the representation theorem (cf. Ramsey, 1990, pp. 73–75). Ramsey goes on to show how the utility representation can then be used to derive a representation theorem for subjective beliefs (that is, a personal probability function). In the 1947 second edition of *Theory of Games and Economic Behavior*,

13. Pareto also questioned the interpersonal comparability of utility scales. The problem he saw is that even if we could find a scale to measure one person's utilities, we could not be sure that the same scale would apply to another person. Interpersonal comparisons are important for those who hold that total utility is to be maximized, because summing over incomparable scales is nonsensical. Pareto argued, however, that any change in social policy that would increase one person's utility without worsening anyone else's could still be seen as good, without relying on any ability to make interpersonal comparisons.

von Neumann and Morgenstern independently proved a representation theorem that used probabilities conceived as given by relative frequencies. Like Ramsey, they recognized that utility and probabilities conceived as subjective beliefs could be interderived (cf. von Neumann and Morgenstern, 1953, p. 19, n. 2), though unlike Ramsey's their proof does not derive both. These cardinal representation theorems require axioms about comparisons of lotteries over events and thus also make stricter requirements on agents' behavior than the corresponding theorem for ordinal utility.

For much of economic theory ordinal rankings are enough, for example, the theory of consumer behavior in perfectly competitive markets. But such situations are characteristically nonstrategic situations and thus not the kinds of situations game theory is uniquely able to analyze. Game theory would not go very far with only ordinal utilities. Its reliance on cardinal utilities has been the source of much of its power and its criticisms. However, to get the cardinal representation theorem, and so to be able to analyze expected values, additional axioms are required, and these require statements about probabilities, the subject of the next section.

2. PROBABILITY AS RATIONAL BELIEF, EXPECTATION AS RATIONAL DESIRE

Contemporary game theory[14] requires a probability theory for two important purposes. First, as we mentioned above, it uses concepts of probability theory to define and construct cardinal utility functions for agents. Many of the solution concepts can be justified only with comparisons of preferences for risky goods, and this is supplied by cardinal utilities. Second, probability theory allows game theory to model situations of risk and uncertainty. Risky situations are ones in which an agent does not know some future state of the world but has an objective probability estimate about the possible future states. A situation is uncertain if the future state is unknown and there is no objective probability estimate. In game theoretic certainty situations, players have full information about their own and others' possible actions, and about the outcomes that would result from any combination of actions, and they have common knowledge of these facts. Playing a game of chess each player knows the other's utility function (winning is better than losing), the available strategies (the possible combinations of moves of the pieces), and, if we assume they have good memories, they know the previous moves of the game. Furthermore, since the moves are fully

14. I have in mind here game theory taken as a whole, that is, noncooperative as well as cooperative game theory and imperfect as well as perfect information games. Cooperative games of perfect information do not require cardinal utilities or probabilistic belief functions, and other special cases can also be treated without probability theory.

public and known to be so, we can assume that all of these facts are common knowledge. Game theoretic models of these situations are called complete and perfect information models. Situations in which there is a chance element that is a result of agents' lack of full information about causal forces relevant to the situation, or of past decisions of other agents, or of the other agents' attributes are modeled by imperfect information games when the possibilities for these variables and their objective probabilities are commonly known and incomplete information when they are not.[15] For example, imagine a game, call it 'chess*', which is much like chess, but each player has some restrictions on the movements of his pieces that are known only to him. If there is a well-defined set of possible restrictions that all the players know, and, in addition, they know the probability with which each restriction would apply, then the game may be modeled as an imperfect information game. In fact, even if the expectations of the players about the uncertainties can themselves be predicted with some commonly known probability, then the game is imperfect. If, on the other hand, the players may not know all the possible restrictions, or the others' expectations about them, then the game may be irreducibly incomplete. Since most realistic situations involve uncertainties, the advances made possible by probability theory are crucially important to the plausibility of game theory as a general theory of rationality.

Probability, as we now conceive it, came into being around the decade of 1660 in the work of several people who approached the idea of probability from different directions.[16] Three ideas about rationality come together in the idea of probability: the idea that rational monetary expectations in games of chance are equal to the product of the probability of the outcome and the monetary value of the outcome; the idea that probability codifies rational belief also in contexts other than games of chance; and the idea that rational beliefs must be conditioned by evidence, and that this means that rational beliefs are conditional probabilities. In 1657, Christiaan Huygens published the first textbook of probability, (translated as) *Calculating in Games of Chance*, in which the notions of expected value and equivalent gambles were invented. The expectation of a gamble is what Huygens defined as the fair price for participating in a lottery. If a fair lottery gives one a chance of winning x dollars, with an equal chance of each ticket being the winning ticket, and there are n tickets and only one winning ticket, then the fair price of the lottery is x/n. This is also the expectation that one has in buying

15. This needs to be qualified somewhat. Harsanyi (1967–68) shows that in many cases incomplete information games may be reduced to imperfect information games. This can be done, however, only if there is common knowledge of the possible future states of the world among the players.
16. I owe much here to Hacking (1975).

any one lottery ticket. He goes on to argue that a fair price can be found for any bet by finding a fair lottery such that the bettor is indifferent between the lottery ticket for the original bet and the fair lottery; such lotteries are called "equivalent gambles."

In a 1738 paper Daniel Bernoulli developed the idea of expectation more fully and connected monetary value with what he called the "moral worth" of money, or what we would call its *utility*. He writes, "the determination of the *value* of an item must not be based on its price, but rather on the *utility* it yields" (Bernoulli, 1954, p. 201). Specifically, Bernoulli claimed that the moral worth of money was proportional to its logarithm, or $U(x) = a \cdot ln(x)$, making its utility a function that increases at a decreasing rate. One example to which he applies his utility function for money is his famous solution to the St. Petersburg paradox, which goes as follows: Suppose that one could pay a fee to enter a lottery which paid one 2^n dollars with a probability of 2^{-n} for each integer n, how much should one be willing to pay? Since the expected payoff of the lottery is infinite as n goes to infinity, it may seem that there is no limit on the fee one ought rationally be willing to pay. But our intuitions tell us otherwise, since, for example, to win \$128 one has only a 1 in 128 chance. To solve the problem, Bernoulli appeals to the "moral worth" of money, arguing that \$1 is worth more to the pauper than to the millionaire, and that one's first \$100 is worth more to anyone than the next \$100. The point is that the moral worth of each new dollar declines as one gets more money, while the moral worth of one's most recently lost dollar becomes greater as one loses money. Thus, although the expectation of the lottery is unlimited, the expected moral worth of the additional gains decreases as n increases, while the expected moral worth of the additional losses increases.

The ideas of expectation and equivalent gambles are the key notions in the development of cardinal utility scales to solve problems about uncertain futures. In order to derive the cardinal representation theorem that allows one to deal with uncertainty, one needs axioms that constrain the preferences of rational agents in trade-offs among different lotteries. Huygens showed what one must rationally expect in such trade-offs. Given what one must rationally expect and the fact that rational agents prefer more to less, this entails that one must rationally prefer those lotteries from which one can expect a greater payoff. Thus, probability theory gives us a principle of rationality in preference over lotteries allowing us to go beyond ordinal to cardinal scales.

While Huygens and Bernoulli forged the connection between monetary expectations in simple games of chance and probability, they did not link probability with rationality in belief more generally. In 1658, Pascal wrote the *Pensées*, in which he describes an argument justifying belief in God that we know now as Pascal's wager. Among other important contributions to the theory of rationality, this argument linked

probabilistic reasoning with beliefs. The wager argument first attempts to show that belief in God is warranted because it is a dominant action. A dominant action is one that leads to an outcome that is as good or better than any other regardless of the actions of others or of the outcome of uncertain events. In this case it is uncertain whether God exists, but Pascal argues that if He does, then an infinite payoff can be realized by believing in Him, and if He does not, then nothing is lost, while not believing in Him leads to at best a zero outcome either way. So either way one does at least as well by believing: "if you win, you win all; if you lose, you lose nothing. Wager then, without hesitation that He does exist" (Pascal, quoted in Rescher, 1985, p. 8). But this argument fails to take account of the cost of believing, and when that is factored in dominance does not apply, since if God does not exist, it is better not to believe (or so one might argue). To answer this objection Pascal does a calculation of expectation. He argues that because there is a positive probability that God exists and because belief in God if there is a God leads to an infinite future payoff, one ought to believe in God. That is, the expectation from this belief is infinite. With this argument Pascal showed that one ought to use probabilistic reasoning to determine the likelihood of an event and that expected value calculations apply to broader contexts of decision.

Rationality in belief also requires that one condition one's beliefs on the evidence one has. In 1665, in his essay "De conditionibus," Leibniz independently developed a theory of probability in order to give an analysis of rationally conditioning belief on evidence in the law. On his view legal cases are to be decided by taking all of the evidence into account and figuring the probability that one claim is correct by analyzing the states of affairs under which it is correct and incorrect. If the claim is correct under all states of affairs, then its probability is 1; and if it is correct under none, then its probability is 0. The probability in intermediate situations is determined by the number of cases in which it is true divided by the number in which the claim is true or false. This analysis introduces a couple of important ideas: Like Pascal's wager argument, it shows that probability has to do with the statements about states of affairs broadly construed and that the probability of a statement is relative to the known facts or, as we might say now, that all probabilities are conditional probabilities.

The next great advance in the theory of rational belief was the discovery of the rule for conditioning probability estimates on new evidence that grew out of work by Thomas Bayes. The rule, now called Bayes's rule or Bayes's theorem, stems from an essay on probability by Bayes, posthumously published in 1763. In this work he solves the problem of "inverse probabilities," that is, of finding the conditional probability that A will occur given that B has occurred from the probability that B will occur given A, which is a necessary calculation for inferring from ob-

served frequencies to associated probabilities. Bayes was actually trying to solve a more difficult application of what has come to be known as Bayes's rule. He states the problem: "Given the number of times in which an unknown event has happened and failed: Required the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named" (Bayes, 1940, p. 376). But the important observation that he made to solve the problem was that the inverse probability could be figured from the rule for compound probabilities. This is done in propositions 4 and 5 for the case where A and B are independent events. We would write today: $P(A|B) = P(B|A) = P(A\&B)/P(B)$ (Bayes, 1940, pp. 379–81). Where A and B are not independent events, the theorem is somewhat more complicated. The conclusion reached from Bayes's theorem is that one can rationally condition one's prior beliefs on new evidence by using a simple relationship from the probability calculus. For example, suppose we want to know what the probability is that someone gets heart disease given that she is a runner, $P(H|R)$, and we know the probability that someone in the general population gets heart disease, $P(H)$ (the "prior probability") and the conditional probability that someone is a runner from the set of people with heart disease, $P(R|H)$. Then applying Bayes's rule, the formula is $P(H|R) = P(H)P(R|H)/[P(H)P(R|H) + P(\sim H)P(R|\sim H)]$.

The eighteenth-century work formalizing inductive evidence in a probability calculus gradually became known to philosophers as well. Bishop Butler was led to conclude that probabilistic reasoning represented the most important kind of reasoning for mere mortals. He wrote: "But to us, [in contrast to the deity] probability is the very guide of life" (Butler, 1850, p. xlix). Hume, in his *Treatise* of 1739, criticized attempts to rationalize beliefs about the future using probability. He argued that the strength of beliefs varies by the vivacity of the impressions that cause them. A belief about a future event, however, cannot be based on an impression, since we cannot have an impression from an as yet unrealized event. Furthermore, there is nothing in present or past events from which we can infer, logically, future events. Our beliefs about future events, then, must be formed by nonrational processes,[17] foremost

17. He provides us with an impressive list about the ways beliefs can vary from nonrational processes: "When we have not observ'd a sufficient number of instances, to produce a strong habit; or when these instances are contrary to each other; or when the resemblance is not exact; or the present impression is faint and obscure; or the experience in some measure obliterated from the memory; or the connexion dependent on a long chain of objects; or the inference deriv'd from general rules, and yet not conformable to them: In all these cases the evidence diminishes by the diminution of the force and intenseness of the idea. This therefore is the nature of the judgment and probability. . . . nor will it ever be possible upon any other principles to give a satisfactory and consistent explication of them. Without considering these judgments as the effects of custom on the imagination, we shall lose ourselves in perpetual contradiction and absurdity (Hume, 1981, pp. 154–155).

among them the force of habit, or custom, of inferring cause and effect from constant conjunctions of similar events. Altering Butler's phrase, he writes, " 'Tis not, therefore, reason which is the guide of life, but custom" (Hume, 1981, p. 652).

Hume's problem of induction is a powerful criticism of the frequency view of probability, which interprets probability as the long-run frequency of events of the same type. To say that there is a probability of 1/2 that this coin will come up heads is to say that in trials in the past the coin or coins like it have come up heads 50 percent of the time. Clearly it commits the "error" of assuming that future trials will resemble the frequency of past trials.

In this century, Frank Ramsey, Bruno deFinetti, and then Leonard Savage developed the notion of subjective probability. They argued that agents have (nonrational) subjective prior degrees of belief for all statements, which can be rationally updated by evidence using Bayes's rule, without making the illegitimate inductive move criticized by Hume. The trick to avoiding the inductive move is to require one's probability estimates about the future to be internally consistent with each other and with one's knowledge of the present and the past, and then to use the resulting probabilities as measures of one's subjective beliefs and not as objective facts about the world. They showed that the probability calculus is consistent with this interpretation of probability founded entirely on subjective guesses about statements. We can illustrate the basic idea with the following thought experiment of deFinetti. Suppose that agents place odds, or probabilities, including conditional probabilities, on the truth of a set of statements that is closed under the logical operations 'and', 'or', and 'not'. Now suppose a bookie is allowed to take any of the bets or combinations of bets given those odds. Then using the rationality principle that one should prefer more of whatever one wants to less, it can be shown that only certain combinations of probabilities are rational to set and these are the ones consistent with the probability calculus. Any other probabilities will lead one inevitably to lose to the bookie. Thus, the agent's beliefs must be coherent, in the sense demanded by the probability calculus, in order to be rational.[18]

Subjectivists, who are called *Bayesians* because of the importance of Bayes's rule to their theory, also derive from Bayes's rule the principle that all information is evidence upon which we can improve our beliefs about the world. In rational choice generally this led to the principle, sometimes taken to be the central insight of Bayesianism, that all information is to be used optimally in decisionmaking. Bayesians also justify their coherence view by showing that if one uses evidence to update one's beliefs with Bayes's rule, then in the long run the subjective prob-

18. This justification of subjective probability has been justly criticized as applying only to those who actually make all the bets. See Schick (1986).

abilities will converge to the same values as frequency probabilities. Not all game theorists are Bayesians, however, because Bayesianism assumes that agents have prior probabilities for all possible states of the world.

Economists in this century typically have distinguished between risky and uncertain situations (cf. Knight, 1921/1972, chaps. 7–8). A risky situation is a situation where the outcomes of actions or the contingencies of nature that bear on the situation are not perfectly known, but there is some basis for figuring an objective probability estimate for them. A situation of uncertainty is one in which there is no such basis. Bayesians deny that there is any clear distinction between risk and uncertainty, because they think that we are never in a situation in which there is no basis for postulating a prior probability. (Of course, they agree that there are better and worse sources of data.) So, for Bayesians the game theoretic treatments of the two situations become identical.

The two essential aspects of probability theory for game theory that I have discussed in this section, its use in formulating cardinal utilities and its use in risk situations, dovetail neatly. Game theory reduces some imperfect information situations to perfect information models by replacing utilities under risks by expected utilities. Under the right circumstances, incomplete information situations may be reduced to imperfect information situations by replacing commonly known objective probabilities with subjective probability estimates.[19]

Utility theory and probability theory, together with the notion of dominance, form the foundations of individual decision theory. By the early twentieth century, neoclassical economists used these principles to make economic arguments, though many rejected cardinal utility in much of their analyses and concentrated on certainty situations, in which ordinal utility is sufficient for decisionmaking. But decisionmaking in interaction raises new problems that cannot be solved within individual decision theory when one realizes that the agents with whom one is interacting are just as rational and capable of optimally using information. This gives one a new kind of information for predicting the actions of others, information that is not simply reducible to inductive or probabilistic evidence.

3. RATIONAL INTERACTION IS STRATEGIC

Game theory brings the concept of *strategic* interaction to the rational choice model of rationality. While individual decision theory confronts the rational decisionmaker with a nonrational environment, game theory

19. By "commonly known" I mean that the objective probabilities are known by all the relevant persons, and they are known to be known, and it is known that they are known to be known, . . . If this common knowledge is missing, there may not exist game theoretic solutions.

confronts her with a number of other rational agents, each of whom may have an effect on the outcome of her actions. Thus, game theory formalizes the notion that rational interaction requires agents to consider the others' reasoning in choosing actions. In particular, if one believes that the other agents are also rational, then one ought to attribute to them the ability to reason in very similar ways as oneself about the choice situation. This attribution of rationality then gives one added information about how the others might behave, and one must see the others as also realizing this, and so on. Thus, from the Bayesian rationality principle that one take into account all one's information about the situation and from the assumption that others are equally rational, one arrives at a new kind of decision situation known as *strategic interaction*. In this section we shall examine conceptions of strategic rationality and the special problems that strategic thinking poses the theorist of rationality. First among these is the necessity of modeling the nested levels of mutual beliefs or expectations of the agents. Another crucial consequence of this strategic thinking is that groups of three or more present problems that are qualitatively different from groups of two. In groups of three or more the possibility of coalition formation, that is, of two or more agents forming a group within the whole group, arises.

Parametric (as opposed to strategic) models of interaction assume that agents take the others' actions to be fixed according to a small set of predeterminable parameters. Individual decision theory models the environment as this sort of parametric actor: One decides on the probability with which one's own actions will lead to each possible outcome and then makes the decision. But this does not model the others as responding to each other by modeling the thinking of each; it takes the others' actions as given. For some situations we must imagine that this is the best we can do. If the others are nonhuman, or have very foreign values and motivations, then perhaps it is impossible or not worth the trouble to model their thoughts. But, in those cases where the motivations and values of the others are clear, and where it can be assumed that everyone knows what the options are, and this is all known to be known, then information that makes possible a game theoretic model is available.

An example here may help illustrate how strategic modeling of interaction differs from parametric modeling. You want to meet me on a train in which we are both passengers. I want to avoid meeting you. There are only three different cars, numbered 1, 2, and 3, and we get to choose only one, which we will have to remain in for the whole trip. Suppose that all of this information is commonly known to both of us. Then what you will do should rationally depend on what you think I will do, which in turn should depend on what you think I think you will do, which of course depends on what you think I think you think I will do, and so on. Likewise, what I will do ought rationally to depend

on what I think you will do, and that will depend on what I think you think I will do, which depends on what I think you think I think you will do, and so on. Given certain rather strict assumptions about the players' utility functions and information, game theory allows us to show that the iterated considerations may come to an end in an equilibrium solution, which is a best response to everyone else's best response, though they need not in many cases.[20] Considerations about the mutual responses of the agents are paradigm strategic considerations. If, on the other hand, I say to myself, "you usually take 3, so I will take 1," then I am modeling your behavior parametrically, I am taking your action to be known or merely stochastically varying.

Although, as we shall see, game theory understood itself from the beginning as adding the strategic element to modeling rationality, game theorists have only recently begun to model the mutual knowledge and beliefs explicitly. But the view that iterated mutual expectations are important for strategic rationality can be found in earlier philosophical investigations of rational social interaction and even earlier examinations of military strategy. For example, Thucydides in his *History of the Peloponnesian War*, refers many times to how one army considered the future actions of another by trying to model their reasoning. In the dialogue between the Melians and the Athenians, the Melians argue that the Lacedaemonians will come to their aid against the Athenians:

> But it is for this very reason that we now trust to their respect for expediency to prevent them from betraying the Melians, their colonists, and thereby losing the confidence of their friends in Hellas and helping their enemies . . . . We believe that they would be more likely to face even danger for our sake, and with more confidence than for others, as our nearness to Peloponnese makes it easier for them to act, and our common blood ensures our fidelity. (Thucydides, 1952, p. 506)

Their strategies are determined by their beliefs about the Lacedaemonians' probable course of action, which in turn is determined by what the Melians think the Lacedaemonians believe about them, namely, the Melians believe that the Lacedaemonians will believe that they (Melians) are trustworthy because of their "common blood." It is likely that one can find similar kinds of reasoning in almost any good commentary on war, or for that matter, chess or other games of strategy. Steven Brams (1980) has applied game theory to the stories of the Old Testament. Although he does not show that the writers of the Bible themselves recognize the strategic nature of the interactions, he finds enough evidence to attribute to the actors' strategic thought and motivation.

20. Bacharach (1987) discusses the definition of a *solution* for games and argues that there is little reason for us to be optimistic that there will be a solution in a wide variety of games, namely those with multiple Nash equilibria, or with none at all.

Beginning in modern times, strategic analysis has been applied to everyday social interaction. Hobbes's state of nature arguments contain some early strategic analysis. In many contemporary reconstructions, Hobbes's state of nature problem is equated with the prisoner's dilemma, the solution to which is found using dominance reasoning. When a dominant strategy is available, one need not ever consider the moves of the other player in the game. That is, even if one knew precisely what the other was going to do, one rationally should still play the game the same way: always defect. So it cannot pay one to consider the thoughts of the other player in that game; the situation is, in effect, a parametric situation for players with dominant strategies. But Hobbes's analysis of the state of nature also contains paradigm strategic elements, even if dominance ultimately rules. Hobbes claims that there are "three principle causes of quarrell. First, Competition; Secondly, Diffidence; Thirdly, Glory" (Hobbes, 1982, p. 185). The first of these does not obviously require strategic thought, if it is a prisoner's dilemma situation, but the latter two must involve some mutual expectations. About the second he writes: "And from this diffidence of one another, there is no way for any man to secure himselfe, so reasonable, as Anticipation" (Hobbes, 1982, p. 184). To see the usefulness of anticipation requires one to try to think one step ahead of one's opponent, realizing that he has followed one's own reasoning to that point. In rationalizing glory as a motivation, Hobbes bids us consider the long-run effect of having the power and reputation that glory causes: "if others, that otherwise would be glad to be at ease within modest bounds, should not by invasion increase their power, they would not be able, long time, by standing only on their defence, to subsist" (Hobbes, 1982, p. 185). Finally, Hobbes argues that because rational persons can come to see that all are so motivated, they ought to institute an absolute sovereign. Thus, rational interaction is strategic for Hobbes, and this rational consideration of mutual expectations makes avoiding war possible.

Hume was more clearly aware of the importance of mutual expectations in his remarks on conventions in the *Treatise*. A convention solves one of the commonest kinds of strategic situations we face, situations in which we have to coordinate our actions to bring about what each of us most wants. Hume himself gives several examples: the two persons who must pull together on the oars of a boat, property rights or "abstinence in the possessions of others," languages, use of a common measure of exchange, and promise-making. According to Hume, the convention for respecting others' property rights arises "by a slow progression, and by our repeated experience of the inconveniences of transgressing it. . . . This experience assures us still more, that the sense of interest has become common to all our fellows, and gives us a confidence of the future regularity of their conduct: And 'tis only on the expectation of this, that our moderation and abstinence are founded" (Hume, 1981,

p. 490). We judge from experience that everyone sees the value in abstaining from others' possessions, and from this we build common knowledge of the interest we each have in maintaining the growing convention not to steal. So we come to expect that the others will abstain from ours, and hence we abstain from the others. One might interpret this passage as suggesting that we arrive at property right conventions through a myopic trial-and-error kind of process. Perhaps Hume is saying that we begin by trying various ways of getting along and eventually stumbling on the convention of property rights. Such myopic search has recently been modeled game theoretically by evolutionary stable strategies, in which the equilibria are reached by nonrational processes, for example, whole species are modeled as agents "seeking" survival. (Cf. Maynard Smith, 1982.) This interpretation would have Hume saying that we repeat many experiences of not having property rights and a few of preserving it, and eventually we come to see simply that the latter is better than the former and we continue to do that. In that case our guide, as it is for induction, is custom, not reason. Thus, it would not be true to say from this passage that Hume recognizes rational strategic interaction. I think that myopic equilibrium search is a plausible story of how many human conventions are initially stumbled upon, especially Hume's two rowers example. However, rational agents can fashion their conventions quite self-consciously at times, with an eye to the justification of the convention. Legislated laws are perfect examples of this. Hume seems to be describing here a situation in which it is important that people recognize that there are levels of mutual knowledge of interest, which is necessary for the forward-looking strategic thought and not for the myopic equilibrium search.

Rousseau recognizes the importance of using oneself as a model of others to discern mutual expectations in his *Discourse on the Origin of Inequality*. In discussing the primitive situation of humans he attempts to explain how early prelinguistic interactions took place. In interaction it is most important to know what the others will do in order to secure one's own best outcome. To figure this out, a primitive human, he supposed, would reason as follows: "Seeing that they all acted as he would have done under similar circumstances, he concluded that their way of thinking and feeling was in complete conformity with his own. And this important truth, well established in his mind, made him follow, by a presentiment as sure as dialectic and more prompt, the best rules of conduct that it was appropriate to observe toward them for his advantage and safety" (Rousseau, 1983, p. 141). If we model others as like ourselves, then we can use ourselves as analogue indicators of others' beliefs and desires. And if we model them as rational like ourselves, then we can make predictions about how they respond to us. But modeling others as rational also leads us to conclude that they will model us

as we model them, and this raises for us the issue of common knowledge or mutual belief.

In his seminal (1928) paper, "On the Notion of Games of Strategy," John von Neumann addressed the problem of how players must behave in games of strategy in order to achieve their best outcome and presented the first formal treatment of strategic games. He began by stating the problem as a circularity: "the fate of each player depends not only on his own actions but also on those of the others, and their behavior is motivated by the same selfish interests as the behavior of the first player" (von Neumann, 1928, p. 13). The presence of more than one player makes this problematic because the outcome of any one player is due not only to her action and the outcome of any risky events, but also the decisions of others. von Neumann formalized the notion of strategy by first reducing games of chance, that is, games in which there is a risky event, to games of pure strategy by calculating the expected outcome for each player and for each possible outcome of the risky event. Then a strategy for each player consists in a set of decisions that he makes, one action for each possible decision point contingent upon the information that he has at that point. Thus, in effect, once players' strategies are chosen, the outcome of the game is determined, "the player knows beforehand how he is going to act in a precisely defined situation: he enters the play with a theory worked out in detail" (von Neumann, 1928, p. 18). Now von Neumann is able to prove his important maximin theorem, showing that there exists a unique solution that maximizes the minimum gain for each player in 2-person zero-sum[21] games.

Oskar Morgenstern was independently interested in the nature of the *foresight* required rationally to solve decision problems. In his (1935) "Perfect Foresight and Economic Equilibrium," Morgenstern considers the "Moriarty and Holmes problem": Holmes is riding on a train from London to Dover, being pursued (as he knows) by Moriarty. Holmes gets off at an intermediate station because he surmises that Moriarty took a faster train to catch Holmes in Dover. Suppose that Moriarty and Holmes have "full foresight," that is they know the probability with which events in the future will happen, including "foresight about the probable behavior of others" (Morgenstern, 1935, p. 173). Then Moriarty could have considered this possibility and gotten off too. But then Holmes could have expected that, and so forth. Morgenstern writes: "Always, there is exhibited an endless chain of reciprocally conjectural

21. Zero-sum games are games in which the gains and losses to the players for every outcome sum to zero. This is obviously a very restricted set of games. In particular, a very important class of games for modeling social conventions, coordination games, are positive sum. Schelling (1980; first published 1960) first argued persuasively for the extension of game theory to nonzero-sum games.

reactions and counter-reactions. This chain can never be broken by an act of knowledge but always only through an arbitrary act – a resolution . . . . The paradox still remains no matter how one attempts to twist or turn things around. Unlimited foresight and economic equilibrium are thus irreconcilable with one another" (Morgenstern, 1935, p. 174; italics omitted). Morgenstern does not solve this problem in this paper, and indeed it may not be soluble in general, but he is clearly thinking about how decisions are made in contexts in which agents consider each others' iterated considerations about each other.

Later, von Neumann and Morgenstern (1953), clarify the nature of the complexity that *strategic* interaction introduces to decisionmaking. They write: "each participant attempts to maximize a function of which he does not control all variables. This is certainly no maximum problem, but a peculiar and disconcerting mixture of several conflicting maximum problems . . . . those 'alien' variables cannot . . . be described by statistical assumptions. This is because the other are guided, just as he himself, by rational principles" (von Neumann and Morgenstern, 1953, p. 11). A person in social interaction "will be influenced by his expectation of [the other participants' actions and volitions] and they in turn reflect the other participants' expectation of his actions" (von Neumann and Morgenstern, 1953, p. 12). Thus, they argue, social interaction presents us with a conceptual difficulty, not merely a technical one, and this is what they devised game theory to solve.

von Neumann and Morgenstern presented first their conception of a rational solution for the simplest possible games: 2-person zero-sum games, that is, games in which the loss of one is equal to the gain of the other, so that the sum of the total gain is zero. A solution is a set of strategies, one for each player. To be called rational on the instrumental theory we have been developing, it should maximize the expected utility of the players. But as Morgenstern demonstrated, in strategic contexts there is no straightforward way of characterizing the expected actions of the others, as there would be in a parametric context. Instead, choosing an optimal strategy involves considering the mutual reactions to one anothers' strategies. There is a condition that we can set on the choice of strategies, as Luce and Raiffa (1957) write: "if a theory offers a_{i0} and b_{i0} as suitable strategies, the mere knowledge of the theory should not cause either of the players to change his choice" (Luce and Raiffa, 1957, p. 63). The reason for this should be clear, if players want to change their strategies, there must be a preferred strategy, and so the original one cannot be optimal. But if all the players are satisfied that they cannot do better, then the optimal strategy, given the others' strategies, has been reached. One limitation of this idea that equilibria are solutions is that they are appropriate only in static contexts, as von Neumann and Morgenstern clearly recognized. (Cf. von Neumann and Morgenstern, 1953, p. 45.) A solution to a (static) noncooperative game,

then, is an equilibrium, a set of strategies, one for each player, such that no player can do better by deviating from her equilibrium strategy provided the others are playing their equilibrium strategies.[22] As I reported earlier, for zero-sum games von Neumann proposed that each player ought rationally to play that strategy that guarantees her the maximum of all the minimum gains that she might realize, the "maximin" solution. He proved that if one allows strategies that are probability combinations of more than one strategy, or "mixed strategies," there is a maximin solution for every zero-sum game. But many interesting games are non-zero-sum, and for some such games there is no maximin strategy, even in mixed strategies. John Nash (1951) generalized this conception of an equilibrium to what is now called the Nash equilibrium, and showed that there is such an equilibrium for every noncooperative game, perfect or imperfect, zero-sum or non-zero-sum, provided that the structure of the game and the payoffs of the players in the game is common knowledge among them.[23]

To illustrate the kind of thinking involved in finding an equilibrium, consider the following situation. Suppose that Kim and Tim want to meet but cannot communicate with each other before they are to meet. Suppose that it is common knowledge that there are two places to meet, in Kim's office or in Tim's office, because they always meet in one of them, but there is no rule for where they should meet this time. Which one should they choose? A game theoretic analysis would normally begin with a matrix representation of their options and their utility payoffs, as follows.

		Tim: Kim's	Tim: wait
Kim	wait	2, 1	0, 0
	Tim's	0, 0	1, 2

FIGURE 1.

Each entry in the matrix represents Kim's and Tim's payoff in utilities, respectively, for the corresponding combination of strategies. For example, if Kim waits and Tim goes to Kim's office, then Kim's payoff is

22. But see Bacharach (1987) for some serious criticisms of the game theoretic concept of solution.
23. Nash made no explicit reference to the common knowledge necessary, though it is clear from the proof that he is assuming it.

2 and Tim's is 1. Now to find their equilibrium strategies, Kim would begin by reasoning that if she were to wait in her office, then Tim would want to go to her office, and if he were to go to her office, then she would want to wait. Thus her waiting and Tim going to Kim's office is an equilibrium. Unfortunately, one can go through the same reasoning for Tim waiting and Kim going to Tim's office. Thus, there are two pure-strategy Nash equilibria. Additionally, there is a mixed-strategy Nash equilibrium, where each waits with probability 2/3 and goes to the other's office with probability 1/3. So any of these three actions make sense provided each thinks that the other is doing the other part of the equilibrium strategy she or he has chosen, in the sense that one could not do any better by doing something else, and one may well do worse.

This example also points out two of the biggest conceptual problems for game theory. First, in many games there are multiple equilibria, and then it is not clear what an agent ought to do.[24] While it may seem rational to pick one, if the others choose their equilibrium strategies for different equilibria, it no longer has the characteristic of being the best response one can make. The second problem is that mixed strategies lack a clear strategic rationale as well. If one agent follows her mixed strategy, then it does not matter what the other does – he has the same expected outcome. But if he fails to play his corresponding mixed strategy, then her mixed strategy no longer is the best response to his action. These problems are particularly poignant for game theory as a normative theory of rationality, because the solutions seem to lose their logical connection to rationality, that is, to the principle of expected-utility maximization. John Harsanyi (1973) proved that we can often model these games in such a way that the mixed strategy equilibria are stable, so that players may not deviate from their equilibrium strategies without penalty. But this requires that we transform the original game into a game where the players' knowledge of other players' outcomes is imperfect, that is, the payoffs randomly fluctuate with a known probability. This kind of transformation poses a puzzle for those of us interested in game theory as a model of rationality. Why should players use the transformed game with uncertain payoffs when they are certain of the payoffs? Why should they in effect throw out some of their information, when by doing so they do not change their payoffs? Harsanyi's solution may be quite plausible as a rationale for explaining why players played their equilibrium mixed strategies, but it does not offer any new principle of rationality.

24. For that matter, there are some reservations about whether it is clear that one ought to play one's Nash equilibrium strategy when it is the unique solution. See note 20 above. I take game theory to require rational agents to play their unique Nash equilibrium strategies when they exist.

While much game theory can be done without explicitly considering the background common knowledge assumed by the standard model, once it was recognized just how much and what sort of knowledge is necessary for the agents to find equilibria on the standard model, new possibilities and new problems arose. The immediate predecessor of the explicit formalizaton of the notion of common knowledge and belief in game theory was David Lewis's (1969) book *Convention*. Lewis introduced the term *common knowledge* (meaning that not only all know what is common knowledge, but that they all know that they know it, and they know that they know that they know it, . . .), in order to give a formal rational choice account of the origin of conventions. On his account, if agents are aware of a need for a conventional solution to some interaction situation, and they have common knowledge of this, and there is some particularly salient solution, which is itself common knowledge, then they will perform that action called for by the solution without any prior agreement. They can do this because, being rational, they are able to reason to the solution, and if the situation is common knowledge in the way just specified, they can reason that the others reason to the same solution, and that the others reason that they all reason to that solution, and so on. Aumann's (1976) paper "Agreeing to Disagree," which introduced common knowledge as a topic to game theorists, relies heavily on Lewis's work, and provides an algebra of states of affairs for distinguishing common knowledge from various levels of mutual beliefs, which are finite levels of iterated knowledge or probabilistic belief claims. Jaakko Hintikka's (1962) book *Knowledge and Belief* on epistemic logic has made it possible to explicitly model levels of mutual knowledge that are less than common knowledge. The existence among players of common knowledge, or at least several levels of mutual knowledge, of some facts of the interaction situation is now seen as a crucial assumption which makes possible strategic interaction, that is, which makes it possible for agents to model each other as rational agents like themselves. Explicit models of common knowledge in game situations have been useful in solving some problems, such as finding cooperative solutions to finitely repeated prisoner's dilemmas (cf. Kreps, Milgrom, Roberts, and Wilson, 1982), and it holds promise for future research.

An important consequence of strategic rationality is that rational interaction among three or more individuals holds the possibility of coalition formation. In other words, it may be possible for a subset of the whole to cooperate to achieve a better outcome than they would acting individually. In economic discussions such cooperation is sometimes called collusion, or monopoly or oligopoly formation, and in political theory it is often referred to as collective action. Adam Smith noted the effect of collusive action on markets and the advantages it gives to those who can easily combine in order to collude.

> What are the common wages of labour depends every where upon the contract usually made between those two parties, whose interests are by no means the same. The workmen desire to get as much, the masters to give as little as possible. The former are disposed to combine in order to raise, the latter in order to lower the wages of labour.
>
> It is not, however, difficult to foresee which of the two parties must, upon all ordinary occasions, have the advantage in the dispute, and force the other into a compliance with their terms. The masters, being fewer in number, can combine much more easily; and the law, besides, authorises, or at least does not prohibit their combinations, while it prohibits those of the workmen. (Smith, 1976, pp. 83–84)

While both sides are "disposed" to collude, there are social forces pushing apart laborers more than capitalists. Aside from laws rigged in their favor, what helps the capitalists is their relatively small number, since there is less temptation to collude within the group in sharing the proceeds of their combining against the laborers. von Neumann (1959) also recognized this feature of strategic interaction in games of more than two persons, but he had little formally to say about how to approach the problem.

Since von Neumann and Morgenstern's book, it has become common to distinguish two different kinds of situations that form the major division in the classification of game theoretic models: cooperative and noncooperative games. In noncooperative games one assumes that players cannot make binding agreements, while in cooperative games they may make binding agreements. A binding agreement is a promise to perform certain actions in the future, regardless of any other considerations that might intervene between the promise and the time for performance. For example, I may offer my house for sale at a certain price, and if I agree with a buyer on that price both of us are bound to honor that agreement, regardless of the opportunities that seem to arise after the handshake. Inability to make binding agreements effectively precludes coalition formation on purely rational grounds, since coalitions rationally should form only if the members of the coalition can determine how they will share the gains from forming the coalition before the game situation is settled.

To unite, then, members of potential coalitions need the opportunity to communicate their collective interest and the ability to negotiate an agreement about how they will share the gains from their cooperation. In addition, they may need an enforcement mechanism to punish any violators of the agreement; and the greater the temptation for individuals to cheat on the agreement, the stronger the mechanism needs to be. For rational individuals, this enforcement mechanism constitutes another

part of the cost of entering the agreement; and this cost is reflected in the model by the utility assignments to the various potential coalitions.

Whenever the outcome for a subset of the players is better for them if they form a coalition and they can enforce that coalition, it rationally ought to form. The issue among them then becomes, "how shall we divide the resulting payoff?" But there are often competing coalitions that can enforce their outcomes, and so players ought rationally choose that coalition that guarantees them the highest individual payoff. The main rationality concept of cooperative game theory, the core solution, derives from dominance reasoning and addresses the question of which coalitions can form. To define the core, we need to define the notion of dominance reasoning in a cooperative situation. To say that an outcome is dominated if any group of players could guarantee itself a better outcome by acting as a coalition. The core is the set of all undominated outcomes. Given the instability of coalition formation, any of the outcomes of the core would seem possible, though none outside the core would. This is, of course, not the only solution concept of cooperative game theory, but it is the best general illustration of the sort of claims about rationality that are made by the theory, most of which hinge on dominance reasoning. But the core solution does not tell us how, exactly, winning coalitions will divide their gains. The most influential of the solutions to this problem that have been proposed is the Shapley value, which weights each players' contribution to the coalition according to how much more the coalition makes in her presence rather than in her absence (cf. Shapley, 1953; Roth, 1988). However, all the solutions to this problem rely on some problematic assumptions, arguably arbitrary from the point of view of rationality.

Readers sophisticated in game theory will notice that I have omitted discussion of many topics in contemporary game theory, including bargaining theory and the theory of matching. While bargaining theory is very important for theories of rationality in action, it is largely derivative on the ideas we have developed here, and there exist accessible discussions of it as an account of rationality (Gauthier, 1986; Resnik, 1987; or for a more noncooperative account of bargaining theory, Binmore and Dasgupta, 1987). Matching theory, on the other hand, seems to me to have little to add to a normative theory of individual rationality and to be of value as a descriptive theory and as a theory for making policy prescriptions (cf. Roth and Sotomayor, 1990).

4. RECENT DEVELOPMENTS OF THE CONCEPT OF RATIONALITY

The common knowledge discussions in the game theoretic literature have opened up new flexibility in modeling rational interaction, but have also raised some problems for the traditional account of rationality we

have developed in this paper. It has been shown that common knowledge of some aspects of the interaction situation is necessary for there to exist Nash equilibrium solutions in general. Since common knowledge is a strong assumption, it limits the theory's application, as presently conceived. Several responses to this problem have been sketched. One is to return to parametric modeling of other agents' behavior in situations where common knowledge is not available. This is, in effect, to return to individual decision theory and reject the possibility of modeling the thought processes of others. But it is not a good alternative; as long as there is some information about others' beliefs, it is inconsistent with the Bayesian ideal arbitrarily to put aside that information. Furthermore, a parametric agent is open to exploitation by the agent who goes one step further in modeling opponents' actions using higher levels of mutual knowledge. Another response has been to argue that the necessary common knowledge is not a strong assumption after all, because it is just the background information that forms the very foundation of interaction. One might argue this in the way that Davidson (1985) argued against exclusive "conceptual schemes," but this kind of response has only been hinted at in the literature and awaits a full development. Common knowledge assumptions might be justified by an account of the communal nature of rationality, though this would have to be done carefully to avoid any conflict with the rational individualism of game theoretic models. Most of the work in response to the common knowledge problem has been to push back the limitations of game theory without common knowledge, by defining new solution concepts that require common beliefs, that is, common probability estimates. What remains to be explored is the extent to which the *commonness* of the knowledge can be weakened, that is, what sort of requirements rationality puts on interaction with fewer levels of mutual knowledge or belief. When the commonness of knowledge and belief is more seriously considered, another of the basic assumptions of game theory will also have to come under scrutiny, namely, the assumption that rationality inheres in individuals, not in groups.

Another set of problems that has received a great deal of attention recently is the refinement of the Nash equilibrium and, more recently still, attempts to abandon equilibria in favor of a more general notion of "rationalizability." The refinement program addresses the problem that there is often no unique rational solution, as in the Tim and Kim example of Section 3. Ideally one would like a solution that is unique and exists for every game, but none has been found, and there are other problems with each of the solutions that has been proposed (cf. van Damme, 1987). For instance, the refinement known as "trembling hand perfect" equilibria claims that a rational strategy choice is one that is a Nash equilibrium that remains an equilibrium when players are allowed to "tremble" in their choice of strategies, that is, to make

mistakes in their play with some small probability (cf. Selten, 1975). But this means that to find the rational action one has to consider that mistakes are possible while assuming that it is common knowledge that all the players act rationally. Furthermore, trembling hand perfect equilibria are also not always unique. Harsanyi and Selten (1988) take a slightly different tack by offering a way to select one equilibrium in each situation by pruning out successively less desirable equilibria. The problem remains, though, when different equilibria strategy combinations lead to the same outcomes. Finally, Bernheim (1984) and Pearce (1984) offer a more general conception of rational solution. Their idea is to specify a whole set of behaviors that possess some plausible rationale for players. These will not necessarily lead to equilibria; in most cases, in fact, there will be no equilibrium for the game, because of some unresolvable uncertainty. The point is that they will then have to find some other way to rationalize a behavior, and they have some mutual knowledge of this fact. The refinements to the Nash equilibrium solution and the concept of rationalizability each make claims about the nature of rationality. Game theorists constantly push forward the understanding of rationality through new solution concepts, and it is an important future project for philosophers to assess the claims that these "solutions" make about the nature of rationality.

Important work is also being done on utility theory. There are several criticisms of the univocality of the concept of utility. Sen (1977) argues that we have preferences that are motivated by commitment and other moral notions, and that these motivations cannot be compared on a single scale with mere desires. Etzioni (1986) proposes a utility function that has two components in the range of the function, one to account for the moral component of satisfaction and one for the satisfaction of inclinations. Schick (1984) argues that there is an aspect of "sociality" or "responsiveness" in our motivations that is unlike the self-interested desires of utility theory and that must be included in a theory of rational behavior in order for that theory to account for a great range of cooperative behavior. Multivalued utility functions, though, pose at least three problems for an account of rationality. First, practical rationality requires that there be a solution, a rational thing to do. But multivalued functions are ambiguous about what is important – is the "socially concerned" value to take precedence or the individual one? This question can be decided by the rationality principle one applies, but then it is not clear why having an ambiguous utility representation is preferable, as a choice still has to be made. Second, multivalued utility functions are less mathematically tractable. Third, they presuppose a moral theory that then cannot be constructed from utilities on pain of regress. None of these theories has convinced the game theorists yet, but they may represent important challenges to the concept of rationality underlying it.

5. CONCLUSION

In this essay I have presented the main links between game theory and older philosophical ideas about rationality. Rationality, in the tradition we have discussed, inheres in individuals. Rational individuals make choices that they rationally believe will best bring about their most desired outcomes, and these choices can be modeled by expected utility calculations. While their beliefs and desires are held subjectively, those beliefs and desires are subject to rationality conditions. The desires of rational agents are transitive, complete, and relatively stable over time, and so can be represented by single-valued utility functions. The beliefs of rational agents are consistent with the probability calculus and formulated from all of the information the agent has. In interactions among rational agents, they consider each others' best responses to their strategies in formulating their strategies and try to find an equilibrium where everyone is making a best response, and they consider the possible coalitions they might form with others.

Game theory brings this model of rationality together in a formal package, and theorists use it in a wide variety of contexts. Its formalization and applications have led to further articulation of the theory of rationality and have also allowed us to see some internal contradictions and conceptual problems of that theory. A better understanding of common knowledge and belief made possible by game theory leads us to question the idea that rationality is fundamentally individualistic. The problems of multiple equilibria and of the rationale of mixed strategies leads us to question the idea that rational individuals are equilibrium seekers, and hence also the idea that they seek to maximize expected utility. And the broad range of application of the formal model of utility theory brings into question whether it can really be single-valued, or in other words, whether "desire" as it has been applied in the traditional model of rationality is a univocal concept.

Game theory is a theory of rational interaction that has its roots firmly in the history of philosophical conceptions of instrumental rationality. It is an evolving theory that formalizes the lessons gleaned from this history and is in a continual state of self-examination and correction for internal inconsistency. In these respects it has been enormously successful. It has allowed the formalization of a whole tradition of thought on rationality, which has led to rapid development and extensions of the theory. Furthermore, the formalization of the theory has increased the ability to find inconsistencies in the conceptions of rationality that inspired it. These successes are important for the modeling of rationality in its own right. The formal theory has also lent itself to a wide variety of applications in moral, social, and political theory. And, these applications in turn give us arenas in which to further test the theory, as well as to make new extensions to the model. A discussion of these applications, however, will await another occasion.

REFERENCES

Aristotle. 1941. *Nicomachean Ethics*. Translated by W. D. Ross, in *The Collected Works of Aristotle*, edited by Richard McKeon, pp. 935–1112. New York: Random House.

Aumann, Robert. 1976. "Agreeing to Disagree." *The Annals of Statistics* 4:1236–38.

Bacharach, Michael. 1987. "A Theory of Rational Decision in Games." *Erkenntnis* 27:17–55.

Battalio, Raymond, John Kagel, and Don MacDonald. 1985. "Animal Choices over Uncertain Outcomes: Some Initial Experimental Results." *American Economic Review* 74:597–613.

Bayes, Thomas. 1940. "An Essay Toward Solving a Problem in the Doctrine of Chances." In *Facsimiles of Two Papers by Bayes*. Washington: The U. S. Dept. of Agriculture.

Becker, Gary. 1976. *The Economic Approach to Human Behavior*. Chicago: University of Chicago Press.

Bentham, Jeremy. 1983. "The Article on Utilitarianism." In *Deontology together with A Table of the Springs of Action and The Article on Utilitarianism*, edited by Amnon Goldworth. Oxford: Clarendon Press.

Bernheim, B. D. 1984. "Rationalizable Strategic Behavior." *Econometrica* 52:1007–28.

Bernoulli, Daniel. 1954 (first ed. 1738). "Exposition of a New Theory on the Measurement of Risk." *Econometrica* 27:23–36. Translated by Louise Sommer from "Specimen Theoria Novae de Mensura Sortis." Reprinted in Page (1968, pp. 199–214).

Binmore, Ken. 1987–88. "Modeling Rational Players." Parts I and II. *Economics and Philosophy* 3:179–214; 4:9–56.

Binmore, Ken, and Partha Dasgupta. 1987. *The Economics of Bargaining*. New York: Basil Blackwell.

Borel, Emile. 1924. "Sur les jeux ou interviennent l'hasard et l'habileté des joueurs." In *Théorie des probabilités*, edited by J. Hermann, pp. 204–24. Paris: Librarie Scientifique. Translated as "On Games that Involve Chance and the Skill of the Players" by L. J. Savage in *Econometrica* 21(1953):101–15.

Brams, Steven. 1980. *Biblical Games*. Cambridge: MIT Press.

Buchanan, James M., and Gordon Tullock. 1962. *The Calculus of Consent*. Ann Arbor: University of Michigan Press.

Butler, Bishop Joseph. 1850 (first ed. 1736). *The Works of Joseph Butler*. New York: Robert Carter & Brothers.

Coleman, James. 1986. *Individual Interests and Collective Action*. New York: Cambridge University Press.

Davidson, Donald. 1985. "The Very Idea of a Conceptual Scheme." In *Essays on Truth and Interpretation*, pp. 183–98. New York: Oxford University Press.

Edgeworth, F.Y. 1881. *Mathematical Psychics*. London: Paul.

Ellsberg, Daniel. 1968 (first ed. 1954). "Classic and Current Notions of 'Measurable Utility.' " *Economic Journal* 64:528–56. Reprinted in *Utility Theory: A Book of Readings*, edited by A. N. Page, pp. 269–96. New York: Wiley.

Etzioni, Amitai. 1986. "The Case for a Multipe-Utility Conception." *Economics and Philosophy* 2:159–84.

Fishburn, Peter. 1989. "Retrospective on the Utility Theory of von Neumann and Morgenstern." *Journal of Risk and Uncertainty* 2:127–58.

Gauthier, David. 1986. *Morals By Agreement*. Oxford: Oxford University Press.

Hacking, Ian. 1975. *The Emergence of Probability*. New York: Cambridge University Press.

Halevy, Eli. 1972. *The Growth of Philosophic Radicalism*. Clifton, NJ: Augustus M. Kelley Publishers.

Harsanyi, John. 1967–68. "Games with Incomplete Information Played by Bayesian Players." Parts I–III. *Management Science* 14:159–82, 320–24, 486–502.

———. 1973. "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed Strategy Equilibrium Points." *International Journal of Game Theory* 5:61–94.

———. 1984. "Morality and the Theory of Rational Behaviour." In *Utilitarianism and Beyond*,

edited by Amartya Sen and Bernard Williams, pp. 39–62. Cambridge: Cambridge University Press.

Harsanyi, John, and Reinhard Selten. 1988. *A General Theory of Equilibrium Selection in Games*. Cambridge: MIT Press.

Hintikka, Jaakko. 1962. *Knowledge and Belief*. Ithaca: Cornell University Press.

Hobbes, Thomas. 1982. *Leviathan*. New York: Penguin Books.

Hume, David. 1981. *A Treatise of Human Nature*. Oxford: Oxford University Press.

Keynes, John Maynard. 1929. *A Treatise on Probability*. London: MacMillan and Co., Ltd.

Knight, Frank. 1971 (first ed. 1921). *Risk, Uncertainty and Profit*. Chicago: University of Chicago Press.

Kreps, David. 1988. *Notes on the Theory of Choice*. Boulder: Westview Press.

Kreps, David, Paul Milgrom, John Roberts, and Robert Wilson. 1982. "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma." *Journal of Economic Theory* 27:245–52.

Kyburg, Henry., Jr. 1961. *Probability and the Logic of Rational Belief*. Middletown, CT: Wesleyan University Press.

Lewis, David. 1969. *Convention*. Cambridge: Harvard University Press.

Luce, R. Duncan, and Howard Raiffa. 1957. *Games and Decisions*. New York: Wiley.

Maynard Smith, John. 1982. *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

Morgenstern, Oskar. 1976 (first ed. 1935). "Perfect Foresight and Economic Equilibrium." In *Selected Writings of Oskar Morgenstern*, edited by A. Schotter, pp. 169–83. New York: New York University Press.

Moulin, Herve. 1982. *Game Theory for the Social Sciences*. New York: New York University Press.

Nash, John. 1951. "Noncooperative Games." *Annals of Mathematics* 54:286–95.

Page, Alfred N. 1968. *Utility Theory: A Book of Readings*. New York: Wiley.

Pareto, Vilfredo. 1971. *Manual of Political Economy*. Translated from the French edition of 1927 by Ann Schwier; edited by Ann Schwier and Alfred N. Page. New York: Augustus M. Kelley Publishers.

Pearce, David. 1984. "Rationalizable Strategic Behavior and the Problem of Perfection." *Econometrica* 52:1029–50.

Ramsey, F.P. 1990 (first ed. 1926). "Truth and Probability." In *Philosophical Papers*, edited by D. H. Mellor, pp. 52–109. Cambridge: Cambridge University Press.

Rescher, Nicholas. 1985. *Pascal's Wager*. Notre Dame: University of Notre Dame Press.

Resnik, Michael. 1987. *Choices*. Minneapolis: University of Minnesota Press.

Roth, Alvin. 1988. *The Shapley Value: Essays in Honor of Lloyd S. Shapley*. New York: Cambridge University Press.

Roth, Alvin, and Marilda Sotomayor. 1990. *Two-Sided Matching: A Study in Game Theoretic Modeling and Analysis*. New York: Cambridge University Press.

Rousseau, Jean-Jacques. 1983. *On the Social Contract and Discourses*. Translated by D. Cress. Indianapolis: Hackett.

Samuelson, Paul. 1983. *Foundations of Economic Analysis*. Enlarged edition. Cambridge: Harvard University Press.

Savage, Leonard. 1972. *The Foundations of Statistics*. Second revised edition. New York: Dover Publications.

Schelling, Thomas C. 1980 (first ed. 1960). *The Strategy of Conflict*. Cambridge: Harvard University Press.

Schick, Frederic. 1984. *Having Reasons: An Essay on Rationality and Sociality*. Princeton: Princeton University Press.

———. 1986. "Dutch Bookies and Money Pumps." *Journal of Philosophy* 83:112–19.

Schumpeter, Joseph. 1954. *History of Economic Analysis*. New York: Oxford University Press.

Selten, Reinhard. 1975. "Re-examination of the Perfectness Concept for Equilibrium in Extensive Games." *International Journal of Game Theory* 4:22-25.

Sen, Amartya. 1977. "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory." *Philosophy and Public Affairs* 6:317–44.

Shafer, Glenn. 1989. "The Unity and Diversity of Probability." Ronald G. Harper Distinguished Professor of Business Inaugural Lecture, University of Kansas.

Shapley, Lloyd S. 1953. "A Value for N-person Games." In *Contributions to the Theory of Games, II, Annals of Mathematics Studies* 24:307–17, edited by H. W. Kuhn and A. W. Tucker. Princeton: Princeton University Press.

Skyrms, Brian. 1990. *The Dynamics of Rational Deliberation*. Cambridge: Harvard University Press.

Smith, Adam. 1976 (first ed. 1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. Indianapolis: Liberty Press.

Spiegel, Henry. 1983. *The Growth of Economic Thought*. Revised edition. Durham, NC: Duke University Press.

Stigler, George. 1950. "The Development of Utility Theory, I and II." *Journal of Political Economy* 58:307–27, 373–96.

Thucydides. 1952. *The History of the Peloponnesian War*. Translated by Richard Crawley. Chicago: University of Chicago Press.

van Damme, Eric. 1987. *Stability and Perfection of Nash Equilibria*. Heidelberg: Springer-Verlag.

von Neumann, John. 1959 (first ed. 1928). "On The Theory of Games of Strategy." In *Contributions to the Theory of Games*, Vol. IV, edited by A. Tucker and R. D. Luce, pp. 13–42. Princeton: Princeton University Press. Translated by Sonya Bargmann from "Zur Theorie der Gesellschaftsspiele." *Mathematik Annalen* 100:295–320.

von Neumann, John, and Morgenstern, Oskar. 1953. *Theory of Games and Economic Behavior*. Third edition. Princeton: Princeton University Press.