4

Rational choice in the context of ideal games

EDWARD F. MCCLENNEN

1. INTRODUCTION

Traditionally, the problem for the theory of two-person games has been to establish the solution to an ideal type of interdependent choice situation characterized by the following background condition.

(1) **Common knowledge.** There is full common knowledge of (a) the rationality of both players (whatever that turns out to mean), and (b) the strategy structure of the game for all players, and the preferences that each has with respect to outcomes.

The force of this condition is that if a player *i* knows something that is relevant to a rational resolution of *i*'s decision problem, then any other player *j* knows that player *i* has that knowledge. This is typically taken to imply (among other things) that one player cannot have a conclusive reason, to which no other player has access, for choosing in a certain manner. That is, there are not hidden arguments for playing one way as opposed to another.

In addition, one invariably finds that the analysis proceeds by appeal to the following (at least partial) characterization of rational behavior for the individual participant.

(2) **Utility maximization.** Each player's preference ordering over the abstractly conceived space of outcomes and probability distributions over the events that condition such outcomes can be represented by a utility function, unique up to positive affine transformations, that satisfies the expected-utility principle.

(3) **Consequentialism.** Choice among available strategies is strictly a function of the preferences the agent has with respect to the outcomes (or disjunctive set of outcomes) associated with each strategy.

Following Hammond (1988), condition (3) can be taken to imply that strategies are nothing more than neutral access routes to outcomes (or disjunctions of outcomes); the latter are what preferentially count for the agent. In particular, then, if two strategies yield exactly the same probabilities of the same outcomes occurring, then the agent will be indifferent between those strategies.

2. The conceptual problem and its traditional solution

At the very outset of the formal study of games, one finds von Neumann and Morgenstern (1953) suggesting that one faces a conceptual as distinct from a technical difficulty when moving from the study of the isolated individual (the proverbial Robinson Crusoe) who faces an "ordinary maximum problem" to the study of interacting persons. With regard to the former:

> Crusoe is given certain physical data (wants and commodities) and his task is to combine and apply them in such a fashion as to obtain a maximum resulting satisfaction. There can be no doubt that he controls exclusively all the variables upon which this result depends – say the allotting of resources, the determination of the uses of the same commodity for different wants, etc. (1953, p. 10)

To be sure, outcomes may be conditioned by such "uncontrollable" factors as weather; but these, they go on to indicate, can be registered by appeal to statistical assumptions and mathematical expectations.

What happens when one shifts to the case of a participant in a social exchange economy? In this case, the authors argue,

> the result for each will depend in general not merely upon his own action but on those of the others as well. Thus each participant attempts to maximize a function . . . of which he does not control all the variables. (1953, p. 11)

This is a problem, they suggest, that cannot be overcome simply by appeal to probabilities and expectations:

> Every participant can determine the variables which describe his own actions but not those of the others. Nevertheless these "alien" variables cannot, from his point of view, be described by statistical assumptions. This is because the others are guided, just as himself, by rational principles – whatever that may mean – and no *modus procedendi* can be correct which does not attempt to understand those principles and the interactions of the conflicting interests of all participants. (1953, p. 11)

Kaysen, in a very early review of von Neumann and Morgenstern's *Theory of Games and Economic Behavior,* takes this to mean that the theory of

> . . . games of strategy deals precisely with the actions of several agents, in a situation in which all actions are interdependent, and where, in general, there is no

possibility of what we have called parametrization that would enable each agent (player) to behave as if the actions of the others were given. In fact, *it is this very lack of parametrization which is the essence of a game.* (1946–7, p. 2; emphasis added)

What is interesting about this, of course, is that starting with von Neumann and Morgenstern and continuing on in more or less unbroken fashion ever since, the basic strategy has been to solve the game problem by showing that, contrary to Kaysen's suggestion, parametrization of one sort or other is possible. In very general terms, the notion is that if (1)–(3) govern the behavior of both players then this (together with certain additional assumptions) will enable each player to frame expectations about the behavior of the other player, expectations that are sufficiently determinate to enable each to treat the decision problem as a single maximization problem.

Within this sort of framework, the typical first move has been to defend the following as a necessary condition of rational choice.

Rejection of dominated strategies. An agent should never choose a strategy whose associated set of possible outcomes is strictly dominated by the set of outcomes associated with some other strategy.

The rationale for the dominance condition is thought to be clear enough. To say that one strategy *strictly dominates* another is to say that, regardless of one's expectation concerning how the other player will choose, the utility of the expected outcome of the former is strictly greater than the utility of the corresponding expected outcome of the latter. But then, (2) and (3) together would seem to require that one reject the latter in favor of the former.

Dominance, even when invoked iteratively, does not get one very far. For most game theorists, however, conditions (1)–(3) have been thought to be sufficiently strong to provide a grounding for Nash's (1951) equilibrium concept. Luce and Raiffa (1957) is the *locus classicus* here. The argument is indirect in form, and provides a model of how to show that parametrization will be possible within the framework of assumptions (1)–(3). Luce and Raiffa begin by assuming that there exists a theory of rational interdependent choice sufficiently determinate that, under conditions of common knowledge, each player will be able to predict what each other rational player will do; they then proceed to explore what conclusions can be drawn about the content of that theory:

It seems plausible that, if a theory offers A_{i_0} and B_{j_0} as suitable strategies, the mere knowledge of the theory should not cause either of the players to change his choice: just because the theory suggests B_{j_0} to player 2 should not be grounds

for player 1 to choose a strategy different from A_{i_0}; similarly, the theoretical prescription of A_{i_0} should not lead player 2 to select a strategy different from B_{j_0}. Put in terms of outcomes, if the theory singles out (A_{i_0}, B_{j_0}), then:

(i) No outcome O_{ij_0} [i.e., one that 1 could realize, given that 2 plays B_{j_0}, by playing some strategy other than one picked out by the theory] should be more preferred by 1 to $O_{i_0 j_0}$.

(ii) No outcome $O_{i_0 j}$ [i.e., one that 2 could realize, given that 1 plays A_{i_0}, by playing some strategy other than one picked out by the theory] should be more preferred by 2 to $O_{i_0 j_0}$.

And A_{i_0} and B_{j_0} satisfying conditions (i) and (ii) are said to be in *equilibrium*, and the *a priori* demand made on the theory is that the pairs of strategies it singles out shall be in equilibrium. (1957, p. 63)

One can mark in this an implicit appeal to all three of the conditions delineated. In accordance with (1), not only is the strategy and payoff structure of the game presumed to be common knowledge, it is also presumed that each player is aware of what the (postulated) theory of rational choice instructs each other player to do. But this information, when coupled with (2) and (3), provides that player with a basis for choice.

To be sure, the equilibrium condition has more recently come under a certain amount of attack, as evidenced in Bernheim (1984, 1986), Pearce (1984), and Aumann (1987). Bernheim (1986) nicely sorts out a crucial issue – namely, whether it is possible to develop a theory determinate enough for each player to be able to predict the specific choice that each other player will make. Given such predictability, the equilibrium condition is a necessary condition on the solution to any ideal game: that is, the solution to any game will have to be a refinement of the set of Nash equilibria. If predictability does not hold then it appears that one must retreat to some weaker condition, such as rationalizability. Bernheim (1986) also provides a useful exploration of what assumptions, in addition to (1)–(3), suffice to characterize the various alternative solution concepts that have emerged. For my purposes here, however, what is significant is that all of these revisionist moves have in common with the original equilibrium perspective both a commitment to the principle of iterated dominance and all three of the conditions listed in Section 1. My concern is with the implications of any theory of this type, so it will usually suffice, when a representative example is needed, to refer to the equilibrium theory.

3. Problems with conceptualizing interactive choice as a species of parametrized choice

What characterizes the theories just discussed is that each rational player is presumed to face the task of framing some sort of "reasonable" estimate as to how the other players will choose, and then, per conditions (2)

and (3), responding to that estimate by choosing among personal strategies *so as to maximize his or her (subjectively defined) expected utility* over associated outcomes. However, it is a matter of considerable interest, from both an analytic and a historical point of view, that von Neumann and Morgenstern adopted a quite distinct way of conceptualizing the task facing the individual player in the zero-sum (i.e., perfectly competitive) game.

Von Neumann and Morgenstern begin by explicitly appealing to an indirect argument, parallel to but quite distinct from the one employed by Luce and Raiffa:

> Let us now imagine that there exists a complete theory of the zero-sum two-person game which tells each player what to do, and which is absolutely convincing. If the players knew such a theory then each player would have to assume that his strategy has been "found out" by his opponent. The opponent knows the theory, and he knows that a player would be unwise not to follow it. Thus the hypothesis of the existence of a satisfactory theory legitimatizes our investigation of the situation when a player's strategy is "found out" by his opponent. (1953, pp. 147–8)

What is the implication of a player expecting that his choice will be found out? In the case of the zero-sum, two-person game, that the other player will correctly anticipate the given player's strategy choice, and maximize from that player's own perspective, implies that a given player must expect to end up receiving the minimum utility associated with whatever strategy is chosen. In the light of that expectation, (2) and (3) imply that the player should choose a strategy whose associated minimum-valued outcome takes on a maximum value; that is, the player should *maximin*. In this context, then, their indirect argument directly ratifies not the equilibrium requirement but rather the principle that each player should employ a maximin strategy.

Of course, within the context of zero-sum, two-person games, it is easy to establish that pairs of maximin strategies are also equilibrium pairs and vice versa. However, von Neumann and Morgenstern's indirect argument is quite general; nothing in its formulation limits its application to zero-sum games. In a more general setting, this indirect argument implies that a player must still expect that the other player will choose a utility-maximizing response to what the first player chooses. But this in turn implies that the first player should choose so as to maximize expected utility, computed on the presupposition that whatever choice is made, the other player will maximize (from that player's own perspective) in response. Let us designate any strategy that satisfies this condition a *maxilor* strategy. Correspondingly, the expected return from playing a given strategy, on the assumption that the other player responds in a utility-maximizing fashion, can be characterized as the *maxilor return* for that

strategy. Given these distinctions, one can now formulate the following condition on a rational solution.

Maxilor condition. If a theory prescribes that player 1 select strategy A_{i_0}, then A_{i_0} should maximize expected utility for player 1 in the light of a maximizing response on the part of player 2; similarly, if a theory prescribes that player 2 select strategy B_{j_0}, then B_{j_0} should maximize expected utility for player 2 in the light of a maximizing response on the part of player 1.

Here, then, is a quite distinct way to arrive at the notion that each player is in a position to take the behavior of the other player as a given. To be sure, the choice behavior of the other player is in this instance a dependent variable. The net effect, however, is the same: for each available strategy, the agent need only specify the expected value of the consequence of selecting that option; thus, once again, the agent faces an ordinary maximization problem.

Note that both the equilibrium argument and the maxilor argument turn on the assumption that rational players satisfy (2) and (3). The former argument invites one to think of oneself as an outcome maximizer against an estimated choice on the part of the other player; the latter invites one to think about the requirements of outcome maximizing when one must contend with a counterpart player who is maximizing against oneself, under conditions of common knowledge. The former line of reasoning proceeds, in effect, to trace out the implications for a utility maximizer of being actively able to find out what the other player will do; the latter traces out the implications of passively anticipating that the chosen strategy will be found out by the other player. Both implications are seemingly forced upon us by the consideration that, under ideal conditions, each will be able to anticipate what the other will do.

It turns out that for the zero-sum, two-person game, a solution will satisfy the equilibrium condition if and only if it also satisfies the maxilor condition. In this context, the existence of rival conceptions of parametrizing the choice situation poses no problem. But in the non–zero-sum case, the two indirect arguments will typically yield conflicting requirements. In the much-discussed game of "chicken," for example, the maxilor solution has the two players crash head-on; equilibrium solutions exclude that case and include the two cases where one player swerves. If both arguments can be sustained, then one has an impossibility result with respect to the rational solution for non–zero-sum, two-person games.

This conflict can be resolved, of course, by dropping the maxilor condition. But there remains a problem, and it is one that arises even within

the context of zero-sum games. As I sought to argue in McClennen (1972), because equilibrium strategies are not necessarily unique best replies to the choice of an equilibrium strategy by the other player, the equilibrium solution concept cannot be squared with the implications of consequentialism and expected-utility theory, specifically with respect to strategies with the same expected utility.

It should also be noted that von Neumann and Morgenstern's maxilor model at the very least serves as a striking reminder that even the strict dominance condition cannot be directly derived from the framework of conditions (1)–(3). Most game theorists have taken strict dominance as bedrock, but in the presence of a belief that the other player's choice is probabilistically dependent upon one's own choice of a strategy – precisely the sort of belief that characterizes the maxilor perspective – regimentation to the principle of strict dominance is not necessarily rational. Yet even in the recent revisionist work of Bernheim (1984, 1986), Pearce (1984), and Aumann (1987), this issue seems to be begged: certain logically possible forms of probabilistic dependence apparently have been ruled out of court.

Another (and much more remarked upon) problem that arises in connection with the standard approach for non–zero-sum games is that "rational" solutions will typically be suboptimal. The problem of suboptimality is usually introduced in connection with what is known as the prisoners' dilemma game. But the dilemma is, of course, endemic to the whole class of non–zero-sum, n-person games. If rational choice for interactive situations under the ideal conditions specified in (1) must satisfy conditions (2) and (3), as those have usually been interpreted, then one is forced to the unhappy conclusion that fully rational players who have common knowledge of each other's rationality, and common knowledge of the strategy and payoff structure of the game, must each nonetheless deliberately choose in a manner that leaves both with a less preferred outcome than they could have achieved if only they had coordinated their choices. It is also clear that there is nothing to be gained in this respect by shifting from an equilibrium to a maxilor perspective.

The response has typically been to hold fast to some version of the standard theory and insist that the suboptimality of solutions to most games is simply an unavoidable anomaly of an adequate theory of rational choice. Once this point is made, one usually finds a remark to the effect that the problem can be circumvented if provision is made for binding agreements. The suggestion is that it will be rational to ensure that agreements are binding, since both parties stand to gain thereby. But all of this carries with it the rather curious implication that rational agents will be willing to expend resources on restructuring their environment to ensure that agreements will be binding (and thereby increase their return), and

yet be unwilling to agree to and act on a self-policing approach – despite the fact that typically, in virtue of the costs of making agreements binding, the latter approach would result in even greater returns to each.

4. Diagnosing what fuels these various problems

Consider the following very simple game, a demi-version of the standard prisoners' dilemma:

		Player 2	
		B_1	B_2
Player 1	A_1	3, 4	1, 3
	A_2	4, 1	2, 2

Player 1 has a dominant strategy, A_2, whereas player 2 does not. Player 2 would prefer to choose B_1 and thereby cooperate so as to realize the outcome (3, 4), but only if player 2 were convinced that player 1 would be cooperative. On the standard way of reasoning, however, that player 1 has a dominant strategy suffices to determine the rational outcome of this game. The rational choice for player 1 is to play this dominant strategy, and in turn player 2, given conditions of full information and common knowledge of the rationality of each, will expect player 1 to behave just so. However, in light of this expectation, player 2's best response will be B_2. On the usual account, then, barring some way to make binding agreements with one another, rational agents who know each other to be such must settle for the outcome (2, 2), despite the fact that each would prefer the outcome (3, 4).

What characterizes the standard way of thinking about interactive rationality is the manner in which it anchors the choice of a strategy. As suggested in Section 1, there is invariably (if often only implicitly) an appeal to a consequentialist perspective, according to which strategies are merely neutral access routes to consequences. Within the framework of this assumption, any preference with respect to strategies must be accounted for by reference back to preferences for expected consequences.

Yet consequentialism so interpreted is worrisome. In particular, it is precisely player 1's (allegedly rational, and consequentialist based) disposition to choose the dominant over the dominated strategy that precludes coordination with player 2 to jointly implement a plan the consequences of which are preferred to the consequences of choosing A_2 and player 2 responding with B_2. That is, by hypothesis, player 1 prefers the outcome of coordination to the outcome of what is alleged to be rational interaction. Notice, moreover, that player 1 cannot rationalize having to

settle for a utility payoff of 2 rather than 3 by reference to the dispositions of agent 2. Agent 2 would be quite willing to coordinate on a plan that will realize the joint payoff (3, 4) – once assured that player 1 will really cooperate – so what stands between agent 1 and the larger payoff is just agent 1's own disposition.

This suggests that perhaps more is involved in the standard arguments than simply an appeal to conditions (2) and (3), utility maximization with respect to abstractly considered outcomes, and consequentialism. Adapting an argument offered in McClennen (1988, 1990) with respect to dynamic foundations for expected-utility reasoning, one can suggest that there is, in addition, an implicit appeal to a separability assumption. In the agent's deliberation concerning the choice of neutral means to preferred outcomes, it is invariably presupposed that the agent can separate out the piece of the problem that pertains just to the choice to be made – that is, consider it *in abstraction from the context of the interactive problem* itself – and consider how to evaluate just those options if it were the case that, instead of interacting with another rational player who is also deliberating about what choice to make, the agent needed only to take into account some parameter (about whose value the agent may, of course, be uncertain). The choice made under these transformed conditions is then the one that should be made in the context of the interactive situation (the game) itself. This may be stated somewhat more formally as follows.

Separability. Let G be any game, and let D be the problem that a given player in G would face, were the outcomes of the available strategies in G conditioned not by the choices of another player but rather by some "natural" turn of events in the world, so that the player faces (in effect) a classic problem of individual decision making under conditions of risk or uncertainty. Suppose further that the player's expectation with regard to the conditioning events corresponds to the expectations held with regard to the choice that the other player will make in G. Then the first player's preference ordering over the options in G must correspond to that player's preference ordering over the options in D.

For the decision situation D_{dpd}, corresponding to the demi-prisoners' dilemma game G_{dpd}, consequentialism appears to unproblematically imply that the player's preference ordering of outcomes in D_{dpd} determines the ordering of the alternatives in D_{dpd}. In particular, on the assumption that the turn of events in D_{dpd} is not linked (causally or probabilistically) with choices made, agent 1's preferences for the corresponding outcomes – together with the dominance principle – imply that agent 1 must choose

A_2. The point is that if it is certain that B_1 would occur then agent 1 would obviously choose A_2, and if it is certain that B_2 would occur then agent 1 would also choose A_2; so A_2 is the best choice, regardless of the turn of events in the world. In this case, then, no matter what the expectations with regard to conditioning events, there is a clear choice. But separability, in turn, requires that player 1's ordering of the alternatives in D_{dpd} determines the ordering of the alternatives faced in G_{dpd}. Thus, these two principles taken together yield the standard conclusion.

Notice, more generally, that expected-utility maximization, consequentialism, and separability together imply that if a given agent can anticipate the choice to be made by the other agent (or at least assign a probability distribution over an exclusive disjunction of possible choices), then the first agent should maximize (expected) return, given this anticipation. Correspondingly, they also imply that if the agent believes that the other player will correctly anticipate whatever choice is made, and maximizes in response, then the agent should choose a maxilor strategy. In this case, however, the appropriate model is one of decision making against nature under conditions where states of nature are causally dependent upon the agent's choice of an action. The separability assumption, then, plays a key role in both of the ways (equilibrium and maxilor) in which, contrary to Kaysen's suggestion, ideal forms of interactive choice can be treated as presenting each player with a parametrized choice problem.

Note the logic of the evaluation of choices in any such separable framework. An agent who is committed to separability will choose so as to maximize with respect to preferences for expected consequences, given the agent's expectations as to how the other agent will choose. This has the further (and most important) implication that the evaluation of any proposed coordination plan proceeds from the evaluation of what the plan calls upon a given agent to choose, holding all other features of the plan fixed; that is, it proceeds from the evaluation of each segment of that plan to the whole plan. A plan must be judged as not acceptable if it calls for some agent to make a particular choice that the agent would not be disposed to make if the decision problem were viewed as separable in the sense just introduced. Returning to game G_{dpd}, the plan calling for agent 1 to choose A_1 and agent 2 to choose A_2 must, from this perspective, be rejected: It calls upon agent 1 to make a choice that would not be made were that same set of outcomes presented in abstraction from the interactive setting of G_{dpd}.

Thus, separability places substantial restrictions on the capacity of an agent to coordinate choices with others. Indeed, separability in this context precludes coordination in any meaningful sense of that term. What is left to the agent who is committed to such a separability principle is not

coordination but strategic interaction: The agent's task is to estimate how the other agent will choose and then to make unilateral adjustments in choice so as to maximize expected return.

5. The case for nonseparability in the non-zero-sum context

What recommends separability as a necessary condition for non-zero-sum rational interactive choice? As already remarked, it might well seem that, within the framework of conditions (2) and (3), it is one and the same whether the outcome of choosing an action is conditioned by choices that another agent makes or by natural events. That is, the rational agent is to conceive the problem here as no different from that faced in the case of statistical decision making against nature. This problem is one that calls for the agent to make independent adjustments in choice, against independently or dependently fixed values of the other variables, so as to achieve (by means of such an adjustment) a maximum expected return.

But this is *consequentially* costly. In the case of our demi-prisoners' dilemma, adoption of such a separable perspective precludes agent 1 from being able to agree to a cooperative scheme with agent 2 and then adhere to it. Effective cooperation between two rational agents in this situation requires agent 1 to be willing to refrain from an independent readjustment of choice (i.e., switching to A_2), given the expectation that agent 2 will choose B_1. But this is precisely what an agent who is committed to separability cannot do. To be sure, agent 1's commitment to viewing things from a separable perspective supplies a motive for persuading agent 2 to believe that agent 1 will cooperate, but it is equally certain that cooperation is not rational for agent 1. Under conditions of common knowledge, agent 2 must then expect that agent 1 will not cooperate, and so on. Thus, the agent who views rationality from such a separable perspective must forego the gains that coordinated choice would make possible.

This not only serves to undercut the plausibility of separability as a criterion of rational choice, but it does so, interestingly enough, by reference to a consequentialist consideration. In very general terms, the notion is that a condition C cannot be taken as a criterion of a consequentially oriented theory of rational choice for a given class of games if acceptance of C works to the agent's own disadvantage. For the case in question, an agent committed to a separable perspective ends up with an outcome (consequence) less preferred to one that could have been realized had that agent been disposed to coordinate choices with the other agent. For this class of cases, then, the consequence of a commitment to separability is that the agent does less well in interaction with other rational agents than

it would be possible to do. But this renders suspect the claim that separability is a necessary condition of rational choice.

One can attempt to ground the separability condition in some other way, by appeal (say) to some intuitive notion of "consistency." But granting that, what requires us to take any such intuitive basis as overriding when it comes into conflict with the revised version of consequentialism? That is, what sense can we make of a consistency requirement the imposition of which implies that rational persons under conditions of common knowledge must settle for less than they could otherwise obtain?

I do not expect, of course, that all will be converted by this argument. But it does seem at the very least that there is a need to sort out more carefully the presuppositions that characterize the modern theory of rationality. Ever since the prisoners' dilemma pattern was first identified, theorists have persisted in treating a consequentialist perspective as requiring mutual defection, even though the consequence is that both players do less well.

It is interesting to note that von Neumann and Morgenstern partially abandon the separable perspective when they move to the theory of *n*-person (rather than two-person) zero-sum games. In the case of a game between three or more players, there can be a parallelism of interests that makes cooperation desirable; this will, in at least some cases, lead to an agreement between some of the players involved. If the game is zero-sum then of course it cannot be in the interests of *all* the players to join in a grand coalition, but smaller coalitions may still form. When this happens, von Neumann and Morgenstern imagine that the coalition will coordinate to secure the maximum payoff possible for members of that group, thereby ensuring that between the coalition and those who remain outside there will be a strict opposition of interest (1953, Section 25). This, then, provides a real place for full cooperation within their theory of *n*-person, zero-sum games.

Von Neumann and Morgenstern also sketch a theory of non–zero-sum games that retains the presupposition that rational agents will be disposed to coordinate when there are gains to be secured thereby. In particular, they suppose that any non–strictly competitive game involving n agents can be embedded in a strictly competitive game in which there is one additional "fictional" player – might not one think of this as nature? – whose payoff is simply the negative of the payoff that the n players can achieve if they form a coalition of all n players. The suggestion is that in a game such as G_{dpd}, the two agents can think of themselves as jointly playing a strictly competitive game against "nature," where their best strategy is to fully cooperate with one another and thereby force the maximum joint payoff possible from nature (1953, Section 56).

6. Nonseparable interactive rationality

I am not at all sure what a full theory of rational interactive choice would look like within a nonseparable framework. I suggest, however, that a theory that is prepared to reject the separability condition for non-strictly competitive games would take the familiar principle of collective rationality – the Pareto optimality condition – as a necessary condition of rational interaction under conditions of common knowledge. I have sought to say something about how one might motivate this view in McClennen (1985). On such an alternative conception of rational interaction, rational agents – who are able to communicate with one another (or who can tacitly bargain) and who have common knowledge of each other's rationality, and so forth – will not face the classical prisoners' dilemma problem. Such agents will be able to reach an agreement on, and then implement, a plan that satisfies the Pareto optimality condition. In a corresponding manner, models of suboptimal equilibrium outcomes are best understood as models of interaction under nonideal conditions, that is, those of imperfect rationality or imperfect information.

What, in addition to the Pareto optimality condition, is likely to figure in a theory of interaction between rational agents under circumstances of common knowledge? Recent work in the theory of bargaining and negotiation is clearly relevant here. Unfortunately, since the prevailing view has been that interactive situations in general are best understood in terms of models of strategic (noncooperative) rather than cooperative choice, bargaining theory is somewhat less than fully developed. However, important contributions to such a theory are to be found, for example, in Nash (1953), Kalai and Smorodinsky (1975), and Gauthier (1986, chap. V). One might hope, moreover, that appreciating the unsatisfactory implications of the standard approaches just surveyed will lead theorists to consider reintroducing the concept of coordination into an area from which it has been systematically banished – namely, the theory of the non-zero-sum game between ideally rational players – and that this will, in turn, spur increased interest in the subject of bargaining theory.

References

Aumann, R. J. (1987), "Correlated Equilibrium as an Expression of Bayesian Rationality." *Econometrica* 55: 1–18.

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

Bernheim, B. D. (1986), "Axiomatic Characterizations of Rational Choice in Strategic Environments." *Scandinavian Journal of Economics* 88: 473–88.

Gauthier, D. (1986), *Morals by Agreement*. Oxford: Clarendon Press.

Hammond, P. (1988), "Consequentialist Foundations for Expected Utility." *Theory and Decision* 25: 25–78.

Kalai, E., and Smorodinsky, M. (1975), "Other Solutions to Nash's Bargaining Problem." *Econometrica* 43: 513–18.
Kaysen, K. (1946–7), "A Revolution in Economic Theory?" *Review of Economic Studies* 14(1): 1–15.
Luce, R. D., and Raiffa, H. (1957), *Games and Decisions.* New York: Wiley.
McClennen, E. F. (1972), "An Incompleteness Problem in Harsanyi's General Theory of Games and Certain Related Theories of Non-Cooperative Games." *Theory and Decision* 2: 314–41.
McClennen, E. F. (1985), "Prisoners' Dilemma and Resolute Choice." In R. Campbell and L. Sowden (eds.), *Paradoxes of Rationality and Cooperation.* Vancouver: University of British Columbia Press.
McClennen, E. F. (1988), "Dynamic Choice and Rationality." In B. R. Munier (ed.), *Risk Decision and Rationality.* Dordrecht: Reidel, pp. 517–36.
McClennen, E. F. (1990), *Rationality and Dynamic Choice: Foundational Explorations.* Cambridge: Cambridge University Press.
Nash, J. (1951), "Non-Cooperative Games." *Annals of Mathematics* 54: 286–95.
Nash, J. (1953), "Two-Person Cooperative Games." *Econometrica* 21: 128–40.
Pearce, D. G. (1984), "Rationalizable Strategic Behavior and the Problem of Perfection." *Econometrica* 52: 1029–50.
Von Neumann, J., and Morgenstern, O. (1953), *Theory of Games and Economic Behavior,* 3rd ed. New York: Wiley.