The Journal of Socio-Economics 37 (2008) 2155–2167

The Journal of Socio-Economics

www.elsevier.com/locate/soceco

Cooperative games and cooperative organizations

Roger A. McCain *

Drexel Univerasity, Philadelphia, PA, United States

Received 21 June 2007; received in revised form 21 January 2008; accepted 6 February 2008

Abstract

It is well known that game theory has two major branches, cooperative and noncooperative game theory. Noncooperative game theory is the better known and more influential of the two. A difference is that cooperative game theory admits of binding agreements to choose a joint strategy in the mutual interest of those who agree. Cooperative organizations, too, are seen as being in the mutual interest of the members, but there has been little contact between the two bodies of thought. This paper surveys cooperative game theory and explores the extent to which cooperative game theory may help us to understand (and perhaps extend) cooperative organizations. In particular, reciprocity motives are introduced into the cooperative game analysis, and this may provide a link between cooperative game theory and cooperative organizations.

JEL classification: C71; P31; J59

Keywords: Game theory; Cooperative games; Cooperatives

In the 60 years since the publication of *The Theory of Games and Economic Behavior*, game theory has become an interdisciplinary field of research with applications in economics, political science, management, and philosophy, encompassing a wide range of models and methods. In particular, there is a broad dichotomy between noncooperative and cooperative game theory. Of the two, noncooperative game theory has been more successful and influential, if only in that the five men who have been honored with the Nobel memorial prize in economics for their work in game theory have all been honored primarily for work in noncooperative game theory. Nevertheless, the literature of cooperative game theory is extensive and important, and has some applications in economics and other fields. Cooperative game theory is applicable whenever the players in a game can form "coalitions," groups that choose a common strategy to improve the payoffs to the members of the group.

According to the International Cooperative Alliance (1995), "A co-operative is an autonomous association of persons united voluntarily to meet their common economic, social, and cultural needs and aspirations . . .." There seems a clear parallel between these definitions. However, the definition of a cooperative organization continues "... through a jointly-owned and democratically-controlled enterprise." These elements do not occur in the concept of a cooperative game. Is there more to this parallel than the word "cooperative?" Has cooperative game theory anything to teach the Cooperative Movement?

* Tel.: +1 215 895 2176.
E-mail address: mccainra@drexel.edu.

1053-5357/$ – see front matter
doi:10.1016/j.socec.2008.02.010

1. The emergence of cooperative game theory

Much as economics springs from Smith's (1937) *An Inquiry into the Nature and Causes of the Wealth of Nations*, game theory springs from a single influential book, *The Theory of Games and Economic Behavior*, by von Neumann and Morgenstern. Von Neumann and Morgenstern adopted the mathematician's research strategy: to find a solution for the problem in its simplest form, and then to extend the solution step by step to more complex (and perhaps realistic) cases, with applications to "serious" interactions such as business competition and war. Schelling (1960) and Aumann (2003), who shared the 2005 Nobel Memorial Price, have both suggested that game theory can be better understood as "interactive decision theory."

1.1. Founders

The simplest cases of interactive decisions are those with only two decision-makers, that is, two-person games. Von Neumann and Morgenstern simplified still further, considering two-person games in which anything gained by one must be lost by the other, that is, two-person zero sum games. Their solution to this simplest class of games – that each person will choose the strategy that maximizes his minimum payoff – is still recognized as the canonical solution of this simple case. But in a zero-sum game, there is no potential for cooperation in any sense, as there is (by assumption) no possibility of mutual gain. Thus, this was the beginning of *noncooperative game theory*. Later, Nash (1951) would revise the von Neumann and Morgenstern (1944) solution in a form that could be extended to nonconstant sum games (that is, interactions with win–win and lose–lose potentialities) and demonstrate that all two-person games would have such solutions. This, with Alfred Tucker's famous Prisoner's Dilemma example, set the stage for the growth of noncooperative game theory.

However, this was not the direction that von Neumann and Morgenstern took. They observed that in any game with more than two players, or with win–win or lose–lose potentialities, players could benefit by forming groups for mutual benefit, that is, coalitions. They first studied games without win–win or lose–lose potentialities, that is, constant-sum games with three or more players. In these games the sum of the winnings of all players is constant, but it may be that one group of players can increase their earnings by "ganging up" on the others. A solution for a game in coalition function form should tell us (1) which coalitions will form, if any, and (2) how each coalition will divide its winnings among the members. In order to answer that, we must make some judgment as to what payoffs "rational" game-players will demand or expect or offer. The judgment that von Neumann and Morgenstern relied on was that a "rational" game player would not reject a take-it-or-leave it offer that is better than he could get individually in the absence of any cooperative agreement and (drawing on their solution for two-person zero sum games) that what he could get individually is the minimum payoff. They define a concept of "dominance" that can lead to a very large number of solutions, and this was seen as a major drawback of the von Neumann and Morgenstern analysis.

In any case, they extended their analysis to nonconstant sum games by treating these games as having a "dummy" player. The "dummy" player does not play but has gains or losses that make up the losses or gains of all the "real" players. Thus, the nonconstant sum game is equivalent to a constant-sum game with one more player. In general, it would be to the advantage of the "real" players to gang up and extract as much as possible from the dummy, who, after all, has neither defenses nor pains of deprivation. Still, in general, there could be a very large number of solutions.

The ongoing research of cooperative game theorists in the following years mostly focused on narrowing the field—on refining or modifying the von Neumann and Morgenstern solution to reduce the number of possible solutions. Ideally the mathematician would like to have exactly one solution. Shapley (1953) proposed a solution, the Shapley value, that had that desirable property. We shall have a bit more to say about the Shapley value later. Another proposed solution came from Nash (1950), who proposed a solution based on bargaining. Nash supposes that the bargainer will balance the risk of failing to come to an agreement against the benefit of demanding a higher share of the payoff. Like Shapley's solution, it had some attractive properties, and indeed has continued to be used in studies of bargaining; but it is difficult to extend Nash's bargaining theory to more than two parties in any common-sense way.[1] What neither Shapley nor Nash considered was this possibility: a subgroup might be able to organize themselves to gain total payoffs more than

[1] Harsanyi (1963) attempted to do so, but his contribution has not been much developed or applied in the subsequent literature, and in any case is presented as a modification of Shapley's value concept (p. 203 et. seq.).

they are offered, and if so, would refuse the offer even if they could not do better individually. If we rule out all such offers of individual payments, we have the core of the game, an idea due to Gillies (1953), and extended by Shapley and Shubik (1969), and others. This concept of solution seems to have been more widely applied in economics than the other cooperative solution concepts. There are some other cooperative game solution concepts as well; we shall not go into them here. The number of different solution concepts is probably one reason why cooperative game theory has had less influence than noncooperative game theory.

1.2. Concepts of solution

Even more than noncooperative game theory, cooperative game theory is a mathematical field, and relies especially on mathematical set theory for many of its basic ideas. The usual assumption is that *any* group of agents in the "game" can form a coalition, and a coalition among agents A, B, and C would be denoted as {A, B, C}. The brackets {} are conventional in set theory to indicate the "elements" of a "set," or in alternative ordinary language, the individuals making up a grouping. Now suppose A, B, and C coalesce,[2] that is, form a coalition. The expectation is that by working together and choosing a joint strategy they will be able to improve their results overall. It may be that one member, let us say C, bears a special cost for this, or another agent, such as A, gets most of the benefit. An example might be the modification of a river-course, so that those downstream benefit (with water supplies for irrigation) but those upstream lose (as some of their land is flooded). Then A is downstream and C is upstream. To enlist C in the coalition it may be necessary for A to pay C some compensation. It is common to assume "transferable utility,[3]" which means that a simple transfer of some of A's winnings to C can fully compensate C. In that case all that matters is the total payoff to the group {A, B, C}. Therefore, it is common in cooperative game theory to ignore all the details and to focus on the total values the various coalitions can obtain.

Most studies in cooperative game theory will then begin by an enumeration of all possible coalitions. Suppose there are N "players in the game." An individual agent can be indicated by A_i with $i = 1, \ldots, N$. Common sense would see that any group of agents with more than one and less than N could form a coalition (with more or less difficulty). That's right, but it is not complete, and in cooperative game theory mathematical completeness is important. Therefore, in addition to those groupings, we also enumerate all singleton coalitions, $\{A_i\}$, that is, "coalitions" with just one member, and also the grand coalition of all N agents in the game and the null coalition, {Ø}, a "coalition" with no members. (By convention in set theory Ø means a set with no members.) This mathematical completeness allows us to make mathematical statements about the game relatively concisely, and thus with less difficulty. Now, as suggested above, focus on the total value that each coalition can attain, and assign to each coalition a number expressing that value. It is commonly called the value of the coalition.[4] (Of course, the value of {Ø} is zero). This assignment is called a "characteristic function" in mathematical set theory and is sometimes called the "coalition function" in cooperative game theory.

Now return to the example of a coalition of {A, B, C} and suppose we have the values for each of the singleton coalitions, coalitions with just two members, and the value of the grand coalition v{A, B, C}. At the least, a cooperative agreement among the three will have to be efficient, in the Paretian sense, so that none of the three can be made better off without making another worse off. Nothing goes to waste. We recall that von Neumann and Morgenstern's proposed solution could have a very large number of solutions, but that proposals by Nash and Shapley were designed to find a single solution among these many. The theory of the core interprets dominance somewhat differently than von Neumann and Morgenstern. Suppose that A and B can secede together and realize a value of v{A, B}. They may insist that $y_A + y_B$ be at least v{A, B} to give them the incentive to remain in the grand coalition. Other subgroups who might secede can make similar demands. But suppose that $v\{A, B, C\} < v\{A, B\} + v\{C\}$. Then there is no way to assign all three agents payoffs large enough to persuade them to stay in the grand coalition, and we say that the coalition {A, B} *dominates*

[2] There is no commonly accepted verb form for "to form a coalition," but "to coalesce" seems apposite.

[3] This phrase recognizes that the ultimate benefits of economic activity are subjective, that is, in the economist's language, "utility," but assumes that utility is nearly enough proportional to money that money side payments will be sufficient to assure that everybody in a coalition benefits on net from the coalition's activity.

[4] We are so accustomed to speaking of the value of an object that it may seem strange to speak of "the value" of a group of people. However, note that cooperators have long held that value is created (to use a common business phrase) by people working together, so speaking of the value of (created by) the group seems a common point between the cooperative movement and cooperative game theory.

Table 1
Part of Game 1 in strategic normal form

Order of payoffs: A, B	B	
	L	R
A		
U	1, 5	2, 4
D	0, 3	20, 2

any payments that can be made by {A, B, C}. If we narrow down the von Neumann and Morgenstern solutions to allow only those that are nondominated in this sense, we have the *core* of the game. But there may be no payoffs for any coalition sufficient to give the members reason to continue with the coalition, that is, the core may correspond to the empty set, Ø. On the other hand, as with the von Neumann and Morgenstern solutions, there may be many payoff schedules in the core. These are considered disadvantages of the core as a solution concept.

The Shapley value can always be computed and is unique, and has, as well, other attractive mathematical properties. The Shapley value assigns payoffs to the members of the grand coalition and tells us nothing else about the joint action of coalitions. To obtain the Shapley value, first suppose that the agents in the game are approached in the order A, B, C. Each is offered his *marginal contribution*: that is, A is offered $v\{A\} - v(\emptyset) = (A)$, B is offered $v\{A, B\} - v\{A\}$, and C is offered $v\{A, B, C\} - v\{A, B\}$. But this is somewhat arbitrary: the marginal contributions might have been different if they had been approached in a different order. Now suppose an arbitrator was to approach A, B, and C with the following proposition: "You can expect to receive your marginal contribution, but your marginal contribution depends on the circumstances, specifically the order in which you are recruited. Will you sign a contract that gives you the average of your marginal contribution over all possible circumstances, that is, all possible orders?" Being rational beings, but perhaps a little risk averse, we might expect them to accept. Anyway, this average of the marginal contributions is the Shapley value. It has the advantage that it is straightforward to compute (indeed a little easier than this example suggests in simple cases), and that it adds up to the total value, as well as the other advantages that have been mentioned.

2. The dichotomy of cooperative and noncooperative game theory

Cooperative solution concepts differ from those in noncooperative game theory, presumably because the two theory types make different assumptions about the nature of the game or about the character of rational human behavior. A common interpretation is that cooperative game theory is applicable when enforceable agreements can be made, for example, by contract. However, a reconstruction of the early history of cooperative game theory points in a different direction. This section will proceed by example, using well-known conventions of noncooperative[5] game theory. We will first introduce the examples, then consider the noncooperative solutions, then the cooperative solutions, then compare and conclude.

2.1. Examples

First consider Game 1, which is shown in extensive form in Fig. 1. In Game 1, a sequence of decisions is made by players A and B and their decisions are denoted by the oval decision nodes and arrows. Conventionally, the oval decision node labeled "B" indicates that B must make his decision between L and R without knowing what decision A has made between U and D. Accordingly, the first two stages of Game 1 can be represented in strategic normal form as shown in Table 1. This highly unsymmetrical game might represent a case in which B supplies an unprocessed raw material that is processed by A. (Of course, the numerical payoffs are arbitrary and not *derived* from any such example.) The second stage of Game 1 could represent a "side payment" from A to B to equalize their benefits from the game.

[5] The elementary concepts of noncooperative game theory used here may be found in any standard introduction to game theory, such as McCain (2004, cch. 1–2).

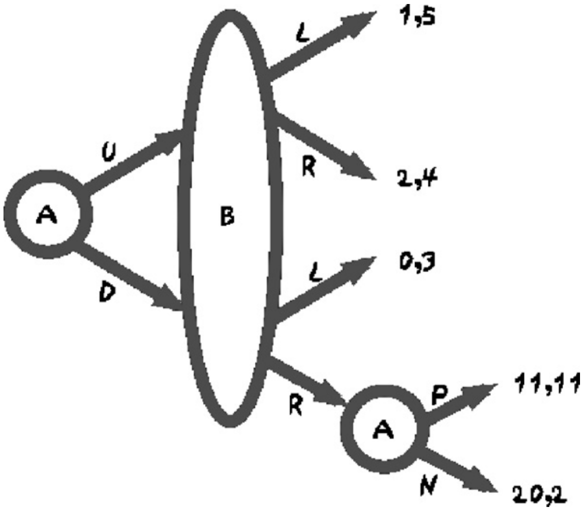

Fig. 1. Game 1 in extensive form about here.

Now consider Game 2, again shown in extensive form as Fig. 2. This game takes place in a maximum of three stages, although it can be cut short by either player. Fig. 3 shows the first two stages without the third.

This is a version of the "centipede game" that has played an important role in some recent experimental studies (McKelvey and Palfrey, 1992). It might represent a simple case of division of labor, where agent A may choose to operate independently without division of labor (D1), while efficient division of labor (R1, R2) triples the productivity of the pair (along lines suggested by Smith). However, if A initiates the division of labor process, B may opportunistically (Williamson, 1964) exploit his possession of the product at the second stage (D2) to his advantage but A's disadvantage. In Fig. 2, player A's second decision node introduces a "punishment" or "threat strategy," P, that gives A the option of reducing B's payoff at the cost of some reduction in his own, or of not doing so (arrow N). Since a decision for P leaves A worse off than he would be otherwise, P would be an instance of "altruistic punishment" (Boyd et al., 2003; Hayden, 2005; Fehr and Fischbacher, 2004).

Now consider Game 3, the three-person game shown in Fig. 4. This is similar to Game 2 except that the punishment stage is under the control of a third party, agent C, who is not involved in the earlier game. The payoffs to C are shown

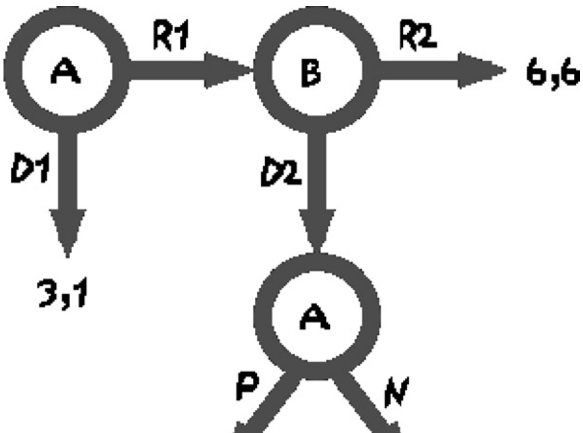

Fig. 2. Game 2 in extensive form about here.

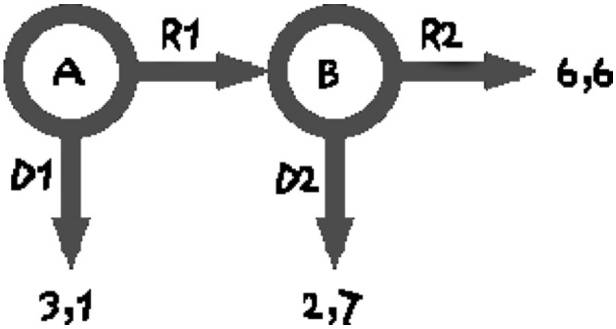

Fig. 3. Part of Game 2 about here.

third. This is an example of third-party punishment and, as the choice of P makes agent C worse off, is also an instance of altruistic punishment.

2.2. Noncooperative solutions

Following Kuhn, U, D, L, R, and P ("pay") and N ("no pay") are called "behavior strategies." In von Neumann and Morgenstern, a "strategy" is a contingency plan laying out the decisions to be made at each stage (that is, the behavior strategies chosen), conditioned on the information that the agent may have at that point. In what follows strategies in this second sense will be called "contingent strategies." In Game 1, for example, A now has three contingent strategies: (1) U; (2) D, and if R then P; (3) D, and if R then N. However, following work by Nash, Kuhn, and Selten, noncooperative game theory is often conducted in terms of behavior strategies, using "backward induction" to determine the best response at each stage.

First, however, consider the abbreviated version of Game 1 in Table 1. The Nash equilibrium for this game is at behavior strategies U, L; and since it is unique there will be no other noncooperative solutions. While this is not a social dilemma (since it is unsymmetrical and A's best response strategy depends on B's strategy choice) it reproduces a key element of the Prisoner's Dilemma: the noncooperative equilibrium is necessarily inefficient. Now consider the complete game in Fig. 1. For this game we apply "backward induction": we start with A's second decision point, and treat it as a subgame. It is not in A's interest at that point to pay the side payment, reducing A's own payoff from 20 to 11. But B can anticipate that, so will choose L for 5 or 3 rather than 4 or 2 respectively, and A in turn can anticipate that and will choose U for 1 rather than D for 0. It is never in A's interest to make a side payment, and in general, in

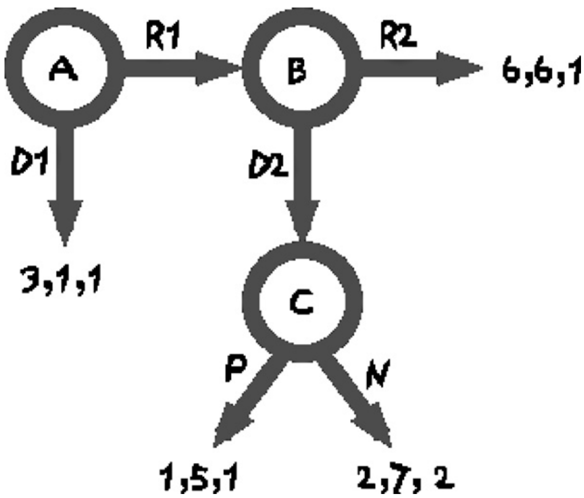

Fig. 4. Game 3, a game of third-party punishment about here.

noncooperative games that have a payment as a final stage, it is always in the payer's best interest to make the payment as small as possible.

Now consider the abbreviated version of Game 2, as shown in Fig. 3. Once again we may apply "backward induction": B will choose behavior strategy D2 for 7 rather than 6, and A will anticipate this and choose D1 for 3 rather than 2.

Now consider the complete game as shown in Fig. 2. To find all Nash equilibria for this game, it is necessary to analyze the game in terms of contingent strategies. A's contingent strategies are (1) D1, (2) R1 and, if D2, N, and (3) R1 and, if D2, P. B's contingent strategies are (4) if R1 then R1 and (5) if R1 then D2. This game has two Nash equilibria: 1, 5 and 3, 4. The latter of these, however, is a Nash equilibrium only because it is never put into practice: if agent B chooses strategy 4, agent A has no opportunity to choose between behavior strategies P and N. If we apply "backward induction" in this case, A will never choose P, since it leaves him with 1 rather than 2; B can anticipate this and so chooses D2, and A anticipates this in turn and chooses D1, reproducing Nash equilibrium 1, 5. Thus, 1,5 is the only "perfect equilibrium" (Selten, 1975)—the only meaningful noncooperative solution for this game.

Finally, consider Game 3, as shown in Fig. 4. Once again, applying "backward induction," we see that agent C will never punish, since he faces a cost if he does; and anticipating this, A and B will play just as they would in the abbreviated game in Fig. 3.

These results illustrate some common aspects of noncooperative solutions to games: side payments more than the minimum allowed are never made, altruistic punishments never occur, in particular third-party punishment never takes place, and agents anticipate this and choose their own strategies accordingly. However, extensive experimental evidence indicates that these predictions are not true statements about human behavior.[6] One very common suggestion for reconciling the noncooperative analysis with the experimental evidence is to posit reciprocity motives, which will be discussed in detail in a later section.

2.3. Cooperative solutions

Again, consider the abbreviated form of Game 1 shown in Table 1. Any cooperative solution will require the choice of behavior strategies D, R. Clearly, though, it will be necessary for A to offer B a "side payment" in order to induce B to enter a cooperative agreement. In principle, payment of a side payment is yet another strategic move, and the full game in Fig. 1 allows for this. For simplicity, we consider only a side payment that equalizes payoffs for the two agents, if the strategies D, R are chosen. Accordingly, agent A now has three contingent strategies: (1) U; (2) D, and if R then P; (3) D, and if R then N. If agent A can commit himself to conditional strategy 2, the best response for B is to choose R, and the cooperative outcome is realized. It is in A's interest to do so if A can. If A can assume a contractual obligation to strategy 2, this can support a cooperative solution to the game.

Again, beginning with the short version of Game 2, Fig. 3, we find that any cooperative solution requires the behavior strategy sequence R1, R2. For the complete game shown in Fig. 2, it will be helpful again to return to the contingent strategies. If agent A can commit himself to contingent strategy 3, or B to contingent strategy 4, either is sufficient to assure the cooperative strategy pair R1, R2.

It will be useful to derive the values of the coalitions for the two games. For Game 1, of course the value of the grand coalition $v\{A, B\}$ is 22. For coalitions other than the grand coalition, the convention in cooperative game theory is that each coalition obtains its security value, that is, the largest it can get assuming that the coalition of other players chooses the most damaging strategy that it can—not necessarily the one that makes it best off. Thus $v\{A\} = 1$, since $\{B\}$ can reduce $\{A\}$ to 1 by choosing strategy L, and $v\{B\} = 3$, since $\{A\}$ can reduce $\{B\}$ to 3 by choosing D.[7] In Game 2, $v\{A, B\} = 12$, $v\{A\} = 3$, and $v\{B\} = 1$, since A can, by choosing D1, both secure 3 regardless of B's commitments and reduce B to his lowest payoff, 1.

In Game 1, however, to reduce B to the minimum payoff 3, agent A must choose D, reducing his own payoff from 1 to zero. The principle founder of noncooperative game theory, Nash (1953, p. 130), wrote "Supposing A and B to be rational beings, it is essential for the success of the threat that A be *compelled to carry out his threat* ... if B fails

[6] See e.g. McKelvey and Palfrey (1992), Fehr and Fischbacher (2004), Fehr and Gachter (2002), and Hoffman et al. (1998).

[7] As Telser notes, the strategies that are supposed to be chosen by a coalition are different when we are determining the value of that coalition than when we consider it as an opposition coalition in determining the value of the coalition of the rest of the participants in the game. We will return to this inconsistency later.

to comply. Otherwise it will have little meaning. For, in general, to execute the threat will not be something A would want to do, just of itself." (Italics added). Here we have an example. Similarly, the threat strategy in Game 2 is such a threat, and for Nash it would be *irrational* for agent A to carry out a commitment to put such a threat into action. By contrast, in *The Theory of Games and Economic Behavior*, the assumption is that it would be irrational for A to fail to commit himself to the strongest threat he can make.

3. Concepts of rationality

3.1. Two concepts of rationality

These are conflicting concepts of rationality. It will help to have different terms to distinguish between them. In what follows we will call the rationality of noncooperative game theory "perfect" rationality, as it corresponds to Selten's (1975) "perfect equilibrium," while rationality in the sense of von Neumann and Morgenstern will be called "ideal rationality." Nash' own noncooperative equilibrium concept did not quite capture the reasoning in his quote. Kuhn (1997, original publication 1953) extended the ideas of both von Neumann and Morgenstern and Nash, clarifying the role of information in extensive games and arguing that analysis in terms of behavior strategies would be advantageous. He showed (for an important class of games[8]) that any contingent strategy could be expressed as a sequence of behavior strategies and that if the behavior strategies chosen are locally best responses then the sequence will be a Nash equilibrium. This leads to the procedure of backward induction. Kuhn's conclusion is sometimes expressed by the slogan "behavior strategies suffice"; but the slogan goes too far. First, the converse is not so. As we have seen (and note Selten, 1975) Nash equilibria exist that are not produced as a sequence of locally optimal behavioral strategies. Second, Kuhn's argument does not apply at all to cooperative games, as Selten (1964) established. Schelling (1960) had observed that "the power to constrain an adversary may depend on the power to bind oneself . . .." His contributions to noncooperative game theory assumed that weakness of will would be a problem for a person who might wish to choose a punishment strategy as in Game 2, and explored the means by which this difficulty might be resolved. It should be said that Schelling, unlike most game theorists and neoclassical economists, did not assume unbounded rationality and in some later work (1980) explicitly treated weakness of will as an aspect of bounded rationality. However, in one of the very few papers on the relation of games in extensive form to cooperative solutions, Selten (1964) both proved that backward induction would not be applicable to cooperative games and noted (in an afterword responding to Schelling) that cooperative game analysis differs from noncooperative analysis in presupposing a capacity for commitment, that is to say, strength of will. In any case, noncooperative game theory since (including Selten's) has consistently adopted Nash' concept of rationality, with its implied weakness of will (as also neoclassical economics has).

Perfect rationality is a rather odd concept of rationality, in application to a situation like Game 2a. It says in effect that if agent A makes a commitment (to strategy 3), and is made better off as a result, agent A has acted irrationally! Strength of will can only be irrational. In respect of a game like 1, perfect rationality supposes a failure to act opportunistically is also irrational. As we have observed, these predictions do not agree well with experimental evidence.

This is not to say that people are always ideally rational. While there is no equally explicit test of the hypothesis of ideal rationality, common experience suggests that weakness of will and opportunism are fairly common human phenomena (whether infinitely or boundedly rational, or both). There is experimental evidence for multiple types of rationality. In particular, a rational agent who knows that different people act rationally in different senses – some perfect, some ideal – will base his own decisions on a judgment as to the types of rationality he is most likely to encounter, as it may depend on the proportion of the people who are of each type and on his best judgment on the particular situation. This is sophisticated rationality. (See Stahl and Wilson (1995) for some related concepts that have suggested this term.)

3.2. Reciprocity

Experimental studies in game theory have been primarily based on models in noncooperative game theory. One game that has been extensively studied is the "ultimatum game" (Henrich et al., 2005, e.g.). The Ultimatum Game is a

[8] Essentially these are games in which the players are individuals, not teams. Note Selten (1975), on this point.

two-person game along the following lines: the two agents may be able to share a fixed amount, such as $100. The first agent, the proposer, suggests a payment to go to the second agent, the responder. If the responder accepts the payment, he receives it, and the balance is paid to the proposer. However, if the responder rejects the payment, neither agent gets anything. The noncooperative equilibrium is one in which the proposer makes the smallest possible positive offer and the responder accepts it. However, experimental evidence disagrees with this prediction. If the proposer makes a very small offer, the responder is sometimes observed to reject the proposal despite sacrificing the small positive payment. Moreover, offers are often more than the minimum needed to avoid a rejection, and 50–50 offers are fairly common.[9] This appears to be an instance of reciprocity motives.

Traditional game theory proceeds from strong assumptions about human rationality to strong conclusions about the nature of equilibrium. One can ask whether either the assumptions or the conclusions are empirically valid, and indeed there is a long history of experimental studies in game theory that explore this point. Social psychologists and others quite early provided evidence that people facing a Prisoner's Dilemma-like game do not always act as neoclassical maximizers (Lave, 1965; Rapoport and Chammah, 1965; Morehouse, 1967; Kreps and Wilson, 1982). Evidence of nonrational (in the sense of perfect rationality) or non-equilibrium behavior in other games is equally plentiful.[10]

Some of the early experiments on the Prisoner's Dilemma were interpreted as evidence that altruism is an element in human behavior. Unfortunately, altruism is not always well-defined. Altruism was inferred, however, from a tendency to choose the cooperative strategy even when it is not a best-response strategy, for example, in Prisoner's Dilemma games. Some more recent studies have often focused instead on fairness or reciprocity. Berg et al. (1995, p. 139) prefer the term "reciprocity," and say their ". . . results suggest that both positive and negative forms of reciprocity exist and must be taken into account . . . [and] provide strong support for current research efforts to . . . integrate reciprocity into standard game theory. . .." In general, reciprocity motives take the form of negative and positive reciprocity, with positive reciprocity meaning that the agent intends to sacrifice her self-regarding interests to return a favor for a favor and negative reciprocity meaning that the agent intends to sacrifice her self-regarding interests to return a punishment for a wrong. Reciprocity motives cannot generally be characterized except by reference to norms or framing, consistently with what Gintis (2007) calls the strong reciprocity hypothesis. We would say that an act is framed as a favor when the other agent has contributed more than the norm and that the act is framed as a wrong when the other agent has contributed less than the norm. The ultimatum game provides a good example of the role that negative reciprocity can play. If the proposer makes a very small offer, the responder may perceive this as an aggressive act and sacrifice even the small payment he is offered in order to retaliate by refusing the offer, leaving the proposer with nothing. This is negative reciprocity. On the other hand, if the proposer's offer is generous, the responder may perceive this as a generous act. The responder sacrifices nothing by accepting and thus rewarding the generous offer. In other game experiments, however, a player might have to sacrifice something in order to reward the other's generous action. (For brevity examples will not be given here.) This would be positive reciprocity.

Once again, let us continue with the example of the ultimatum game. As we have noted, there are several concepts of "the cooperative solution"; what they have in common is the idea that a person will never refuse a take-it-or-leave-it offer that makes him better off than the noncooperative alternative. Of the common solution concepts, only the Shapley value offers a precise prediction, and it predicts equal division. However, the evidence is negative in that (1) refusals do occur, and (2) while extreme offers – close either to the Nash equilibrium or 50–50 – are less common than offers in between these limits, equal division is also uncommon. The Shapley value predictions on the two different approaches are clearly rejected. It seems that something is missing from all those hypotheses. That something may be reciprocity.

Continuing with the ultimatum game, suppose that the responder has internalized a social norm that a participant in such a game should receive at least $x\%$ of the payoff. An offer of less than $x\%$ will then be perceived as aggressive and ungenerous. (This perception is part of what we mean by the term "a social norm.") Suppose, for example, that the norm is that each person should receive at least 25% of the amount to be distributed. Then the responder will reject offers substantially less than 25%. Conversely, the proposer, anticipating the responder's acceptance (from which the

[9] E.g. Guth et al. (1982), Henrich et al. (2005), and note also Roth (1995), Stanley and Tran (1998), Roth et al. (1991), Andreoni and Blanchard (2006), Oosterbeek et al. (2004).

[10] Because of the number of references – without any representation of completeness – it seems best to reserve them to a footnote. See Lieberman (1960), Wolfe and Shubik (1975), Guth et al. (1982, 1993), Selten and Stoecker (1986), O'Neill (1987), Rapoport and Boebel (1992), Roth et al. (1991), Cho and Kreps (1987), Diekmann (1993), Berg et al. (1995), Schotter et al. (1994), Roth and Erev (1995), Friedman (1996), and Stahl and Wilson (1995).

proposer stands to benefit) may want to reward that friendly act (in advance) by offering something more than the norm of 25%, but the offer still will probably not approach 50%. These predictions would be similar for any social norm well within the interval from 0 to 50%, and agree with the experimental results on the ultimatum game.

The ultimatum game is of interest primarily for its evidence of the importance of norms and reciprocity. Since it is a constant-sum game, it offers no possibility of a mutual benefit from working together. For cooperatives, and particularly worker cooperatives, the effort dilemma will be of more direct interest. This is addressed in McCain (2007).

4. Cooperative games and organizations

4.1. Some problems in applying cooperative game theory

Cooperative game theory as presented here, and in most of the received literature, makes use of powerful simplifying assumptions, but these simplifying assumptions create difficulties when we attempt to apply it to any real world phenomenon, such as cooperative organizations. As von Neumann and Morgenstern noted, the problem is to find the numbers that correctly express "the total value that the coalition can realize." Suppose, for example, we have a large game consisting of two merchants and a large group of customers. If the merchants compete on price, the customers can realize benefits with a relatively good value of (let us say) ψ for each customer. Call the profits of the two merchants in this case π_1 and θ_1. In this example all are acting individually, that is, as singleton coalitions. Now suppose the two merchants form a coalition, that is, a cartel, and raise the prices to the consumers. The money value of consumer benefits is reduced to ν, less than ψ, while the profits of the merchants are increased to π_2 and θ_2. Now suppose the customers respond by forming a consumers' cooperative, that is, a coalition of all (or a large part) of the customers to supply their own needs rather than relying on the merchants. This raises the money value of their benefits to ζ, at least as large as ψ and perhaps greater, and reduces the profits of the merchants to zero.[11] Now, then, which of these values should we assign to these various coalitions? Following von Neumann and Morgenstern, we would assign the minimum—zero for a coalition of merchants and ν for customers.

An additional simplifying assumption that is usually made is that the game is "superadditive," that is, a coalition formed by the merger of two or more coalitions will realize a value at least as great as the sum of the values of the coalitions merged. We see this to some extent in the example: the merchants increase their profits by forming a cartel, and the customers increase their standard of living by forming a cooperative. Superadditivity would mean that a grand coalition of all merchants and customers together would have a value no less than $\pi_2 + \theta_2 + \zeta$ per customer. Since that is greater than the values of the coalitions separately, $0 + 0 + \nu$ per customer, assigned as von Neumann and Morgenstern do, it seems that there is a large surplus that could be realized by a socialist ministry of distribution.

It should be noted that superadditivity is more than an assumption. As Aumann and Dreze (1974, p. 233) note, there are arguments for superadditivity that are quite persuasive, but, as they also note, superadditivity is quite problematic in many economic applications (including the one in the previous paragraph).

Von Neumann and Morgenstern were writing before Nash's work began the formation of noncooperative game theory. An alternative would be to assign the values of coalitions according to the Nash equilibrium in play among the coalitions.[12] Remarkably, this idea does not seem to have been developed until the 1990s (Zhao, 1992), although Telser (1978) considered it in passing. In general, market equilibria can be identified as instances of Nash Equilibrium. Therefore in our example, the values of the coalitions in the first instance – in which all agents operate as singleton coalitions – would be π_1 and θ_1 for the merchants and ψ for each of the customers. Notice, however, how the formation of a cartel by the merchants changes the value of a singleton consumer, and again how the formation of a cooperative changes the value of the merchants' cartel. These changes in the value of a coalition, as a result of the formation of

[11] This example is suggested by *Paddy the Cope* (Gallagher, 1942), which is still a very good read and a valuable biography of a life in the cooperative movement.

[12] It could be argued that there is an inconsistency in this, since Nash equilibrium assumes that commitments *cannot* be made as cooperative game theory assumes they *can*. One interpretation that reconciles them would be that the commitments can be made, but only within a coalition, while commitments to others outside the coalition, and to other coalitions, will not be honored if opportunistic behavior is more profitable. This is not unreasonable if we regard commitments as explicit promises and is consistent with a social norm of promise-keeping. In any case, since market equilibria are noncooperative equilibria in which coalitions interact noncooperatively, these assumptions are implied by ideas such as cartels, firms, employment relations, and cooperatives.

another coalition, are called *externalities*[13] in recent cooperative game theory (Carraro, 2003). When the coalition function is taken as the only important information about the game, as so much cooperative game theory has done, this amounts to the unstated assumption that externalities are unimportant. This, too, seems quite problematic for economic applications. As early as 1963 Thrall and Lucas proposed a more complex way of assigning values to coalitions, the partition function, that allows for externalities in this broad sense. However, as Aumann and Dreze (1974, p. 233, note) remark, it creates technical (mathematical) difficulties. Perhaps for this reason it was not widely used in cooperative game theory until the 1990s.

As Telser observes, the Nash equilibrium is ambiguous in the coalition function framework, since the Nash equilibrium will depend on the way in which the nonmembers of the coalition being valued will sort themselves into smaller coalitions. However, the partition function takes this dependency into account, and if (as with Zhao) we express the game in partition function form and then assign the values by Nash equilibrium play, valuation of coalitions is both consistent and free of the ambiguity Telser notes.

4.2. Cooperative games and cooperative organizations

What this suggests is that in order to understand cooperative *organizations* – and, as Gintis argues, a wide range of other social formations – we need to extend game theory with a conception of human motivation than has been part of game theory to date. Reciprocity motives and social norms would be a part of that conception. If we return to the general notion of a cooperative solution to the game as a "common strategy to improve the payoffs to the members of the group," social norms that support the common strategy would be among the dimensions of a "cooperative solution" of the game. To capture the idea, let us define the "extended cooperative solution" of a game as comprising (1) a common strategy that is efficient from the point of view of the members of the group, (2) a set of social norms that, in the presence of strong reciprocity motives, would result in the choice of the common strategy, without further enforcement, and (3) a set of guidelines for the sharing of the benefits of that common strategy, that is stable in the presence of a mixture of self-interest and reciprocity motives. The social norms for strategy choice and equable distribution are commitments, and to the extent that people are rational in the sense of cooperative game theory they will be carried out.

For an example of an extended cooperative solution in this sense, see McCain (2007). This is a "game" of effort determination in a cooperative enterprise, in which productive efficiency requires an effort commitment by each individual that is greater than the commitment the person would choose on the basis of pure self-interest in a noncooperative solution of the "game." Suppose that there is a norm of effort commitment, and suppose that the norm is the effort commitment required for efficient production. If one member makes effort less than the norm then other members perceive this as an unfriendly act and retaliate (perhaps by shunning or "putting in Coventry," though McCain is not explicit on this). Reciprocity motives lead to one or the other of two "solutions": in one, everyone obeys the norm, production is efficient, and there is no retaliation. The other solution recapitulates the noncooperative one: everyone chooses the effort commitment on the basis of self-interest, without regard to social norms, and, each being equally a slacker, there is again no retaliation.[14] The "cooperative" norms, in this case, are the ones required for efficient production, which also defines the common strategy. In this model, equal sharing of work time and pay are assumed throughout. As a result, one employee's effort is a favor to the other employees, and reciprocity motives lead them to reciprocate it. If, however, the enterprise is organized on the profit principle, then the benefit of one worker's increased effort goes to the proprietor, not to the other employees, and reciprocity motives would not lead them to increase their own effort, and the extended cooperative solution breaks down.

Of course, this is not a new idea. We find it in Mill (1987): "[C]ooperation tends . . . to increase the productiveness of labour, . . . by placing the labourers, as a mass, in a relation to their work which would make it their principle and their interest – at present it is neither – to do the utmost, instead of the least possible, in exchange for their remuneration." The

[13] The term "externalities" is used more narrowly in economic theory. When a cartel is formed and thus imposes costs on customers, this would have been called a "pecuniary externality" in e.g. Scitovsky (1954); but modern economic theory does not regard "pecuniary externalities" as externalities.

[14] In the noncooperative perspective, both solutions are equally valid, and one implication is that the social norms will be effective in bringing about a cooperative solution only if there is a wide expectation that they will be followed, at least approximately. But we would hardly consider as norms a set of demands that no-one expects ever to fulfill or be fulfilled.

key point for present purposes is the link it establishes among reciprocity, cooperative organizations, and cooperative solutions to games.

More generally, we began by observing that cooperative games and cooperative organizations have a common beginning in the idea of a group of people who choose a common course of action for their mutual benefit. The cooperative principles go beyond this by requiring that the group constitute "... a jointly-owned and democratically-controlled enterprise." It is the presence of reciprocity motives that makes those standards necessary, in many social contexts, to achieve a cooperative solution to the game. Minority ownership, and the remission of most of the benefits of efficient common action to the proprietor, in themselves seem to offend against reciprocity. As we have seen they interfere with the working of reciprocity among the workers in production. Monopolization of decision-making by a small minority also seems in itself to offend against reciprocity. Minority rule and profit may be stable and efficient so long as most people believe in the social norms expressed by the hymn: "The rich man in his castle, The poor man at his gate, He made them, high or lowly, And ordered their estate." That will be an easy norm for the rich man to believe in, but as ordinary people learn that they have alternatives, other and more equalitarian norms are likely to arise, and have arisen in most of the world.

5. Summary and conclusions

Cooperative game theory and cooperative organization share the idea that agents join together and work together or choose a joint strategy for mutual benefit, but cooperative organizations are in addition jointly owned and democratically controlled. In the framework of cooperative game theory, profit-seeking firms (for example) are coalitions that manifest cooperative solutions just as cooperative organizations are. Like noncooperative game theory, cooperative game theory is best thought as a rational action hypothesis, but a hypothesis based on a different concept of rationality than is noncooperative game theory. Indeed neither of these rational action hypotheses corresponds well with experimental evidence. However, when we modify the self-interest hypothesis of neoclassical economics and game theory, allowing for reciprocity motives and social norms, the experimental evidence can largely be accounted for. The reciprocity motives eliminate the most extreme predictions of self-interest theory in the experimental studies of the ultimatum game, and support the idea that in enterprise, a cooperative (in the sense of the cooperative movement) organization may be required in order to fully realize a cooperative (in the sense of game theory) solution to the interactive decision problem all group enterprises create.

References

Andreoni, J., Blanchard, E., 2006. Testing subgame perfection apart from fairness in ultimatum games. Experimental Economics 9 (4), 307–321.

Aumann, R.J., 2003. Presidential address. Games and Economic Behavior 45, 2–14.

Aumann, R.J., Dreze, J.H., 1974. Cooperative games with coalition structure. International Journal of Game Theory 3, 217–237.

Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. Games and Economic Behavior 10 (1), 122–142.

Boyd, R., Gintis, H., Bowles, S., Richerson, Peter J., 2003. The evolution of altruistic punishment. Proceedings of the National Academy of Science 100 (6 (March 18)), 3531–3535.

Carraro, C., 2003. The Endogenous Formation of Economic Coalitions. Edward Elgar, Cheltenham, U.K.

Cho, In-Koo, Kreps, D.M., 1987. Signaling games and stable equilibria. Quarterly Journal of Economics 100 (May (2)), 179–222.

Diekmann, A., 1993. Cooperation in an asymmetric volunteer's dilemma game theory and experimental evidence. International Journal of Game theory 22, 75–85.

Fehr, E., Fischbacher, U., 2004. Third-party punishment and social norms. Evolution and Human Behavior 25, 63–87.

Fehr, E., Gachter, S., 2002. Altruistic punishment in humans. Nature 415, 137–140.

Friedman, D., 1996. Equilibrium in evolutionary games: some experimental results. Economic Journal 106' (January), 1–25.

Gallagher, P., 1942. Paddy the Cope: An Autobiography. The Devin-Adair Company (2005 edition published by Kessinger Publishing).

Gillies, D.B., 1953. Some theorems on *n*-person games. Ph.D. Thesis. Department of Mathematics, Princeton University.

Gintis, H., 2007. A framework for the unification the behavioral sciences. Behavioral and Brain Sciences 30 (1).

Guth, W., Schmittberger, R., Schwartze, B., 1982. An experimental analysis of ultimatum bargaining. Journal of Economic Behavior and Organization 3, 376–388.

Harsanyi, J., 1963. A simplified bargaining model for the n-person cooperative game. International Economic Review 4 (May (2)), 194–220.

Hayden, Thomas, 2005. "Why We Need Nosy Parkers," U. S. News and World Report (June 13).

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Henrich, N.S., Hill, K., Gil-White, F., Gurven, M., Marlowe, F.W., Patton, J.Q., Tracer, D., 2005. 'Economic man' in cross-cultural perspective: behavioral experiments in 15 small-scale societies. Behavioral and Brain Sciences 28 (December (6)), 795–815.

Hoffman, E., McCabe, K., Smith, V.L., 1998. Behavioral foundations of reciprocity: experimental economics and evolutionary psychology. Economic Inquiry 36 (July (3)), 335–352.

International Cooperative Alliance, 1995. Statement on the Co-Operative Identity, available at: http://www.ica.coop/coop/principles.html, as of June 20, 2006, last modified on 2/9/2006.

Kreps, D.M., Wilson, R., 1982. Sequential equilibrium. Econometrica 50 (July (4)), 863–894.

Kuhn, H.W., 1997. Extensive Games and the Problem of Information Classics in Game Theory. Princeton University Press, Princeton, pp. 46–68.

Lave, L.B., 1965. Factors affecting cooperation in the prisoner's dilemma. Behavioral Science 10, 26–38.

Lieberman, B., 1960. Human behavior in a strictly determined 3x3 matrix game. Behavioral Science 5, 317–322.

McKelvey, R.D., Palfrey, T.R., 1992. An experimental study of the centipede game. Econometrica 60 (4), 803–836.

McCain, R., 2004. Game Theory: A Nontechnical Introduction to the Analysis of Strategy. Southwestern Press, Mason, Ohio.

McCain, R., 2007. Cooperation and effort, reciprocity and mutual supervision in worker cooperatives. In: Novkovic, S. (Ed.), Advances in the Economic Analysis of Participatory and Labor-Managed Firms. JAI Press, Greenwich, CT.

Mill, J.S., 1987. In: Kelley, A.M. (Ed.), Principles of Political Economy (Reprint of 1909 edition).

Morehouse, L.G., 1967. One-play, two-play, five-play and ten-play runs of prisoner's dilemma. Journal of Conflict Resolution 11, 354–362.

Nash, J., 1950. The bargaining problem. Econometrica 18, 155–162.

Nash, J., 1951. Non-cooperative games. Annals of Mathematics 2 (September), 286–295.

Nash, J., 1953. Two-person cooperative games. Econometrica 21 (January), 128–140.

O'Neill, B., 1987. Nonparametric test of the minimax theory of two-person zerosum games. Proceedings of National Academy of Sciences 84, 2106–2109.

Oosterbeek, H., Sloof, R., van de Kuilen, G., 2004. Cultural differences in ultimatum game experiments: evidence from a meta-analysis. Experimental Economics 7 (June (2)), 171–188.

Rapoport, A., Boebel, R., 1992. Mixed strategies in strictly competitive games: a further test of the minimax hypothesis. Games and Economic Behavior 4, 261–283.

Rapoport, A., Chammah, A.M., 1965. Prisoner's Dilemma. University of Michigan Press.

Roth, A., Erev, I., 1995. Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. Games and Economic Behavior 8, 164–212.

Roth, A.E., Prasnikar, V., Okuno-Fujiwara, M., Zamir, S., 1991. Bargaining and market behavior in Jerusalem, Ljubliana, Pittsburgh, and Tokyo: an experimental study. American Economic Review 91 (December (5)), 1068–1095.

Roth, A., 1995. Bargaining experiments. In: Kagel, J., Roth, A. (Eds.), Handbook of Experimental Economics. Princeton University Press, Princeton, pp. 253–348.

Schelling, T., 1960. The Strategy of Conflict. Harvard University Press, Cambridge.

Schelling, T.C., 1980. The intimate contest for self-command. The Public Interest 60 (Summer), 94–118.

Schotter, A., Weigelt, K., Wilson, C., 1994. A laboratory investigation of multiperson rationality and presentation effects. Games and Economic Behavior 6, 445–468.

Scitovsky, T., 1954. Two concepts of external economies. The Journal of Political Economy 62 (April (2)), 143–151.

Selten, R., 1964. Valuation of *n*-person games. In: Dresher, M., Shapley, L.S., Tucker, A.W. (Eds.), Advances in Game Theory (Annals of Mathematics Studies, Number 52). Princeton University Press, Princeton, pp. 577–626.

Selten, R., 1975. Reexamination of the perfectness concept for equilibrium points in extensive games. International Journal of Game Theory 4, 25–55.

Selten, R., Stoecker, R., 1986. End behavior in sequences of finite prisoner's dilemma supergames. Journal of Economic Behavior and Organization 7, 47–70.

Shapley, L.S., 1953. A value for *n*-person games. In: Kuhn, H.W., Tucker, A.W. (Eds.), Contributions to the Theory of Games II. Princeton University Press, Princeton, pp. 305–317.

Shapley, L.S., Shubik, M., 1969. On market games. Journal of Economic Theory 1, 9–25.

Smith, A., 1937. The Wealth of Nations. The Modern Library, New York.

Stahl, D.O., Wilson, P.W., 1995. On players' models of other players: theory and experimental evidence. Games and Economic Behavior 10, 218–254.

Stanley, T.D., Tran, U., 1998. Economics students need not be greedy: fairness and the ultimatum game. Journal of Socio-Economics 27 (6), 657–663.

Telser, L., 1978. Economic Theory and the Core. University of Chicago Press, Chicago.

Thrall, R.M., Lucas, W.F., 1963. *n*-Person games in partition function form. Naval Research in Logistics Quarterly 10, 281–298.

von Neumann, J., Morgenstern, O., 1944. Theory of Games and Economic Behavior. Princeton University Press, Princeton.

Williamson, O., 1964. The Economics of Discretionary Behavior: Managerial Objectives in the Theory of the Firm. Prentice-Hall, Englewood Cliffs, NJ.

Wolfe, G., Shubik, M., 1975. Teams compared to individuals in duopoly game with an artificial player. Southern Economic Journal 41 (4), 635–648.

Zhao, J., 1992. The hybrid solutions of an *n*-person game. Games and Economic Behavior 6 (3), 145–160.