Game-Theoretic Models and the Role of Information in Bargaining

Alvin E. Roth and Michael W. K. Malouf
Department of Business Administration, University of Illinois

This article examines the behavioral implications and empirical testability of the game-theoretic models of bargaining that follow in the tradition begun by Nash. The classical game-theoretic models are reviewed with particular attention to the assumptions they make concerning the kinds of information shared by the bargainers. The experimental literature concerned with testing these models is also reviewed, with particular attention to the extent to which these experiments have conformed to the assumptions of the models they are designed to test. Some of the principal differences between the theoretical predictions and the results reported in these experiments seem to be due to questions concerning the information shared by the bargainers. A new game-theoretic model is presented, based on assumptions about information that more closely correspond to the conditions under which many of these experiments have been conducted. An experiment manipulating the appropriate kind of information supported the hypothesis that Nash's bargaining model has predictive value in situations that conform to its assumptions about information, but the results also suggest that only a relatively narrow range of situations may conform completely to these assumptions.

Bargaining is a pervasive and important social phenomenon that has been studied from a multitude of perspectives by investigators from a variety of disciplines. The primary purpose of this article is to examine the behavioral implications and empirical testability of the game-theoretic models of bargaining that follow in the tradition begun by Nash (1950) and that provide the theoretical basis for much of the work on bargaining done by economists. These models are intimately associated with expected utility theory, and they present formidable obstacles to laboratory experimentation. Largely for this reason,

these models have not entered into the mainstream of bargaining research conducted by social psychologists. For instance, this family of models is not mentioned at all in the encyclopedic survey of experimental bargaining research by Rubin and Brown (1975), whereas Chertkoff and Esser (1976) stated that

although these models may have been ignored to a certain extent because many social psychologists are not mathematically inclined, there are undoubtedly other reasons. First, the theories contain constructs, like utility, that are difficult to measure. Second, the models usually deal with broad concepts, like utility of payoffs, without specifying in any detail what variables are related to them, and in what way. (p. 466)

Morley and Stephenson (1977) go one step further, and state that "these theories not only have not been subject to empirical tests but also do not have any obvious behavioral implications." (p. 86)

This article begins with a review of the classical game-theoretic models of Nash (1950) and Raiffa (1953) with particular attention to the mathematical condition specifying the kind of information that the models assume

Game-Theoretic Models and the Role of Information in Bargaining

Alvin E. Roth and Michael W. K. Malouf

can potentially influence the outcome of bargaining. Included in this review are recent mathematical results emphasizing the extent to which the predictions of the models depend on this informational condition.

The experiments that have been conducted to test these models are then reviewed, with particular attention to the degree to which the conditions of the models have been successfully implemented and to the kinds of information that the various implementations make available to the bargainers. Although the game-theoretic models are defined in terms of the expected utility of the bargainers, the experiments designed to test these models have usually equated utility payoffs with monetary payoffs. We argue that bargaining situations in which the players have common knowledge of one another's monetary payoffs make available a different quality of jointly shared information than situations in which the players have common knowledge of one another's utility payoffs. This is because the players' utility payoffs are uniquely defined only up to independently normalized interval scales, whereas monetary payoffs are defined on a common absolute scale.

Situations in which the players know one another's monetary payoffs permit them to make comparisons that they otherwise could not make. A new mathematical result is presented, characterizing a model of bargaining that differs from the classical models in its assumptions about information. This new model reflects the comparisons that can influence the outcome of bargaining when players know one another's monetary payoffs.

A new experiment is then reported, which is designed to investigate the effect of this difference in the quality of the shared information. This experiment permits both the expected utility available to the players and the information they share as common knowledge to be controlled and manipulated. The (preliminary) results suggest that Nash's classical game-theoretic model is a good predictor of behavior in bargaining situations that make available precisely the right kind of information to the bargainers. These results also suggest, however, that relatively few bargaining situations provide precisely this

information, and that the classical models require modification in situations that make available either too little or too much of various kinds of information.

The Classical Game-Theoretic Models[1]

Before entering into a discussion of the bargaining models, we briefly consider the elements of expected utility theory on which these models depend. (The new experimental design presented later also depends critically on the elements of utility theory reviewed here.) von Neumann and Morgenstern (1944) were the first to demonstrate conditions on an individual's preferences which are sufficient so that the individual's choice behavior over risky events is the same as if the individual were maximizing the expected value of a real valued function called this individual's *utility function*.[2] Given sufficient information about an individual's preferences, a corresponding utility function can be constructed, although this function is uniquely defined only up to an interval scale.

For simplicity, consider the case in which the set of Alternatives A contains elements a and c such that, for any Alternative b in the set A, the individual in question likes Alternative a at least as well as b and b at least as well as c. Then if u is a utility function representing this individual's preferences over the set of alternatives A, it must have the property that $u(a) \geq u(b) \geq u(c)$. Since u is defined only up to an interval scale, we may arbitrarily choose its unit and zero point, and in particular we may take $u(a) = 1$ and $u(c) = 0$. The problem of determining $u(b)$ then becomes the problem of finding the appropriate value between 0 and 1 so that all

---

[1] A more complete and technical account of this subject is to be found in a recent monograph (Roth, 1979).

[2] Although there is experimental evidence (e.g., Tversky, 1969) that individuals' preferences may not always obey the conditions necessary for them to be modeled by a utility function, these conditions still retain normative significance. Furthermore, the violations of these conditions have generally been observed in multiattribute decision situations, rather than in decisions involving a single commodity (e.g., money) of the sort involved in the experiments to be considered here.

those lotteries over alternatives that the individual prefers to b have a higher expected utility, and all those lotteries to which b is preferred have a lower expected utility. If we denote by $L(p) = [pa; (1 - p)c]$ the lottery that with probability p yields Alternative a and with probability $(1 - p)$ yields Alternative c, then the utility of participating in the lottery $L(p)$ is its expected utility, $pu(a) + (1 - p)u(c) = p$. If p is the probability such that the individual is indifferent between b and $L(p)$, then their utilities must be equal, and so, $u(b) = p$. Thus when we say that the utility of Alternative b to a given individual is known, we mean that the probability p is known such that the individual is indifferent between having Alternative b for certain or having the risky alternative $L(p)$.

For instance consider an individual who is faced with a choice of receiving half a million dollars for certain or participating in a lottery that will yield a million dollars with probability p and otherwise yield zero dollars. Then if we set the individual's utility function for zero dollars at 0 and set the utility for a million dollars at 1, determining the individual's utility for half a million dollars means determining the probability p that would leave this individual indifferent between the lottery and the half-million dollars. Most of us would require p to be considerably greater than one-half before we would take the lottery over the assured half million, which is to say that our utility function is not linear in money, and our utility for half a million dollars is more than half-way between our utility for zero dollars and our utility for a million dollars. In what follows, when we say that one individual knows another's utility for a given event (e.g., a particular reward), we are not requiring that he or she know any utility *theory*, but rather that the individual has sufficient knowledge of the other's preferences to be able to determine an equivalent lottery of the sort just described. (For a more complete discussion, see Herstein & Milnor, 1953; Krantz, Luce, Suppes, & Tversky, 1971; von Neumann & Morgenstern, 1953.)

Following Nash (1950) we can now define a *pure bargaining game* between two players as a situation in which there is a set of fea-

sible outcomes, any one of which will be the final outcome of the game if it is agreed to by both players. In the event that they fail to reach agreement, some fixed disagreement outcome becomes the final outcome of the game. That is, the rules of the game give each player a veto over any outcome other than the disagreement outcome. The players are assumed to be fully informed both about the rules of the game and about the set of feasible outcomes.

Nash (1950) modeled such a game by a pair (S, d), where S is a subset of the plane which contains d. He further required that the set S be compact and convex, and that it contain at least one point x such that $x > d$ (i.e., $x_1 > d_1$ and $x_2 > d_2$).

The interpretation Nash (1950) gave is that the set S is the set of feasible expected utility payoffs available to the players; that is, if the players agree on a point $x = (x_1, x_2)$ in S, then Players 1 and 2 receive a utility of x_1 and x_2, respectively, whereas $d = (d_1, d_2)$ denotes the utility payoffs to the players in the event of disagreement. Thus, according to Nash's interpretation, the assumption that the players know the set of outcomes means that they know one another's utilities for each potential agreement.[3] The condition that S is convex is thus appropriate if, for instance, any lottery between feasible agreements is also a feasible agreement.[4] The condition that S is compact is justified if, for instance, the set of feasible outcomes consists of a finite number of distinct alternatives and the lotteries between them, whereas the requirement that S contains at least one point x such that $x > d$ ensures that both players prefer some feasible agreement to the consequences of failing to reach agreement. That is, this latter requirement confines our attention to games in which both players can gain from reaching an agreement.

---

[3] This is a standard assumption in game theory, referred to as the assumption of "complete information."

[4] A convex set of vectors in one that contains any weighted average of any two elements. Since the utility of a lottery is its expected utility (i.e., the weighted average of the utilities over which the lottery is conducted), S is convex if lotteries are feasible.

Having thus modeled the bargaining *situations* to which his theory is addressed, Nash modeled the bargaining *process* by a function (called a solution) that selects a feasible outcome for every bargaining game. That is, letting B denote the class of bargaining games (and R^2 denote the plane), then a *solution* is a function $f: B \rightarrow R^2$ such that for every game (S, d), $f(S, d)$ is an element of S. Thus a solution models a bargaining process by specifying its outcome. Nash proposed that a solution should possess the following properties:

Property 1. Pareto optimality: If $f(S, d) = z$, then the set S contains no element x distinct from z such that $x \geq z$.

The first property specifies that the bargaining process will not yield any outcome that *both* players find less desirable than some other feasible outcome.

Property 2. Symmetry: If $f(S, d) = z$ and if (S, d) is a symmetric game (i.e., if $d_1 = d_2$ and if for every (x_1, x_2) in S, (x_2, x_1) is also contained in S), then $z_1 = z_2$.

The second property specifies that if the game (S, d) makes no distinction between the players, then neither should the solution.

Property 3. Independence of irrelevant alternatives: If (S, d) and (T, d) are bargaining games such that S contains T, and if $f(S, d)$ in an element of T, then $f(T, d) = f(S, d)$.

The third property can perhaps best be understood to specify that the bargaining process in question yields, in some sense, the "best" outcome z in the feasible set S, and that the best outcome in S would also be the best outcome in any smaller set T.

The fourth property relates to the fact that the game (S, d) is interpreted in terms of the expected utility functions of the players, which are defined only up to an interval scale, that is, only up to an arbitrary choice of origin and unit. It states that if a game (S', d') is derived from (S, d) by transforming the utility functions of the players to equivalent representations of their preferences, then the same transformations applied to the outcome of the game (S, d) should yield the outcome selected in (S', d'). That is, it states that the solution should depend only on the preferences of the players, and not on any arbitrary features of the utility functions that represent these preferences.

Property 4. Independence of equivalent utility representations: If $f(S, d) = z$, and (S', d') is a bargaining game related to (S, d) by the transformations $d' = (a_1 d_1 + b_1,\ a_2 d_2 + b_2)$, $S' = \{(a_1 x_1 + b_1,\ a_2 x_2 + b_2 \mid (x_1, x_2) \in S\}$, where $a_1, a_2 > 0$, then $f(S', d') = (a_1 z_1 + b_1, a_2 z_2 + b_2)$.

Sometimes Property 4 goes by the alternate name *independence of positive linear transformations of the payoffs*. Since the choice of origin and scale for each player's utility function is unrelated to that of the other player's, this property essentially specifies that the numerical *levels* of utility assigned to each outcome have no standing in the theory, and no *comparisons* of the numerical utility payoffs received by the two players can play any role in determining the outcome of bargaining.

Nash (1950) proved the following remarkable theorem.

Theorem 1: There is a unique solution that possesses Properties 1–4. It is the solution $f = F$ defined by $F(S, d) = z$ such that $z > d$ and $(z_1 - d_1)(z_2 - d_2) > (x_1 - d_1)(x_2 - d_2)$ for all points x in S distinct from z such that $x > d$.

That is, in the region of S yielding positive gains to both players, Nash's solution D selects the unique point z that maximizes the geometric average (i.e., the product) of the gains available to the players, as measured against their disagreement payoffs. Furthermore, the solution F possesses Properties 1–4, and it is the only solution that does.

Although the four properties that Nash considered are sufficient to characterize the solution F, there are, of course, other properties of the solution that may be useful to consider. One of the most elementary of these is that, as we have noted, Nash's solution always yields each player a positive gain over the disagreement point; that is, it possesses the following property.

Property 5. Strong individual rationality: $f(S, d) > d$. This property is elementary in the sense that it might reasonably be expected to be descriptive of any bargaining process that results in agreement, since neither player has any incentive to accept an agreement that does not yield the player a higher utility than can be achieved without agreement. (Recall that the payoffs are assumed to be in

terms of utility, which completely summarize a player's preferences.)

It has recently been shown (Roth, 1977b, 1978) that if Property 5 is incorporated into the characterization of a solution, then the elimination of Property 1 causes no change in Nash's conclusion, whereas the elimination of Property 2 leads to a straightforward generalization of Nash's result. Specifically, we have the following two theorems.

Theorem 2: Nash's solution F is the unique solution that possesses Properties 2–5 (Roth, 1977b).

Theorem 3: Any solution that possesses Properties 3–5 is of the form $f(S, d) = z$ such that $z > d$ and $(z_1 - d_1)^p(z_2 - d_2)^q > (x_1 - d_1)^p(x_2 - d_2)^q$ for all points x in S distinct from z such that $x > d$, where p and q are fixed positive numbers that sum to 2 (Roth, 1978).

Theorem 3 states that if we drop the requirement of symmetry, then the resulting solution must be one that maximizes the *weighted* geometric average of the gains available to the players, with weights p and q. Thus the only role of symmetry in Nash's theorem is to set $p = q = 1$, causing the solution to weight equally the gains of both players. Dropping symmetry permits the solution to reflect factors such as differential bargaining ability between players that are outside the formal description of the game, but it leaves the form of the solution unchanged. (Asymmetric Nash solutions are also considered, from somewhat different perspectives, by Harsanyi and Selten, 1972; and Kalai, 1977a.)

Taken together, Theorems 2 and 3 indicate that Nash's result depends hardly at all on Pareto optimality or symmetry but relies almost entirely on Properties 3 and 4, the two independence properties. It is therefore toward these two properties that we will direct our attention.

As already noted, Property 4, independence of equivalent utility representations, specifies that the solution may not make use of any information about the payoffs, to the players, which depends on the origin or unit of those payoffs. The other independence property, independence of irrelevant alternatives, imposes restrictions of a different sort. Rather than reflecting limitations on the information contained in the payoffs, it restricts the kinds of comparisons that the bargaining process may employ. Given a fixed disagreement point, a bargaining process that possesses Property 3 essentially proceeds by establishing a means to make binary comparisons between pairs of alternatives and choosing the "best" alternative, in terms of this comparison, from any feasible set. That is, the comparison between any two particular alternatives is determined independently of the feasible set in which they are contained.

Of the four properties that Nash suggested, Property 3 has been the subject of the most criticism and discussion (Luce & Raiffa, 1957; Roth, 1977a). Other solutions have been proposed that possess all of Nash's properties with the exception of Property 3, and of these alternative solutions, the one that has received the most attention is one of several originally proposed by Raiffa (1953). Raiffa's solution G is defined as follows:

For any game (S, d), let the *ideal point* $I(S, d) = \bar{x} = (\bar{x}_1, \bar{x}_2)$ be defined by $\bar{x}_1 = \max\{x_1 | x \geq d$ and $x = (x_1, x_2)$ is contained in $S\}$, $\bar{x}_2 = \max\{x_2 | x \geq d$ and x is contained in $S\}$. Then $\bar{x}_1$ and $\bar{x}_2$ represent the maximal payoffs available to Players 1 and 2, respectively, in the individually rational region of S. The solution G selects the maximal feasible point on the line joining d to $\bar{x}$; that is, $G(S, d) = z$ is the Pareto optimal point such that $(z_1 - d_1)/(z_2 - d_2) = (\bar{x}_1 - d_1)/(\bar{x}_2 - d_2)$. That is, the solution G chooses the point z that maximizes the gains of the players subject to the restriction that the players' actual gains are in the same proportion as their maximum potential gains. It is easy to verify that the solution G possesses Properties 1, 2, and 4 but does not possess Property 3. That is, the feasible points that determine $I(S, d)$ are not "irrelevant" to Raiffa's solution in the sense in which they are to Nash's solution.

Kalai and Smorodinsky (1975) characterized G as the unique solution that possesses Properties 1, 2, 4 and an additional property called individual monotonicity. The solution G has also been independently proposed and studied by other investigators; for example, Crott (1971) anticipated the solution on the basis of an experimental study (reviewed later),

and Butrim (1976) studied the solution axiomatically. Unlike Nash's solution, which can be generalized in a straightforward way to the case of pure bargaining games among more than two players, it can be shown (Roth, in press) that the solution G does not generalize to the case of more than two players.

Before proceeding to a review of the experimental literature concerning these models, it should be mentioned that subsequently Nash (1953) and Raiffa (1953) also analyzed certain specialized two-stage bargaining situations, in which it is possible in the first stage for bargainers to choose threats and make irrevocable commitments that bind them to carry out their threats should no agreement be reached in the second stage. The analysis of these "variable-threat" situations concentrates on the question of what threats should be chosen by the players in the first stage, under the assumption that these threats will determine the disagreement outcome in the second stage for a bargaining process of the kind we have already considered, modeled by Nash's (or Raiffa's) solution. The conclusions reached concerning the optimal choice of threats in such situations depend critically on the assumption that the commitment to carry out a threat is irrevocable, as well as on the assumption that bargaining in the second stage can be modeled by the solutions considered earlier in this section. The pure bargaining games we have been considering are often called "fixed-threat" games, to distinguish them from the two-stage variable threat situations by indicating that the disagreement outcome is fixed.

Previous Experimental Research

The scope of this review is limited to investigations that explicitly sought to test propositions based on one of the game-theoretic models of bargaining we have been considering. It does not include results of experiments that were conducted without any reference to game-theoretic models but that nevertheless might have some relevance to them. Most of the experiments considered here have not previously been considered in

reviews either of the more general literature on bargaining (Chertkoff & Esser, 1976; Morley & Stephenson, 1977; Rubin & Brown, 1975) or of game-theoretic models in general (Murnighan, 1978).

Although a number of the experiments testing game-theoretic models of bargaining have been concerned not only with the original fixed-threat pure bargaining games but also with the variable-threat games, we will concentrate on the former models only, for two reasons. First, the fixed-threat models are more basic than the variable-threat models, since the latter models depend in a critical way on the validity of the former. Second, most of the experiments devoted to testing the variable-threat models do not implement any procedures to make threats irrevocable in the required way, and so these experiments tend not to be closely related to the theoretical models they are designed to test.

Since any bargaining situation having a disagreement outcome that each bargainer can enforce can be described as a pure bargaining game, the conditions of the formal models leave room for a wide variety of experimental implementations. However, since the bargaining theories under consideration are designed to model cooperative bargaining behavior, it seems reasonable to expect that they would be most descriptive of the result of bargaining conducted under conditions that permit at least limited communication. Indeed, with few exceptions, the experiments designed to test these theories have been conducted under what Chertkoff and Esser (1976) call conditions of *explicit* bargaining, in that they permit communication, compromise, and the exchange of provisional offers and counteroffers.

The experiments considered in this section involved bargaining for monetary payoffs, sometimes through the intermediary step of bargaining for "points" or "chips" with monetary value. These experiments will be organized here according to the information that was made available to the bargainers: full information to both bargainers about the payoffs to both bargainers, or full information to each bargainer about his or her own payoffs and partial information about the other bar-

gainer's payoffs.[5] Most but not all of the experiments to be considered can be classified into one of these categories, and the exceptions will be noted.

Full Information

A straightforward experiment was conducted by Nydegger and Owen (1975), who proposed to test each of Nash's properties by observing the results of a series of simple bargaining situations. A secondary purpose of the Nydegger and Owen experiment was to compare the predictive value of Nash's solution with Raiffa's solution. In their experiment, 30 pairs of undergraduates each participated in a single bargaining encounter involving the distribution of monetary payoffs, about which they were fully informed. Nydegger and Owen interpreted the monetary payoffs in these games as being identical to the utility received by the bargainers. In each of the three conditions of the experiment, a different game was played, chosen to test Nash's properties. The bargaining was conducted verbally and face-to-face, with the bargainers seated together at a table.

Nydegger and Owen interpreted their results as supporting the proposition that bargaining behavior is symmetric, Pareto optimal, and independent of irrelevant alternatives while contradicting the proposition that it is independent of equivalent utility representations. Specifically, in each of their games, the bargainers reached agreements that gave them equal monetary payoffs. This supports the conclusion that in bargaining for money with full information about payoffs, the scale of the monetary payoffs available has an effect on the agreements reached, and that comparison of the monetary payoffs received by each player plays a role in determining the outcome. A remarkable feature of these results is the fact that within each condition the identical agreement was reached by each pair of bargainers. Even if this is primarily due to the social forces acting on undergraduate subjects when placed in an artificial situation to bargain face-to-face for negligible amounts of money, it at least indicates that the information available to the bargainers was

sufficient to enable them to identify a particular salient outcome.

Another experiment designed to test Nash's properties was conducted by Rapoport, Frenkel, and Perner (1977). Like the previous experiment, it was conducted for monetary payoffs about which the bargainers were given full information, and for which they bargained face-to-face. This experiment was designed to test three of Nash's four properties: symmetry, Pareto optimality, and independence of equivalent utility representations. Participants were randomly paired and bargained in 24 games against the same opponent in games chosen to test the three properties. Independence of irrelevant alternatives was not tested. One goal of this experiment was to consider differences between the variable-threat bargaining situation and the classical fixed-threat situation, but we are concerned here only with the condition concerning the classical bargaining situation with a fixed disagreement outcome.

Bargaining was conducted with the bargainers seated face-to-face on either side of a partition that allowed them to see one another and talk freely but also allowed them to write privately. Following a period of unstructured verbal bargaining, each bargainer separately wrote down his or her demand. There were two iterations of this procedure, after which, if the demands were compatible, each player received the weighted average he or she had demanded. Otherwise, each player received his or her disagreement payoff.

The principal conclusion drawn by the authors is that the bargaining process does not obey the property of independence of equivalent utility representations. Rapoport et al. (1977) stated that "invariance of the solution with respect to positive linear transformations on the payoffs is grossly violated" (p. 89). As in the previous study, they account for this by observing that to a large extent the bargainers make interpersonal comparisons of the payoffs.

---

[5] The distinction between "full" and "partial" information made here should not be confused with the technical game-theoretic terms "complete" and "incomplete" information (cf. Footnote 3) or "perfect" and "imperfect" information.

Another conclusion drawn from this study is that the bargaining process shows "significant departures from Pareto optimality." The difference between this and the previous study in this respect may be primarily due to the different procedure by which agreements were reached. Although both studies permitted unrestricted verbal bargaining, the actual final demands in this study were made independently. This seems to have resulted in an increase both in the number of disagreements and in the amount of hedging of demands to ensure compatibility, both of which contribute to departures from Pareto optimality.

Some departures from symmetry were observed as well, although these did not reach significance. As an ordinal measure, it was found that Nash's solution was a good predictor of which player would come out ahead in a given game, even though the agreements reached could be quite far from those predicted by Nash's solution.

The results of this study are consistent with others conducted by Rapoport and his colleagues, reported in Rapoport and Perner (1974), and Rapoport, Guyer, and Gordon (1976). These studies differed from Rapoport et al. (1977) in that they did not attempt to test Nash's properties separately, nor did they employ actual monetary rewards. Also, the rules under which bargaining was conducted were somewhat more restrictive than in the study just considered. However, players in both studies were given full information about the payoffs. The principal conclusion of these studies is that the outcome of bargaining is heavily influenced by various salient outcomes involving interpersonal comparisons, and that, although Nash's solution is useful as a predictor of which player gets the larger payoff, there is a strong tendency for agreements to be closer to an equal division than would be predicted by Nash's solution.

Both the procedures used and the conclusions reached in the studies by Rapoport and his colleagues are similar to those in the study conducted by O'Neill (Note 1), who also investigated both fixed- and variable-threat bargaining without, however, attempting to test Nash's properties separately. In O'Neill's study, participants played 11 games against the same opponent, all involving the division of a fixed sum of money, with full information. In each game a 3-minute period of face-to-face negotiation was allowed, after which the players separately wrote down their bids. Comparison of the bids for compatibility followed, with the players each receiving either their share of the sum to be divided if agreement had been reached or receiving their disagreement payoff. O'Neil concluded that the Nash solution is highly correlated with the outcome of these games, so that it serves as a good predictor of the relative advantage that a player enjoys in one game as opposed to another. However, as in the previous studies, O'Neill (Note 1) also noticed "a systematic shift for all games toward the equal-split outcome" (p. 100).

Other studies have observed the same kind of shift in bargaining over monetary payoffs with full information. For instance, in a famous series of experiments in which the amount of information available to the players was an experimental variable, Siegel and Fouraker (1960; see also Fouraker & Siegel, 1963) found that agreements in a bilateral monopoly (one buyer, one seller) tended toward equal profit, with agreements approaching equality most closely under the conditions of fullest information about payoffs. Their studies (concerned only incidentally with Nash's solution) involved bargainers trying to reach price and quantity agreements on the sale of some commodity, communicating via written numerical bids under conditions of full information, partial information, or asymmetric information (one player informed, one uninformed) about the monetary payoffs that each player would receive from any agreement. A similar conclusion about the effect of varying the amount of available information was reached by Crott and Montmann (1973) and by Felsenthal (1977), who studied a bargaining situation in which there was a discrete (rather than a convex) set of feasible monetary payoffs. Messé (1971) also considered bargaining with full information over a discrete set of feasible monetary payoffs, and in the symmetric condition of his experiment, he found a strong tendency toward equal payoffs. Komorita and Kravitz (1979) reached a somewhat similar conclusion in a study in which experimental variables were

differences in disagreement payoffs, magnitude of the prize, and group size. Komorita and Kravitz studied pure bargaining (each bargainer could enforce the disagreement outcome) involving groups of two, three, or four bargainers negotiating over the distribution of a fixed number of "points."

Two more studies involving bargaining games with full information about payoffs are not directly comparable to the other studies considered here because they were conducted in such a way as to permit no communication of any sort between players. Both Stone (1958) and Crott (1971) presented participants with a set of games in which the feasible set was portrayed graphically. Each participant was instructed to independently mark a demand on the graph with the understanding that the graphs marked by the player and his or her opponent would subsequently be compared to see if their demands were compatible; this would determine whether agreement had been reached. In Stone's experiment, participants were told to make their demands on a series of games, and were told that their opponent had already made his demands. In fact the opponent was imaginary. The games were defined in terms of monetary payoffs, but the monetary awards were imaginary, although participants were told that those whose imaginary winnings fell in the top 25% would receive a prize of $1. Stone was primarily interested in the degree to which players hedged their demands, and he reported a correlation between "cautious" behavior in one game and in another, across a variety of games. In Crott's (1971) experiment, the games were also defined in terms of the feasible set of monetary payoffs. After each participant had indicated his or her demand in a series of games, a random pairing of bargainers in a subset of these games determined the actual cash payoffs. Crott (1971) found that the ratio of the players' maximum feasible payoffs was a good predictor of the ratio of their demands. (For games in which the disagreement point is the origin, the solution G predicts payoffs in the same ratio.) Crott (1972a) reported a related investigation into different aspects of bargaining.[6]

Taken together, the results of the bargaining experiments in which the bargainers had full information about one another's monetary payoffs are consistent with the proposition that comparison of players' payoffs plays a part in determining the outcome of bargaining. In view of the fact that the game-theoretic models under consideration are stated in terms of utility rather than monetary payoffs, the question arises whether the experimentally observed comparisons involve money only, or whether they involve comparisons of the players' utility in some deeper way.

Partial Information

Two experiments have been reported that attempt to address this issue by studying bargaining situations in which the bargainers negotiate over the distribution of chips with only partial information about the value of those chips (each bargainer knows the value of the chips to himself but not to his or her opponent). In situations of this sort, bargainers cannot directly compare the value of their payoffs, since this information is not common knowledge. The two studies (Heckathorn, 1978; Nydegger, 1977) that have been conducted using this procedure illustrate the difficulty of implementing a study of this kind in such a way that it conforms to the assumptions of the models it is intended to test.

Nydegger's (1977) study, also reported in Nydegger (1979), was modeled on the study by Nydegger and Owen (1975) and was intended both to test Nash's four properties and to compare the predictive value of Nash's and Raiffa's solutions in bargaining with partial information about the payoffs. To establish the payoff scales, each individual was asked to select 10 items he would like to have from a list of 20 items from the campus store and to rank them in order of preference. He was then told to assign the

---

[6] Further investigations by Crott and his co-workers into various aspects of bargaining are reported in Crott (1972b), and in Crott, Simon, and Yelin (1974); Crott, Lumpp, and Wildermuth (1976); Crott and Muller (1976); Crott, Muller, and Hamel (1978). These studies fall outside the scope of this review, but they are mentioned here for completeness, since they have not been included in the reviews of the English language literature.

most preferred item a value of "100 units," and to assign lower values to the other items "with the only constraint being to preserve the ordinality of the rankings." Participants were then brought together to bargain for chips that were assigned values in "units," with the understanding that a player who earned a certain number of units in bargaining would then receive the most preferred item that he had assigned not more than that number of units.

The difficulty in interpreting the results of this experiment arises from the fact that the actual payoffs available to the participants consisted of a discrete set of prizes. Thus, for instance, if one of the bargainers had valued one of the potential prizes at 55 units and the next most desirable prize at 40 units, then he would presumably be indifferent to receiving anywhere from 40 to 54 units, since any agreement in this range would yield the same prize. Thus if the players are assumed to be motivated by the prizes, the units cannot be taken to represent the utilities of the players.

Heckathorn's (1978) study was also intended to study bargaining under conditions of partial information and to compare Nash's solution with Raiffa's.[7] However, the rules under which the bargaining was conducted did not correspond to those of a pure bargaining game, since there was no disagreement outcome that could be enforced by both players. A further difficulty encountered in this article is that Raiffa's solution is identified by an incorrect formula that is not independent of changes in the origins of the players' payoff scales.

Partial information of a different sort was considered in an experiment by Hoggatt, Selten, Crockett, Gill, and Moore (1978), motivated by a theoretical model based on Nash's solution, proposed by Harsanyi and Selten (1972). In the game played in this experiment, participants bargained over the distribution of 20 "money units" whose value (10¢) was known to both of them. In the event of disagreement, each player received 0 units. What each player did not know was whether his or her opponent had a "high cost" of 9 units or a "low cost" of 0 units that would be subtracted from his or her

winnings following any agreement. Before the game commenced, each player's cost was determined by a random process that gave a 50% chance of having high or low cost. Each player was informed only of the result of his or her own random process, so that when the bargaining commenced, each player knew his or her own cost and the probability distribution that determined the opponent's cost. (Situations of this type are called games of *incomplete* information.) After equating the utility of the players with their monetary payoffs, it was found that there was substantial qualitative agreement between aspects of the experimental results and predictions of the theory, although agreements at an equal division of the 20 units were the most common, even in the case in which a player with low cost bargained against a player with high cost.

Modification of the Classical Models

The experimental evidence considered in the previous section reveals some considerable discrepancies between observed experimental results and the predictions of the game-theoretic models. In particular, the property of independence of equivalent utility representations was consistently violated in studies in which each player knew the monetary value of his or her opponent's payoffs.

There is thus ample support for the proposition that comparison of the monetary payoffs available to the bargainers plays a role in determining the outcome of bargaining in situations that make available to both bargainers the information necessary to make such comparisons. To design a descriptive theory of bargaining for situations in which the participants know each other's monetary payoffs (rather than their utility payoffs), it is therefore necessary to reexamine those aspects of the classical models that deal with the nature of the payoffs and the information they convey. In this section we consider a new model that allows for the comparison of the payoffs between players.

---

[7] Heckathorn refers to Raiffa's solution as the "Smorodinsky-Kalai" solution.

When bargaining games (S, d) defined in terms of the feasible utility payoffs were considered, it was natural to impose the restriction that the set S be convex. However, when the game (S, d) is defined in terms of monetary payoffs, this restriction is no longer natural, and so this section is concerned with games from a larger class than the class B, considered so far. Let B* denote the class of games (S, d) where S is a compact (but not necessarily convex) subset of the plane containing the point d and at least one point x such that $x > d$. For simplicity we will also require that the Pareto optimal subset of S be connected, as would be the case in any game in which the bargaining concerns the distribution of a divisible commodity. When a game (S, d) is considered as a member of the class B*, the set S is to be interpreted as the set of feasible monetary payoffs and the point d as the monetary payoffs resulting from disagreement.

A model of bargaining in such games will be a solution f defined on the class B*. To be consistent with the available experimental evidence, this solution should possess Properties 1–3 but not Property 4. Specifically, it should incorporate information (of a kind that is precluded by Property 4) about the relative payoffs the players receive at any agreement. It is instructive to consider the effect of replacing Property 4 with the following property that in certain respects can be considered almost the opposite of Property 4.[8]

Property 4′: Independence of ordinal transformations preserving interpersonal comparisons: Let (S, d) and (S', d') be games in B* such that $d'_i = t_i(d_1, d_2)$ for $i = 1, 2$, and $S' = \{ (t_1(x_1, x_2), t_2(x_1, x_2)) \mid (x_1, x_2) \in S \}$ where $t = (t_1, t_2)$ is a transformation $t : S \to R^2$ such that for all x, y in S and $i = 1, 2$;

(i) $t_i(x_1, x_2) \geq t_i(y_1, y_2)$ if and only if $x_i \geq y_i$, and

(ii) $t_1(x_1, x_2) - t_1(d_1, d_2) \geq t_2(x_1, x_2) - t_2(d_1, d_2)$, if and only if $x_1 - d_1 \geq x_2 - d_2$.

Then

$$f_i(S', d') = t_i(f(S, d)).$$

This property states that if a game (S', d') is derived from (S, d) via a transformation t (of feasible payoff vectors) that (a) preserves each player's ordinal preferences and (b) preserves information about which player makes larger gains at any given payoff, then the same transformation applied to the final agreement of the game (S, d) should yield the final agreement of the game (S', d'). That is, it states that the solution should depend only on the ordinal preferences of the players and on the ordinal comparison of the payoffs to the players at any given agreement, and not on any other features of the payoffs. A solution possessing Property 4′ thus differs from a solution possessing Property 4 in two ways: (a) It treats the feasible payoffs to a player as ordinal rather than interval data (i.e., it reflects the direction of each player's preferences but not how much the player prefers one payoff to another) and (b) it permits ordinal comparisons of one player's payoffs to those of the other player, rather than precluding all such comparisons.

One solution that possesses Property 4′ is the solution $f = E$, that selects the outcome $E(S, d) = z$, where z is the Pareto optimal point in S such that $\min\{z_1 - d_1, z_2 - d_2\} > \min\{x_1 - d_1, x_2 - d_2\}$ for all Pareto optimal points x in S distinct from z. The solution E picks the Pareto optimal point in S that maximizes the minimum gains available to the players. The letter E was chosen to reflect the fact that this solution selects the outcome that gives both players equal gains, whenever there exists a Pareto optimal outcome with this property. In any event, the solution E always selects the Pareto optimal point that comes closest to giving the players equal gains.

Just as Nash's solution is uniquely characterized by Properties 1, 2, 3, and 4, it turns out that the equal gains solution E can be characterized by Properties 1, 2, 3, and 4′. That is, we can state the following new result.

Theorem 4: The equal gains solution E is the unique individually rational solution that possesses Properties 1, 2, 3, and 4′.

---

[8] An alternative approach is investigated by Kalai (1977b), Myerson (1977), and Roth (1979). These and other related approaches are considered in Roth (in press).

The proof of this theorem is given in the Appendix.

The equal gains solution E can thus be thought of as differing from Nash's solution only in the kind of information that is assumed to determine the outcome of bargaining. Property 4 permits Nash's solution to be sensitive to the *intensity* of each player's preferences for the various potential outcomes (as measured by von Neumann & Morgenstern, 1953, utility functions), but requires it to be insensitive to any comparison between players. Property 4', on the other hand, permits the equal gains solution E to be sensitive to comparisons of the payoffs the bargainers get at any given outcome, and to be sensitive to each players' ordinal preferences over different outcomes, but prevents it from being sensitive to the intensity of their preferences over different outcomes.

The experiment described in the following section is designed to help distinguish the effects of those two kinds of information.

Experiment

The experiment reported here is designed to test the hypothesis that the information that influences the outcome of bargaining is to a large extent the information shared by both bargainers, as opposed to information available to only one of the bargainers. We examine the hypothesis that Nash's solution is descriptive of bargaining situations in which each player does know his or her opponent's von Neumann–Morgenstern utility for each outcome but does not know the opponent's monetary payoff. We will also examine the related hypothesis that when the players know both their opponents' monetary payoffs as well as their utilities, the outcome of bargaining will be influenced by interpersonal comparisons, in the direction of equal gains. Simply put, we will be examining the hypothesis that Nash's solution is descriptive of bargaining when the players share the sort of information assumed by Property 4, and that Property 4 (and consequently Nash's solution) fails to be descriptive when the shared information is of a different sort. (The specific implications of these hypotheses for our experiment

will be described in detail following the description of the experimental design.)

None of the experiments reviewed earlier necessarily permit the players to know one another's von Neumann–Morgenstern expected utility for an agreement, since even the experiments conducted under conditions of full information revealed only the monetary payoffs available. To the extent that a player has a utility function that is not linear in money, this utility function is not known to the player's opponent (nor to the experimenter). At the same time, in giving each player information about the opponent's monetary payoffs, these experiments gave the players information about their opponent's payoffs that is *not* contained in the players' utility functions, and that is defined on a common absolute scale rather than on independently normalized interval scales. The fact that observed agreements tended toward equal payoffs makes clear that this information played a role in the bargaining, since the comparisons necessary to determine equality are not well-defined on independently normalized interval scales.

Thus the preceding experiments provided *less* information of one kind (about the utilities) and *more* information of another kind (about the monetary awards) than is assumed to be relevant by the classical game-theoretic models being tested.[9] The experiment described next is designed to provide participants with the required information about their opponent's utility for the available payoffs, and to permit information about the underlying monetary payoffs to be provided or withheld as an experimental variable.

Design

Recall that knowing an individual's expected utility for a given agreement is equivalent to knowing what lottery he or she thinks is as desirable as that agreement. Thus in a bargaining game in which the feasible agreements are the appropriate kind of lotteries,

---

[9] That is, knowing a player's monetary payoffs does not permit his or her utility function to be known except in the special case in which utility is known to be linear in money. Even in this case, however, knowing the monetary payoffs conveys additional information not conveyed by the utility function alone.

knowing the utilities of the players at a given agreement is equivalent to simply knowing the lottery they have agreed on.

In each game of this experiment, therefore, players bargained over the *probability* that they would receive a certain monetary prize, possibly a different prize for each player. Specifically, in each of four games played under two information conditions, players bargained over how to distribute "lottery tickets" that would determine the probability that each player would win his or her personal lottery (i.e., a player who received 45% of the lottery tickets would have a 45% chance of winning his or her specified monetary award and a 55% chance of winning nothing). In the event that no agreement was reached, each player received nothing. In the full information condition, each player was informed of the value of his or her own potential prize and of his or her opponent's potential prize; in the partial information condition, each player was informed only of the value of his or her own prize.

Each player played four games, in random order, under one of the information conditions, against different opponents. Players were allowed to communicate freely by teletype, but they were unaware of the identity of their opponents. In Game 1, no restriction was placed on the percentage of lottery tickets that each player could receive, and both players had the same potential payoff of $1. Game 2 was played with the same potential payoffs as Game 1, but one of the players (Player 2) was restricted to receive no more than 60% of the lottery tickets. Game 3 was played with the same rules as Game 1, but with different monetary payoffs for the two players: $1.25 for Player 1 and $3.75 for Player 2. Game 4 was played under the same rules as Game 2, with the same prizes as Game 3 (see Table 1).

To interpret the set of feasible outcomes in each of these games in terms of each player's utility function for money, recall that if we consider each player's utility function to be normalized so that the utility for receiving his or her own prize is 1, and the utility for not receiving it is 0, then the player's utility for any lottery between those two alternatives is the probability of winning the lottery.

Note that we are considering the feasible set of utility payoffs to be defined in terms of the utility function of each player for the lottery that the player

receives, independently of the bargaining that has taken place to achieve this lottery, and even independently of the lottery that the opponent receives. In doing so, we are taking the point of view that although the progress of the negotiations may influence the utilities of the bargainers for the agreement eventually reached, the description of any effect that this has on the agreement reached belongs in the model of the bargaining process rather than in the model of the bargaining situation. Considerable confusion in the literature has resulted from attempts to interpret bargaining models in terms of the player's utilities for outcomes *after* the bargaining has ended, since no bargaining model can be falsified by experimental evidence if, after an outcome has been chosen, the utilities of the players can be interpreted as having changed in whatever way is necessary to be consistent with the model. To have predictive value, bargaining theories must be stated in terms of parameters that can be measured independently of the phenomena that the theories are designed to predict, and it is for this reason that we consider the utilities that define the game, in either information condition, to be simply each player's utility for money. (See the passage from Chertkoff and Esser, 1976, quoted on p. 574 of this article.)[10]

Predictions of the Models

Because the monetary awards available to the players in each game are the same under both information conditions, so are the feasible utility payoffs. Since the classical game-theoretic models depend only on the feasible set of utility payoffs (i.e., since they are defined on the class B of games), both Nash's and Raiffa's solutions predict no difference between the two information conditions.

---

[10] An alternative approach (particularly in the full information condition) would have been to try to assess the utility of each player for all possible *divisions* of the lottery tickets; that is, to incorporate into the utility functions each player's preferences for his or her standing relative to the opponent, as well as for the monetary payoff that the player receives. In this particular experiment, this would have no effect on the utility functions of the players for any potential Pareto-optimal agreement, as long as each player prefers, of any two agreements the one that gives him or her the higher percentage of lottery tickets. (This is because each agreement is a lottery between the same two outcomes—the most preferred and the least preferred Pareto-optimal agreements.) However, there could, in general, be an effect on the utility of the players for the disagreement outcome and for the non-Pareto-optimal agreements. For the reasons indicated, we will consider the set of feasible utilities associated with each game to be defined by each player's utility for money.

Table 1
*Prizes and Feasible Distributions
for Games 1–4*

Game	Prize for Player 1	Prize for Player 2	Maximum % allowed Player 1	Maximum % allowed Player 2
1	$1	$1	100	100
2	$1	$1	100	60
3	$1.25	$3.75	100	100
4	$1.25	$3.75	100	60

Property 1, Pareto optimality, predicts that in all four games agreements will be reached and will divide all of the lottery tickets. Property 2, symmetry, predicts that in Game 1 the players will receive equal percentages of the lottery tickets. (Thus Properties 1 and 2 together imply a 50–50 split in Game 1.) Property 3, independence of irrelevant alternatives, then predicts that Game 2 will reach the same outcome as Game 1 (since the restriction in Game 2 does not exclude the 50–50 split). Property 4, independence of equivalent utility representations, predicts that Game 3 will have the same outcome as Game 1 and Game 4 will have the same outcome as Game 2, since these games differ only in the size of the monetary prizes (which affects only the scale of the utility functions).

Taken together, Properties 1–4 thus predict a 50–50 split in all four games. Since Nash's solution F possesses all four properties, this is the prediction of Nash's solution. In what follows, it will sometimes be convenient to discuss the predictions of the models in terms of the quantity D, defined as the percentage of lottery tickets received by Player 2 minus the percentage of lottery tickets received by Player 1. The prediction of Nash's solution is that D will equal 0 for all four games under both information conditions.

Raiffa's solution G does not possess Property 3, and so it predicts a different outcome for Game 2 than for Game 1; like Nash's solution, it predicts that Game 1 will result in a 50–50 split, but in Game 2 it predicts that Players 1 and 2 will receive 62.5% and 37.5% of the lottery tickets, respectively. Since it possesses Property 4, it predicts that Game 3 should have the same outcome as Game 1, and Game 4 should have the same outcome as Game 2. Thus Raiffa's solution predicts that D will equal 0 for Games 1 and 3, and that it will equal -25 for Games 2 and 4 under both information conditions.

Our principal experimental hypothesis is that only in the partial information condition will the observed outcomes be consistent with Property 4, since Property 4 specifies that only the utility functions of the players should be relevant to the outcome, and this is precisely the information shared by the players in the partial information condition. Thus Games 1 and 3 should yield the same outcome in this condition, as should Games 2 and 4. It would be consistent with the results of most of the studies reviewed earlier if the observations were consistent with Properties 1–3 as well, and so our hypothesis for this condition is that Nash's solution will be descriptive.

In the full information condition, the principal hypothesis is that the observations will continue to be consistent with Properties 1–3, but will no longer be consistent with Property 4: that is, Game 1 will yield a different outcome than Game 3, and Game 2 will yield a different outcome than Game 4. In this condition, the players share both the information about one another's utility specified by Property 4 as well as the information about each other's monetary payoffs needed to make the kind of comparisons specified by Property 4'. Since Property 4 yields Nash's solution F when combined with Properties 1–3, whereas Property 4' yields the equal gains solution E, our hypothesis is that, in this condition, the observations will tend to fall between those predicted by the two solutions. (Note that the results of the earlier studies conducted under conditions of full information also tended to fall between the predictions of these two solutions.) The equal gains solution defined on the expected monetary payoffs (rather than the expected utility payoffs) predicts that the players will each receive 50% of the lottery tickets in Games 1 and 2, and that in Games 3 and 4 Players 1 and 2 will receive 75% and 25%, respectively. (This would result in an expected monetary payoff of 50¢ to each player in Games 1 and 2, and 95¢ to each player in Games 3 and 4.)

The experimental hypotheses are, therefore, that in the full information condition, D will equal 0 for Games 1 and 2, and be between 0 and -50 for Games 3 and 4. In the partial information condition, D should equal 0 for all four games.

Method

Subjects. The subject pool consisted mainly of college sophomores from an introductory business administration course. No special skill or experience was required for participation. Pretests were run with

the same subject pool to ensure that the instructions to participants were clear and easily understandable.

Procedure. Each participant was seated at a visually isolated terminal of a computer-assisted instruction system (called PLATO) developed at the University of Illinois, whose features include advanced graphic displays and interactive capability. The experiment was conducted in a room containing over 70 terminals, most of which were occupied at any given time by students uninvolved in this experiment. No more than 9 of the terminals were used for the experiment at any time (8 terminals occupied by participants, and 1 terminal used by the experimenter to monitor the proceedings). Participants were seated by the experimenter in order of their arrival at scattered terminals throughout the room, and for the remainer of the experiment they received all of their instructions and conducted all communication through the terminal.

Background information such as a brief review of probability theory was first presented. The main tools of the bargaining were then introduced. These consisted of sending messages or sending proposals. A proposal was a pair of numbers, the first of which was the sender's probability of receiving his/her prize and the second was the receiver's probability. The use of the computer enabled any asymmetry in the presentation to be avoided. PLATO also computed the expected value of each proposal and displayed the proposal on a graph of the feasible region. After being made aware of these computations, the bargainer was given the option of canceling the proposal before its transmittal. Proposals were said to be binding on the sender, and an agreement was reached whenever one of the bargainers returned a proposal identical to the one he or she had just received.

Messages were not binding. Instead, they were used to transmit any thoughts that the bargainers wanted to convey to each other. To ensure anonymity, the monitor intercepted any messages that revealed the identity of the players. In the partial information condition, the monitor also intercepted messages containing information about the available prizes. The intercepted message was returned to the sender with a heading indicating the reason for such action.

To verify their understanding of the basic notions, the subjects were given some drills followed by a simulated bargaining session with the computer. As soon as all the participants finished this portion of the experiments, they were paired at random and the bargaining started.

At the end of 12 minutes or when agreement was reached (whichever came first), the subjects were informed of the results of that game and were asked to wait until all the other bargainers were finished. For the subsequent game there were new random pairings, and the bargaining resumed. The cycle continued until all four games were completed. At no point in the experiment were the players aware of what the other participants were doing, or of the identity of their opponents.

Table 2
Means and Standard Deviations for D

	Game			
Statistic	1	2	3	4
	Full information (11 pairs)			
M	0.0	−1.909	−34.600	−21.636
SD	0.0	12.169	19.277	22.482
	Partial information (8 pairs)			
M	0.0	1.325	2.500	−2.500
SD	0.0	8.335	4.629	4.106

Note. The mean and standard deviation are reported after the removal of an outlier ($D = +98$ resulting from a $(1, 99)$ agreement) that is 6.8 standard deviations from the mean.

The bargaining process consisted of the exchange of messages and proposals, and participants were instructed that "your objective should be to maximize your own earnings by taking advantage of the special features of each session." Only if the bargainers reached agreement on what percentage of the "lottery tickets" each would receive were they allowed the opportunity to participate in the lottery for the particular game being played. All transactions were automatically recorded.

The lotteries were held after all four games were completed, and each player was informed of the outcomes and the amount of his winnings. A brief explanation of the purpose of the experiment was then given, and the subjects were offered the opportunity to type any comments, questions, and so on, and were directed to the monitor who paid them.

Results

The 76 games played yielded 72 agreements (95%) of which 71 (99%) were Pareto optimal, so that 93% of the bargaining encounters ended in a Pareto optimal agreement.

Because the players did not know who they were bargaining with and since a different random pairing of the subjects was performed in each session, we shall assume the replication effect due to games to be negligible. This assumption of independence is consistent with the assumptions made by Kahan and Rapoport (1974).

Table 2 gives the means and standard deviations for *D*. (An outlier in Game 3 of the full information condition has been removed.) The zero variance for Game 1 precludes conventional analysis. However, Games 2, 3,

and 4 in the partial information condition were not significantly different from zero, $t(7) = .45$, $t(7) = 1.528$, and $t(7) = 1.722$, respectively. In the full information condition also, Game 2 was not significantly different from zero: $t(10) = .520$.

Both Games 3 and 4 were then compared across information conditions. A t test was performed yielding a significant difference $t(10.28)$[11] $= 5.88$, $p < .001$ for Game 3, and $t(10.91) = 2.76$, $p < .02$ for Game 4. (The Mann-Whitney U test substantiated this finding.)

In comparisons within conditions, t tests showed that in the partial information condition, Game 1 was not significantly different from Game 3, $t(14) = -1.53$; and Game 2 was not significantly different from Game 4, $t(14) = -1.16$. In the full information condition, however, Game 1 was significantly different from Game 3, $t(19) = 5.97$, $p < .001$; and Game 2 was significantly different from Game 4, $t(20) = -2.56$, $p < .02$.

We then compared Game 1 to Game 2 finding no significant difference in either the full or the partial information condition, $t(20) = .52$, and $t(14) = -.45$, respectively. Comparison of Games 3 and 4 did not yield a significant difference in the full information condition, $t(19) = 1.41$, whereas, in the partial information condition, there was a significant difference, $t(14) = -2.29$, $p < .04$.

Thus the results are consistent with the hypothesis that Games 3 and 4 in the full information condition are different from all the other games. Inspection of the data in Table 2 also clearly affirms the presence of the predicted effect. (The unaggregated data are given in Table 3.)

The high percentage of Pareto optimal agreements lends support to the proposition that the bargaining process observed here can be described by Nash's Property 1, under both information conditions. Nash's Property 3, independence of irrelevant alternatives, is descriptive in the full information condition but is not supported in the case of Games 3 and 4 of the partial information condition. Under the partial information condition, the similarity of the means for all four games, and the fact that they are not significantly different from zero, support the proposition that

the bargaining process under this condition possesses Properties 2 and 4. In the full information condition, the differences between the data for Games 1 and 3 and the differences between the data for Games 2 and 4 suggest that comparison of the expected monetary payoffs to each player played a role in determining the agreements reached. Informal examination of the transcripts containing the messages exchanged by the players also supports this conclusion. Of the 22 outcomes observed in Games 3 and 4 under full information, 19 resulted in agreements and 17 (87%) of these lie between the predictions of Nash's solution and the equal gains solution. The principal hypotheses that the experiment was designed to test are thus supported by the results.

Summary and Conclusions

The principal issue with which this article has been concerned is the effect that the quality of the information commonly shared by the bargainers has on the outcome of bargaining. We saw that Nash's model of the bargaining process depends primarily on only two of the properties he proposed, one of which is intimately connected with the assumption that the utility function of a bargainer constitutes the only information available to his or her opponent about the bargainer's payoff at any agreement. A review of the experimental literature showed that when the utility of the players is taken to be identical to their monetary payoffs, this property is violated.

A new experiment was designed to investigate whether this divergence between the theoretical prediction and the results observed in previous experiments might be due to the fact that the information made available in those experiments differed from the assumptions of Nash's model. By having players bargain over lotteries, it was possible to produce experimental conditions in which the players knew one another's utilities. In the partial information condition, this was the only information the players had about their

---

[11] Due to the unequal variances in the two samples, a noninteger degree of freedom is reported.

opponent's payoffs, so the information shared by the bargainers conformed to the assumptions of Nash's model. The experimental results supported the hypothesis that Nash's solution is descriptive of the bargaining process under this condition. In the full information condition in which the players knew each other's monetary awards as well as their

Table 3
Percentage of Lottery Tickets

		Full information			Partial information			
		Player				Player		
Game no.	Group no.	1	2	D	Group no.	1	2	D
1	1	50.00	50.00	—	2	50.00	50.00	—
		50.00	50.00	—		50.00	50.00	—
		50.00	50.00	—		50.00	50.00	—
	3	50.00	50.00	—	4	50.00	50.00	—
		50.00	50.00	—		50.00	50.00	—
		50.00	50.00	—				
	5	50.00	50.00	—	6	50.00	50.00	—
		50.00	50.00	—		50.00	50.00	—
	7	50.00	50.00	—	8	50.00	50.00	—
		50.00	50.00	—				
	9	50.00	50.00	—				
2	1	50.00	50.00	—	2	40.00	60.00	20.00
		50.00	50.00	—		—	—	—
		50.00	50.00	—		50.00	50.00	—
	3	50.00	50.00	—	4	50.00	50.00	—
		50.00	50.00	—		49.70	50.30	.60
		50.00	50.00	—				
	5	68.00	32.00	−36.00	6	55.00	45.00	−10.00
		25.00	40.00	15.00		50.00	50.00	—
	7	50.00	50.00	—	8	50.00	50.00	—
		50.00	50.00	—				
	9	50.00	50.00	—				
3	1	73.00	27.00	−46.00	2	45.00	55.00	10.00
		1.00	99.00	98.00		50.00	50.00	—
		65.00	35.00	−30.00		50.00	50.00	—
	3	—	—	—	4	50.00	50.00	—
		70.00	30.00	−40.00		45.00	55.00	10.00
		75.00	25.00	−50.00				
	5	—	—	—	6	50.00	50.00	—
		70.00	30.00	−40.00		50.00	50.00	—
	7	75.00	25.00	−50.00	8	50.00	50.00	—
		70.00	30.00	−40.00				
	9	75.00	25.00	−50.00				
4	1	75.00	25.00	−50.00	2	50.00	50.00	—
		50.00	50.00	—		54.00	46.00	−8.00
		55.00	45.00	−10.00		50.00	50.00	—
	3	50.00	50.00	—		51.00	49.00	−2.00
		74.00	26.00	−48.00	4	50.00	50.00	—
		50.00	50.00	—				
	5	—	—	—	6	55.00	45.00	−10.00
		60.00	40.00	−20.00		50.00	50.00	—
	7	80.00	20.00	−60.00	8	50.00	50.00	—
		60.00	40.00	−20.00				
	9	65.00	35.00	−30.00				

utilities, the results confirmed the hypothesis that in games in which the monetary awards to the players differed, the agreements reached would show a shift in the direction of equal monetary gains.

The full information condition permitted the players to make comparisons of the expected monetary payoffs each player would receive from any agreement, and the results strongly support the hypothesis that these comparisons played a role in the bargaining process. It was shown that if Nash's Property 4 was replaced with a Property (4') that permitted such comparisons, then a solution could be derived that predicted agreements in which the players received equal gains. (This could be viewed as a formal derivation of some of the "equity"-related predictions that have been made in bargaining contexts). Both Nash's Property 4 and Property 4' are rather extreme, in that Property 4 permits only the intensity of players' preferences but not comparison of their payoffs to affect the bargaining process, whereas Property 4' allows only for the effect of comparisons but not for the intensity of preferences. Since the full information condition provides the necessary information both for comparisons of payoffs and for judgments about intensity of preferences, it is not too surprising that the results in this condition largely fell between the predictions of Nash's solution and the equal gains solution. This suggests that in this condition, both kinds of information influenced the bargaining.

This phenomenon seems likely to be quite general, since even in many conventional bargaining situations in which the bargainers have no direct knowledge of each others utilities it is still probable that each bargainer can form some estimate of the intensity of his or her opponent's preferences over various agreements. The fact that the results of many of the earlier experiments also fall between the predictions of Nash's solution and the equal gains solution tends to bear this out. It may also be the case that bargaining situations in which the bargainers share the most information will also offer the most scope for individual bargaining ability. The variances presented in Table 2 lend support to this hypothesis, since the disper-

sion of the results is greater under the full information condition than under the partial information condition. Further experimental work will be needed to explore how different kinds of information are incorporated into the bargaining process and how the distribution of this information between the bargainers influences the outcome of bargaining. More sophisticated mathematical models will be needed to deal simultaneously with the different kinds of information available to the bargainers, and to indicate how different kinds of information may affect different aspects of the bargaining.

Much research in the social psychology literature has concentrated on what focal points become "salient" in the course of negotiations. The experimental results presented here, together with the consequences of Properties 4 and 4' as reflected by Nash's solution and the equal gains solution, make it clear that "salience" in negotiations depends on the information shared by the bargainers. Insight into the mechanism by which outcomes become salient in bargaining may shed light on more general questions concerning how mutual expectations are formed in social situations.

A more general conclusion supported by this article is that game-theoretic models of bargaining provide a powerful theoretical framework, with testable empirical content and sufficient flexibility, to permit the study of a wide range of bargaining situations. The practice of deriving a solution from its characteristic properties permits the design of experiments that test these properties, and the results of such experiments suggest new properties that, because of the deductive nature of formal mathematical models, permit the derivation of new solutions. There thus appear to be good prospects for considerable interaction between further development of theory and continued experimental investigation in this area.

Reference Note

1. O'Neill, B. *Variable threat bargaining: A test of the Raiffa-Nash theory.* (Tech. Report MMPP76-7) Ann Arbor: Michigan Mathematical Psychology Program, University of Michigan, 1976.

References

Butrim, B. I. A modified solution of the bargaining problem (Russian). *Zhurnal Vychislitel'noi Matematiki Matematicheskoi Fiziki*, 1976, *16*, 340–350.

Chertkoff, J. M., & Esser, J. K. A review of experiments in explicit bargaining. *Journal of Experimental Social Psychology*, 1976, *12*, 464–486.

Crott, H. W. Experimentelle untersuchung zum verhandlungsverhalten in kooperativen spielen. *Zeitschrift für Sozialpsychologie*, 1971, *2*, 61–74.

Crott, H. W. Der einfluss struktureller und situativer merkmale auf das verhalten in verhandlungssituationen: Teil I. *Zeitschrift für Sozialpsychologie*, 1972, *3*, 134–158. (a)

Crott, H. W. Der einfluss struktureller und situativer merkmale auf das verhalten in verhandlungssituationen: Teil II. *Zeitschrift für Sozialpsychologie*, 1972, *3*, 227–244. (b)

Crott, H. W., & Montmann, V. Der effekt der information über die verhandlungsmöglichkeiten des gegners auf das ergebnis einer verhandlung. *Zeitschrift für Sozialpsychologie*, 1973, *4*, 209–219.

Crott, H. W., Simon, K., & Yelin, M. Der einfluss des anspruchsniveaus auf den verlauf und das ergebnis von verhandlungen. *Zeitschrift für Sozialpsychologie*, 1974, *5*, 300–314.

Crott, H. W., Lumpp, R. R., & Wildermuth, R. Der einsatz von bestrafungen und belohnungen in einer verhandlungssituation. In H. Brandstatter & H. Schuler (Eds.), *Zeitschrift für Sozialpsychologie*, Beiheft 2, Entscheidungsprozesse in Gruppen. Bern: A. Huber, 1976.

Crott, H. W., & Muller, G. F. Der einfluss des anspruchsniveaus und der erfahrung auf ergebnis und verlauf dyadischer verhandlungen bei vollstandiger information über die gewinnmöglichkeiten. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 1976, *4*, 548–568.

Crott, H. W., Muller, G. F., & Hamel, P. L. The influence of the aspiration level, of the level of information and bargaining experience on the process and outcome in a bargaining situation. In H. Sauermann (Ed.), *Contributions to experimental economics*. (Proceedings of the second conference on experimental economics) Tubingen: J. C. B. Mohr, 1978.

Felsenthal, D. S. Bargaining behavior when profits are unequal and losses are equal. *Behavioral Science*, 1977, *22*, 334–340.

Fouraker, L. E., & Siegel, S. *Bargaining behavior.* New York: McGraw-Hill, 1963.

Harsanyi, J. C., & Selten, R. A generalized Nash solution for two-person bargaining games with incomplete information. *Management Science*, 1972, *18*, 80–106.

Heckathorn, D. A paradigm for bargaining and a test of two bargaining models. *Behavioral Science*, 1978, *23*, 73–85.

Herstein, I. N., & Milnor, J. W. An axiomatic approach to measurable utility. *Econometrica*, 1953, *21*, 291–297.

Hoggatt, A., Selten, R., Crockett, D., Gill, S., & Moore, J. Bargaining experiments with incomplete information. In H. Sauermann (Ed.), *Contributions to experimental economics*. (Proceedings of the second conference on experimental economics) Tubingen: J. C. B. Mohr, 1978.

Kahan, J. P., & Rapoport, A. Test of the bargaining set and kernal models in three-person games. In A. Rapoport (Ed.), *Game theory as a theory of conflict resolution*. Dordrecht, Holland: D. Reidel, 1974.

Kalai, E. Nonsymmetric Nash, solutions and replications of 2-person bargaining. *International Journal of Game Theory*, 1977, *6*, 129–133. (a)

Kalai, E. Proportional solutions to bargaining situations: Interpersonal utility comparisons. *Econometrica*, 1977, *45*, 1623–1630. (b)

Kalai, E., & Smorodinsky, M. Other solutions to Nash's bargaining problem. *Econometrica*, 1975, *43*, 513–518.

Komorita, S. S., & Kravitz, D. A. The effects of alternatives in bargaining. *Journal of Experimental Social Psychology*, 1979, *15*, 147–157.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. *Foundations of measurements: Vol. 1. additive and polynomial representations.* New York: Academic Press, 1971.

Luce, R. D., & Raiffa, H. *Games and decisions: Introduction and critical survey.* New York: Wiley, 1957.

Messé, L. A. Equity in bilateral bargaining. *Journal of Personality and Social Psychology*, 1971, *17*, 287–291.

Myerson, R. B. Two-person bargaining problems and comparable utility. *Econometrica*, 1977, *45*, 1631–1637.

Morley, I., & Stephenson, G. *The social psychology of bargaining.* London: Allen & Unwin, 1977.

Murninghan, J. K. Models of coalition behavior: game theoretic, social psychological, and political perspectives. *Psychological Bulletin*, 1978, *85*, 1130–1153.

Nash, J. F. The bargaining problem. *Econometrica*, 1950, *28*, 155–162.

Nash, J. F. Two person cooperative games. *Econometrica*, 1953, *21*, 129–140.

Nydegger, R. V. Independent utility scaling and the Nash bargaining model. *Behavioral Science*, 1977, *22*, 283–289.

Nydegger, R. V. The use of independently scaled utility functions in the experimental applications of game theory. In V. Smith (Ed.), *Research in experimental economics* (Vol. 1). Greenwich, Conn.: J.A.I. Press, 1979.

Nydegger, R. V., & Owen, G. Two-person bargaining: An experimental test of the Nash axioms. *International Journal of Game Theory*, 1975, *3*, 239–249.

Raiffa, H. Arbitration schemes for generalized two-person games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the theory of games. II: Annals of mathematics studies.* Princeton, N.J.: Princeton University Press, 1953.

Rapoport, A., Frenkel, O., & Perner, J. Experiments with cooperative 2 × 2 games. *Theory and Decision*, 1977, *8*, 67–92.

Rapoport, A., Guyer, M. J., & Gordon, D. G. *The 2 × 2 game*. Ann Arbor: University of Michigan Press, 1976.

Rapoport, A., & Perner, J. Testing Nash's solution of the cooperative game. In A. Rapoport (Ed.), *Game theory as a theory of conflict resolution*. Dordrecht, Holland: D. Reidel, 1974.

Roth, A. E. Independence of irrelevant alternatives and solutions to Nash's bargaining problem. *Journal of Economic Theory*, 1977, *16*, 247–251. (a)

Roth, A. E. Individual rationality and Nash's solution to the bargaining problem. *Mathematics of Operations Research*, 1977, *2*, 64–65. (b)

Roth, A. E. The Nash solution and the utility of bargaining. *Econometrica*, 1978, *46*, 587–594, 983.

Roth, A. E. Proportional solutions to the bargaining problem. *Econometrica*, 1979, *47*, 775–778.

Roth, A. E. *Axiomatic models of bargaining: Lecture notes in economics and mathematical systems* (No. 170): Springer, in press.

Rubin, J. Z., & Brown, B. R. *The social psychology of bargaining and negotiation*. New York: Academic Press, 1975.

Siegel, S., & Fouraker, L. E. *Bargaining and group decision making*. New York: McGraw-Hill, 1960.

Stone, J. J. An experiment in bargaining games. *Econometrica*, 1958, *26*, 286–296.

Tversky, A. Intransitivity of preferences. *Psychological Review*, 1969, *76*, 31–48.

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

von Neumann, J., & Morgenstern, O. *Theory of games and economic behavior* (3rd ed.). Princeton: Princeton University Press, 1953.

Appendix

Proof of Theorem 4

Theorem 4: The equal gains solution E is the unique solution that is strongly individually rational, strongly Pareto optimal, independent of irrelevant alternatives, and independent of ordinal transformations preserving interpersonal comparisons.

Proof: It is straightforward to verify that the solution E is well defined and possesses the properties specified by the theorem. We need to show that it is the unique solution with those properties; that is, if f is a solution possessing the specified properties, then $f(S, d) = E(S, d)$ for any (S, d) in B*. For any game (S, d), let $\bar{P}(S)$ denote the Pareto optimal subset of S, and let $S_d{}^+ = \{x \in S | x \geq d\}$ be the individually rational subset of S.

First observe that if f is individually rational, strongly Pareto optimal, and independent of irrelevant alternatives, then for any $(S, d) \in$ B*, $f(S, d) = f(T, d)$, where $T = \bar{P}(S_d{}^+) \cup \{d\}$. (This follows since individual rationality implies that $f(S, d) \in S_d{}^+ = \{x \in S | x \geq d\}$, strong Pareto optimality implies $f(S, d) \in \bar{P}(S_d{}^+)$, and so independence of irrelevant alternatives implies $f(S, d) = f(T, d)$ for any subset T of S that contains $\bar{P}(S_d{}^+)$.) We will sometimes denote the game (T, d) by $(\bar{P}(S_d{}^+), d)$, a slight abuse of our notation, since $d \notin \bar{P}(S_d{}^+)$. However, it will be understood that in this case the set of feasible payoff vectors is $\bar{P}(S_d{}^+) \cup \{d\}$. So it will be sufficient to show that f and E coincide on games of the form $(\bar{P}(S_d{}^+), d)$, and Property 4' further ensures that it will be sufficient to show this when $d = (0, 0) \equiv \bar{0}$.

Let $\bar{A}$ be the convex hull of the points $(2, 0)$; $(0, 2)$; and $\bar{0}$; we will show that $f(\bar{P}(\bar{A}), \bar{0}) = E(\bar{P}(\bar{A}), \bar{0}) = (1, 1)$, where $\bar{P}(\bar{A})$ is of course the line segment joining the points $(2, 0)$ and $(0, 2)$). To show this it is sufficient to note that the set $\bar{P}(\bar{A}) \cup \{\bar{0}\}$ can be mapped into itself by the transformation $t = (t_1, t_2)$ given by

$$t_1(x_1, x_2) = 2x_1/(1 + x_1) \qquad \text{for } 0 \leq x_1 \leq x_2$$
$$= x_1 + x_2 - t_2(x_1, x_2) \qquad \text{for } 0 \leq x_2 < x_1;$$

$$t_2(x_1, x_2) = x_1 + x_2 - t_1(x_1, x_2) \qquad \text{for } 0 \leq x_1 \leq x_2$$
$$= 2x_2/(1 + x_2) \qquad \text{for } 0 \leq x_2 < x_1.$$

This transformation t defined on the set $\bar{P}(\bar{A}) \cup \{\bar{0}\}$ satisfies the conditions of Property 4' and leaves only the points $\bar{0}$, $(2, 0)$, $(1, 1)$, and $(0, 2)$ fixed. But, since t transforms the game $(\bar{P}(\bar{A}), \bar{0})$ into itself, Property 4' requires that $f(\bar{P}(\bar{A}), \bar{0}) = t(f(\bar{P}(\bar{A}), \bar{0}))$; that is, $f(\bar{P}(\bar{A}), \bar{0})$ must be a fixed point of t. The unique fixed point of t in $\bar{P}(\bar{A})$ that is strongly individually rational is the point $(1, 1)$, and so $f(\bar{P}(\bar{A}), \bar{0}) = (1, 1)$ as required.

Next, observe that it will be sufficient for our proof to show that f and E coincide on games $(\bar{P}(S_0{}^+), \bar{0})$ such that $\bar{P}(S)$ is a subset of $\bar{P}(\bar{A})$. To see this, consider an arbitrary game of the form $(\bar{P}(S_0{}^+), \bar{0})$, and let $(\bar{P}(T), \bar{0})$ be a game such that $\bar{P}(T)$ contains $\bar{P}(S)$, and $\bar{P}(T) = \{(x_1 \phi(x_1)) | 0 \leq x_1 \leq \bar{x}_1\}$, where ϕ is a continuous decreasing function such that $\phi(\bar{x}_1) = 0$. (Thus $\bar{P}(T)$ touches both axes, i.e., it contains points of the form $[0, \bar{x}_2]$ and $[\bar{x}_1, 0]$.) Then there is a (unique) point x^* in $\bar{P}(T)$ that gives the players equal gains; let $x^* = (c, c)$,

where c is a positive real number. The transformation $t = (t_1, t_2)$ given by

$$t_1(x_1, x_2) = x_1 \qquad \text{for} \quad 0 \leq x_1 \leq c,$$
$$\qquad\qquad = 2c - x_2 \quad \text{otherwise};$$
$$t_2(x_1, x_2) = 2c - x_1 \quad \text{for} \quad 0 \leq x_1 \leq c,$$
$$\qquad\qquad = x_2 \qquad \text{otherwise};$$

transforms the game $(\bar{P}(T), \bar{0})$ into the game $(c\bar{P}(\bar{A}), \bar{0})$. So the transformation $t' = t/c$ transforms $(\bar{P}(T), \bar{0})$ into $(\bar{P}(\bar{A}), \bar{0})$, and thus transforms the arbitrary game $(\bar{P}(S_0^+), \bar{0})$ into a game whose strong Pareto set is a subset of $\bar{P}(\bar{A})$. Property 4' thus assures that if f and E coincide on subsets of $\bar{P}(\bar{A})$ then they coincide everywhere.

But if $\bar{P}(S)$ is a subset of $\bar{P}(\bar{A})$ that contains the point $f(\bar{P}(\bar{A}), \bar{0}) = (1, 1)$, then independence of irrelevant alternatives implies $f(\bar{P}(S), \bar{0}) = (1, 1)$, as well. If $(1, 1)$ is not an element of $\bar{P}(S)$, then $\bar{P}(S)$ is contained either in a line segment joining $(0, 2)$ to $x = (\bar{x}_1, x_2)$, such that x maximizes Player 1's payoff in $\bar{P}(S)$ and $\bar{x}_1 < x_2$, or else in a line segment joining $(2, 0)$ to the point $x = (x_1, \bar{x}_2)$ that maximizes Player 2's payoff, and for which $\bar{x}_2 < x_1$.

Either of these line segments can be transformed into itself leaving only its endpoints fixed, so that Property 4' together with independence of irrelevant alternatives and strong individual rationality implies that $f(\bar{P}(S), \bar{0}) = x = E(\bar{P}(S), \bar{0})$, which completes the proof.