AN ECONOMIC APPROACH TO THE STUDY OF BARGAINING

Alvin E. Roth

INTRODUCTION

Much of game theory, which has become a central part of economic theory, deals with what we might call bargaining or negotiation. In ordinary usage these two terms are interchangeable, but in the economics literature it has become the predominant practice to reserve the term "bargaining" for bargaining between two parties, or bargaining between many parties in cases in which unanimous agreement is required to change the status quo. This distinction recognizes the increased complexity of negotiations between more than two parties when there is a possibility for subcoalitions to form.

In order to keep this paper to a manageable size, and because coalition formation is the topic of another paper in this volume, I concentrate here on two-party bargaining, in which two agents may allocate some valuable resource between themselves in any way they like, provided they both agree. If they fail to agree on how to allocate the resource, they each receive nothing. This is an example of what is sometimes called "pure" bargaining. This relatively simple kind of bargaining is representative of a reasonably rich class of observable bargaining situations, although in many cases of interest to economists (e.g., when the two parties to the bargaining may themselves be

Research on Negotiation in Organizations, Volume 3.
Handbook of Negotiation Research, pages 35-67.

ISBN: 1-55938-249-X.

composed of individuals whose interests do not completely coincide), it remains to assess whether the approximations necessary to describe them as pure bargaining games are tolerable for the purposes at hand. But even apart from the direct empirical interest in pure bargaining situations, an additional reason why the pure bargaining problem has been a subject of interest to economists is that it can be thought of as the opposite of the idealized case of "perfect" competition: whereas (infinitely many, infinitesimally small) individual agents have negligible influence under perfect competition, in a pure bargaining game each agent has an absolute veto over every division of the resource between the two bargainers.

The particular theme of this paper is the manner in which economists have started to combine game-theoretic modelling with laboratory experimentation to advance the study of bargaining behavior. Loosely speaking, the experimental results are that the traditional game-theoretic theories are terrible point-predictors, but that some of their qualitative predictions are supported by the evidence, and that much of the observed behavior shows regularities of a kind that game theory ought to be able to accommodate.

When I present results such as these to audiences of economists, I tend to emphasize how the relatively poor performance of these theories reflects the paucity of information we allow to play a role in the theories: this is a point that is often surprising to economists, because there are respects in which these theories are informationally quite demanding. However, for the audience of noneconomists to whom this present paper is directed, it will probably be more useful to emphasize the virtues of tractable formal models which concentrate on specific aspects of a problem, while explicitly excluding other aspects from consideration.[1] Not least of these is that, combined with careful experimentation, such models provide a disciplined way to make incremental progress in understanding what remains a complex and little understood subject. Often even the effects of factors excluded from the formal models can be most clearly identified by examining data which differ from the predictions of those models.

Given that the aim of this paper is to convey something of the *method* by which these tools permit us to investigate questions about bargaining, I will depart from the usual model of an introductory survey, which aims to acquaint its readers briefly with a broad range of conclusions. Instead, I will concentrate in more detail on several particular lines of investigation, which will better allow me to illustrate the cumulative nature of this kind of research, while still introducing some of the main issues of concern in the economics literature.[2]

To this end, the remainder of this section briefly introduces the two main lines of game-theoretic bargaining theory, cooperative and strategic, both of which have their roots in the work of John Nash. I first discuss a few experiments in a series motivated by the cooperative theory, and will then consider a recent series of experiments motivated by a part of the strategic theory that has been the object of considerable interest. Both kinds of theory are discussed in their simplest form, which is under the assumption that the bargainers have "complete information" about one another and about the bargaining situation. The penultimate section considers some directions of investigation suggested by the experimental results and, in particular, considers how some of the observed behavior that appears anomalous from the point of view of the theories that assume complete information might be accounted for in the more complex and realistic models of incomplete information. The models of incomplete information we employ in game theory date from the work of John Harsanyi, and in recent years important insights about bargaining have been obtained through the kind of analysis he proposed. Notable in this regard is the work of Myerson and Satterthwaite, concerning the prevalence of disagreements in bargaining with incomplete information.

The Historical Background

Theories of bargaining that depend on purely ordinal descriptions of bargainers' preferences tend to predict large sets of outcomes, and for this reason many economists (at least since Edgeworth, 1881) have argued that bargaining is fundamentally indeterminate. In the language of cooperative game theory, the problem is that the "core" of a pure bargaining game is the large set of outcomes corresponding to the entire set of agreements that leave no part of the resource unallocated. And this same set of outcomes can be achieved as the equilibria of the appropriately defined game in strategic form. The predictions that observed outcomes will be in the core, or will be equilibria, are therefore fairly weak (although certainly not empty, since in some circumstances they are false). Thus, there has been considerable sustained interest in developing theories that attempt to predict specific outcomes in the core, or specific equilibria. Such theories typically attempt to make use of information concerning some measure of the *intensity* of agents' preferences, and most often do so by using the kind of information contained in a von Neumann-Morgenstern expected utility function representing those preferences.

The most influential single theory of this sort is one due to Nash (1950) whose work led to the development of a family of related theories in the tradition of cooperative game theory, which have become known as axiomatic theories.[3] Nash (1953) also initiated the complementary line of work, in the tradition of noncooperative game theory, by which bargaining is studied in terms of the detailed rules through which offers are made and agreements reached. More recently, the body of work in this latter tradition has grown markedly. Some of this work combines information about the mechanics of bargaining with information about time preferences rather than risk preferences of the bargainers, in a manner exemplified by the work of Stahl (1972) and Rubinstein

(1982). This work makes essential use of the concept of a "subgame perfect equilibrium" due to Selten (see Selten, 1975), which will be discussed later.[4] Although there has been an increasing convergence between the theoretical literature concerned with strategic and cooperative models of bargaining (see, e.g., Roth, 1989), the bargaining environments for which their predictions can be most clearly derived are rather different.

EXPERIMENTAL TESTS OF AXIOMATIC MODELS

The models discussed in this section are formulated in terms of the expected utilities of the bargainers. As this presents some critical problems in experimental design, it will be helpful to begin by reviewing the essential elements of expected utility theory.

Expected Utility Theory

von Neumann and Morgenstern introduced, in their seminal 1944 book, not only the outlines of a theory of interactive behavior, but also a model of goal-oriented, "rational" behavior that has become the dominant model of individual choice behavior in economics. They modeled individual choice by means of a binary preference relation defined on the set of alternatives, and established conditions on preferences that, if satisfied, implied that the corresponding choice behavior could be viewed as the result of maximizing an *expected utility* function. That is, a utility function represents the preferences by assigning every alternative α a number $u(\alpha)$, with $u(\alpha)$ being greater than $u(\beta)$ if and only if alternative α is preferred to alternative β. Lotteries between alternatives—that is, probability distributions over alternatives—are themselves alternatives over which preferences are defined, and von Neumann and Morgenstern showed how utility representations could be constructed so that the utility of a lottery was equal to the expected value of the utility of the outcome of the lottery. That is, if p is a probability, and $\mathrm{L} = [p\alpha;(1-p)\beta]$ is the lottery that yields the alternative α with probability p and the alternative β with probability $(1-p)$, then $u(\mathrm{L}) = pu(\alpha) + (1-p)u(\beta)$ is the utility of participating in the lottery L.

For preferences that obey the regularity conditions they proposed, their method of construction involves scaling the utility of any alternative in terms of an arbitrarily chosen origin and unit. Consider any two alternatives α and β such that α is preferred to β, and set the utilities $u(\alpha) = 1$ and $u(\beta) = 0$. Now consider an alternative γ such that α is preferred to γ which is in turn preferred to β. Then finding the utility $\mathrm{u}(\gamma)$ consists of finding the probability p such that the preferences are indifferent between γ and the lottery $\mathrm{L}(\gamma) = [p\alpha;(1-p)\beta]$, so that $u(\gamma) = u(\mathrm{L}(\gamma)) = p$. A utility function constructed in this way conveys not only an individual's preferences among nonrisky alternatives, but also his willingness to undertake risky ventures. This latter property has come to be called the individual's "risk posture."

Bargaining Models and Experimental Design

For various reasons, it was convenient to represent the feasible outcomes of multiperson decision problems—"games"—as numerical outcomes representing the utilities of the players. Nash (1950) followed in this tradition when he modelled a pure bargaining problem by a pair (S,d), where S is a subset of the plane, and d a point in S. The set S represents the feasible utility payoffs to the bargainers—that is, each point $x = (x_1,x_2)$ in S corresponds to the utility payoffs to players 1 and 2 from some alternative α in the set of feasible alternatives A, and $d = (d_1,d_2)$ corresponds to the utility payoffs to the players from the disagreement alternative δ. (Recall that the rules of the game are that any alternative in A—and hence any utility payoff in S—will be the outcome if both bargainers agree, but otherwise the outcome will be δ, with utility payoffs given by d.)

Nash proposed to model the bargaining process by a function f that would associate with each pair (S,d) a point in S. That is, for each bargaining problem represented in terms of the utilities of the bargainers, such a function f would predict what agreement would be reached, also in terms of the utilities of the bargainers. In fact, Nash characterized a particular function f as the unique such function possessing certain properties (axioms) that he proposed. However, for our purposes here, it will be sufficient to note that any of this class of functions constitutes a theory of bargaining that takes as its data the set (S,d). That is, such a function f embodies a theory of bargaining that predicts that the outcome of bargaining will be determined by the preferences of the bargainers over the set of feasible alternatives, together with their willingness to tolerate risk.[6]

Because of the difficulty of attempting to capture the information contained in bargainers' expected utility functions, there were some claims in the experimental literature that the theory was essentially untestable.[7] To get around this difficulty, the earliest experiments designed to test Nash's theory assumed, for the purpose of making predictions about the outcome, that the utility of each bargainer was equal to his monetary payoff.[8] That is, they assumed that the preferences of all bargainers were identical and risk neutral. Important aspects of the predictions of the theory obtained in this way were inconsistent with the experimental evidence. This disconfirming evidence, however, was almost uniformly discounted by game theorists, who felt that the results simply reflected the failure to measure the relevant parameters. Nash's theory, after all, is one that predicts that the preferences and risk aversion of the bargainers exercise a decisive influence on the outcome of bargaining (and, furthermore, that these are the only personal attributes that can influence the outcome when bargainers are adequately informed). If the predictions made by Nash's theory *under the assumption* that bargainers had identical risk neutral preferences were disconfirmed, this merely cast doubt on the assumption. The theory itself had yet to be tested.

It was therefore clear that, in order to provide a test of the theory that would withstand the scrutiny of theorists, an experiment would have to either measure or control for the expected utility of the bargainers.

A class of games that control for the bargainers' utilities was introduced in the experiment of Roth and Malouf (1979). In these *binary lottery games*, each agent i can eventually win only one of two monetary prizes, a large prize λ_i or a small prize σ_i (with $\lambda_i > \sigma_i$). The players bargain over the distribution of "lottery tickets" that determine the probability of receiving the large prize: for example, an agent i who receives 40% of the lottery tickets has a 40% chance of receiving λ_i and a 60% chance of receiving σ_i. Players who do not reach agreement in the allotted time each receive σ_i. Because the information about preferences conveyed by an expected utility function is meaningfully represented only up to the arbitrary choice of origin and scale (and because Nash's theory of bargaining is explicitly constructed to be independent of such choices), there is no loss of generality in normalizing each agent's utility so that $u_i(\lambda_i) = 1$ and $u_i(\sigma_i) = 0$. The utility of agent i for any agreement is then precisely equal to his probability of receiving the amount λ_i, meaning, equal to the percentage of lottery tickets he has received. Thus in a binary lottery game, the pair (S,d) that determines the prediction of Nash's theory is precisely equal to the set of feasible divisions of the lottery tickets.

Let me pause for a moment here to emphasize the role that utility theory plays in interpreting experiments that employ binary lottery games. No assumptions have been made here about the behavior of the experimental *subjects* in binary lottery games. (That is, the subjects might not be utility maximizers, or they might have preferences over distributions of payoffs to both players, rather than over their own monetary payoffs). What binary lottery games do allow us to know is the utility of utility maximizers who are concerned with their own payoffs. Because this is the kind of data required by Nash's theory, experiments using binary lottery games allow us to use the theory to make precise predictions. Thus, we use binary lottery games not to control how the subjects behave, but to control what the theory predicts. It is this that was missing from earlier experiments, and from efforts to analyze field data by inferring ex post what the utility of the bargainers might have been.[9]

Some Experiments

The set of feasible utility payoffs to the players of a binary lottery game is insensitive to the magnitudes of λ_i and σ_i for each agent i (because it equals the set of feasible divisions of lottery tickets). Furthermore, the bargainers have what the game theory literature calls "complete" information whether or not they know the value of one another's prizes, since knowing a bargainer's probability of winning his prize is equivalent to knowing his utility. Thus a theory of bargaining under conditions of complete information, that depends only on the utility payoffs to the bargainers, predicts that the outcome of the game will depend neither on the value of the prizes, nor on whether the bargainers know the value of one another's prizes.

The experiment of Roth and Malouf (1979) was designed in part to test this prediction, and determine whether changes in the size of the prizes, and the bargainers' knowledge of one anothers' prizes would influence the outcome.[10] All games were played by bargainers seated at separated computer terminals that enabled them to send text messages to each other but prevented them from identifying themselves to one another or determining with whom they were bargaining. Each bargainer played games with different prizes against different opponents in one of two information conditions. In the "full information" condition, each bargainer knew both his own prize and that of his counterpart's; bargainers in the "partial information" condition knew only their own prize value. (In each of these games, under both information conditions, the prediction of Nash's theory is that the bargainers would each receive 50% of the lottery tickets.)

The results were that, in the partial information condition, and also in those games of the full information condition in which the two bargainers had equal prizes, observed agreements clustered very tightly around the "equal probability" agreement that gives each bargainer 50% of the lottery tickets. In the full information condition, in those games in which the bargainers' prizes were unequal, agreements tended to cluster around two 'focal points': the equal probability agreement, and the "equal expected value" agreement that gives each bargainer the same expected value. The mean agreement in these games fell approximately half way between the equal-probability and equal expected value agreements. That is, in these games the bargainer with the lower prize tended to receive a higher share of the lottery tickets. Thus, contrary to the prediction of the theory, the monetary values of the bargainers' prizes were clearly observed to influence the agreements reached when the bargainers knew each other's prizes.[11]

The Fine Structure of Information

In the above experiment, either each bargainer knew his opponent's prize or neither bargainer knew his opponent's prize, and each player always knew what information his counterpart possessed in this regard. The experiment of Roth and Murnighan (1982) was conducted to separate the observed effect of information into components that could be attributed to the possession of specific information by specific individuals. Each game of the experiment was a binary lottery game in which one player had a \$20 prize and the other a \$5 prize. In all eight conditions of the experiment, each player knew at least his own prize. The experiment used a 4 (information) × 2 (common knowledge) factorial design (see Table 1). The information conditions were: (1) neither knows his opponent's prize; (2) the *\$20 player knows* both prizes, but the \$5 player knows only his own prize; (3) the *\$5 player knows* both prizes, but the \$20 player knows only his own prize; and (4) *both players know* both prizes. The second factor made this information common knowledge for half the bargaining pairs, but not common knowledge for the other half.[12] For example, when the \$20 player is the only one who knows both prizes, then the (common) instructions to both players in the common knowledge condition reveal that both players are reading the same instructions, and that after the instructions are presented, one player will be informed of only his own prize, and the other will be informed of both prizes. In the not-common knowledge condition, the instructions simply state that each player will be informed of his own prize, and may or may not be informed of the other prize.

The results of this experiment permitted three principal conclusions. First, the equal expected value agreement became a focal point if and only if the player with the smaller prize knew both prizes. When the \$5 player knew that the other player's prize was \$20, this was reflected not only in his messages and proposals, but also in the mean agreements (i.e., mean percentage of lottery tickets obtained by each player) when agreement was reached, and in the shape of the distribution of agreements. In the four conditions in which the \$5 player did not know his opponent's prize, the distribution of agreements had a single mode, corresponding to the 50-50 equal probability agreement. However, in the four conditions in which the \$5 player did know that the other player had a \$20 prize, the distribution of agreements was bimodal, with a second mode corresponding to the 20-80 equal expected value agreement.[13] The mean agreements reached when neither player knew both prizes and when both players knew both prizes replicated the results of Roth and Malouf (1979), both in direction and magnitude.

Second, whether the information possessed by the bargainers is common knowledge or not influences the frequency of disagreement. The frequency of disagreement in the two noncommon knowledge conditions, in which the \$5 player knew both prizes, was significantly higher than in the other conditions. The highest frequency of disagreement (33%) occurred when the \$5 player knew both prizes, the \$20 player did not, but the \$5 player didn't know that the \$20 player didn't know both prizes. (In this situation the \$5 player could not accurately assess whether or not the \$20 player's (honest) skepticism that his opponent's prize was only \$5 was just a bargaining ploy.)

Third, in the noncommon knowledge conditions, the relationship among the outcomes is consistent with the hypothesis that the bargainers are rational

Table 1. Mean Outcomes to the $2D and $5 Players in each formation

	Common Knowledge		*Non-Common Knowledge*	
Information	*$20 Player*	*$5 Player*	*$20 Player*	*$5 Player*
Neither player knows both prizes	41.6_{ab}	43.3_{c}	43.5_{a}	48.2
Only the $20 player knows both prizes	34.9_{bc}	45.1_{bc}	40.9_{a}	42.4
Only the $5 player knows both prizes	27.2_{c}	53.6_{ab}	25.0_{b}	42.0
Both players know both prizes	27.2_{c}	56.4_{a}	25.5_{b}	48.8

Notes: Common Knowledge condition over all interactions (disagreements are included as zero outcomes). Within a column, means with common subscripts are *not* significantly different from one another using the Mann-Whitney *U* test ($a = .01$); none were significantly different in the non-common knowledge conditions for the $5 player.

utility maximizers who correctly assess the tradeoffs involved in the negotiations. That is, in the noncommon knowledge conditions there is a tradeoff between the higher payoffs demanded by the $5 player when he knows both prizes (as reflected in the mean agreements), and the number of agreements actually reached (as reflected in the frequency of disagreement). One could imagine that, when $5 players knew both prizes, they might have tended, as a group, to persist in unrealistic ambitions about how high a percentage of lottery tickets they could expect to get. The mean overall (utility) payoffs (i.e., percentage of lottery tickets) given in Table 1 (which include both agreements and disagreements) indicate that this was not the case. The increase in the number of disagreements just offset the improvement in the terms of agreement when the $5 players knew both prizes, so that the overall expected payoff to the $5 players did not change. This means that the behavior that $5 players were observed to employ in any one of these conditions could not have been profitably substituted for the behavior observed in any other condition.

Consider, for example, a $5 player who knows his opponent's prize is $20, but does not know if his opponent knows both prizes. (In general, $20 players who did know their opponent's prize tried to conceal this knowledge.) Suppose the $5 player thinks it is equally likely that his opponent does or does not know his prize. Then, looking at Table 1 we see that he faces a 50-50 gamble between 48.8 or 42.0 if he acts as if he knows both prizes, and a 50-50 gamble between 42.4 and 48.2 if he acts as if he does not know both prizes. Because the expected values of these two gambles do not significantly differ, the $5 players who knew both prizes could not have profited if they had behaved as if they did not know.

The same is true of the $20 players: In particular, the expected payoff of $20 players who knew both prizes does not differ from that of those who knew only their own prize (although it is significantly affected by what the *$5 player* knows). Therefore, a $20 player who knew both prizes, for example, could not have profited from behaving as he would have if he knew only his own prize. The situation facing $20 players is a little different from that facing $5 players, because $5 players who knew both prizes were virtually always quick to say so (often in their very first message). As such, a $20 player who knew both prizes should not have been in much doubt about whether his opponent knew both prizes also. Looking at Table 1 we see that a $20 player whose opponent knew both prizes had an expected payoff of 25.5 if he knew his opponent's prize and 25.0 if he did not; these payoffs are not significantly different. A $20 player whose opponent knew only his own prize had an expected payoff of 40.9 if he knew his opponent's prize, and 43.5 if he did not; again, these payoffs are not significantly different. Therefore, like a $5 player, a $20 player who knew both prizes could not profit by behaving as he would have if he had known only his own prize.[14]

This kind of tradeoff is exactly what is seen at what game-theorists call strategic equilibrium, in which each player does as well as he can given how the other is behaving, and so the results suggest that these unpredicted effects of information can be modelled with game-theoretic tools. We will return to this question later.

Risk Aversion

The experiments discussed above involved variables that the theories in question predict will not influence the outcome of bargaining. They revealed ways in which the theories systematically fail to be descriptive of observed behavior. As such, the experimental results demonstrate serious shortcomings of the theories. However, in order to fully evaluate a theory, we also need to test the predictions it makes about those variables it predicts *are* important. For theories based on bargainers' expected utilities, risk posture is such a variable.

The predictions of these theories concerning the risk posture of the bargainers were developed in a way that lent itself to experimental test in Roth (1979), Kihlstrom, Roth, and Schmeidler (1981), and Roth and Rothblum (1983). A broad class of apparently quite different models, including all the standard axiomatic models, yield a common prediction regarding risk aversion. Loosely speaking, they all predict that risk aversion is disadvantageous in bargaining, except when the bargaining concerns potential agreements that have a positive probability of yielding an outcome worse than disagreement.

Intuitively, we may understand this prediction as saying that, in most situations, the risk of failing to reach a profitable agreement will cause a highly risk averse bargainer to settle for less favorable terms than he might obtain if he were less risk averse. Thus in most situations you would prefer to bargain with my more risk averse (but otherwise identical) twin than with me. But now suppose that the question at hand is how to divide up the profits from a new business we are thinking of opening by pooling our savings. Although the prospects for a profitable business are good, there is some risk that the business will fail, in which case we will both be worse off than if we had kept our savings in the bank. In this situation you prefer to bargain with me, rather than with my more risk averse twin, since you will have to give him better terms to overcome his reluctance to take the risk.

Three closely related experimental studies exploring the predicted effects of risk aversion on the outcome of bargaining were reported in Murnighan, Roth, and Schoumaker (1988). Whereas binary lottery games were employed in the earlier experiments precisely in order to control for the individual variation due to differences in risk posture, these studies employed *ternary* lottery games having three possible payoffs for each bargainer *i*. These are large and small prizes λ_i and σ_i obtained by lottery when agreement is reached, and a disagreement prize δ_i obtained when no agreement is reached in the allotted time. (In binary lottery games, $\sigma_i = \delta_i$.)

The bargainers' (local) risk postures (for choices involving λ_i, σ_i, and δ_i) were first measured by having them make a set of risky choices. Note that, in contrast to the experiments just discussed, the strategy in this experiment was to measure preferences rather than to control them. Statistically significant differences in risk aversion were found among the population of participants, even on the relatively modest range of prizes available in these studies (in which typical choices involved choosing between receiving $5 for certain or participating in a lottery with prizes of $\lambda_i = \$16$ and $\sigma_i = \$4$).

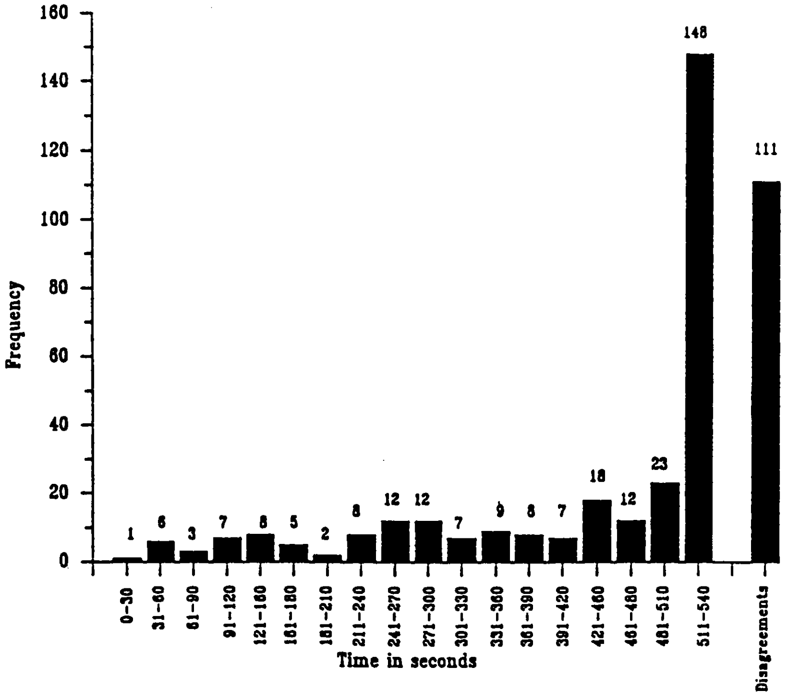
Those bargainers with relatively high-risk aversion bargained against those with relatively low-risk aversion in pairs of games such that the disagreement prizes were larger than the small prizes in one game and smaller than the small prizes in the other. The prediction of game theoretic models such as Nash's is that agreements reached in the first game should be more favorable to the more risk averse of the two bargainers than agreements reached in the second game.

Let me be precise. The theory actually makes a stronger prediction, but only the weaker form is confirmed by the experiments, and the reasons for this illuminate not only the design and analysis of these experiments, but also the many experiments designed to test economic theories. When the prizes of both bargainers are all equal (i.e., $\lambda_1 = \lambda_2 = \lambda$, $\sigma_1 = \sigma_2 = \sigma$, and $\delta_1 = \delta_2 = \delta$) the theories in question predict that the more risk averse player will get more than 50% of the lottery tickets when $\delta > \sigma$, and less than 50% of the lottery tickets when $\delta < \sigma$. Thus the prediction is not only that the more risk averse player should do better in the first game than he does in the second, but also that he should do better than the less risk averse player in the first game, and worse than the less risk averse player in the second.

Now, as had already been established by the earlier experiments, these axiomatic theories fail to predict the effects of the bargainers' information about one anothers' prizes. Among the earlier observations was the very high concentration of (50%, 50%) agreements in games with equal prizes or in which bargainers know only their own prizes, and a shift in the direction of equal expected values in games with unequal prizes known to both bargainers. The strongest form of the predictions about risk aversion concern games in which the bargainers have equal prizes, and so the first experiment of Murnighan, Roth, and Schoumaker (1988) used such a symmetric game. However, a test of the predictions requires data from pairs of agreements between the same subjects, and it was quickly observed that a high percentage of pairs reached (50%, 50%) agreements in the game with $\delta < \sigma$, and ended in disagreement in the game with $\delta > \sigma$. Although there was a weak effect of risk aversion in the predicted direction, it was not significant. One way to read this, of course, is as a rejection of the prediction, but in view of the relatively small scale of the prizes it was thought that any effect of risk aversion might simply be overpowered by the "focal point" effect already observed in connection with the equal probability agreement. It was therefore decided to run a subsequent experiment in which the prizes were unequal, in order to give any effect of risk aversion a wider range on which to be observed.[15] But, as had already been noted, this meant that the player with the smaller prize could be expected to receive the higher percentage of lottery tickets, irrespective of the relative risk aversion of the two bargainers. Consequently, only the weaker form of the risk aversion prediction could be tested on such a game, and it is this prediction that was ultimately confirmed by the data. That is, the results of these experiments support the predictions of the game theoretic models that more risk averse bargainers do better when the disagreement prize is high than when it is low.[16]

The 'Deadline Effect'

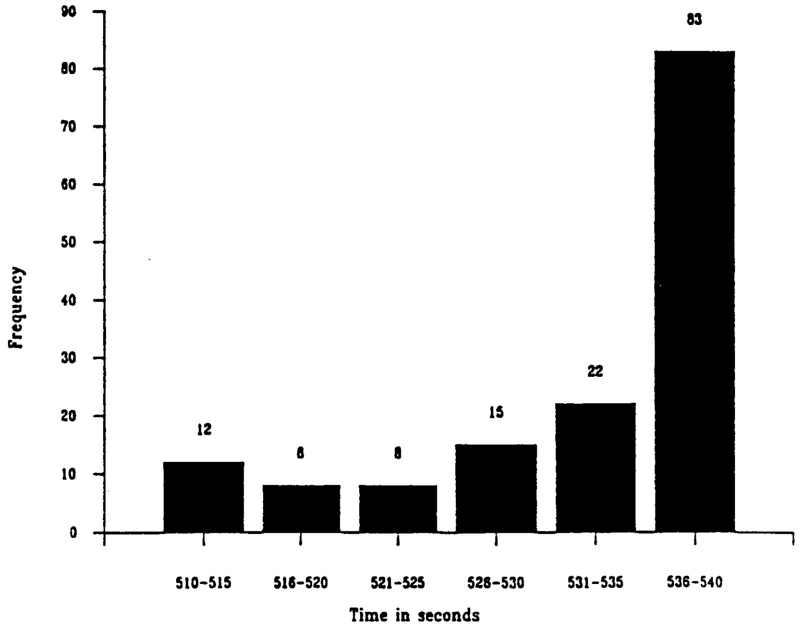
In looking over a series of experiments including those discussed above, two other phenomena stand out. First, there was a non-negligible frequency of disagreements, to which we shall return later. Second, there was a clear "deadline effect."[17] Across all experiments, which varied considerably in the terms and distribution of agreements, the data reveal that a high proportion of agreements were reached in the very final seconds before the deadline. (Figure 1 is representative: 1a shows the distribution of agreements over time in 30 second intervals, while 1b shows the distribution of agreements in the last 30 seconds broken into 5 second intervals.) Four experimental studies considered in Roth, Murnighan, and Schoumaker (1988) all show a substantial number of late agreements, even as late as the last second.



Source: Murninghan, Roth, and Schoumaker (1988).

Figure 1a. The Frequency of Agreements and Disagreements

Because last-minute agreements are widely believed to occur frequently in naturally occurring negotiations,[18] it may be helpful to state clearly just what it is that laboratory investigations have to contribute to the study of deadline phenomena. Although there is a great deal of anecdotal information about the frequency of "eleventh hour" agreements in naturally occurring negotiations, it has proved difficult to collect reliable data. And, being able to study deadline phenomena in the laboratory will allow us to distinguish between alternative hypotheses in a way that the study of field data does not permit. For example, labor negotiators often attribute a tendency to reach agreements just before contracts expire to the difficulty of selling any agreement to a diverse constituency if there is still time for continuing negotiations. However, the deadline effect observed in our laboratory environment cannot be attributed to this because each bargainer is bargaining strictly on his own behalf.



Source: Murninghan, Roth, and Schoumaker (1988).

Figure 1b. The Frequency of Agreements Reached in the Last 30 Seconds of Bargaining

EXPERIMENTAL TESTS OF STRATEGIC MODELS

Recently, a good deal of attention has been given to models of bargaining in which two bargainers, 1 and 2, alternate making offers over how to divide some amount k (of money). Time is divided into periods, and in odd numbered periods t (starting at an initial period $t=1$) player 1 may propose to player 2 any division (x, k-x). If player 2 accepts this proposal then the game ends and player 1 receives a utility of $(\delta_1)^{(t-1)}x$ and player 2 receives a utility $(\delta_2)^{(t-1)}(k—x)$, where δ_i is a number between 0 and 1 reflecting player i's cost of delay. (That is, a payoff of y dollars to player i at period t gives him the same utility as a payoff of δ_i y dollars at period $t—1$.) If player 2 does not accept the offer, and if period t is not the final period of the game, then the game proceeds to period $t+1$, and the roles of the two players are reversed. If an offer made in the last period of the game is.refused, then the game ends with each player receiving 0. A game with a maximum number of periods T will be called a T-period game.[19]

A *strategy* in such a game is a decision rule that tells a player what to do at each point at which he has a decision to make, given the history of the game up to that point. An *equilibrium* is a pair of strategies, one for each player, such that each player's strategy maximizes his payoff, given the other player's strategy. A *subgame perfect equilibrium* (cf. Selten, 1975) is an equilibrium in which the strategies are also an equilibrium in each "subgame" of the original game, that is, in which the strategies maximize each player's payoff given the other player's strategy, *starting from any point that could arise in the game*, not merely from the initial point, and not merely from those points that actually do arise when the players use particular equilibrium strategies.[20]

A subgame perfect equilibrium can, therefore, be computed by working backward from the last period (as long as the payoffs in which the game is represented are utility payoffs; that is, as so long as they adequately summarize what it is that the players are maximizing). An offer made in period T is an ultimatum, and so at such an equilibrium player i (who will receive 0 if he rejects the offer) will accept any non-negative offer when payoffs are continuously divisible.[21] So at a subgame perfect equilibrium, player j, who gets to make the proposal in period T, will receive 100 percent of the amount k to be divided, if the game continues to period T. Consequently, at period $T-1$ player j will refuse any offer of less than $(\delta_j)k$ but accept any offer of more, so that at equilibrium player i receives the share $k-(\delta_j)k$ if the game goes to period $T-1$, and so at period $T-2$ he must be offered $(\delta_i)(k-(\delta_j)k)$, and so forth. Working back to period 1 in this way, we can compute the equilibrium division: that is, the amount that the theory predicts player 1 should offer to player 2 at period 1, and player 2 should accept. (When payoffs are continuous this equilibrium division is unique.) So, when payoffs are continuous, subgame perfect equilibrium in a two-period game calls for player 1 to offer player 2 the amount $\delta_2 k$ in the first period (and demand $k-\delta_2 k$ for himself), while in a three-period game player 1 offers player 2 $\delta_2(k-\delta_1 k)$ in the first period, and demands $k-\delta_2(k-\delta_1 k)$ for himself. For example when $\delta_1 = \delta_2 = .6$, the subgame perfect equilibrium proposal by player 1 is $(.4k, .6k)$ in the two-period game, and $(.76k, .24k)$ in the three-period game.

Recent experimental studies of this kind of bargaining have reported markedly different results. Their authors have drawn quite different conclusions about the predictive value of perfect equilibrium models of bargaining, and about the role that experience, limited foresight, or bargainers' beliefs about fairness might play in explaining their observations. (Questions of fairness arise because in some of these experiments also many observed agreements give both bargainers 50% of the available money.) In reviewing these experiments my aim is to show how, even in the earliest stages of a program of experimental research when there is room for substantial disagreement about what is being observed, early experiments suggest later ones, subsequent experimental results suggest reinterpretations of earlier ones, and the process of experimentation offers the prospect of steadily (if somewhat slowly) narrowing the areas of potential disagreement.

In each of the following experiments, the predictions tested involved only the ordinal utilities of the bargainers, not their risk posture. Following standard practice in the experimental literature when only ordinal utilities are of concern, the utility of the bargainers was assumed to be measured by the amount of money they receive.

Guth, Schmittberger, and Schwarz (1982) examined one-period ("ultimatum") bargaining games. Player 1 could propose dividing a fixed sum of k Deutsche Marks any way he chose, by filling out a form saying "I demand DM x". Player 2 could either accept, in which case player 1 received x and player 2 got $k-x$, or he could reject, in which case each player received 0 for that game.

The perfect equilibrium prediction for such games is that player 1 will ask for and get (essentially) 100% of k. However, the average demand that players 1 were observed to make was for under 70%, both for players playing the game for the first time and for those repeating the game a week later. About 20% of offers were rejected. The authors conclude that " . . . subjects often rely on what they consider a fair or justified result. Furthermore, the ultimatum aspect cannot be completely exploited since subjects do not hesitate to punish if their opponent asks for 'too much'."

Binmore, Shaked, and Sutton (1985) write: "The work of Guth et al. seems to preclude a predictive role for game theory insofar as bargaining behavior is concerned. Our purpose in this note is to report briefly on an experiment that shows that this conclusion is unwarranted . . . ." Their experiment studied a 2-period bargaining game, in which player 1 makes a proposal of the form $(x, 100-x)$ to divide 100 pence. If player 2 accepts, this is the result. Otherwise, player 2 makes a proposal $(x', 25-x')$ to divide 25 pence. If player 1 accepts, this is the result, otherwise each player receives 0. Thus in this game $\delta_1 = \delta_2 = .25$, and (because proposals are constrained to be an integer number of pence) at any subgame perfect equilibrium player 1 makes an opening demand in the range 74-76 pence, and player 2 accepts any opening demand of 74 pence or less. Subjects played a single game, after which player 2 was invited to play the game again, as player 1. In fact, there was no player 2 in this second game, so only the opening demand was observed.

The modal first demand in the first game was 50 pence, and 15% of the first offers were rejected. In the second game (in which only first demands were observed), there was a mode around a first demand near 75 pence. There was thus a clear shift between the two distributions of first demands, in the direction of the equilibrium demand. The authors conclude

> "Our suspicion is that the one-stage ultimatum game is a rather special case, from which it is dangerous to draw general conclusions. In the ultimatum game, the first player might be dissuaded from making an opening demand at, or close to, the 'optimum' level, because his opponent would then incur a negligible cost in making an "irrational" rejection. In the two-stage game, these considerations are postponed to the second stage, and so their impact is attenuated."

Guth and Tietz (1987) responded with an experiment examining two two-stage games with discount factors of .9 and .1 respectively. So the subgame perfect equilibrium predictions (in percentage terms) for the two cases are (10%-90%) and (90%-10%) respectively. They say "Our hypothesis is that the consistency of experimental observations and game theoretic predictions observed by Binmore et al. as well as by Fouraker and Siegel is solely due to the moderate relation of equilibrium payoffs which makes the game theoretic solution socially more acceptable." Subjects played one of the two games twice, each with a randomly chosen other bargainer. Subjects who played the first game as player 1 played the second game as player 2. One difference from the sequential bargaining games discussed above was that disagreement automatically resulted if player 2 rejected an offer from player 1 but made a counterproposal that would give him less than player 1 had offered him.[22]

In the first game, the average first demand in games with a discount factor of .1 was 76%, and in the second game 67%. For games with a discount factor of .9, the average first demand in the first game was 70%, and in the second game 59%. (Recall that when the discount factor is .9, the equilibrium first demand is only 10%.) Guth and Tietz conclude "Our main result is that contrary to Binmore, Shaked and Sutton 'gamesmenship' is clearly rejected, i.e., the game theoretic solution has nearly no predictive power."

Neelin, Sonnenschein, and Spiegel (1988) also responded to Binmore, Shaked, and Sutton (1985). They reported two experiments involving 2-period, 3-period, and 5-period bargaining games. Neelin et. al. observe that the data for all their (2, 3, and 5 period) games are near the perfect equilibrium prediction for 2 period games. They conclude " . . . the strong regularity of the behavior we observed is one of the most noteworthy aspects of our results and lends power to our rejection of both the Stahl/Rubinstein theory and the equal-split model."[23]

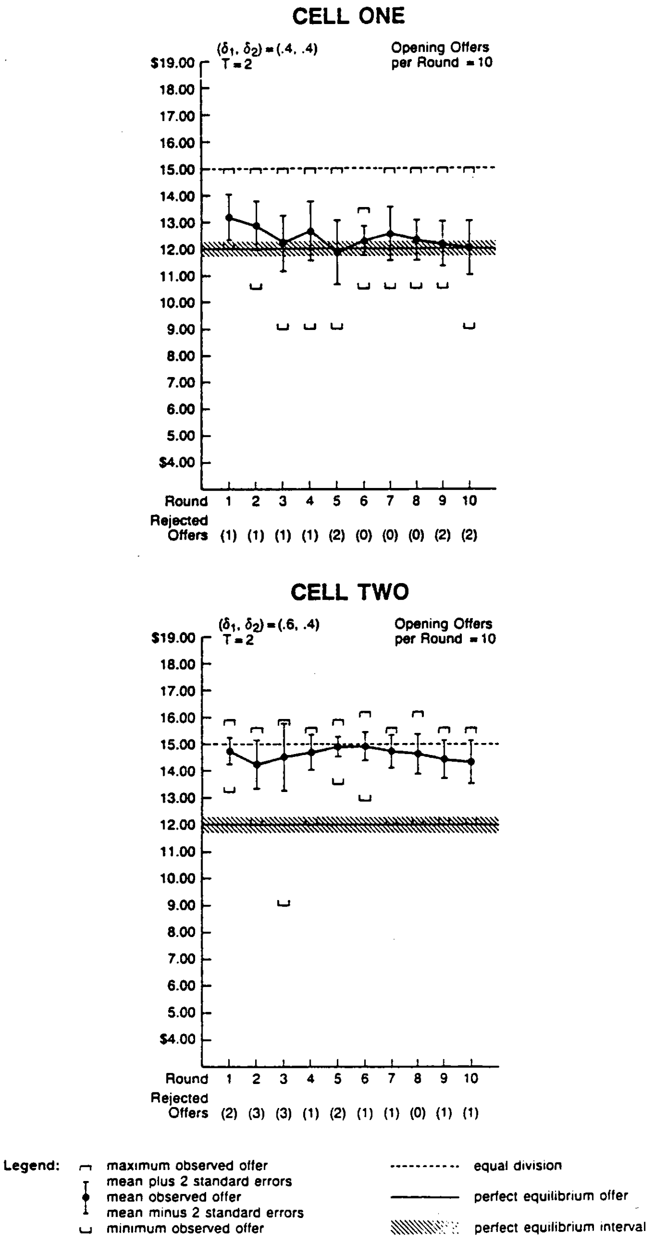
Following most of this exchange, Ochs and Roth (1989) noted that the prior analyses had focused on the accuracy of the perfect equilibrium as a point predictor, that is, on whether the observed outcomes were distributed around the perfect equilibrium division or around some other division of the available money. Their experiment was designed to test the predictive accuracy of some of the *qualitative* predictions of the perfect equilibrium in sequential bargaining, and was designed to detect whether changes in the parameters of the game influence the observed outcomes in the predicted direction, even in the case that there might be a systematic error in the point predictions (recall the discussion of experiments concerned with the effects of risk aversion). To this end, the experiment was implemented in a way that allowed the discount factors δ_1 and δ_2 of the two bargainers to be varied independently.[24] In order to compare games like those considered in the earlier experiments, the experimental design allowed comparisons between different combinations of discount factors for games of fixed length, as well as between games of different length for given discount factors. The eight cells of the experiment constitute a 4×2 design, with the two treatment variables being the discount rates δ_1 and δ_2 (the 4 way variable, with values $(\delta_1,\delta_2) = (.4,.4)$, $(.6,.4)$, $(.6,.6)$ and $(.4,.6)$) and the number of periods T (with values $T = 2,3$). Each subject participated in ten consecutive bargaining games, with the same parameters, against different individuals. In each round (i.e., in each game) the amount available to the bargainers if they reached agreement in the first period was $30.

Figure 2a and 2b (from Ochs & Roth, 1989) display the first-period offers, round by round, for each of the cells of the experiment, along with the number of first-period offers that were rejected. Also displayed for comparison are lines showing the perfect equilibrium prediction for each cell, and the 50-50 offer of $15.

The perfect equilibrium predictions do poorly both as point predictions and in predicting qualitative differences between cells, such as mean first period offers. Although parts of the data appear to be consistent with similar observations made in the earlier experiments, the larger experimental design allows more comparisons to be made, so that observations which, piecewise, appear contradictory, emerge as part of a larger picture.[25] But overall, the data reveal some striking regularities, some of which we can summarize as follows.

1. A consistent first-mover advantage was observed in all the cells of this experiment. (In fact, in most cells a 50% share was the *maximum* ever offered to player 2.)
2. The discount factor of player 1 was observed to influence the outcome even in the two-period games, contrary to the perfect equilibrium prediction.
3. A substantial percentage (16%) of first offers were rejected.
4. The observed mean agreements deviate from the equilibrium predictions in the direction of equal division. (Although the mean offer is virtually always significantly less than equal division.)
5. A substantial percentage of rejected offers were followed by "disadvantageous counterproposals" (which are discussed next).

Briefly, 125 (out of 760) first offers met with rejection,[26] and of these, 101 (81%) were followed by counterproposals in which player 2 demanded *less* cash



Source: Ochs and Roth (1989).

Figure 2a: Opening Offers to Player 2

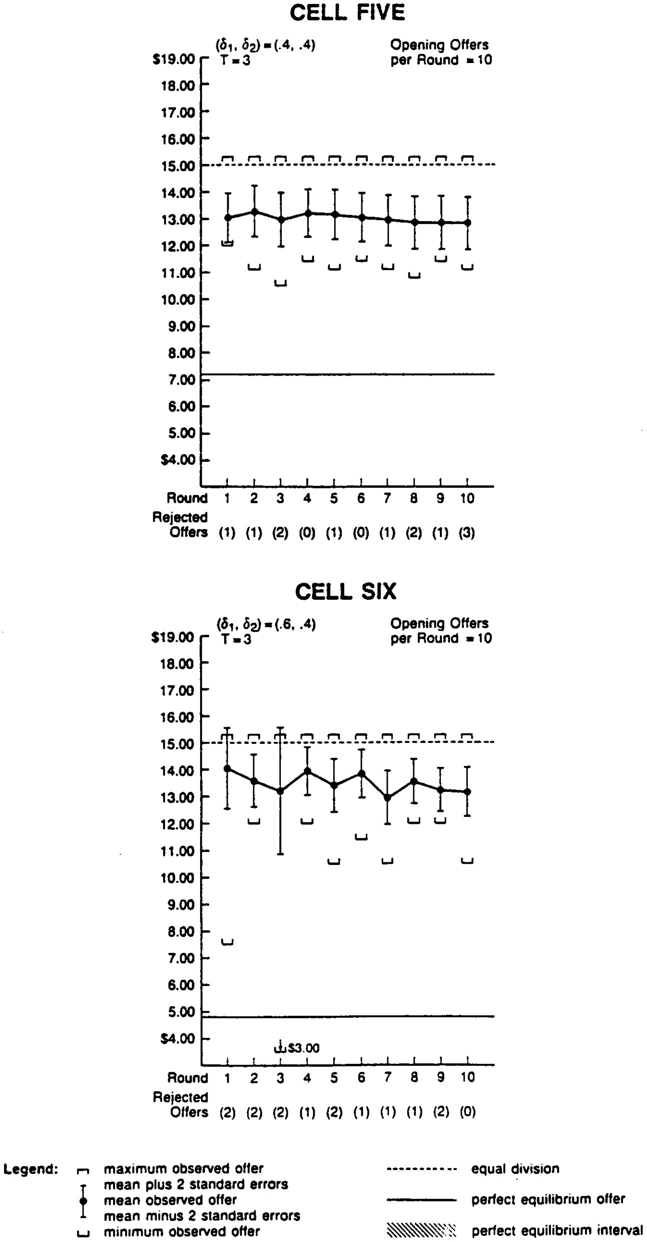


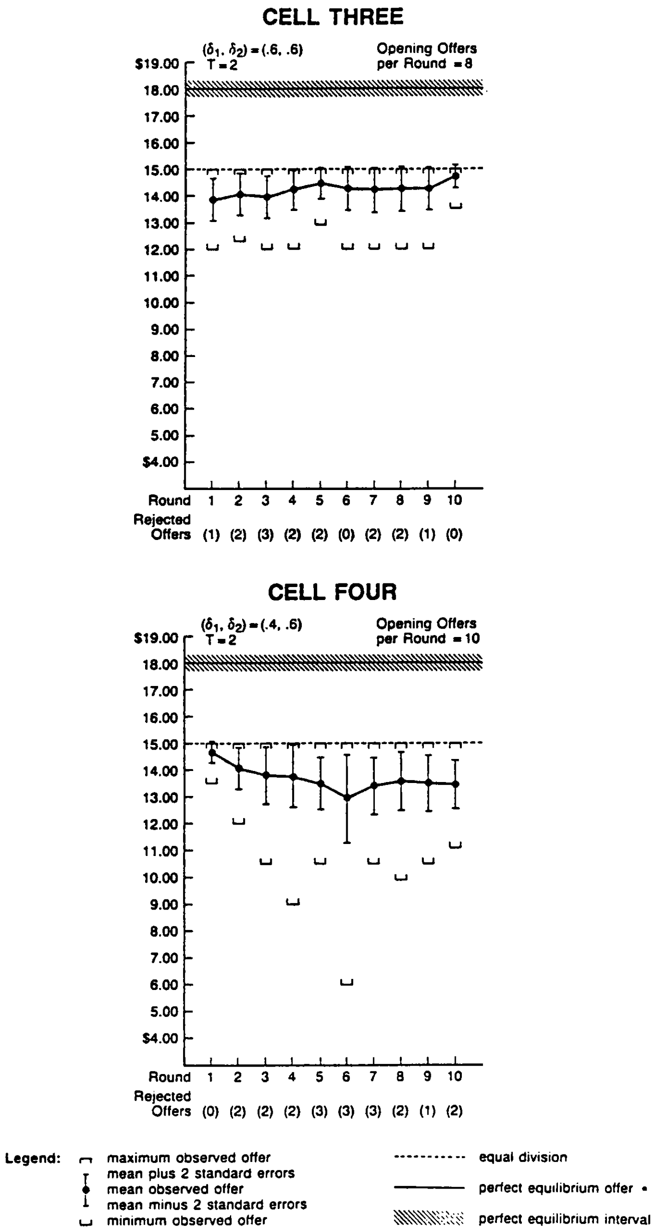
Figure 2a: (Cont'd)

than he had been offered. That is, a significant number of players 2 were rejecting small shares of the relatively large gains available in the first period in favor of larger shares of the much smaller gains available in the second period. Because, after player 1 has made a proposal, player 2 is faced with an individual choice problem, we can conclude by revealed preference that these player 2's utility is *not* measured by their monetary payoff, but must include some nonmonetary component. When the data of the previous experiments were reanalyzed with this in mind, it turned out that this pattern of rejections and counterproposals was strikingly similar in all of these experiments.[27]

Ochs and Roth (1989) go on to argue that this and other patterns in the data can plausibly be explained if the unobserved and uncontrolled components of utility in these experiments have to do with subjects' perceptions of "fairness," which involve comparing their share of the available wealth to that of the other bargainer. They note that in most cases agents propose divisions that give them more than half of the proceeds, and say " . . . we do not conclude that players 'try to be fair.' It is enough to suppose that they try to estimate the utilities of the player they are bargaining with, and . . . at least some agents incorporate distributional considerations in their utility functions." That is, if agents' preferences are such that they will refuse "insultingly low" offers, then this must be taken into account in making offers.

Note how this differs from the interpretation in parts of the psychology literature that the outcome of various kinds of interactions can best be understood as reflecting the common beliefs of the participants about what constitutes a fair outcome. If this were what accounted for the data graphed in Figure 2, we would presumably expect to see the equal split offer (of $15) made quite frequently, but instead we see that equal splits mostly lie outside of the 95% confidence interval (as approximated by the interval of plus or minus two standard errors from the mean). This seems less consistent with the notion that players 1 are 'trying to be fair' than with the explanation that they are trying to get as much as the traffic will bear. (Of course, how much the traffic will bear may depend on player 2's ideas about what constitutes an unfair offer.)[28]

A subsequent experiment by Forsythe, Horowitz, Savin, and Sefton (1988) directly investigates these alternative hypotheses about the role of fairness, by comparing one-period ultimatum games of the kind described above, with "dictator games" in which player 1 simply decides on the division between the two players (i.e., in dictator games player 2 is passive, he need not accept the offer and cannot reject it). The basis for the comparison rests on two observations. First, if we assume that the players are simple income maximizers, then the perfect equilibrium prediction for both games is that player 1 will get essentially 100% of the funds to be divided. Second, if we assume the player 1's are simply 'trying to be fair' then the outcomes for both games should also be the same. In any event, because players in position 1 of the dictator game



Source: Ochs and Roth (1989).

Figure 2b: Opening Offers to Player 2

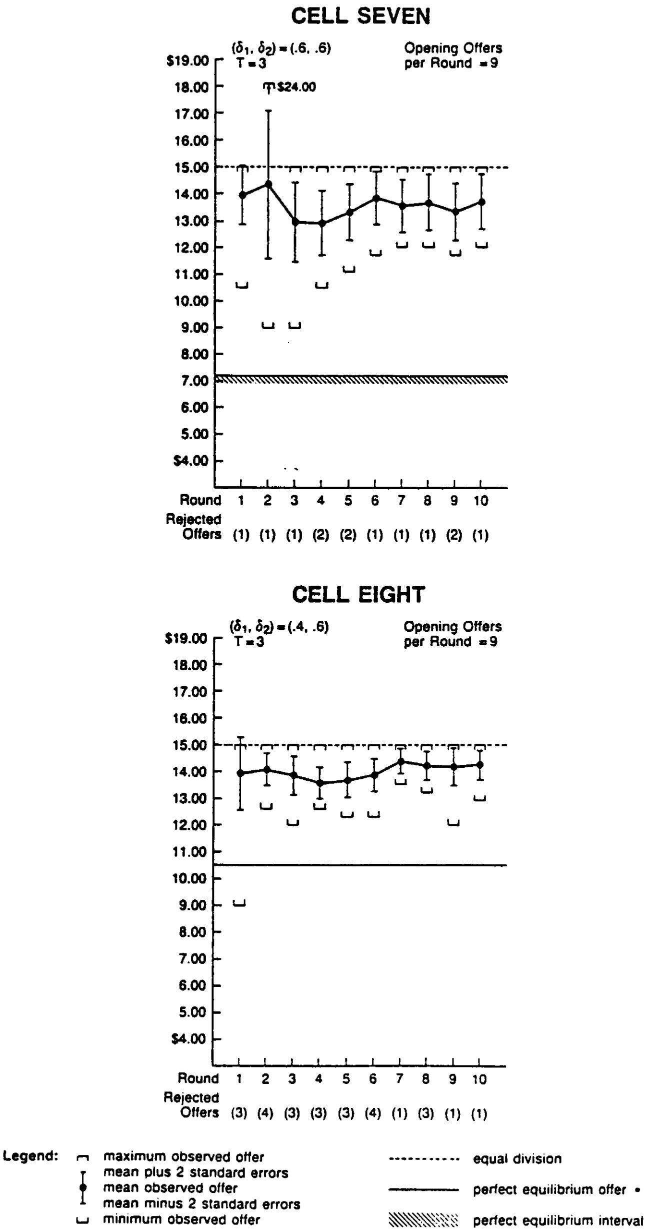


Figure 2b: (Cont'd)

are faced with an individual choice problem, the divisions observed in that game will reflect the pure preference of the players for divisions; in the ultimatum game, any differences from the divisions proposed in the dictator game can be attributed to the strategic calculations needed to assess 'what the traffic will bear.'

Forsythe et al. observed important differences between the two games. In the ultimatum game, the overwhelming majority of proposals were for 50-50 divisions, and there were no proposals for 100%-0% divisions. But in the dictator games, a substantial proportion (36%) of players 1 decided on 100%-0% divisions. These results thus provide strong support for the hypothesis that the outcomes of the bargaining games do not merely reflect the preference of the proposer for equal divisions, but rather reflect his assessment of the risk of receiving nothing if his proposal is rejected.

Note finally that uncontrolled elements in the bargainers' utility in these experiments suggests that none of them can be easily interpreted as tests of perfect equilibrium per se, because to compute a perfect equilibrium we need to know the preferences of the players (and so do they).[29] However, the uniformity with which 'disadvantageous counterproposals' have appeared in the experiments to date, in contrast to their otherwise quite varied results, suggests that bargaining may be an activity that systematically gives bargainers motivations distinct from simple income maximization. One natural direction in which to continue this (still young) series of experiments is to attempt to directly observe or manipulate these so far uncontrolled motivating factors.[30]

SOME DIRECTIONS FOR FURTHER WORK

This section briefly considers how some of the patterns that begin to emerge from the data may be consistent with more complex game—theoretic models, and how these will provide testable predictions that suggest new experiments. To begin, consider one of the regularities observed in all of the experiments discussed above; namely, that there is a significant proportion of disagreements. This was the case both in the relatively free-form bargaining observed in the experiments motivated by axiomatic models, and in the highly structured alternating-offer bargaining motivated by strategic theories. Because of the simplicity of the bargaining in each case, the frequency of agreements cannot be explained as resulting from the failure of the bargainers to recognize that joint gains were possible.

However, in spite of their simplicity, we may not be able to assume that these experimental environments constitute games of complete information. This was particularly clear in the tests of strategic models, where the evidence suggests that the bargainers incorporate distributional concerns into their utility functions. The argument there focused on the point that, in the experiments conducted to date, the bargainers' utility functions were only indirectly observed by the experimenters. But, if bargainers have unobserved components in their utility functions, then each bargaining encounter will be a game of *incomplete* information, because the bargainers themselves will be uncertain about one anothers' preferences. This, it turns out, suggests a reason for the nonnegligible frequency of disagreements.

Whereas the complete information literature suggests that all bargaining outcomes will be efficient (i.e., no disagreements will occur when joint profits are possible), various forms of inefficiency emerge as equilibrium behavior in incomplete information bargaining models (see Chatterjee, 1985, for a survey[31]). In single period models, this inefficiency takes the form of a positive probability of disagreement at equilibrium (see, e.g., Myerson & Satterthwaite, 1983). In multiperiod models in which time is discounted, it takes the form of delays in the time at which agreement is reached (see, e.g., Cramton, 1985). In multiperiod models in which there is some probability that each period will be the last, it also takes the form of delays, which now also imply a positive probability of disagreement, because there is a positive probability that the bargaining will terminate before agreement is reached.

Perhaps the seminal result concerning disagreements is the result of Myerson and Satterthwaite (1983) that, in a wide class of games of incomplete information, a positive probability of inefficiency is an *inevitable* consequence of equilibrium behavior. That is, when each bargainer is doing as well as he can in an expected utility sense, given how the other bargainer is behaving, then sometimes there will be disagreements even when there would have been gains from trade, or costly delays prior to reaching agreement.[32] Loosely speaking, the intuition is that if you reach agreement whenever an agreement is possible, you must be settling for too little.

It is easy to see how this might apply to the sequential bargaining games discussed earlier. For example, if each player 2 has some unknown threshold, such that he will reject offers that are below this threshold, then player 1, even if he is simply trying to maximize his own payoff, is left with a decision problem under uncertainty. If there is even a little variance among the population of potential player 2's, then unless player 1 is extremely risk averse, the optimal solution to his decision problem will be to make an offer which has some positive probability of being rejected.

Although game-theoretic models admit other causes of disagreement aside from incomplete information,[33] the incomplete information literature offers some related predictions about the patterns of behavior that may be associated with disagreements. For example, the "deadline effect" discussed earlier is powerfully suggestive of what are called "separating equilibria."

To get some intuition about separating equilibria, consider a model of bargaining over time in which (at least) one of the bargainers has some private information about something germane to the outcome of bargaining.[34]

Suppose for the moment that the information is binary, and indicates that the informed bargainer is either in a "strong" or "weak" position such that if the information were to become common knowledge, the informed bargainer would obtain a more favorable agreement if he is in the strong position. Even though only one bargainer is informed about whether his (own) position is strong or weak, there may still be equilibria at which a bargainer in a strong position will obtain a more favorable agreement than one in a weak position. These equilibria are called *separating equilibria* (because they separate strong from weak bargainers). At such an equilibrium there must be a cost to a bargainer in a strong position that he would be unwilling to pay if he were in a weak position. (Otherwise a bargainer in a weak position would simply do whatever the equilibrium would have called for him to do had he been strong.) In models in which delay is costly, the cost to the strong bargainer of demanding the more favorable agreements is that these are reached later than agreements reached by weak bargainers. When the costliness of delay arises from a positive probability that each period will be the last, this means that at a separating equilibrium strong bargainers face a higher percentage of disagreements than weak bargainers. Therefore, even in the relatively unstructured bargaining environments in which the deadline effect was observed in Roth et al. (1988), the high percentage of disagreements (recall Figure 1a) suggest that bargainers were waiting "until it hurt," that is, until the probability of disagreement was very real. The separating equilibrium hypothesis is that they were doing so because only in the last moments would the wolves be separated from the sheep.[35]

CONCLUDING REMARKS

In closing, the point I would like to emphasize is the usefulness of models that make precise (and therefore falsifiable) predictions, in combination with empirical investigations that can be tailored to precisely test those predictions.

In this regard, I have concentrated on experiments motivated by game-theoretic models of complete information. We saw that one of the principle predictions (or assumptions) of these models is clearly falsified,[36] because information other than utility information was seen to influence the outcome. In view of the fact that the bargainers can draw on commonly held notions of fairness in bolstering the credibility of their positions (an interpretation supported by the evidence of the experiments concerned with strategic models), this may not be surprising, but notice how the evidence speaks clearly against the idea advanced in some quarters that the bargainers simply "try to be fair." On the contrary, it seems clear that in the experiments considered here, notions of fairness are used to a large extent in a strategic and self-interested way (see note 28).[37] So even when we reject the predictions of a well specified model, we can sometimes suggest clear dimensions in which the model is lacking, and in which further exploration seems likely to be fruitful.

At the same time, we have seen that some of the most subtle predictions of these models, namely those concerning the influence of bargainers' risk aversion on the outcome of bargaining, are largely supported by the data. It is difficult to think how these predictions could have been made without the use of formal models.

Finally, if we are prepared to incorporate (as I think we must be) some of the empirical observations concerning, for example, the strategic uses of fairness into more complex models (such as the models of incomplete information), which I have argued seem promising, then the data so far gathered appear once more to be broadly consistent with the predictions of such models. So the further predictions of these models will suggest further experiments, and further modifications of the theory. They thus give us one way to proceed in unravelling some of the puzzles that bargaining behavior presents.

NOTES

[1] The fact that the economics profession may invest disproportionately in this kind of work does not diminish its real virtues.

[2] In concentrating on just a few lines of investigation, which combine bargaining theory and experiments in economics, I will perforce also devote more space to work in which I have personally been involved than would be seemly in a broader survey. I hope I will be forgiven.

[3] A notable contribution to this literature is that of Kalai and Smorodinsky (1975); see Roth (1979) for a full account of this literature.

[4] Attention in what follows will be focused on one-time bargaining situations, rather than repeated interactions between the same or different parties. The more complex case has also been of great interest to economists: see, for example, Wilson (1985) for a survey of some of the work that has been done concerning how reputations are built and maintained.

[5] Quite early on (cf. Allais, 1953), experiments by economists and others revealed that individuals' choice behavior could be shown to systematically deviate in a variety of ways from the assumptions underlying utility theory, which can therefore only be regarded as an approximation of observable individual choice behavior (see, e.g., Machina, 1987, for a discussion of some contemporary alternative approximations). However, expected utility theory remains a fruitful source of insights into and predictions about economic phenomena. Of course, in view of the fact that individuals are known to deviate from utility maximizing behavior, special care must be taken in designing and interpreting tests of predictions made by economic theories that are stated in terms of utility theory.

[6] This was the traditional assumption in cooperative game theory. Indeed, games in which the players know both the rules of the game and one another's preferences and risk posture are referred to as games of "complete information." The tacit assumption underlying Nash's work (although it plays no part in the mathematics) is that the games he considers are played under conditions of complete information.

[7] For example, Morley and Stephenson (1977) state "these theories . . . do not have any obvious behavioral implications" (p. 86).

8. The experiments of Nydegger and Owen (1975) and Rapoport, Frenkel, and Perner (1977) were designed to test specific assumptions underlying Nash's solution. Related experiments by Rapoport and Perner (1974) and Rapopor, Guyer, and Gordon (1976) used only hypothetical payments. The early and influential study of Siegel and Fouraker (1960) was concerned only incidentally with Nash's solution. These experiments are briefly reviewed in Roth and Malouf (1979).

9. Note, however, that because the payoffs facing each individual consist only of two monetary payoffs and the lotteries between them, many of the systematic deviations from utility maximizing behavior that have been observed on more complex domains cannot be observed here.

10. That this experiment was published in a psychology journal, while the subsequent experiments I will discuss appear in economics journals, is a symbol of how quickly experimental methods have started to establish themselves in the mainstream of modern economics. (That is not to say that experiments are themselves new in economics: the earliest informal account I know of is in Bernoulli [(1738) 1954], in connection with the St. Petersburg paradox.)

11. A subsequent experiment (Roth, Malouf, & Murnighan, 1981) showed that this cannot be simply explained as an artifact of the experimental procedures.

12. A piece of information is common knowledge between us if not only do we both know it, but I know you know it and you know I know it, and I know you know I know it, and so on. Knowledge of an event can be thought of as becoming common knowledge when the event occurs in public, so not only do we each see it, but we each see each other seeing it, and so on. The notion of common knowledge, which is now common in economic theory, seems to have been first formally considered by the philosopher David Lewis (1969) in his treatment of social conventions.

13. A subsequent replication of the common knowledge condition in which both players know both prizes, conducted with a variety of prizes, reproduced the shift in mean agreements, but not the bimodality of the distribution of agreements. Roth, Murnighan, and Schoumaker (1988) includes a brief account.

14. The situation is different in the common knowledge conditions, in which the ability of players to misrepresent what they know is more limited. But the strategies available to the $20 player when it is common knowledge that only he knows both prizes are the same as those availabe to the $5 player when it is common knowledge that he is the only one to know both prizes, and yet the expected payoff to the $20 player in this situation is only 34.9% of the lottery tickets compared to 53.6% for the $5 player in the corresponding situation. To understand this, a more detailed analysis of the record of negotiations was undertaken in Roth and Murnighan (1983). Both the proposals and messages generated by the bargainers were analyzed and it was concluded that $5 players who knew both prizes received a higher mean payoff than any other players in these conditions primarily because they demanded more. No other players made demands of near 80%, as did the informed $5 players. This is not to say that other players did not demand more than their "fair share." However, when informed $20 players knew that the $5 player was uniformed and chose to mispresent their own prize, they did not claim that their own prize was only one quarter of their opponents', and they did not stick to their demand with the tenacity of the informed $5 players.

15. That is, it was thought that the high concentration of agreements around a "focal point" such as (50-50) might reflect forces at work that made it unprofitable for bargainers to try to achieve small deviations from equal division, but that, once the bargaining had shifted away from such a compelling focal point (into a region in which previous experiments had shown agreements would have greater variance), the influence of risk aversion on the precise terms of agreement might be greater.

16. One lesson that can be drawn from all this is that it is possible to design experiments to investigate the qualitative predictions of theories that may already be known not to be good point predictors.

17. In all the experiments discussed above, there was a fixed time limit, typically from 9 to 12 minutes, by which time any agreements must be concluded. Three minutes before the deadline, a clock came on the screen.

18. Some representative quotations:

> Disputes over the terms of collective bargaining agreements are frequently settled only at the last minute. The last-ditch all-night parleys are as familiar to newspaper readers as they are wearing on reporters . . . The frequency of these photo-finishes suggests that something may be involved fundamental to the process of collective bargaining. (Dunlop & Healy, 1955, p. 57).
>
> Few of us, even those least involved in bargaining activities, are unfamiliar with the 'eleventh hour' effect widely publicized in mass media accounts of collective bargaining. (Rubin & Brown, 1975, p. 120).

19. Much of the recent theoretical work using this kind of model follows the treatment by Rubinstein (1982) of the infinite horizon case. An exploration of various aspects of the finite horizon case is given by Stahl (1972). This literature considers the cost of delay in more general form rather than only the discounting discussed here. An experiment motivated by this literature which considers a fixed cost per period of bargaining is Rapoport, Weg, and Felsenthal (1988).

20. For example, a strategy for player 1 in the one-period game (T=1) is simply the offer he makes, while a strategy for player 2 is the rule which associates with each offer he receives either the decision "accept" or "reject." One equilibruim which is not a subgame perfect equilibrium is the strategy pair in which player 2 accepts only offers giving him 80% or more of the money to be divided, and in which player 1 offers player 2 80%. This is an equilibrium since neither player can get a higher payoff by changing his strategy, given the strategy of his opponent. But it is not a subgame perfect equilibrium, since *if* player 1 were to offer player 2 some smaller amount, say 10%, then player 2 would not be maximizing his payoff if he followed his strategy and rejected it, since in this case he would get nothing.

21. If payoffs are discrete, such that offers can only be made to the nearest penny, for example, then there are subgame perfect equilibria at which i refuses to take 0 but accepts the smallest positive offer, for example, one cent.

22. Note that this rule makes the games more like ultimatum games, since some demands of player 1 (e.g., demands of less than 90% in games with discount factor of .1) can only be rejected at the cost of disagreement.

23. In a reply, Binmore, Shaked, and Sutton (1988) decline to attribute the same significance to these results, and conjecture that the various differences described among these experiments may be due to the various differences in experimental procedures employed.

24. Each of the earlier experiments was designed to correspond to the case that the players have equal discount factrs, that is, $\delta_1 = \delta_2 = \delta$, with the costlines of delay implemented by making the amount of money being divided in period t+1 equal to δ times amount available at period t. Because half the cells of the experimental design of Ochs and Roth require different discount rates for the two bargainers, the discounting could not be implemented in this way. Instead, in each period, the commodity to be divided consisted of 100 "chips." In period 1 of each game, each chip was worth $0.30 to each bargainer. In period 2, each chip was worth δ_1 ($0.30) to player 1 and δ_2 ($0.30) to player 2, and in period 3 of the three period games each chip was worth $(\delta_1)^2$ ($0.30) and $(\delta_2)^2$($0.30), respectively. That is, the rate at which subjects were paid for each of the 100 chips that they might receive depended on their discount rate and the period in which agreement was reached.

25. In this regard, the paper notes "... if we had looked only at Cell 1 our conclusions might have been similar to those of Binmore et al., since the data for that cell looks as if after one or two periods of experience, the players settle down to perfect equilibrium proposals ... And if we had looked only at Cells 1 and 5, our conclusion might have been similar to those of Neelin et al., since in those two cells both the two and three period games yield observations near the two period predictions ... And if we had looked only at cells 5 and 6, we might have concluded, like Guth and Teitz, that the phenomena observed here was closely related to the relatively extreme equilibrium predictions in those cells."

26. And the rate did not decline in games with experienced subjects who had played ten games against different opponents.

27. For example, Ochs and Roth (1989) report that on reanalyzing the data from Binmore et al. (1985) and Neelin et al. (1988), they find percentages of first offer rejections of 15% and 14%, and of these rejections, the percentages followed by disadvantageous counteroffers are 75% and 65%, respectively. These figures are quite comparable both to one another and to the corresponding rates of 16% disagreements and 81% disadvantageous counterproposals observed by Ochs and Roth.

28. In general, I think that one of the contributions of the experimental bargaining results is that they reveal that subjects may possess multiple, different notions of fairness, and employ them selectively, for strategic purposes. Thus, for example, the agreements reported in the binary lottery games of Roth and Murnighan (1982) had modes at both the equal probability and the equal expected value agreements, and the transcripts of the bargaining reveal that notions of "fairness" were employed by both bargainers in arguing for their claims. But the bargainer who saw the equal expected value agreement as "fair" was invariably the one with the smaller ($5) prize, while the player with the larger ($20) prize was the champion of the fairness of equal division of the lottery tickets.

29. However, Ochs and Roth (1989) do report consistency across subgames, which could be interpreted as indirect evidence supporting the subgame perfectness hypothesis with repect to the unobserved preferences.

30. And to the extent that these other motivations may reflect some element of bargainers' perceptions of fairness, this may help explain why the results of these various experiments may have been more sensitive than might have been expected to details of the experimental environment. Studies of fairness (based on survey questions) suggest that peoples' ideas about what is "fair" may be both clear and very labile, subject to dramatic change in response to how the issue is presented. (This is particularly clear in the study reported by Yaari and Bar-Hillel [1984]. See also the related work of Kahneman, Knetch, and Thaler, [1986a, 1986b] and of Bazerman [1985] and Farber and Bazerman [1986].)

31. Also see, for example, Chatterjee (1982); Chatterjee and Samuelson (1983); Cramton (1985); Fundenberg and Tirole (1983); Fundenberg, Levine and Tirole (1985); Myerson and Satterwaite (1983); Myerson (1985); Rubinstein (1985a, 1985b); Sobel and Takahashi (1983).

32. The argument of Myerson and Satterthwaite (1983) was phrased (by way of the "revelation principle") in terms of one-period games, in which the only inefficiency that could be observed was disagreement. Ausubel and Deneckere (1988) have recently observed that the same argument can be used directly to show that in multi-period games the same unaviodable kind of ineefficiency may appear as costly delays.

33. For example, Roth (1985) considers the experimental evidence in the light of the disagreements which occur at mixed strategies of coordination games of complete information.

34. This is often taken to be information about the size of the potential profits to be divided, but that is not the only kind of information that could enter a model in this way. For the purpose of thinking about the deadline effect, it might, for example, be useful to think of the private information as concerning the bargainer's own subjective probability distribution over the time at which a final proposal can be made.

35. Nevertheless, there are aspects of the data that appear difficult to reconcile with such a model. The substantial fraction of the agreements observed well before the deadline (when the probability of sudden termination is still zero) may be difficult to explain with this kind of model, particularly since the terms of early agreements do not seem to be distributed differently than late agreements. The idea is that, if bargainers in strong positions generally get better agreements by holding out longer, then late agreements should look different than early agreements. However, the absence of systematic correlations between time of agreement and terms of agreement in the data of Roth et al. (1988) is not too surprising, in view of the fact that the experimental designs make all the observable features of the bargaining common knowledge between the bargainers in all the experiments discussed except that of Marnighan, Roth, and Schoumaker (1987). (And in that experiment, there is no significant correlation between the time of agreement and the terms achieved by the less risk averse bargainer.) Maybe if we could observe players' prior expectations about exactly how much time was left we'd see some correlation, with players who were more relaxed about the deadline doing better in last minute agreements. This is, of course, an idea that can be investigated with an appropriately designed experiment.

36. At least under the interpretation that players are utility maximizers concerned with their own consumption, as operationalized through binary lottery games.

37. I would like to throw out for discussion the idea that some of the "framing" phenomena that we are beginning to understand in individual choice contexts may have to be similarly reinterpreted in strategic contexts. For example, if different "frames" are advantageous for different sides in a bargaining encounter, we may expect to see each side vigorously try to frame the issues in the most favorable way. In this case the influence of framing effects on the outcome of bargaining may be very different than in choice situations in which the frame is specified exogenously.

REFERENCES

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americane." *Econometrica, 21*, 503-546.

Ausubel, L. M., & Deneckere, R. J. (1988). *Stationary seuential equilibria in bargaining with two-sided incomplete information* (Discussion paper no. 784). Center for Mathematical Studies in Economics and Management Science, Northwestern University.

Bazerman, M.H. (1985). Norms of distributive justice in interest arbitration. *Industrial and Labor Relations, 38*, 558-570.

Bernoulli, D. ([1738] 1954). Specimen theoriae novae de mensura sortis. *Econometrica, 22*, 23-36.

Binmore, K., Shaked, A., & Sutton, J. (1985). Testing noncooperative bargaining theory: A preliminary study. *American Economic Review, 75*, 1178-1180.

——— (1988). A further test of noncooperative bargaining theory: Reply. *American Economic Review, 78*, 837-839.

Chatterjee, K. (1982). Incentive compatibility in bargaining under uncertainty. *Quarterly Journal of Economics, 96*, 717-726.

——— (1985). Disagreement in bargaining. In A.E. Roth (Ed.), *Game-theoretic models of bargaining*. Cambridge: Cambridge University Press.

Chatterjee, K. & Samuelson, W. (1983). Bargaining under incomplete information. *Operations Research, 31*, 835-851.

Cramton, P. C. (1985). Sequential bargaining mechanisms. In A. E. Roth (Ed.), *Game-theoretic models of bargaining*. Cambridge: Cambridge University Press.

Dunlop, J.T., & Healy, J. J. (1955). In Richard D. Irwin (Ed.), *Collective Bargaining: Principles and Cases*. Homewood, IL:

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

Farber, H. S., & Bazerman, M. H. (1986). The general basis of arbitrator behavior: An empirical analysis of conventional and final-offer arbitration. *Econometrica, 54*, 1503-1528.

Forsythe, R.. Horowitz, J. L., Savin, N.E., & Sefton, M. (1988). *Replicability, fairness and pay in experiments with simple bargaining games* (Working Paper 88-30). Department of Economics, University of Iowa.

Fudenberg, D., Levine D., & Tirole, J. (1985). Infinite-horizon models of bargaining with one-sided incomplete information. In A. E. Roth (Ed.), *Game-theoretic models of bargaining.* Cambridge: Cambridge University Press.

Fudenberg, D., & Tirole, J. (1983). Sequential bargaining under incomplete information. *Review of Economic Studies, 50*, 221-247.

Guth, W., Schmittberger, R., & Schwarz, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization, 3*, 367-388.

Guth, W., & Tietz, R. (1987). *Ultimatum bargaining for a shrinking cake–An experimental analysis.* (Mimeo)

Harsanyi, J. C. (1967). Games with incomplete information played by "Bayesian" players, I: The basic model. *Management Science, 14*, 159-182.

________ (1968a). Games with incomplete information played by "Bayesian" players, II: Bayesian equilibrium points. *Management Science, 14*, 320-334.

________ (1968b). Games with incomplete information played by "Bayesian" players, III: The basic probability distribution of the game," *Management Science, 14*, 486-502.

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

Kahneman, D., Knetsch, J.L., & Thaler, R.H. (1986a). Fairness and the assumptions of economics. *Journal of Business, 59*, 285-300.

________ (1986b). Fairness as a constraint on profit seeking: Entitlements in the market. *American Economic Review, 76*, 728-741.

Kihlstrom, R., Roth, A.E., & Schmeidler, D. (1981). Risk aversion and solutions to Nash's bargaining problem. In O. Moeschlin & D. Pallaschke (Eds.), *Game theory and mathematical economics*, Amsterdam: North-Holland.

Lewis, D. K. (1969). *Convention: A Philosophical Study.* Cambridge, MA: Harvard University Press.

Machina, M. J. (1987). Choice under uncertainty: Problems solved and unsolved. *Economic Perspectives, 1*, 121-154.

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

Murnighan, J. K., Roth, A.E., & Schoumaker, F. (1988). Risk aversion in bargaining: An experimental study. *Journal of Risk and Uncertainty, 1*, 101-124.

Myerson, R. B. (1985). Analysis of two bargaining problems with incomplete information. In A. E. Roth (Ed.), *Game-theoretic models of bargaining* (pp. 115-147). Cambridge: Cambridge University Press.

Myerson, R. B., & Satterthwaite, M. A. (1983). Efficient mechanisms for bilateral trading. *Journal of Economic Theory, 29*, 265-281.

Nash, J. (1950). The bargaining problem. *Econometrica, 18*, 155-162.

________ (1953). Two person cooperative games. *Econometrica, 21*, 129-140.

Neelin, J., Sonnenschein, H. & Spiegel, M. (1988). A further test of noncooperative bargaining theory. *American Economic Review, 78*, 824-836.

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

Ochs, J., & Roth, A. E. (1989). An experimental study of sequential bargaining. *American Economic Review, 79*, 355-384.

Rapoport, A., Weg, E., & Felsenthal, D. S. (1988). Effects of fixed costs in two-person sequential bargaining. Department of Psychology, University of North Carolina. (Mimeo)

Rapoport, A., Frankel, O., & Perner, J. (1977). Experiments with cooperative 2x2 games. *Theory and Decision, 8*, 67-92.

Rapoport, A., Guyer, M.J., & Gordon, D.G. (1976). *The 2x2 Game.* Ann Arbor, MI: University of Michigan Press.

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

Roth, A. E. (1979). *Axiomatic models of bargaining.* (Lecture Notes in Economics and Mathematical Systems *v*170). New York: Springer Verlag.

________ (1985). Toward a focal-point theory of bargaining. In A. E. Roth (Ed.), Game-theoretic models of bargaining, (pp. 259-263). Cambridge: Cambridge University Press.

________ (1989). Risk aversion and the relationship between Nash's solution and subgame perfect equilibrium of sequential bargaining. *Journal of Risk and Uncertainty, 2*, 353-363.

Roth, A. E., & Malouf, M.W.K. (1979). Game-theoretic models and the role of information in bargaining. *Psychological Review, 86*, 574-594.

Roth, A. E., Malouf, M. W.K., & Murnighan, J.K. (1981). Sociological versus strategic factors in bargaining. *Journal of Economic Behavior and Organization, 2*, 153-177.

Roth, A. E., & Murnighan, J. K. (1982). The role of information in bargaining: An experimental study. *Econometrica, 50*, 1123-1142.

________ (1983). Information and aspirations in two person bargaining. In R. Tietz (Ed.), *Aspiration Levels in Bargaining and Economic Decision Making.* New York: Springer.

Roth, A. E., Murnighan, J.K., & Schoumaker, F. (1988). The deadline effect in bargaining: Some experimental evidence. *American Economic Review, 78*, 806-823.

Roth, A.E., & Rothblum, U.G. (1982). Risk aversion and Nash's solution for bargaining games with risky outcomes. *Econometrica, 50*, 639-647.

Rubin, J.Z., & Brown, B.R. (1975). *The Social Psychology of Bargaining and Negotiation.* New York: Academic Press.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica, 50*, 97-109.

________ (1985a). Choice of conjectures in a bargaining game with incomplete information. In A. E. Roth (Ed.), *Game-theoretic models of bargaining.* Cambridge University Press.

________ (1985b). A bargaining model with incomplete information about time preferences. *Econometrica, 20.*

Selten, R. (1975). Re-examination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory, 4*, 25-55.

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

Stahl, I. (1972). *Bargaining theory.* Stockholm: Economic Research Institute.

Sobel, J., & Takahashi, I. (1983). A multi-stage model of bargaining. *Review of Economic Studies, 50*, 411-426.

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

Wilson, R. (1985). Reputations in games and markets. In A. E. Roth (Ed.), *Game-theoretic models of bargaining.* Cambridge, MA: Cambridge University Press.

Yaari, M. E., & Bar-Hillel, M. (1984). On dividing justly. *Social Choice and Welfare, 1*, 1-24.