PART II

CHAPTER 8

GAMING WITH FAIRNESS

SOME CONJECTURES ON BEHAVIOR IN ALTERNATING-OFFER BARGAINING EXPERIMENTS

RAMI ZWICK AND VINCENT MAK

Adam Smith wrote in The Theory of Moral Sentiments (1759), "How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it." Rami Zwick and Vincent Mak show how a model of alternating-offer games initially predicated on a fully self-interested man became a vehicle for exploring his interest in the fortune of others. They highlight the research questions considered settled and unsettled.

Rami Zwick is professor of marketing at the University of California Riverside's School of Business Administration. His research interests include consumer behavior, neuromarketing, and negotiation and auctions. Vincent Mak is University Lecturer in marketing at the University of Cambridge's Judge Business School. He studies pricing, strategic consumer behavior, and experimental game theory.

* * *

If I am not for myself, who will be for me?
If I am only for myself, what am I?
If not now, when?
Rabbi Hillel (Avoth 2:19, first century B.C.)

Since Rubinstein (1982) proposed a model of alternating-offer bargaining, an experimental literature has emerged that examines behavior in laboratory settings of his model, with apparently disparate conclusions. In this chapter, we attempt to understand when and why theorizing succeeds or fails to predict behavior in this type of experiment and, in more general terms, the complex interaction between fairness considerations and bargainers' gaming inclinations. In particular, we conjecture that the success or failure of predicting behavior using game theoretical reasoning can be explained by the following three principles:

1. The same bargainer could be sometimes self-centered ("gamesman") and sometimes inequity averse ("fairman") depending on the context.
2. Bargaining advantages might be exploitable to different degrees according to the sources of the advantages.
3. Fairness has a price, and the higher its price, the lower the demand for it.

Consider a pie jointly owned by two people who can enjoy a part of it if they reach an agreement on how to divide it. If an agreement cannot be reached, neither party can enjoy any part of it. How should the pie be divided? This is a generic statement of the classic bargaining problem of which two general approaches have been pursued in game theory: the static axiomatic approach, with Nash (1950) as its most well-known representative, and the strategic approach, best represented by Ståhl (1972) and Rubinstein (1982; see also Osborne and Rubinstein, 1990). The axiomatic approach specifies certain desirable properties the division should satisfy, and, in the best case, these properties generate a unique solution to the problem. This method refrains from stipulating any protocol the bargainers should follow. Therefore, by definition, it is independent of the bargaining process. The strategic approach describes the bargaining process explicitly. It does not depend on the need to define, a priori, what properties the allocation should satisfy—properties that are not always endorsed unanimously by bargainers in naturally occurring situations—but rather lets the allocation emerges from the process itself. It was Rubinstein (1982), equipped with the then new concept of the subgame perfect equilibrium (SPE) developed by Selten (1975), who made important contributions to this approach by exploiting what seemed to many to be a natural bargaining scenario. Ideally, the two approaches should reach common grounds in the sense that a bargaining protocol could be found to justify the axiomatic approach to bargaining. This idea has been pursued as part of the "Nash program" (Nash, 1953) and progresses have been made over the past decades (e.g., Binmore, Rubinstein, and Wolinsky, 1986; see also Serrano, 2005). However, it is the strategic approach—more specifically experiments that employ the alternating-offer paradigm—that is explored in this chapter.

Experimentation on alternating-offer bargaining began shortly after Rubinstein proposed his model (see Roth 1995, chapter 4; Weg and Zwick, 1999, chapter 11; and Camerer 2003, chapter 4, for surveys). This stream of studies has been characterized by the observations that behavior often deviates from game theoretical predictions and yet sometimes comes close to them. In this chapter, we contribute a number of thoughts on the conditions under which human behavior in strategic interaction is expected to be well within the game theoretic predictions and the conditions under which it is expected to significantly deviate from it.

The typical features of an alternating offer bargaining experiment include: (1) bargaining is bilateral; (2) bargaining takes place over a succession of more than one time period; and (3) in the first period, one bargainer is the proposer while the other bargainer is the responder, and from then on bargainers alternate roles from period to period. In each period, the proposer suggests how to divide the pie between the two parties and the responder replies by choosing one of the well-defined actions from the response menu, including "accept" and "reject." Once the responder accepts a proposal, the parties reach an agreement, the bargaining ends, and the parties divide the pie according to the proposal. Typically, a game terminates after a period only if the responder in that period accepts the proposal, if it is the final period, or if a randomization procedure that mimics time discounting terminates it. Note that our description of alternating-offer bargaining eliminates protocols in which one of the bargainers has no opportunity to make an offer, so we do not consider the ultimatum game (Güth et al., 1982) or sequential bargaining games in which the offers always come from one of the bargainers (e.g., Rapoport et al., 1995). Nevertheless, excluding the implications of results from ultimatum game experiments is impossible for the present discussion, as they are at the heart of many of the issues that we will examine.

Early controversies

Early experimental studies on alternating-offer bargaining were motivated, in part, by Rubinstein's model and, in part, by the failure of SPE to correctly predict behavior in ultimatum game experiments (Güth et al., 1982; Güth and Tietz, 1990; Güth, 1995). In general, proposers in ultimatum game experiments offered the responders much more than the lowest monetary unit that should have theoretically enticed the responder to accept, while the responders frequently rejected positive offers that were lower than 50 percent of the pie. Naturally, the debate over ultimatum game experiments focused on the extent to which the roles that self-interested, strategic considerations as well as fairness concerns played in determining subject behavior. It quickly inspired a sequence of studies and counterstudies that employed the protocol of alternating-offer bargaining (Binmore et al., 1985, 1988, 1989, 1991; Güth and Tietz, 1988; Neelin et al., 1988; also cf. the aforementioned surveys). Based on

the results of the ultimatum game and early experiments on finite-horizon alternating-offer bargaining games, Güth and his colleagues concluded that "considerations of distributive justice seriously destroy the prospects of exploiting strategic power" (Güth and Tietz, 1990) and that "SPE loses all its predictive power when its payoff implications become socially unacceptable" (Güth and Tietz, 1988).

Yet based on their own experiments, Binmore and his colleagues suggested that, when subjects are faced with a new problem, they simply choose "equal division" as an "obvious" and "acceptable" compromise. However, such considerations are easily displaced by calculations of strategic advantages once players fully appreciate the structure of the game. Such an appreciation can only emerge with extended opportunities to learn and ample incentives to pay close attention to the various procedural details of the situation. Further, the authors argued that if the preconceived rules of thumb (memes) that are embodied with fairness criteria and depend on salient or focal features of the environment in which they are used are not too firmly established, then "they can be displaced by more sophisticated rules that take better account of the strategic realities of the situation. Moreover, there is evidence that subjects are willing to justify their new behavior by asserting that it is fair" (Binmore et al., 1991). Binmore and his colleagues did not claim that fairness and focal considerations are not important or that subjects are natural gamesmen but rather that a theory that ignores strategic considerations is not likely to address what lies at the heart of human bargaining behavior. Bargainers will continue to use the rules of thumbs as long as they produce satisfactory results, but when not, they would be motivated to switch to other rules that are often consistent with game theoretical reasoning. It is worthwhile to note that recently Binmore et al. (2007) carried on this approach and described new findings that supported their attempt to resolve game theoretical predictions, fairness concern, and other behavioral characteristics without denigrating either. Binmore and his colleagues' new experiment began with a conditioning phase in which the subjects knowingly played against robots. In different treatments, the robots were programmed to converge on one of a number of different "focal points" that include what the researchers called egalitarian (equal allocation) outcomes and utilitarian (efficient) outcomes. When the subjects later played one another, they began by playing as they had been conditioned but then gradually adjusted their behavior until they were playing one of the exact Nash equilibria of the game. Binmore et al. (2007) suggested that the behavior in the game replicated in miniature what happens in many laboratory experiments. The conditioning phase mimics the socializing processes of real life through which we learn to operate social norms. When faced with an unfamiliar laboratory game, subjects then begin by simply operating whatever social norm is triggered by the way the game is framed, but as subjects gain experience through repeated play with different partners, their play gradually adapts to the laboratory game they are actually playing. In particular, a simple model of myopic adjustment proposed by Binmore et al. (2007) to explain the data works well mainly because Rubinstein's prediction is a Nash equilibrium with stationary expectations.

Two further studies in the late 1980s and early 1990s contributed directly to the initial debate in the 1980s. Ochs and Roth's (1989) and Kahn and Murnighan's (1993) studies were designed to detect whether changes in the parameters and protocol of play influence the observed outcomes in the predicted direction. Both studies concluded that SPE not only fails as a point predictor of observed behavior, it also fails to account for observed qualitative differences. In particular, the high frequency of disadvantageous counterproposals indicated that there are nonmonetary arguments in the bargainers' utility functions that define what acceptable or unacceptable demands are. The authors of these studies did not conclude that bargainers "try to be fair" but rather that bargainers are averse to being treated unfairly and search for the minimally acceptable offer in the specific environment they are facing, both in making offers and in responding to them. Although the two bargainers' perceptions of what constitute a minimally acceptable offer may differ because of their own egocentric interpretation of fairness, neither bargainer pushes the proposals to the extremes of the equilibrium predictions.

The debate was settled in a draw and is best described by Rabbi Hillel's proverb that provides the epigraph for this chapter. Bargaining is a complex social phenomenon, which gives bargainers systematic motivations distinct from simple payoff maximization. In bargaining, there is a conflict of interest between the parties as well as an internal conflict between self-interest ("If I am not for myself, who will be for me?") and social norms that determine which source of power is "legitimately" exploitable ("If I am only for myself, what am I?"). When both parties recognize that one bargainer has an advantage, the privileged party for the most part will exhibit gamesman tendencies and will try to exploit her advantages. Still, the disadvantaged party will exhibit what looks like fairman propensities, resisting the exploitation (cf. the use of the words *gamesmen* and *fairmen* by Bimore et al., 1988, and Spiegel et al., 1994). Even fairness disposition might be adopted instrumentally and what may appear to be fair is in fact a manifestation of strategic reasoning (Weg and Zwick, 1994; Carpenter 2003). Consequently, it should not come as a surprise that settlements are usually a compromise between the game theoretical solution and equal split. The degree to which the agreement deviates from game theoretical predictions also depends on the extent to which the disadvantaged player can resist exploitation. Note that this argument gives game theory a central role in predicting the settlement. No theory is needed to determine what a numerically equal split is, and yet, recognizing which elements of the environment empowered bargainers with asymmetric advantages requires careful consideration of both situational and environmental (e.g., protocol) factors. Game theory is uniquely qualified for the task. The problem, of course, is that, unless game theoretical analysis yields intuitive results with "face validity," it is hard to expect bargainers to detect such subtle asymmetric advantages right away. Only extensive experience with the environment and suitable learning dynamic, one in which the disadvantaged player learns faster that resisting exploitation is futile than the advantaged player learns that exploiting her advantages is unsuccessful, could asymmetric advantages manifest themselves distinctly in long-run behavior (Cooper et al., 2003).

More puzzling findings

To further develop these ideas, we first describe a number of puzzling findings in alternating-offer bargaining that show that, in some situations, SPE predicts well, even when the predictions are far from an equal split, whereas in other situations the predictions are highly inaccurate. This is in the spirit of Goeree and Holt (2000), who showed that, given a game wherein which behavior conforms nicely to predictions of the Nash equilibrium or relevant refinement, a change in the payoff structure can lead to huge inconsistencies between theoretical predictions and observed behavior. Our examples will center on changes mainly in the bargaining protocol.

Consider, first, the behavior of subjects in alternating-offer bargaining games with time-discounting (geometric depreciation) versus fixed-time cost (arithmetic depreciation). In Weg et al. (1990)'s Experiment 1, subjects negotiated over the division of a pie worth initially 60 Israeli new shekels using an infinite-horizon alternating-offer bargaining protocol with time discounting. There were three different experimental conditions used in a within-subject design: condition S (for strong) with $\delta_1 > \delta_2$ (where 1 and 2 refer to the first and second mover), condition E (for equal) with $\delta_1 = \delta_2$, and condition W (for weak) with $\delta_1 < \delta_2$. The actual discount factors used were (0.90, 0.50), (0.67, 0.67), and (0.50, 0.90), respectively. Experiment 2 used the same procedure but with faster shrinking rates of (0.50, 0.17), (0.17, 0.17), and (0.17, 0.50). The results were clear: According to SPE the agreement should be reached in the first period with the first mover receiving 91 percent, 60 percent, and 18 percent of the pie in conditions S, E, and W, respectively, in Experiment 1 and 91 percent, 86 percent, and 55 percent, respectively, in Experiment 2. These predictions were rejected overwhelmingly by the mean final offers to the first mover. Not only did the SPE model fail as a point predictor, it also failed to account for observed qualitative differences. The model assigns the largest proportion of the pie to the first mover in condition S and the smallest in condition W, with condition E falling in between. In contrast, the mean final offers were ordered in the opposite way. Further, the absence of learning effects and the relatively large proportion of disadvantageous counteroffers deepened the contrast between behavior and the predictions of the SPE model. Surprisingly, a model that is based on equal monetary payoff received strong support. Zwick et al. (1992) reported similar results when the discounting was implemented as the probability of breakdown.

Now contrast the above findings with the results of Rapoport et al. (1990), who used a similar design as Weg et al.'s (1990). Subjects negotiated over the division of a pie worth 30 Israeli new shekels using an infinite-horizon alternating-offer bargaining protocol. However, rather than time discounting, there was a fixed cost per period. Two experiments are reported that are similar in all respects, except that different combinations of fixed costs were used. In Experiment 1, there were three different experimental conditions used in a within-subject design: condition S ($c_1 < c_2$), condition E ($c_1 = c_2$), and condition W ($c_1 > c_2$), with which c_i is player i's cost

per period. The actual costs were (0.1, 2.5), (2.5, 0.1), and (2.5, 2.5). In Experiment 2, the costs were (0.2, 3.0), (3.0, 0.2), and (3.0, 3.0). What makes the fixed-cost scenario attractive is the clear-cut extreme SPE predictions that it implies. If $c_1 < c_2$, the first mover receives the whole pie (similar to the prediction of the ultimatum game). If $c_1 > c_2$, the first mover receives c_2. If $c_1 = c_2$, every partition in the closed interval $[c_1, p]$ is consistent with SPE, where p is the size of the pie. The experimental results strongly support SPE. The raw data is particularly impressive as extreme demands by and offers to the strong players are not buried in the averages. With experience, the cost-based strong players obtain, in general, what is predicted by SPE. Weg and Zwick (1991) and Zwick and Weg (2000) replicated these findings with similar designs for negotiation over both surplus and shortfall.

If, in the case with time discounting, considerations of fairness drive bargaining outcomes away from SPE, why is the same effect absent from the fixed-cost setup?

Another example demonstrates how changing a theoretically irrelevant element of the protocol can make a significant difference in behavior, so that, in one situation, behavior converges to equilibrium, whereas, in the other, it diverges from it. The starting point of Weg and Zwick's (1994) paper was the different behavior exhibited in playing two forms of bargaining games for which theory predicts similar or identical outcomes: ultimatum and infinite-horizon sequential fixed-cost bargaining games. Both games share the characteristic that, according to SPE, the strong player—the proposer in the ultimatum game and the player with the smaller cost in the sequential fixed-cost game—is expected to be apportioned virtually the whole sum in the first period of bargaining. Yet in the ultimatum context, it has been shown consistently that the proposer refrains from taking full advantage of her position, whereas in the infinite-horizon fixed-cost games, a significant proportion of cost-based strong players demand the whole pie and the cost-based weak player frequently accepts these demands. To investigate the underlying source of these differences, Weg and Zwick broadened the response strategy space in the infinite-horizon fixed-cost games, allowing a quit move that terminates the game with both players receiving nothing (null side values) but still liable for any accumulated cost. This extension preserves an important aspect of ultimatum games—the impending breakdown of the bargaining process—while keeping the normative outcomes identical to those of standard infinite-horizon fixed-cost sequential games. This is because the distinction between having or not having access to a null side value is theoretically inessential, since quitting is not a credible move. Psychologically, however, this is surprising. If the first mover can demand the whole pie when she is the strong player in the infinite-horizon fixed-cost game, can she pretend that the other player's threat of quitting is inconsequential in regard to her demand?

Weg and Zwick's (1994) subjects negotiated the division of a $20 pie over an "infinite horizon" (in practice, a game was terminated if negotiations reached the fourteenth period, which occurred only twice) alternating-offer bargaining protocol with unequal costs per period of $2.00 and $0.10. A 2 x 2 x 3 experimental design was used, where the first factor is the game type (with or without a quit move), the second factor is the costs pattern ($c_1 < c_2$, or $c_1 > c_2$), and the third factor

is iteration (whether the subject playing the first mover holds this role for the first, second, or third time). The first factor was a between-subject manipulation, while the last two factors were within-subject manipulations.

Weg and Zwick reported that, in all cases, first and final offers to a cost-based strong player were lower when the quit move was available. The most striking behavior was in games in which the first mover was the cost-based strong player ($c_1 < c_2$). In these games, by the third iteration the first mover's first-period upper-quartile demand was 90 percent ($18.00) of the pie for no-quit games and only 70 percent ($14.00) for quit games. Further, by the third iteration, 59 percent of the games ended with at least 80 percent of the pie allocated to the cost-based strong player when the first mover was the cost-based strong player in the no-quit games; the corresponding percentage was 28 percent when the first mover was the cost-based weak player. This is different from the corresponding figures (0 percent and 11 percent) for the quit games. Although the cost-based weak player seldom exercised his option to opt out in the quit condition (four times out of 108 games), the availability of this option is sufficient to cause the cost-based stronger player to ask less than what she requested in the no-quit game condition. The quit move seemed to play a symbolic function unaccounted for by direct experience.

A similar question as before can be posed: Why can a theoretically empty threat drive behavior away from SPE prediction, when in the absence of this option, behavior converges to SPE?

Three principles of bargaining behavior

The puzzling findings can be explained by three conceptual principles that govern bargaining behavior:

1. The same bargainer could be sometimes self-centered (gamesman) and sometimes inequity averse (fairman), depending on the context.

This first principle furthers the previous discussion. Note that numerous models have been proposed to incorporate inequity aversion into the bargainers' utility function (e.g., Loewenstein et al., 1989; Bolton, 1991; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002; Falk and Fischbacher, 2006). We are not interested here in the exact fashion by which these concerns are integrated into the utility function. For our purpose, the qualitative statement that bargainers are self-centered yet inequity averse is sufficient.

2. Bargaining advantages might be exploitable to different degrees according to the sources of the advantages.

Consider a professional basketball game that has been scheduled to be played on a court that slopes to the side of one of the teams. Obviously, this would give the

team playing on the downhill side an unfair advantage, and as such the disadvantaged team as well as any unbiased outsider would object to such a setting. But at the same time, no one would argue that a team with taller players has an unfair advantage.

The example illustrates two types of advantages that are central to our proposed second principle of bargaining behavior: When examining bargainers' advantages and the degree to which these advantages are exploitable, it is important to distinguish between the different types of advantages that are derived from different sources. These advantages can be broadly categorized into intrinsic and procedural advantages. Intrinsic advantages correspond to a team having taller players in the basketball example. These advantages are so called because they are derived from intrinsic characteristics of the bargainers, independent of the exact bargaining protocol, and are what the bargainers bring to the negotiation table. Meanwhile, procedural advantages correspond to the not level playing field in the example and are derived from the procedural features of the protocol itself. Intrinsic characteristics of the bargainers are what the bargainers bring to the negotiation table, whereas procedural features are analogous to the shape of the table itself.

Intrinsic characteristics are typically expressed as individual-level parameters, whereas procedural features in bargaining are defined with the pair of bargainers as the unit of reference; both could be imposed on subjects in a lab experiment. For example, outside options (i.e., the best alternative to a negotiated agreement or BATNA), time preferences, and the costs of bargaining are potential advantages that are derived from the bargainers' intrinsic characteristics. Still, being the first or the last to propose is a potential advantage that is contingent on the specific procedural features of the negotiation. Both intrinsic characteristics and procedural features can bestow advantages and can create symmetric and/or asymmetric leverages. For the most part, the experimental evidence of alternating-offer bargaining games can be understood in light of the distinction between the advantages derived from these two sources. The two types of advantages are perceived as qualitatively different by bargainers in these games and have different impact on bargainers' behavior along the following major dimensions:

a. *The two types of advantages are assessed lexicographically.* The intrinsic characteristics of bargainers are usually intuitive and easy to assess with simple introspection, thus providing the first input to generating demands or responding to them. This is not to say that bargainers are oblivious to advantages derived from procedural features, but because these advantages are much harder to quantify, they often require at least some experience with the procedural features to be appreciated. It is clearly unreasonable to expect that naïve subjects will perform complicated mathematical operations to assess the strategic nature of the game. Rather, it is more plausible to assume that subjects create simplified representations of the task and look for easily accessible cues (Selten, 1987). Because advantages that rely on intrinsic characteristics of bargainers

are much more accessible than advantages derived from procedural features, we surmise that subjects attend to them before the other type of advantages. If a significant asymmetry is detected at the level of the intrinsic characteristics of bargainers, this asymmetry will dominate and will be reflected in the bargaining outcome. Only if the bargainers are deemed symmetric at the intrinsic characteristic level would procedural sources be attended to—but experience with the procedure is needed for its full effect to be self-evident. In this sense, the two types of advantages are assessed lexicographically. However, even after advantages derived from procedural features become intellectually recognizable, we expect bargainers to attempt resisting the use of those features.

b. *Intrinsic characteristic-based advantages are often considered more "legitimately" exploitable than procedural advantages.* The second dimension along which the two types of advantages are qualitatively different is with respect to the assessment of which type is "legitimately" exploitable. Existing empirical results suggest that, for the most part, it is considered fair play for a bargainer to exploit advantages derived from intrinsic characteristics, whereas exploitation of procedural advantages are commonly resisted as being not fair (compare, for example, the results of fixed-cost bargaining games, in which players' advantages are derived from intrinsic characteristics, and the results of ultimatum games, in which the proposer's advantage is derived from the procedural definition of the game). Two major factors contribute to the lower legitimacy of the exploitation of procedural features:
   i. For the most part, we are exposed on a daily basis to the importance of fair process and level playing fields. Because the playing field is, for the most part, determined and controlled by an indifferent party, we expect it to be level. If not, we might simply refuse to engage in the interaction under the specified rules. In the lab, both rules and intrinsic characteristics are imposed on the subjects. However, people are much more used to and have developed more intuition in dealing with intrinsic characteristics that are not under their direct control (e.g., gender, height, born to a rich parents) than in dealing with rules that they can object to (cf. the literature on the endogenous emergence of rules and protocol in bargaining games, e.g., Güth et al., 1993; McKelvey and Palfrey, 1997; Kambe, 2009; Oz, 2010).
   ii. Bargaining outcomes in the real world are driven for the most part by fairness arguments broadly defined. When an offer is made, for it to be considered in good faith, the proposer has to demonstrate that he or she has reason to believe that the responder has reasons to accept it. The proposer thus needs to demonstrate that the offer is consistent with some objective principles of fairness that apply to the situation at hand and takes both bargainers' circumstances into account. Market forces, other alternatives, or urgency, for example, provide the basis for

> such principles. The difficulty, of course, is that in almost any dispute multiple, objective principles that are relevant to the situation can be advanced, and each one might imply a different agreement. Similarly, for an offer to be rejected in good faith, a counterprinciple that can be reasonably applied to the situation must be advanced. Given the intuitive nature of the leverages derived from intrinsic bargainers' characteristics, verbal arguments based on these types of advantages cannot be easily dismissed (this is related to Raiffa's [1982] principle of "an offer that cannot be readily refused"). Thus they are considered exploitable and play a major role in the final agreement. Advantages that derive their meaning from procedural features are not always intuitive and often require sophisticated analysis. They are not always transparent and often lack face validity. As a result, they are much harder to convey and defend. Their effect needs to be demonstrated and experienced, and only after sufficient experience might both parties consider it fair play for such advantages to be exploited.

Lastly, in some cases, a disadvantaged bargainer—either in terms of intrinsic characteristics or procedural features—can unilaterally act in some ways to eliminate the other bargainer's advantages. For example, in a bargaining game with time discounting without a quit move, a discount-based weak bargainer can simply refuse to trade for many periods so that no matter how the pie is divided, both bargainers' shares approach zero. Whether an advantage has this characteristic depends on its source in the context of the game; any advantage with this characteristic is certainly less exploitable than otherwise.

3. Fairness has a price, and the higher its price, the lower the demand for it.

This principle is based on the results of Zwick and Chen (1999). Previous studies have shown that strategically strong players are sensitive to the availability of a punishment strategy to the weaker players—examples include ultimatum versus impunity games in Bolton and Zwick (1995), x-veto versus no-revenge games in Güth and Huck (1997), and with versus without (nonprofitable) outside options in a fixed-cost sequential bargaining game in Weg and Zwick (1994). However, this sensitivity was only shown in an all-or-none fashion. Zwick and Chen (1999) investigated whether the strategically strong players are also sensitive to the cost (to the weak players) for delivering the punishment and not only to the presence or absence of it. To accomplish their goal, Zwick and Chen studied bargaining behavior in a situation in which one party is in a stronger position than the other. They investigated both the tradeoff the favored party makes, between pursuing one's strategic advantage and giving weight to other players' aversion to inequality, and the tradeoff the disadvantaged player makes between pursuing a fair outcome from a disadvantaged position and the cost of that pursuit. In particular, they hypothesized that the degree to which strategically strong players attempt to

exploit their strategic advantage depends on how much it costs them to do so. The degree to which weak players persist in seeking equity is also a function of how much it costs them to do so.

Bargainers in Zwick and Chen's (1999) experiment negotiated over the division of a pie worth $50 Hong Kong using an alternating-offer bargaining protocol with finite horizon and unequal fixed-cost per rejection. They used a 3 (low-cost) x 3 (high-cost) x 2 (order) x *n* (iteration) partial factorial design. The first three factors were manipulated among subjects and the last within subjects. The low cost of rejection per period was $0.5, $2.0, or $5.0, whereas the high cost was $10.0, $13.0, or $14.5. Order refers to who made the first proposal—the low-cost or high-cost player. Iteration refers to the number of games each subject played. The parameters were chosen such that, based on SPE, the low-cost player should demand and get the whole pie if he moves first and be offered at least 90 percent of the pie if he moves second (independent of the actual cost differences). Zwick and Chen (1999) partially replicated the findings on fixed-cost bargaining games where cost-based strong players routinely demand (and granted) significant portion of the pie, and often, the cost-based weak players make such offers to the strong players. However, contrary to SPE predictions, subjects were sensitive to the magnitude of the cost differential and not only to its presence.

To investigate the hypothesis that demand for fairness is a function of its price, Zwick and Chen (1999) looked at the pattern of disadvantageous counteroffers that were interpreted as an attempt to resist exploitation and to punish such an attempt, and at the pattern of adaptation from game to game. Their examination was characterized by two major findings. First, the percentage of disadvantageous counteroffers declined almost systematically with experience. Because the low-cost players' demands in period 1 did not decrease with experience, the decline in the number of disadvantageous counteroffers with experience indicates that the high-cost players learned to accept unequal division, rather than the low-cost players learning to demand less. However—and more importantly—both the number of subjects who made at least one disadvantageous counteroffer and the proportions of such offers follow a clear pattern: They were monotonically increasing with the cost to the low-cost player and monotonically decreasing with the cost to the high-cost player. Second, Zwick and Chen found that the level of adaptation is also a function of the costs involved. The percentage of first movers who demanded less (in the first period of the next game) after their offer was rejected (strict adaptive behavior) increases monotonically with their cost and decreases monotonically with the cost to the second mover. The same pattern (a mirror image) was detected by looking at the percentages of demanding the same after rejection. This result indicates that the willingness of the low-cost players to demand their "strategic fair share" and not to adapt to the high-cost players' reply is a decreasing function of their own costs of rejection. The willingness of the high-cost players to demand fairness and to persist in their demand for fairness (by not adapting to the low-cost players' reply) is itself a decreasing function of their own costs of rejection.

Low-cost players recognize their strategic advantage and attempt to exploit it. The degree to which they attempt to exploit this advantage depends on their own costs of rejection. The higher their own costs, the less extreme are their own demands. High-cost players recognize the strategic advantage of low-cost players and attempt to resist its exploitation. Their willingness to persist in rejecting extreme divisions of the money is eroded with experience and with persistent extreme demands by low-cost players. Further, the willingness of high-cost players to demand fairness and to persist in their demands for fairness is a decreasing function of their own costs of rejection. This suggests that demand for fairness is subject to cost-benefit evaluation and is, in this sense, deliberate and well thought out.

Implications

The three principles help to explain why SPE predicts behavior well in some experiments whereas in others it fails.

First, environments where asymmetric advantages are derived from intrinsic characteristics of bargainers (e.g., one bargainer, in having a higher time-discount factor or lower time cost than the other bargainer, derives an advantage over the other bargainer) are more likely to result in outcomes that mirror the strategic nature of the asymmetry, but only if the disadvantaged player cannot unilaterally eliminate the inequality. This pattern explains, for example, the different findings in discounting versus fixed-cost conditions. With time discounting, bargainers realize that as the game gets longer due to prolonged bargaining, the monetary payoffs of both parties will approach zero irrespective of how different the discount factors are. If the discount-based disadvantaged bargainer keeps rejecting the other' offers, eventually both players will end up with the same payoff of zero. In contrast, in the fixed-cost case, both players may be forced to pay from their own pocket if bargaining is prolonged. However, and more importantly, as bargaining continues, the difference in the monetary payoffs of the two players grows larger. As a consequence, the cost-based weak player cannot unilaterally eliminate the inequality. Any attempt to do so can only aggravate the inequality.

Similarly, allowing the cost-based weak player to opt out of the negotiation, with both players receiving null side value (as in Weg and Zwick, 1994), provides the weak player with a move that can reduce the inequality (except for the cost of one period). Such a possibility, despite its irrelevance according to conventional game theoretical reasoning, is sufficient for determining much more equal outcome than SPE predicts.

Second, environments where asymmetric advantages are derived from the procedural features of the protocol (e.g., one bargainer, in being assigned to be the first to propose, derives an advantage over the other bargainer) are much less

likely to produce results that mirror the strategic nature of the asymmetry, only after bargainers have obtained sufficient experience with the procedure and only if the disadvantageous bargainers learn faster that resisting exploitation is futile (and can magnify the inequality) than the advantageous bargainers learn that exploiting these advantages are not likely to produce good results. Zwick and Chen (1999) investigated one factor that can be useful in predicting the likely direction of convergence. If the cost to resist exploitation is significantly lower than the cost inflicted on the exploiting party, outcomes are expected to diverge from equilibrium prediction. If, however, the cost to resist exploitation is significantly higher than the cost inflicted on the exploiting party, outcomes are expected to reflect equilibrium prediction.

Endnote: fables and reality

An uneasy relationship exists between theorizing and observed behavior in bargaining experiments, in particular, and in economics, in general. On this point, Ariel Rubinstein interestingly opined that (bargaining) theory should not even be presumed to be about predicting behavior. In a commentary on Binmore et al. (2007), which appears at the end of that paper, and in his 2004 presidential address to the Econometric Society (Rubinstein, 2006), Rubinstein suggested that economic models should be treated as "fables" in which the storyteller draws a parallel to a situation in real life and has some moral he wishes to impart to the reader. "Being something between fantasy and reality, a fable is free of extraneous details and annoying diversions" (Binmore et al., 2007). However, a fable is effective as an education tool (as opposed to sheer entertainment) if the listener can easily draw the parallels to reality and feels that the fable's message provides sound advice or a relevant argument that can be used in the real world. Further, Rubinstein commented that he has "never thought that the alternating offers model (or any other model in economic theory for that matter) is meant to have any predictive power" (Binmore et al., 2007).

Yet what a known fable tells us about the economic inquiry that is supposed to be relevant to it? To draw an analogy, we refer to two characters from *The Little Prince* by Antoine de Saint-Exupéry: the Geographer and the King. The Geographer spends all of his time making maps but never leaves his desk to examine anywhere (even his own planet) to find out if his maps are consistent with reality. According to the Geographer, checking reality is the job of the explorer, but the Geographer does not trust the explorer and would doubt any report indicating inconsistencies between his maps and reality. To those who are familiar with the history of experimental economics, the Geographer is a well-known stock character from days past. However, economics has moved forward, and several Nobel Prizes later, the Geographer breed is an endangered species. All top economic journals are now

routinely publishing experimental papers, and theorists compete to propose models that explain behavioral regularities.

The second character of interest is the King. The King "controls" the stars but only by ordering them to do what they would do anyway. He then relates this to his human subjects and suggests that it is the citizens' duty to obey, but only if the king's demands are reasonable. This should resonate well with any experimentalist who has tested a theory that produces "absurd" conclusions that do seem to square with an intuitive understanding of human behavior. To a great extent, game theory is like the King in that it can successfully predict human behavior only when the theory is not too far from being "reasonable" to subjects from their perspectives in regard to the norm of behavior in an empirical setting of the model. What saves theory from obviousness is that "reasonable," in this sense, can be a complex concept that is not devoid of strategic considerations. It is often dictated by strategic considerations, as human subjects attempt to rationalize and reconcile their self-interest with socially acceptable behavior while trying to learn about the strategic environment they are in through experience. As such, gamesmanship and ideas of fairness are intertwined and inseparable in human bargaining behavior.

References

Binmore, Kenneth, Ariel Rubinstein, Asher Wolinsky. 1986. The Nash bargaining solution in economic modelling. *RAND Journal of Economics* 117 176–188.

Binmore Kenneth, Avner Shaked, John Sutton. 1985. Testing noncooperative bargaining theory: a preliminary study. *American Economic Review* 75(5) 1178–1180.

_____. 1988. A further test of noncooperative bargaining theory: reply. *American Economic Review* 78(4) 837–839.

_____. 1989. An outside option experiment. *Quarterly Journal of Economics* 104(4) 753–770.

Binmore, Kenneth, Joseph Swierzbinski, Chris Tomlinson. 2007. An experimental test of Rubinstein's bargaining model. ELSE, working paper #260, ESRC Centre for Economic Learning and Social Evolution, London.

Binmore, Kenneth, Peter Morgan, Avner Snaked, John Sutton. 1991. Do people exploit their bargaining power? An experimental study. *Games and Economic Behavior* 3(3) 295–322.

Bolton, Gary. 1991. A comparative model of bargaining: theory and evidence. *American Economic Review* 81(5) 1096–1136.

Bolton, Gary, Axel Ockenfels. 2000. ERC: a theory of equity, reciprocity, and competition. *American Economic Review* 90(1) 166–193.

Bolton, Gary, Rami Zwick. 1995. Anonymity versus punishment in ultimatum bargaining. *Games and Economic Behavior* 10 95–121.

Camerer, Colin F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton, NJ: Princeton University Press.

Carpenter, Jeffrey P. 2003. Is fairness used instrumentally? evidence from sequential bargaining. *Journal of Economic Psychology* 24(4) 467–489.

Charness, Gary, Matthew Rabin. 2002. Understanding social preferences with simple tests. *Quarterly Journal of Economics* 117 817–869.

Cooper, David J., Nick Feltovich, Alvin E. Roth, Rami Zwick. 2003. Relative versus absolute speed of adjustment in strategic environments: responder behavior in ultimatum games. *Experimental Economics* 6 181–207.

Falk, Armin, Urs Fischbacher. 2006. A theory of reciprocity. *Games and Economic Behavior* 54 293–315.

Fehr, Ernst, Klaus M. Schmidt. 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114(3) 817–868.

Goeree Jacob K., Charles A. Holt. 2000. Asymmetric inequality aversion and noisy behavior in alternating-offer bargaining games. *European Economic Review* 44(4–6) 1079–1089.

Güth, Werner. 1995. On ultimatum bargaining—a personal review. *Journal of Economic Behavior and Organization* 27 329–344.

Güth, Werner, Peter Ockenfels, Markus Wendel. 1993. Efficiency by trust in fairness?—multiperiod ultimatum bargaining experiments with an increasing cake. *International Journal of Game Theory* 22 51–73.

Güth, Werner, Reinhard Tietz. 1988. Ultimatum bargaining for a shrinking cake—an experimental analysis. In Reinhard Tietz, Wulf Albers, Reinhard Selten (eds.), *Bounded Rational Bargaining Behavior in Experimental Games and Markets, Lecture Notes in Economics and Mathematical Systems,* 314, Berlin: Springer-Verlag. 111–128.

_____. 1990. Ultimatum bargaining behavior: A survey and comparison of experimental results. *Journal of Economic Psychology* 11(3) 417–449.

Güth, Werner, Rolf Schmittberger, Bernd Schwarze. 1982. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization* 3 367–388.

Güth, Werner, Steffen Huck. 1997. From ultimatum bargaining to dictatorship—an experimental study of four games varying in veto power. *Metroeconomica* 48(3) 262–279.

Kahn, Lawrence M., J. Keith Murnighan. 1993. A general experiment on bargaining in demand games with outside options. *American Economic Review* 83(5) 1260–1280.

Kambe, Shinsuke. 2009. Emergence and nonemergence of alternating offers in bilateral bargaining. *International Journal of Game Theory* 38(4) 499–520.

Loewenstein, George F., Leigh Thompson, Max H. Bazerman. 1989. Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology* 57 426–441.

McKelvey, Richard D., Thomas R. Palfrey. 1997. Endogeneity of alternating offers in a bargaining game. *Journal of Economic Theory* 73(2) 425–437.

Nash, John F. 1950. The bargaining problem. *Econometrica* 18 155–162.

_____. 1953. Two-person cooperative games. *Econometrica* 21 128–140.

Neelin Janet, Hugo Sonnenschein, Matthew Spiegel. 1988. A further test of noncooperative bargaining theory: comment. *American Economic Review* 78(4), 824–836.

Ochs, Jack, Alvin E. Roth. 1989. An experimental study of sequential bargaining. *American Economic Review* 79(3) 355–384.

Osborne, Martin, Ariel Rubinstein. 1990. *Bargaining and Markets*, San Diego: Academic Press.

Oz, Shy. 2010. Consistent bargaining. *Economics Bulletin* 30(2) 1425–1432.

Raiffa, Howard. 1982. *The Art and Science of Negotiation*. Cambridge, MA: Harvard University Press.

Rapoport, Amnon, Eythan Weg, Dan S. Felsenthal. 1990. Effects of fixed costs in two-person sequential bargaining. *Theory and Decision* 28(1) 47–71.

Rapoport, Amnon, Ido Erev, Rami Zwick. 1995. An experimental study of buyer-seller negotiation with one-sided incomplete information and time discounting. *Management Science* 41 377–394.

Roth, Alvin E. 1995. Bargaining experiments. In John Kagel Alvin E. Roth (eds.), *Handbook of Experimental Economics*, Princeton, NJ: Princeton University Press, 253–348.

Rubinstein, Ariel. 1982. Perfect equilibrium in a bargaining model. *Econometrica* 50 97–109.

_____. 2006. Dilemmas of an economic theorist. *Econometrica* 74(4) 865–883.

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

_____. 1987. Equity and coalition bargaining in experimental three-person games. In Alvin E. Roth (ed.), *Laboratory Experimentation in Economics—Six Points of View*, Cambridge, UK: Cambridge University Press, 42–98.

Serrano, Roberto. 2005. Fifty years of the Nash program, 1953–2003. *Investigaciones Económicas* 29 219–258.

Spiegel, Matthew, Janet Currie, Hugo Sonnenschein, Arunava Sen. 1994. Understanding when agents are fairmen or gamesmen. *Games and Economic Behavior* 7(1) 104–115.

Ståhl, Ingolf. 1972. *Bargaining Theory*. Stockholm: Economic Research Institute.

Weg, Eythan, Amnon Rapoport, Dan S. Felsenthal. 1990. Two-person bargaining behavior in fixed discounting factors games with infinite horizon. *Games and Economic Behavior* 2(1) 76–95.

Weg, Eythan, Rami Zwick. 1991. On the robustness of perfect equilibrium in fixed cost sequential bargaining under an isomorphic transformation. *Economics Letters* 36(1) 21–24.

_____. 1994. Toward the settlement of the fairness issues in ultimatum games: a bargaining approach. *Journal of Economic Behavior & Organization* 24(1) 19–34.

_____. 1999. Infinite horizon bargaining games: Theory and experiments. In Erev Budescu, Rami Zwick (eds.), *Games and Human Behavior: Essays in Honor of Amnon Rapoport*, Mahwah, NJ: Lawrence Erlbaum Associates, 259–296..

Zwick Rami, Amnon Rapoport, Eythan Weg. 2000. Invariance failure under subgame perfectness in sequential bargaining. *Journal of Economic Psychology* 21(5) 517–544.

Zwick Rami, Amnon Rapoport, John C. Howard. 1992. Two-person sequential bargaining behavior with exogenous breakdown. *Theory and Decision* 32(3) 241–268.

Zwick, Rami, Ching Chyi Lee. 1999. Bargaining and search: an experimental study. *Group Decision and Negotiation* 8(6) 463–487.

Zwick, Rami, Eythan Weg. 2000. An experimental study of buyer-seller negotiation: self-interest versus other-regarding behavior. In Stephen J. Hoch, Robert Meyer (eds.), *Advances in Consumer Research, Vol. 27*, Provo, UT: Association for Consumer Research, 190–195.

Zwick, Rami, Xiao-Ping Chen. 1999. What price fairness? a bargaining study. *Management Science* 45(6) 804–823.