Can Negotiators Outperform Game Theory?

Max H. Bazerman, Robert Gibbons, Leigh Thompson, and Kathleen L. Valley

It is an empirical question whether rational players, either jointly or individually, can actually do better than a purely formal game theory predicts and should consequently ignore the strategic principles produced by such a theory.  

Thomas C. Schelling, *The Strategy of Conflict*

During the past decade, two parallel literatures emerged and flourished: the behavioral literature on negotiation, and the game-theoretic literature on bargaining. Abundant evidence from both literatures suggests that settlement terms predicted by game-theoretic models are not accurate descriptions of actual negotiated outcomes. Furthermore, the behavioral literature provides much evidence that two common assumptions in game-theoretic models, self-interest and economic rationality, also are not accurate. Both of these discrepancies between theory and evidence—in outcomes and in assumptions—primarily imply that actual negotiated outcomes are less efficient than those predicted by game theory. Evidence is beginning to emerge, however, that under certain conditions, negotiators reach more efficient outcomes than predicted by game-theoretic models.

The argument that individuals can outperform game-theoretic predictions is not new. In *The Strategy of Conflict*, Schelling suggests that social norms may lead actual negotiations toward more efficient outcomes than those predicted by formal game theory. Experimental results have supported this thesis. Axelrod (1984), in his work on the prisoner’s dilemma game, focuses on the role of expected future interaction as an explanation for when interdependent parties can outperform the game-theoretic solution that dictates defection in a finite-period prisoner’s dilemma game. Axelrod’s empirical finding is consistent with the theoretical work of Kreps, Milgrom, Roberts, and Wilson (1982), who develop a model that deviates from standard game theory by assuming a small probability that the other party either is not rational or is not self-interested, and so will not follow the rational pure defection strategy. Our goal is to show that the outperformance of game-theoretic predictions is possible in even a one-trial context, and to offer social and cognitive explanations for such outcomes.

This chapter begins with a whirlwind tour of the behavioral and game-theoretic literatures. Our survey reflects the existing literature’s emphasis on inefficient negotiation outcomes. We then present some intriguing recent evidence suggesting that in some settings negotiators may be able to outperform game-theoretic predictions. Finally, we offer several potential explanations of when and why negotiators outperform game theory. We focus on evidence from negotiations with private information, for which game theory predicts inefficient outcomes but behavioral research finds that face-to-face communication promotes more efficient outcomes. Our explanations for these surprisingly efficient outcomes emphasize social-cognitive aspects of face-to-face interaction that game-theoretic models based on self-interest and economic rationality ignore. We conclude by suggesting that these psychological explanations be incorporated into game theoretic models of bargaining.

A Whirlwind Tour of Two Literatures

Game theory offers a mathematical analysis of behavior in tightly specified strategic settings. Bargaining games are extremely simplified relative to the rich context of most negotiations, but the controlled situation provides precise, testable predictions. The behavioral approach to negotiations, in contrast, often considers richer contexts but typically offers less precise predictions, such as the direction but not the magnitude of a predicted difference in behavior.

One fundamental difference between studies of bargaining from experimental economics and behavioral studies from the negotiations literature is the approach taken toward face-to-face interaction between bargaining parties. Face-to-face interaction introduces a multitude of factors that cannot all be identified, isolated or controlled in the manner required by a strict game-theoretic analysis. Therefore, experimental economics studies on bargaining games are usually conducted with anonymous players whose communication is restricted to choices among specified alternatives, or moves. Negotiations research, in contrast, assumes that face-to-face interaction is central to most negotiations, and experimental studies typically allow unrestricted communication, without necessarily identifying the various factors that operate in this condition. What experimental economics studies of bargaining and behavioral studies of negotiations have in
common is that each approaches face-to-face communication in such a way that it eliminates the possibility of studying the effects of that communication.

The vast majority of game-theoretic analyses of bargaining begins with two assumptions: self-interest and economic rationality. Abundant evidence from the behavioral negotiations literature, however, suggests that neither of these assumptions is accurate. Contrary to the self-interest assumption, negotiators systematically fail to maximize their own outcomes because of social concerns such as fairness (Kahneman, Knetisch, and Thaler 1986a) and comparison with others (Loewenstein, Thompson, and Bazerman 1989). Contrary to the economic rationality assumption, it is well accepted that negotiators’ thought processes differ from those typically assumed in game-theoretic (and other economic) models (Arrow, Mooknin, Ross, Tversky, and Wilson 1993). More specifically, negotiators tend to be inappropriately affected by the positive or negative frame in which risks are viewed (Neale and Bazerman 1985b; Bazerman, Magliozi, and Neale 1985), to anchor their number estimates in negotiations on irrelevant information (Neale and Northcraft 1986; Northcraft and Neale 1987), to make errors as a result of inappropriate optimism (Kahneman and Tversky 1993), to rely too heavily on readily available information (Neale, 1984), and to be overconfident about the likelihood of attaining outcomes that favor themselves (Bazerman and Neale 1982; Kahneman and Tversky 1993). Negotiators are also overly disposed to assume that negotiation tasks are fixed-sum, and thereby miss opportunities for mutually beneficial trade-offs between the parties (Thompson and Hastie 1990; Bazerman, Magliozi, and Neale 1985), to escalate commitment to a previously selected course of action when it is no longer the most reasonable alternative (Neale and Bazerman 1991), and to overlook valuable information that is available by considering the opponent’s cognitive perspective (Samuelson and Bazerman 1985).

Complementary work in experimental economics has begun to catalogue discrepancies between negotiated outcomes and game-theoretic predictions in specific bargaining games (see Roth 1993, for a survey). The two leading examples are ultimatum games and two-stage, alternating-offer games. In the former, players 1 and 2 seek to divide a fixed pie. Player 1 makes an offer, which player 2 then accepts or rejects (in which case both players receive nothing). The pie shrinks if player 1’s offer is rejected—say, from $10 to be divided in the first period to $4 to be divided in the second. Game theory predicts that player 2 should offer epsilon in the second period, thereby earning nearly $4, so player 1 should offer $4 in the first period, and player 2 should accept 1’s offer. These predictions are frequently refuted by the experimental evidence, in ways roughly consistent with the evidence on ultimatum games. Once again, if player 2 rejects player 1’s initial offer, the outcome of the game is inefficient, regardless of whether 1 later accepts 2’s offer, because the pie shrinks after 2’s rejection. Furthermore, a new kind of behavior arises that is also inconsistent with the game-theoretic predictions: if player 2 rejects 1’s offer, 2 often then makes a counterproposal that yields less cash for 2 than did 1’s original offer, such as rejecting a 7-3 split and then offering 2-2.

In sum, ample evidence suggests that the great majority of existing game-theoretic bargaining models do not accurately predict negotiator behavior. Such models typically assume both economic rationality and self-interest, and there is substantial evidence that these underlying assumptions are false. One natural next step is to include robust behavioral findings in game-theoretic models (i.e., to depart in a well-specified way from at least one of the economic rationality and self-interest assumptions) and to test the predictions that result in simple bargaining games. Indeed, recent work has followed exactly this path (Bolton 1991; Rabin 1993).

This quick review might lead to the conclusion that human cognition and socialization inevitably impede the negotiation process. The departures from economic rationality described in the behavioral literature frequently lead to inefficient negotiation outcomes, and experiments involving two of the simplest bargaining games not only reject the point predictions from game theory but also find inefficient outcomes where game theory predicts efficiency. But there is also evidence that negotiators can be surprisingly effective, and in some contexts, negotiators systematically outperform the predictions of standard game-theoretic models. Thus, there appear to be functional, adaptive aspects of human behavior not accounted for by these models. We turn next to evidence that negotiators outperform game-theoretic predictions under specifiable conditions. Based on this evidence, we then develop a taxonomy of potential explanations.

**Evidence That People Can Outperform Game Theory**

In order for negotiators to outperform the game-theoretic prediction for a certain bargaining game, it is necessary that game theory’s prediction for that
game be inefficient, otherwise there is no room to improve. Almost every game-theoretic analysis of a bargaining game with complete information (i.e., no private information) predicts an efficient outcome (two notable exceptions are Crawford 1982 and Fernandez and Glazer 1991). In the ultimatum and two-stage, alternating-offer games, for example, game theory predicts efficient outcomes, so while experimental evidence can (and does) refute game theory’s point predictions (i.e., the specific terms of agreement) for these games, evidence from such games can never show both negotiators outperforming game-theoretic predictions.

In bargaining games with private information, on the other hand, typical game-theoretic analysis predicts an inefficient outcome especially when there is two-sided private information. The intuition behind these predictions is that a player with private information will have an incentive to posture (e.g., be strategically deceptive about his or her interests), and will be willing to forego, or at least delay, trades with positive but small value to achieve more advantageous settlement terms in trades with large positive values.

In this section, we describe two classic models of bargaining with private information. The first of these models is Akerlof’s (1970) analysis of “lemons,” or the “winner’s curse,” a game with one-sided private information. The second is Myerson and Satterthwaite’s (1983) study of bargaining with two-sided private information. In both cases, game theory predicts that the outcome will be inefficient. Yet in both cases, intriguing recent evidence suggests that face-to-face (or even telephone) communication may help negotiators achieve efficient outcomes.

**One-Sided Private Information**

Previous studies have suggested that, under asymmetric information, negotiators fail to consider the implications of information held by their opponents. This may result in negative payoffs—the winner’s curse (Samuelson and Bazer- man 1985; Carroll, Bazerman, and Maury 1988). The most common documentation of the winner’s curse in bilateral bargaining involves the “Acquiring a Company” problem, in which one firm (the acquirer) may offer to buy another (the target). While the acquirer knows only that the firm’s value under current management is between $0 and $100, with all values equally likely, the target knows its current worth exactly. The acquirer does know, however, that the firm will be worth 50 percent more under the acquirer’s management than under the current ownership. Thus, it would be efficient for a transaction to take place. What price should the acquirer offer for the target?

If the acquirer offers any positive value, $X$, and the target accepts, the current value of the company is worth between $0$ and $X$, and any value in that range is equally likely. Therefore, the expected value of the target if an offer of $X$ is accepted is $X/2$. Since the company is worth 50 percent more to the acquirer, the acquirer’s expected value is $1.5(X/2)$, which equals only 75 percent of the offer price. Thus, on average, the acquirer obtains a company worth 25 percent less than the price the acquirer pays when an offer is accepted. Considering all possible values of $X$, this analysis reveals the acquirer’s best strategy is not to make an offer (i.e., $X = 0$).

The irony of the situation is that even though the firm is worth more to the acquirer than to the target, any offer above $0$ leads to a negative expected return to the acquirer, because of the high likelihood that the target will accept the acquirer’s offer when the firm is least valuable to the acquirer, i.e., when it is a “lemon” (Akerlof 1970). As a result, a “winning” bidder should expect to lose money, suffering the winner’s curse.

In contrast to this game-theoretic prescriptive analysis, many studies have documented that buyers presented with this problem commonly offer between $50 and $75. This empirical finding is robust across a wide variety of subject populations, including many that have special training, knowledge, and skills (Samuelson and Bazerman 1985; Bazerman and Carroll 1987). In addition, this finding has been extended to contexts in which subjects were paid according to their performance and allowed multiple trials to learn the correct response (Ball, Bazerman, and Carroll 1991). How is this $50 to $75 decision reached? One common explanation is: “On average, the firm is worth $50 to the target and $75 to the acquirer. Consequently, a transaction in this range typically will be profitable to both parties.” The buyers are not considering what the seller’s potential acceptance of a bid implies about the value of the company.

In the studies reviewed above, all subjects act as buyer; that is, subjects are presented with the problem and asked to submit a bid rather than participate in a two-party negotiation. There is no opportunity to outperform game theory (i.e., multi-person decision theory) in this asocial, single-person version of the winner’s curse problem. More recently, Valley, Moag, and Bazerman (1993) created a two-party, social version of this problem that allowed subjects to underperform, meet, or outperform the game-theoretic prediction.

Valley et al. developed the problem into a dyadic situation, with buyer and seller roles. The buyers were given the basic information from the earlier version. The sellers were told the structure of the problem and the true value of the firm to them (randomly assigned between 0 and 100), but were not told the degree to which the buyer’s value for the firm would exceed their own (again 50 percent). The parties were then allowed to negotiate an acquisition price for the firm. Each party was told the objective was to make as much money as possible.

Valley et al. were intending to use this dyadic version of the winner’s curse to study whether a relationship between the parties would protect the buyer from suffering the winner’s curse. Accordingly, in two studies, the type of relationships
between the parties was varied. They anticipated that without a relationship between the parties, the cognitive limitations suggested by the initial studies would again lead ill-informed buyers to be suckers and fall victim to the winner's curse, but that if parties with a pre-existing relationship were allowed to communicate directly, accurate information would be exchanged and the buyer would not fall prey to the winner's curse.

Unexpectedly, the majority of dyads in all conditions (including absence of a relationship) reached agreements that were mutually beneficial for both parties. Generally, information was openly and honestly shared, and this information sharing eliminated the winner's curse. The sellers did not take advantage of their superior information, despite the fact that they easily could have. Thus, the majority of dyads were able to reach an agreement that Pareto-dominated the impasse predicted by game theory: typically, both the buyer and the seller were better off than they would have been had trade not occurred.

After these studies, Valley et al. tentatively hypothesized that negotiators can solve the dyadic winner's curse simply by talking to each other. In a third study, they tested this hypothesis, manipulating whether the communication within the negotiation was verbal or written. In both conditions, the communication time was ample and the potential volume and content of communication unlimited. The dyads allowed to communicate verbally continued to solve the winner's curse. In the written condition, however, the winner's curse reemerged, and significant barriers to agreement were created. Valley et al. speculated that the verbal interchange created a relational bond which was a successful mechanism for increasing the amount of accurate information shared between the parties, thereby reducing or eliminating the winner's curse in bilateral bargaining. These findings suggest that verbal dialogue may have the potential to allow parties to outperform game-theoretic predictions in other settings.

Two-Sided Private Information

Myerson and Satterthwaite (1983) present a striking theoretical result: if a buyer and a seller have valuations (or "types") for a good that are privately known, independently distributed, and continuously distributed over intervals that at least partially overlap, then there is no game that all types of both parties would be willing to play that produces trade if and only if it is efficient. To be more concrete, consider a modified version of the "Acquiring a Company" problem, where the value of the firm under current management, \( V_s \), is uniformly distributed between $0 and $100, and the value of the firm under the acquirer's management, \( V_b \), is independently and uniformly distributed between $0 and $100. (The same impossibility result would hold if \$0 and \$100 become, say, \$50 and \$200 for \( V_s \).) Myerson and Satterthwaite's result implies that whether one party makes a single, final offer, or the parties alternate in making offers until they quit or reach agreement, or the buyer and seller simultaneously make an offer and a demand, respectively; trading if the offer exceeds the demand, none of these games (or any other game) results in trade if and only if it is efficient (i.e., if and only if \( V_s \geq V_b \)).

Note that the intuition behind this result is quite different from that behind the winner's curse. In Akerlof's model, a low valuation for the seller implies a low valuation for the buyer. Here, in contrast, the buyer's and seller's types are independent, so the buyer cannot unknowingly get stuck with a lemon. Instead, the problem is that each side will posture, because sellers have an incentive to present the good as worth more to them than it really is, while buyers are motivated to present the good as worth less, thereby putting small-value trades at risk.

Myerson and Satterthwaite also prove a second result: for the uniform distributions we assumed in the modified "Acquiring a Company" problem (namely, 0 to 100), the game that maximizes the parties' expected gains from trade is a particular equilibrium of a double auction (or the "sealed-bid mechanism"). In this game, the buyer and seller simultaneously make an offer and a demand, and then trade at the average of the two if the offer exceeds the demand. Chatterjee and Samuelson (1983) analyzed many equilibria of this game, including one in which the parties' strategies are linear (e.g., the buyer's offer is a linear function of her valuation). In this linear equilibrium, trade occurs if and only if the acquirer's valuation exceeds the target's by at least \$25. Thus, Myerson and Satterthwaite predict inefficiencies even under the most efficient of the possible bargaining games for this problem.

Experimental evidence suggests Myerson and Satterthwaite's pessimistic predictions are not supported when parties play games of two-sided private information in a face-to-face context. In one of a series of experiments, Radner and Schotter (1989) replaced the sealed-bid mechanism with face-to-face bargaining. Subjects in dyads were told they would bargain with one another for fifteen rounds. Both buyers and sellers drew envelopes dictating their valuations and costs, respectively, over the fifteen rounds. The distribution of possible values was skewed (up for the buyer and down for the seller, symmetrically) to increase the likelihood that the value to the buyer was greater than the value to the seller. In each round, the pair could bargain for a maximum of five minutes. An administrator was present in the room to enforce a set of rules that stipulated that no values or costs could be revealed, no threats could be made.

1 These skewed distributions result in an equilibrium of the sealed-bid mechanism where: buyers bid value if \( V_s < 36 \), and bid \( 20.25 + 0.438(V_s) \) if \( V_s \geq 36 \); sellers bid cost if \( V_s > 64 \), and bid \( 36 + 0.438(V_s) \) if \( V_s \leq 64 \).
no bargains could be contingent on future rounds, and no side payments could be made or promised. The transaction price was that agreed upon by the two parties. The surprising result was that when bargaining face-to-face, negotiators captured 99 percent of all available trades.

The experiments reviewed in this section suggest that face-to-face negotiation in bargaining games with private information includes social and cognitive factors that may lead to better outcomes than game theory predicts. More research is needed to isolate the factors that contribute to the success of face-to-face interactions. This analysis is best directed by developing a framework for considering how social and cognitive contingencies can affect standard game-theoretic predictions.

Potential Explanations for Why People Beat Game Theory

We have reviewed preliminary evidence suggesting that negotiators can systematically reach agreements that are Pareto-superior to those predicted by game theory. This section proposes a number of potential explanations for this evidence, organized into three broad categories. First, we explore non-traditional utility functions. Second, we explore the role of social heuristics based on psychological processes and assumptions. Finally, we look at how cognitive and behavioral limitations affect negotiations.

Non-traditional Utility Functions

Traditional economic models assume self-interest as the sole motive of actors, ignoring “social goals.” More recent behavioral (and even economic) models, however, allow individuals to value other dimensions of outcomes, including concerns for others, disutility for deception, and preference for equality (e.g., Bolton 1991, Rabin 1993). These components appear to be part of the utility functions of individuals and may be incorporated into formal models. We review attempts to include these attributes in utility functions, and discuss how these revised utility functions may account for some of the evidence that individuals outperform game-theoretic predictions based exclusively on self-interest.

The Evidence

Fairness considerations account for some of the limitations of the explanatory power of economic models. Substantial evidence exists that individuals place utility on fairness in ways that are inconsistent with the self-interest assumption.

In a provocative set of experiments, Kahneman, Knetsch, and Thaler (1986a) asked subjects to evaluate the fairness of the action in the following situation:

A hardware store has been selling snow shovels for $15. The morning after a large snowstorm, the store raises the price to $20. Please rate this action as:

Completely Fair – Acceptable – Unfair – Very Unfair

The two favorable and the two unfavorable categories were combined to indicate the proportion of respondents who judged the action acceptable or unfair. Despite the (short-run) economic rationality of raising the prices on the snow shovels, 82 percent of the respondents considered this action unfair. These results may reflect concerns for fairness; if short-run supply and demand were the only considerations, raising prices would have been perceived as fair. Alternatively, the subjects’ responses also are consistent with an entirely economic argument: The subjects may suspect that the firm has deviated from a long-run equilibrium of a repeated game, in which the firm has private information about supply and/or demand and is supposed to keep prices at non-gouging levels on average.

Fairness issues also arise in ultimatum games. A number of researchers have methodically studied how people respond to ultimatums (see Thaler 1988, for a survey). As noted above, game theory predicts that player 1 will offer player 2 only slightly more than zero, and that player 2 will accept any offer greater than zero. Empirical observation suggests, however, that individuals incorporate fairness considerations into their choices. Player 1 demands less than 70 percent of the funds on average, while individuals in the role of player 2 sometimes reject profitable but unequal offers, and accept zero only 20 percent of the time it is offered (Guth, Schmittberger, and Schwarz 1982). However, distinguishing between the effects of fairness and self-interest here is difficult, given the observed behavior of player 2.

In recent work, Straub and Murnighan (1992) manipulated knowledge of the size of the pie in ultimatum games. When player 1 knew the size of the pie but player 2 did not, player 2s accepted lower offers and player 1s made lower offers. The latter evidence suggests that it is 1’s evaluation of 2’s requirement of fairness that drives 1’s observed behavior when the size of the pie is common knowledge, rather than any absolute preference for fairness on the part of player 1. Since the players are assigned their roles at random, this conclusion about player 1’s motives suggest that player 2’s behavior is driven by a distaste for getting the short end of a deal, rather than by a symmetrical taste for fairness. Consistent with this, Loewenstein, Thompson, and Bazerman (1989) argue that the outcomes of others commonly act as a reference point in interpersonal decision settings, and that interpersonal comparisons can overwhelm concern for the absolute value of one’s own outcome in rating potential resolutions of a dis-
pute. For example, typical individuals rated $500 for self and $500 for another person as more satisfactory than $600 for the self and $800 for the other person.

Fairness considerations may be so powerful that players will be willing to pay to punish their opponent if their opponent asks for too much. In two-stage, alternating-offer bargaining games, Ochs and Roth (1989) found that 81 percent of the counteroffers followed by advantageous counteroffers, where the party who rejected the initial offer demanded less than he or she had just been offered. Ochs and Roth argue that the players' utilities for fairness may explain the results. They also note, however, that a simple notion of equality does not explain the data, since in most cases player 1 asks for more than 50 percent of the resources. Rather, as in ultimatum games, parties realize that the other side may very well refuse offers perceived as unfair, despite the economic rationality of accepting them. Thus, an offer closer to 50 percent than 100 percent maximizes player 1's expected payoff.

Ochs and Roth's argument also is consistent with Forsythe et al.'s (1991) results: Parties played either an ultimatum game or a "dictator" game in which player 1 could simply decide without player 2's approval how the resources would be split. Whereas many player 1s chose a fifty-fifty split in the ultimatum game, none proposed a 100 percent-0 percent split. In the dictator game, in contrast, 36 percent of the player 1s took 100 percent. Thus, when approval was required, proposals were more equal, consistent with 1's fear that 2 will judge an offer unfair. Even in the dictator game, however, 64 percent chose to give the other party some portion of the resources. The latter finding may be the best evidence that people place utility on being fair, rather than that they aim to maximize their own outcome by taking into consideration the other party's distaste for the short end of a deal. Even this evidence is open to questions as to whether the subjects anticipated repeated play, or even future interaction with the experimenter.

Modeling Social Utility

Probably the most well-accepted descriptive theory of decision making is Kahneman and Tversky's Prospect Theory (1979). A central idea of Prospect Theory is that people evaluate the utility of alternative courses of action relative to a reference point (Tversky and Kahneman 1981). Prospect Theory examines decision making in what Loewenstein et al. label intrapersonal contexts. In intrapersonal contexts, the outcomes of the decision affect only the decision maker, and the reference point is most frequently modeled as the current state (i.e., wealth) of the decision maker.

Intrapersonal contexts may be contrasted to interpersonal contexts, in which people's decisions affect their own and others' outcomes. In the latter contexts, the reference point does not simply represent the current state of the decision maker. Interpersonal contexts focus the decision maker on the comparison other, and the most likely comparison points are the outcomes obtained by the comparison other. Messick and his colleagues (Messick and Sentis 1985; Messick and Thorngate 1967) incorporate this concern by distinguishing between non-social utility (utility for own outcomes) and social utility (utility for the difference between own and others' outcomes).

Loewenstein et al. identified a utility function for decision making in interpersonal contexts that dealt explicitly with the comparison between outcomes to self and outcomes to others. The precise form of the estimated function, selected by goodness of fit tests over a variety of forms, was:

\[ U = \text{self} + \text{posdif} + \text{negdif} + \text{posdif}^2 + \text{negdif}^2 \]

where:

\[ \text{posdif} = \text{positive differences between one's own and the other's payoff (advantageous inequality);} \]
\[ \text{negdif} = \text{the absolute value of negative differences between one's own and the other's payoff (disadvantageous inequality).} \]

Loewenstein et al. found that disputants preferred equal outcomes over inequalities, so that the signs of both B1 and B2 were negative, but advantageous inequalities were preferred over disadvantageous inequalities, so that the absolute magnitude of B1 was greater than that of B2.

One could extend this type of analysis to predict that people place a disutility on lying. That is, people regulate their behavior in ways that enhance the correspondence between their self-presentation and how they truly think about themselves (DePaulo 1993). For example, the following adaptation of the regression in Loewenstein et al. might better capture the utility of negotiators:

\[ U = \text{self} + \text{posdif} + \text{negdif} + \text{lying} \]

where lying refers to the disutility individuals place on telling a lie. By adding in such non-traditional attributes, a wide variety of utility functions could be specified and tested.

A central point of this section is that fairness, social comparison, equality, and the disutility of lying affect behavior in competitive contexts, and that they can be included in a model of bargaining behavior. The strength of such an approach is that potentially it empowers game-theoretic models to account for current anomalies. To the extent that individuals have utility for equality (Messick 1991) and negative utility for lying, the anomalies of dyads splitting the pie equally in an ultimatum game and people reaching mutually beneficial agreements in the Valley et al. task begin to be rational manifestations of people's true utility functions.
Social Heuristics

The previous subsection argued that alternative formulations of utility, incorporating non-traditional attributes, offer potential explanations for findings that cannot be explained by game-theoretic models based on self-interest. This subsection, in contrast, will suggest that some anomalies result not from utilities placed on potential outcomes, but instead from social heuristics guiding the negotiation process, and thus cannot be explained by game-theoretic models based on individual rationality. In a two-party negotiation, joint outcomes may exceed the game-theoretic prediction because of shared social heuristics guiding behavior. Social heuristics are the beliefs and assumptions about human behavior that individuals use to guide their own behavior when they are interacting with others; we will focus on these, rather than on the more commonly discussed cognitive heuristics that permeate the behavioral decision area (Dawes 1988; Bazerman 1990; Messick 1991).

Our hypothesis is that social heuristics held by both parties lead toward efficient dyadic outcomes in competitive situations, though applying these heuristics is not necessarily the best strategy for the individual negotiator. We believe that people tend to follow social heuristics in competitive, interactive bargaining situations. The use of these heuristics will be augmented to the extent that the bargaining situation allows individuals to interact in an unconstrained fashion. That is, to the extent that individuals may freely communicate and are in full view of one another, the heuristics are more likely to come into play than in situations in which communication is restricted. Thus, we argue that social heuristics help account for the communication effects observed by Valley et al. and by Radner and Schotter, and identify and review below some of the most powerful and pervasive social heuristics.

Assumption of Future Interaction

It is well known that individuals who expect to interact with someone in the future will behave differently than individuals who expect an interaction to be an isolated episode. People who foresee future interaction behave in ways they believe will affect or influence the other’s future behavior. (Indeed, even game theory makes this prediction.) Acting under this heuristic, individuals often behave as if they were going to interact with a person again, even when they know the interaction is a one-shot situation. Thus, the heuristic departs from game theory’s assumption of economic rationality. Studies of tipping behavior, for example, indicate that individuals frequently leave significant tips even when the probability of repeated interaction is extremely low.

In the context of bargaining, we might expect that negotiators will behave in one-shot interactions in ways that are more consistent with rational behavior under the assumption of future interaction. For example, Axelrod (1984), in his discussion of the prisoners’ dilemma task, argued that cooperative behavior is more likely if people perceive they will interact in the future. In short, cooperative behavior operates as a signal to one’s opponent that one will continue to act cooperatively in the future. If both parties act according to a social heuristic assuming future interaction, it becomes easier for the parties to reach agreement with incomplete information in the presence of only a small bargaining zone.

Concern with Reputation

According to this heuristic, people are concerned with maintaining their reputation even in situations where their reputation is not at stake, such as one-shot or anonymous bargaining situations. This concern obviously is closely related to the assumption of future interaction discussed above, but can apply more broadly. One’s behavior in a dyad might affect one’s reputation with a third party, for example. It is sensible and rational to act in ways that preserve and protect one’s reputation when information about the self will be transmitted to others and will affect one’s future dealings with others. However, reputation should not be a concern in bargaining situations where the parties will not meet again and/or the true identity of the opponent will not be revealed.

The concern for reputation arises from the overgeneralization of a larger social motive, known as self-presentation or social desirability (Goffman 1959; Baumeister 1982). Social psychological research suggests that self-presentation concerns are often internalized. That is, it is not only important to the individual that he or she be seen by others in a favorable light, but also that he or she view himself or herself positively. A social heuristic that causes negotiators to maintain a concern for self-presentation or reputation, even in situations where such a concern may be irrational, may lead them to reveal information they otherwise might not reveal in isolated bargaining situations.

Belief in Simultaneous Causality

According to this heuristic, individuals believe that engaging in a behavior or action will increase the likelihood that another person will engage in the same behavior, even in situations where there can be no connection between the actions. That is, individuals behave as if their behavior at time 1 will make it more likely that other individuals also will engage in such behavior at time 1. A classic example of this faulty heuristic is the voter’s belief that if he or she makes the effort to cast a ballot then others are more likely to vote. Note that this heuristic could have some validity if the voter’s belief was that his or her vote at time 1 would affect the likelihood of voting by others in subsequent elections; it is
more problematic to imagine that casting one's vote could simultaneously affect others' decisions to vote. Nevertheless, such beliefs pervade individuals' thinking. This type of thinking is part of a larger aspect of irrational thought, known as metamagical thinking (Hofstaeder 1983).

In bargaining situations, we might expect that individuals who choose to disclose information honestly will believe that a simultaneous disclosure by their opponent also is likely to be honest. Similarly, Shafir and Tversky (1992) found that a substantial portion of subjects would cooperate in a one-shot prisoners' dilemma game if they did not know the other party's decision, but would defect if they knew that the other party had already chosen either to cooperate or to defect. The subjects who knew the other party's decision made the payoff-maximizing choice, while the behavior of subjects who did not know the other party's choice suggests a belief in simultaneous causality.

Agreement-Is-Good Heuristic

According to this heuristic, individuals believe that reaching a mutual agreement is better than not reaching an agreement. (This belief could, of course, also be modeled as a non-traditional utility function.) In a bargaining context, individuals often lack a clear method of evaluating the utility of an outcome, and therefore may reason that if they reach an agreement with another person then the bargaining situation was successful; the failure to reach an agreement is regarded as unsuccessful. The agreement-is-good heuristic will lead negotiators to share more information than prescribed by a standard game-theoretic model, which is consistent with the empirical evidence reviewed above.

Norm of Reciprocity

Perhaps the most pervasive and robust social heuristic that guides human behavior is the norm of reciprocity (Gouldner 1960; Gergen et al. 1975). The generalized norm of reciprocity is extremely powerful and can lead individuals to reciprocate on dimensions that are out of proportion to the original offer. For example, individuals were more likely to support a confederate's request for buying raffle tickets that could help the confederate win a new convertible car when the confederate had previously bought the subject's cola (Regan 1968). Feelings of obligation are influenced by factors in the situation. For example, a larger favor is reciprocated more often than a smaller favor (Greenberg and Frisch 1972). People's attributions about the motives of the helper also matter. People are more likely to return a favor when the original help is perceived to be given intentionally and voluntarily (Goranson and Berkowitz 1966).

The generalized norm of reciprocity applies to the sharing of intimate or personal information among persons, or self-disclosure (Altman and Taylor 1973). Simply, if we share intimate information with other people, they are likely to respond with equally personal information. Extending the norm of reciprocity to a bargaining context, we would expect that an individual who offers information about his or her lowest selling price or highest possible offer will not only expect that such information will be reciprocated in kind, but also will be likely to receive such information from the opponent. This has the potential to result in more information exchange than accounted for by game theory, and this information exchange can account for mutually beneficial agreements that game theory would not predict.

Social Identity

Social identity theory asserts that the perception of the self is defined by social group memberships. Identifying with others reduces social and psychological distance, and increases our positive evaluations of them. Social identity is not a constant; it varies with the salience of the group context (Turner 1987). Orbell, van de Kragt, and Dawes (1988) found that an individual's sense of social identity with a specific group can be enhanced by communication with others in the group. More direct forms of communication that increase the perceptual salience of others, such as face-to-face or oral communication, should be more effective in increasing an individual's identification with the group.

Extended to bargaining situations, as individuals' social identities are heightened, they are less likely to focus on individual gains and less likely to deceive others (Kramer, Pommerenke, and Newton 1993). To the extent that an individual negotiator is evaluating himself or herself in terms of the group, social heuristics such as reciprocity, reputation concerns, and agreement-is-good should increase. Additionally, social identity perceptions of the self should increase the value one party places on joint outcomes, reducing the utility for rewards gained at the expense of the other party.

Norm of Commitment

The norm of commitment is the obligation that a person feels to comply or follow through with promises. Although communication between parties may lead to greater disclosure of truthful information in several ways, such as by enhancing social identity, as noted above, and creating a perceived relational bond (Kramer and Brewer 1984), the effectiveness of such communication in producing efficient agreements may be due in part to the actual verbal promises or commitments made during interaction. There is evidence that such discussion
leads members of groups to make and keep promises to cooperate. For example, in studies of cooperation and defection in social dilemma tasks, Orbell et al. (1988) found that to the extent that members make promises during a discussion period, cooperation greatly increases during subsequent decision periods, even when fellow group members have no way of checking whether such promises are kept. Similarly, Braver and Wilson (1984) found that, controlling for future interaction and common knowledge, subjects who made promises during discussion were more likely to cooperate. Work in experimental economics is also beginning to consider the significant role of commitments in reaching negotiated agreements. Cooper (1990), testing the effects of promises in the Battle of the Sexes game, found that when one subject was allowed to make a single, nonbinding announcement about intended future behavior, subjects almost always reached a coordinated outcome.

Standard game theory treats unenforceable promises as signals that usually should be ignored by rational negotiators (though Farrell and Gibbons 1989 and Myerson 1989 propose that signals should be taken as true when they are consistent with the promisor’s incentives). Yet the studies presented above suggest that when a party makes a verbal promise, she or he is likely to follow through with it, and the other party will assume this is true and adjust behavior accordingly. Thus, the norm of commitment may make promises binding on the parties, and thereby allow the dyad to coordinate behavior and, in some games, outperform expectations.

To summarize: We describe a set of social heuristics that serve humans quite well in most interactive situations. Like any heuristic, however, these social heuristics may lead to irrational behavior in some contexts. We argue that social heuristics lead individuals to engage in suboptimal behavior from an individual perspective, but may cause the quality of joint outcomes in games with private information to exceed game-theoretic predictions. To incorporate social heuristics into game-theoretic models requires that rationality be considered at the dyadic (or group) level, rather than at the individual level.

Cognitive and Behavioral Limitations

Despite the extensive evidence on the negative effects that cognitive biases have on negotiation (Neale and Bazerman 1991), we argue that there are certain cognitive limitations that benefit the dyad in competitive contexts. Two such limitations are especially important in environments with private information: limitations in individuals’ ability to deceive and lie, and the illusory transparency of knowledge.

Limits on Deception and Lying

In discussing deception and lying, we move beyond our earlier suggestion that people place a disutility on lying, to the argument that there are limitations in our effectiveness as liars. A particularly important judgment that people make in many social interactions, especially in bargaining contexts, is evaluating when people are lying or otherwise fostering deception. One argument is that people will reveal deception through nonverbal cues, even when they are successful in lying verbally (Frank 1988). Ekman and Friesen (1974) argue that people attend more to what they are saying than to what they are doing with their bodies. People trying to deceive others, for example, may lie in a calm way, verbally, but reveal their true emotions through nonverbal cues. In Ekman’s terms, there is “nonverbal leakage.” That is, true emotions leak out even if a person tries to conceal them.

Liars often betray themselves through paralinguistic expressions of anxiety, tension, and nervousness. It is sometimes possible to tell when someone is lying by noting the pitch of the voice. Several studies (Ekman, Friesen, and Scherer 1976; Krauss, Geller, and Olson 1976) indicated that the average (or, more technically, fundamental) pitch of the voice is higher when lying than when telling the truth. The difference is small, and often one cannot tell by listening, but electronic vocal analysis reveals lying with considerable accuracy. In addition, shorter answers, longer delays in responding, more speech errors, and more nervous, less serious answers are all characteristic of people perceived as liars, or instructed to tell lies (Apple, Streeter, and Krauss 1979; Kraut 1978; Zuckerman, DePaulo, and Rosenthal 1981).

For our purpose, this research has two important implications: (1) people will be more accurate in detecting deception in face-to-face interactions and, therefore, (2) people will be less inclined to lie in face-to-face interactions with others. In terms of accuracy, subjects quite consistently perceive deceptive messages as less truthful than honest messages (see DePaulo 1993 for a review).

The research on lying and accuracy of lie detection suggests that bargaining contexts provide an arena that cultivates truth-telling and the accurate detection of deception. First, perceivers are more likely to be in a lie-detection mode. That is, in a negotiation situation, people are more “on guard” than in routinized social encounters. Second, a competitive situation provides good information about the likely motives of a potential liar. DePaulo, Stone, and Laslett (1985) found that lies were easier to detect when the sender of the lie had the greatest motivation to lie. In contrast, lies were harder to detect without any specific information about reasons to lie. It is clear how and why someone would use deception in a bargaining situation: sellers are expected to inflate their demands; buyers are expected to deflate their offers. These effects are enhanced when negotiators are in full view of one another and can take advantage of all the available non-verbal cues.
The Illusory Transparency of Knowledge

In this section, we argue that individuals often act as if another party (e.g., the opponent negotiator) has access to information the individual knows the other party does not possess. That is, we predict negotiators do not realize and achieve the full value of their private information. This prediction stems from work on the “curse of knowledge,” which argues that in predicting others’ knowledge, people are unable to ignore knowledge they possess, even when they know that others do not have access to the same information (Camerer, Loewenstein, and Weber 1989). Camerer (1992) argues that this curse explains the difficulty of teaching, since it is hard to imagine how little the students know, and the difficulty of product design, since it is hard for product designers to comprehend how tough it is for consumers to master high-tech devices. In our context, we argue that (1) a negotiator frequently acts in ways that would make more sense if the negotiator’s opponent had the negotiator’s private information, and (2) when both parties act as if the other party is better informed than is actually the case, this helps the parties achieve mutually beneficial agreements.

Keysar (1992) argues that individuals often assume that when they send an ambiguous message to another individual, their communicative intent will be magically understood by the other party, even when the content of the message is based on information that the receiver does not possess. Keysar had people read scenarios that provided them with privileged information about “David.” They read that David followed a friend’s recommendation for a particular restaurant and had dinner there. Half of the participants in the experiment learned that he really enjoyed the dinner, and the other half learned that he really disliked the dinner. In both cases, they read that David left the following note to the friend: “About the restaurant, it was marvelous, just marvelous.” The participants in the study who knew that he enjoyed the dinner had a strong tendency to believe that the friend would take the comment as sincere. In contrast, participants in the study who knew that he disliked the dinner had a strong tendency to believe that the friend would take the comment as sarcastic. The result occurred despite the fact that both groups of participants knew that the friend had access to only the one note. Keysar argues that we believe an ambiguous message will be accurately received if we have the information to unambiguously interpret the message, even if we know that the recipient of the message does not.

Keysar, Ginzel, and Bazerman (1995) have recently extended these arguments to show that individuals ignore the privacy of private information in predicting the behavior of negotiators. In two experiments, subjects played the role of third-party observers of an adapted version of the Acquiring a Company problem. In the adapted version, all subjects were told that the seller sent a memo to the buyer stating that the firm is worth $80/share to the seller, and that the seller will accept $85/share. (Recall that the firm is worth 50 percent more to the seller.) The subjects vary, however, in terms of what they know about the true situation. One group was told that the true value of the firm is $20/share to the seller, and that the seller’s agent (who sent the memo to the buyer) knew this. A second group was told that the value of the firm is again $20/share to the seller, but the seller’s agent thought that the firm was worth $80/share. Finally, a third group was told that the value of the firm is $80/share to the seller, and that the seller’s agent knew this. Keysar et al. found that (1) the subject’s knowledge of the value of the firm to the seller dramatically affected the subjects’ predictions of the buyer’s behavior (even though the buyer had no access to information about the firm’s value to the seller), and that (2) the subject’s knowledge of the seller’s agent’s knowledge also had a dramatic influence on the subject’s predictions of the buyer’s behavior, independent of the true value of the firm to the seller. Thus, subjects acted as if the information that one negotiator had metamagically would be available to the other party.

The central point of this subsection on cognitive limitations is that if a negotiator acts as if his or her private knowledge is common, this will lead to behaviors that are more typical of the behaviors expected under a full-information condition, consistent with the high rate of mutually beneficial agreements found by Valley et al. and by Radner and Schotter. Thus, in contrast to the ample evidence that cognitive biases impede negotiation, this subsection suggests that certain cognitive limitations improve joint outcomes in negotiations.

Summary

Siegel and Fouraker (1960), discussing the vital role of social variables in face-to-face bargaining, conclude that “such variables should either be systematically studied or controlled in experimentation on bargaining. It cannot be assumed, as has often been done, that such variables may simply be neglected” (1960, 22). In the past, face-to-face communication has been treated by both game-theoretic analyses and behavioral negotiation research in ways that eliminated the possibility of identifying and testing the important effects of specific variables within that communication. The evidence presented at the beginning of this chapter suggests that in games of private information, where game theory predicts less than fully efficient outcomes, negotiators bargaining face-to-face outperform the predictions of game theory. We argue that the social, cognitive, and behavioral components of face-to-face communication may be keys to understanding why negotiators are able to outperform these game-theoretic predictions.
This chapter presents a three-part framework from which to build a systematic study of the vital effects of face-to-face communication in bargaining. First, face-to-face interaction affects the utilities of the parties, such that self-interest is only one variable in a utility function that includes preferences for equality, and disutilities for disadvantageous differences and lying. Second, social heuristics, beliefs and assumptions about human behavior that individuals use to guide their own behavior when they are interacting with others, increase tacit understandings between the parties that suggest that certain moves either are more or less acceptable in face-to-face bargaining. While these heuristics appear irrational at the individual level, they may be rational for the negotiating dyad. Finally, certain cognitive and behavioral limitations—limits on deception and lying, and the illusory transparency of knowledge—benefit the dyad in competitive contexts by reducing the ability of the individual parties to capitalize on private information. These conclusions are based on a limited body of research evidence. Future research is needed to replicate and extend these propositions, and to eliminate alternative explanations for efficient performance in games where inefficient agreements are predicted by game theory.

Schelling asserts that if bargainers “can do better than a purely deductive game theory would predict . . . even a normative, prescriptive, strategic theory cannot be based on purely formal analysis” (1960, 164). This chapter suggests Schelling’s conclusion may be overly pessimistic. Game-theoretical analyses may be able to model some of the social variables leading to this superior performance, but they can only hope to do so if they begin to incorporate social and psychological variables into their analyses. Our objective is not to offer an alternative to game theory. Rather, we have developed a social-cognitive framework that provides a guide for future theoretical development and empirical analyses of behavior in competitive contexts. Models that include non-traditional utility functions, social heuristics, and cognitive and behavioral limitations may well be used to build more precise “normative, prescriptive, [and] strategic” theories of bargaining behavior.

Nature has equipped organisms with the ability to regenerate or repair themselves when damaged. If I cut myself, healing takes over without need of conscious direction. If I am hurt badly enough, the body’s needs will make themselves known and tend to preempt voluntary activities. In this way, a balance between short-term interests (e.g., continuing to work or play) and long-term interests (staying healthy) is established on terms that derive from an evolutionary logic. Similarly, nature strikes a balance between parts and whole, for example, by diverting blood supply in cold weather to preserve the critical internal organs, although risking frostbite to the extremities.

Business decisions about short-term and long-term interests, and local and global interests, have no evolutionary logic or built-in process to rely upon. Instead, these decisions are based on the limited understandings, rules of thumb, and traditional practices of organizational actors (March and Shapira 1982). Research suggests that undue attention or decision weight may be given to the short term because the future is ambiguous or discounted (March 1978; Loewenstein and Prelec 1992). It is difficult for actors to understand the global and delayed consequences of local decisions and immediate actions (Senge and Sterman 1991; Sterman 1989a, 1989b), and

The writing of this paper was supported by the MIT International Program on Enhanced Nuclear Power Plant Safety, a research program of the MIT Energy Laboratory sponsored by an international group of nuclear power utilities and relevant institutions. John Sterman acknowledges the financial support of the MIT Organizational Learning Center.
To an obstacle that became
an opportunity

S N O W