From
Negotiation Analysis
edited by H. Peyton Young,
University of Michigan Press,
1991.

CHAPTER 4

Analysis of Incentives in Bargaining and Mediation

Roger B. Myerson

Much of the difficulty of bargaining derives from the fact that the individual parties in a dispute generally have different information. Thus, an individual negotiator may try to behave inscrutably to conceal his (or her) information, or he may signal to try to convey some information to other individuals. Also, he may disbelieve or distrust the signals conveyed by others. In general, each individual must try to learn about his opponents and must try to influence what they believe. The theory of games can provide powerful analytical techniques for understanding these issues and coping with them more effectively. In this chapter, I survey some of these techniques, emphasizing their application to the problem of designing effective mediation procedures.

The method of game theory is to try to understand issues in bargaining by studying hypothetical situations in which the structure of individuals' incentives and information are specified precisely and quantitatively. Such examples may be unrealistically simple in many respects, but we can hope to gain important insights as long as the key issues that we want to understand are well represented. Of course, this is the method of analysis in any field of inquiry: to pose one's questions in the context of a simplified model in which many (irrelevant) details of reality are ignored. Thus, although we may never hope to be involved in a situation in which people's positions are as clearly defined as those studied by game theorists, we can still hope to understand real bargaining situations better by studying these hypothetical examples.

In the game-theoretic analysis of such examples, it is generally assumed that players in any game will behave rationally (each attempting to maximize his own payoff from the game) and intelligently (so that a game theorist never assumes that he understands the game better than the players in it). These assumptions may be quite limiting in circumstances when foolish behavior by some parties can be systematically predicted. However, there are many bargaining situations in which assuming that people will behave rationally and intelligently is at least a reasonable working assumption for gaining basic insights into the problem. Furthermore, game theory has shown that much behavior that might at first seem irrational or foolish can, indeed, be explained by rational decision making. For example, one of the goals of this chapter is to show that there may be bargaining situations in which rational, intelligent individuals may fail to reach an agreement in bargaining, even though potential agreements exist that would be mutually beneficial. That is, disagreement or failure to realize mutually beneficial opportunities is not necessarily evidence of irrationality or foolishness.

In this chapter, I consider one simple example that can serve to illustrate many of the most important insights that can be gained from game-theoretic analysis of bargaining. Specifically, I consider a single seller and a single potential buyer who are trying to agree on a price for some unique and indivisible object. We assume that this object may be worth either \$0 or \$80 to the seller, and it may be worth either \$20 or \$100 to the buyer. When bargaining begins, each individual knows his or her own value for the object, but thinks that the other individual's value is equally likely to be either of the two possible numbers. Thus, we may think of the two individuals' values as being independent random variables, such that each of the four possible combinations of values has probability .25. For simplicity, we assume that neither individual has any opportunity to trade this object with anyone else.

In the terminology of game theory, the information that an individual has at the beginning of the game, which others do not know, is called his *type*. In this example, each individual has two possible types: one for each of the two values that he or she might have for the object. We may say that the seller's type is *strong* if the object is worth \$80 to him, since he then has relatively less need to trade than might otherwise be expected. On the other hand, we may say that the seller's type is *weak* if the object is worth \$0 to him, since he would then be relatively more eager to accept any given offer from the buyer. Similarly, in our terminology, the buyer's type is *strong* if the object is worth \$20 to her, since she then really cannot be compelled to pay any higher price; her type is *weak* if the object is worth \$100 to her, since she then has relatively more price flexibility than might otherwise be expected. We may anticipate that, in bargaining, each individual may try to convince the other that he or she is strong, even when he or she is weak, to force the other to accept a more favorable price. However, no trade can occur unless at least one individual concedes that he or she is weak, since there are no mutually acceptable prices between the two strong types. The probability that mutually acceptable prices actually do exist is very large (.75), so potential gains from trade probably do exist, but the temptation of both parties to bluff may make it hard to realize these potential gains. Thus, although this is a simple example, it can illustrate (and, I hope, clarify) many of the important problems and dilemmas that arise because different individuals have different information.

To begin to analyze this example, we must begin by recognizing what the results of our analysis should be. Specifically, we must recognize that we cannot simply determine the probability that the object will be traded and (if so) at what price, because these probabilities and prices should depend on the individuals' values for the object, which we do not know. Instead, we must try to determine the *plan* (or rule) by which the price and the probability of trade will depend on the values of the object to the two individuals. Such a trading plan can be presented in a matrix, showing the *probability* of trade and the *price* if the trade occurs, for each of the four possible combinations of values that the two individuals might have for the object.

The first theory that one might be tempted to apply to this example is the theory of "splitting the difference," which suggests that, if the buyer's value for the object is higher than the seller's, then they should agree to trade at a price halfway between their respective values, and, if the seller's value is higher, then they should not trade. For this example, this theory can be represented by matrix 1.

		Buyer's Value	
		\$20	\$100
Seller's Value	\$80	(0, *)	(1, \$90)
	\$0	(1, \$10)	(1, \$50)

Matrix 1. The split-the-difference plan

Here and in each of the subsequent matrices, for each of the four possible combinations of values, the probability of trade is shown first and the price of the object if trade occurs is shown second. (The asterisk indicates that the price-if-trade cannot be defined when the probability of trade is zero.)

Unfortunately, the trading plan in matrix 1 cannot be the correct description of how the two rational individuals would trade in this situation. To see why, suppose that the seller's value is actually \$0 (although the buyer does not know this). Matrix 1 says that he will either sell for \$50 if the buyer is also weak, or for \$10 if the buyer is strong, so that the weak seller's expected gains under this plan are $(.5 \times 50) + (.5 \times 10) = \$30$. On the other hand, if he bluffs and pretends that his value for the object is \$80, instead of revealing his true \$0 value, then he will not sell if the buyer is strong, but he will sell for \$90 if the buyer is weak. The weak seller's expected gains from bluffing would be $.5 \times 90 = \$45$. Thus, we cannot expect that the weak type of seller would trade according to the plan shown in matrix 1, since he could get a higher expected gain by following a different bluffing strategy ($\$45 > \$30$).

(For simplicity, we are assuming here that both individuals are risk neutral, so that each seeks to maximize his or her expected gains from trade.)

Incentive Constraints and Mediation

Now that we have shown that the trading plan in matrix 1 cannot be implemented, let us see how to find the plans that actually can be implemented. To do this, we suppose that the buyer and seller have agreed to let a mediator help coordinate their bargaining process. This mediator may plan to use a procedure of the following form. First, the mediator will go to each individual separately and ask for a confidential report stating his or her value for the object. (In this simple example, the mediator can insist that each individual should report one of his or her two possible values, since anything else would obviously be a lie.) For simplicity, let us suppose (for now) that this mediator wants to implement a trading plan that treats the two parties symmetrically, in some sense. We may suppose that if both individuals concede that they are weak, then the mediator will definitely recommend that they should trade (since the range of mutually acceptable prices is as large as can be), and that the price should then be \$50 (halfway between \$0 and \$100). If they both claim to be strong (values of \$20 to the buyer and \$80 to the seller), then the mediator should not recommend trading, since no mutually acceptable prices can be found. If one individual claims to be strong and the other reports that he or she is weak, then things are a bit more complicated, because their positions are not so clearly symmetric. So, for now, let us avoid specifying exact numbers for these cases. Let q denote the probability that the mediator would recommend trading if the buyer reported that she was strong and the seller reported that he was weak, and let y denote the expected price (presumably between \$0 and \$20, the seller's and buyer's values in this case) that the mediator would have the buyer pay if trading were recommended. For symmetry, we may suppose that q would also be the probability of trading if the seller reported that he was strong and the buyer reported that she was weak, and that $100 - y$ would be the expected price if trade occurred in this case. (In each case, y is the profit that a weak individual would get in trading with a strong individual, whether the weak individual is the buyer or the seller.) Then the mediator's trading plan is summarized in matrix 2.

		Buyer's Value	
		\$20	\$100
Seller's Value	\$80	(0, *)	(q, \$100 - y)
	\$0	(q, \$$y$)	(1, \$50)

Matrix 2. A general symmetric trading plan

For this plan to be feasible, y and q must be chosen to satisfy two basic inequalities. First, we need

$$y \leq 20 \quad (1)$$

to assure that the strong type of each individual is getting nonnegative expected gains from trade, since either individual could refuse to participate in the plan (and thus not trade at all) if he or she expected to lose by such participation. A second inequality comes from the need to give a weak individual some incentive to concede that he or she is weak. In this trading plan, a weak individual (say, the seller with value \$0 for the object) gets expected gains from trade equal to $(.5 \times 50) + (.5 \times q \times y)$ if he reports his type honestly (assuming that the buyer will also report honestly to the mediator); but if he claims to be strong then he gets an expected gain of $(.5 \times q) \times (100 - y)$. Thus, to give the weak types an incentive to report honestly, we need

$$(.5 \times 50) + (.5 \times q \times y) \geq (.5 \times q) \times (100 - y), \quad (2)$$

or, equivalently,

$$q \leq 25/(50 - y). \quad (3)$$

These constraints are called *incentive constraints*, because they are constraints on the trading plan that are derived from the need to give each individual an incentive to participate honestly according to the plan. Analysis of incentive constraints can provide important insights into the structure of competitive situations. In this example, the incentive constraints imply, among other things, that q cannot be greater than $25/(50 - 20) = .833$. This maximal level of q is achieved by letting $y = 20$, which gives the trading plan shown in matrix 3.

		Buyer's Value	
		\$20	\$100
Seller's Value	\$80	(0, *)	(.833, \$80)
	\$0	(.833, \$20)	(1, \$50)

Matrix 3. An incentive-efficient trading plan

In this plan, if one individual claims to be strong and the other claims to be weak, then, with probability .833, the mediator may recommend that trade occur at the strong individual's value for the object; and, with probability

.167, the mediator may recommend that no trade take place, even though the mediator knows that there exists a range of prices that would be acceptable to both parties. If both claim to be strong, then no trade is recommended. If both claim to be weak, then trade is surely recommended at a price of $50.

Thus, in a feasible trading plan, the probability of the buyer getting the object cannot be greater than $(.25 \times 1) + (.25 \times .833) + (.25 \times .833) + (.25 \times 0) = .667$, which is strictly less than the probability (.750) that the object is worth more to the buyer than the seller. That is, in any feasible trading plan, the mediator must anticipate that, with some positive probability (at least $.750 - .667 = .083$), the seller may fail to sell the object even though it is actually worth more to the buyer. It is impossible for a mediator to guarantee that the object will always go to the person who values it most.

This sweeping conclusion may require a bit more justification. We have not considered trading plans in which the two individuals are treated asymmetrically, nor have we considered the possibilities of random prices, trade between two strong individuals, or failure to trade between two weak individuals. However, this argument can be extended to prove that trading plans with these features still cannot achieve any higher probability of trade than .667 if they satisfy the relevant incentive constraints. The incentive constraints become a bit more complicated to write when such general plans are considered (because there are more variables to consider), but the analysis remains essentially the same.

A deeper objection might be raised, however. If the mediator simply did not worry about making the individuals want to report their type honestly, could the probability of trade be increased above what can be achieved with trading plans that induce honest behavior? The answer to this question is No, because of an important and general argument known as the *revelation principle.*

To understand the revelation principle, we must begin by reviewing the basic concept of equilibrium in game theory. Given any game, let us say that a *scenario* is any complete theory of how the players may behave in the game, specifying what moves each player would choose (and with what probability, if two or more moves are considered possible) at each stage in the game, for each possible state of the player's information. An *equilibrium* of a game is a scenario such that no player could ever expect to do better by unilaterally deviating from the predictions of the scenario, if the player believed that everyone else would always behave according to the scenario. That is, if a scenario is not an equilibrium, then there is some point in the game where the scenario predicts that a player will make a move that would clearly be a mistake, in terms of the player's own interests and information, if the scenario is understood. Although people do sometimes make mistakes, game theorists recognize that almost any kind of bargaining outcome could be explained under an assumption that negotiators simply behave foolishly. To avoid this assumption, then, we must suppose instead that players' behavior in a game will follow some scenario that is an equilibrium.

When a mediator chooses a mediation plan (or trading plan), the mediator is essentially creating a game that the other parties must play. They choose their strategies for sending messages and signals in the framework defined by the mediator, and for deciding under what circumstances they will accept the mediator's ultimate recommendations (if such decisions are not binding). We say that a mediation plan is *incentive compatible* if it would be an equilibrium for all parties to report all their private information honestly to the mediator and to accept the mediator's final recommendations, when everyone understands that the mediator is using this plan. The incentive constraints that I have discussed are the mathematical conditions that an incentive-compatible plan must satisfy. In general, for each individual who is involved in the mediation and for any ordered pair of his or her possible information types, there is an *informational incentive constraint* that asserts that, if the player were actually the first type and if everyone else was expected to be honest, then the player should not expect to do better by dishonestly reporting the second type than by honestly reporting the first (and actual) type to the mediator. Also, for each possible type of each individual, there is a *participational incentive constraint* that asserts that the player should not expect to gain by refusing to participate in the mediation plan at all when this is his or her type. Each of these constraints can be expressed as a simple mathematical inequality. In the analysis of our example, it was sufficient for us to consider just one participational incentive constraint (1), and one informational incentive constraint (2).

The *revelation principle* asserts that, given any mediation plan that a mediator could design, and given any equilibrium that describes how the parties will behave (perhaps dishonestly) in the game that this mediation plan defines, there is an equivalent incentive-compatible plan in which the outcome (under honest behavior) is always the same as in the given equilibrium of the given mediation plan. (Notice that a mediation plan could include "just letting them haggle face to face.") The proof of this result is simple. For any given equilibrium of any given mediation plan, an equivalent incentive-compatible plan can be constructed as follows. First, the mediator asks each individual to confidentially report all of his or her relevant private information (that is, the player's type, in our game-theoretic terminology). Then the mediator calculates the communication and bargaining moves that would have been used by each individual with his or her reported information under his or her strategy for communicating and bargaining in the given equilibrium. Then the mediator computes (or simulates, if there is any randomness involved) the outcome or agreement that would have been derived from these bargaining moves under the rules of the given mediation plan. This outcome is the final agreement that the mediator actually recommends in the new plan. If any individual could ever expect to gain by lying to the mediator (or by not participating) in the new plan, then that individual could have expected to gain by effectively lying to himself (or by deciding to not participate) before implementing his given equilibrium strategy in the given plan (since the mediator in the new plan is, in effect, implementing that given strategy for the player); but this cannot happen in a rational equilibrium.

Thus, the revelation principle assures us that we can restrict ourselves to considering incentive-compatible mediation plans without any loss of generality. (The revelation principle could be restated as: "without loss of generality, we can assume that the mediator makes honesty the best policy.") This is important, not just because honesty is important per se, but because the set of incentive-compatible mediation plans is much easier to analyze than the set of possible equilibria of all possible mediation plans. In our example and, indeed, in a very general class of situations, the analysis of the relevant incentive constraints is mathematically straightforward and gives a clear characterization of all incentive-compatible mediation plans. Thus, in our example, we were able to compute that the highest probability of trade generated by any incentive-compatible plan is .667. Now, by the revelation principle, we can conclude that it is impossible for any plan involving dishonest behavior in equilibrium to generate a probability of trade that is higher than .667.

To further appreciate the power of the revelation principle, let us consider what would happen if a mediator tried to violate it and increase the probability of trade. Specifically, let us suppose that the mediator's plan is to implement the "split-the-difference" trading plan shown in matrix 1. When the mediator asks the two individuals to report their values and uses these reports to determine the price of the object according to the split-the-difference plan, the mediator creates a game in which honesty is not an equilibrium (as we have already observed). There are, however, three other equilibria of this game. In one equilibrium, the seller always reports his value honestly but the buyer always claims that her value is $20 (even if it is really $100), so that the object is either sold for $10 or not at all. In another equilibrium, the buyer always reports her value honestly but the seller always claims that his value is $80, so that the object is either sold for $90 or not at all. In the third equilibrium, each individual lies with probability .60 if the player is weak and is honest otherwise. With each of these equilibria, the trading plan from matrix 1 becomes equivalent to a different incentive-compatible plan, as shown in matrices 4, 5, and 6. (When the price is random, I show the expected price if trade occurs between the given types.)

Most impartial arbitrators or mediators would probably consider mediation plans described in matrices 4, 5, and 6 to be very undesirable. Matrix 4

		Buyer's Value	
		$20	$100
Seller's Value	$80	(0, *)	(0, *)
	$0	(1, $10)	(1, $10)

Matrix 4. A trading plan that is equivalent to an equilibrium of the split-the-difference game

		Buyer's Value	
		$20	$100
Seller's Value	$80	(0, *)	(1, $90)
	$0	(0, *)	(1, $90)

Matrix 5. A trading plan that is equivalent to an equilibrium of the split-the-difference game

		Buyer's Value	
		$20	$100
Seller's Value	$80	(0, *)	(.4, $90)
	$0	(.4, $10)	(.64, $50)

Matrix 6. A trading plan that is equivalent to a randomized equilibrium of the split-the-difference game

seems biased unfairly against the seller, and matrix 5 seems biased unfairly against the buyer. Matrix 6 treats the two individuals more symmetrically, but it is very inefficient, because the probability of trade in this plan is only $(.25 \times .4) + (.25 \times 4) + (.25 \times .64) = .36$; agreement to trade is unlikely in this plan.

Thus, in general, it may be better for a mediator to respect the incentive constraints and try to find a plan that satisfies these constraints at the lowest social cost (according to whatever social-cost criterion he or she uses) than for the mediator to try to ignore the incentive constraints. When a mediator ignores the incentive constraints, he or she must expect that the cost of satisfying these incentive constraints will not go away, but instead will be derived from the strategies of misrepresentation and bluffing that the various parties will use. If only one party bluffs, as in the first two equilibria discussed above, then the result may be to transform a seemingly fair plan into one that is very unfair to the nonmanipulating individual (as in matrices 4 and 5). If both parties bluff, then the result may be that agreement is rarely reached at all, as in matrix 6.

The revelation principle assures us that any trading plan that can be implemented by any bargaining procedure, including unmediated face-to-face bargaining, can also be implemented by a mediator who gives the parties no incentive to lie. On the other hand, many plans that can be implemented by such a mediator cannot be implemented by unmediated bargaining (in which the parties must talk directly to each other, rather than to a mediator). For example, consider the plan shown in matrix 3. Because the price depends on the information types of both parties, they would need to share their information with each other before a final price was agreed on, if they were to implement this trading plan by unmediated bargaining. But, under this trading plan, either individual, if his or her type was weak, would want to pretend that he or she was strong once he or she got any information from the other individual suggesting that the probability that the other individual was weak was greater than .5. For example, if the buyer announced her type first and revealed that she was weak, in an attempt to implement matrix 3 without any mediation, then the seller would prefer to claim that the object was worth \$80 to him, even if it were really worth \$0 (because selling for \$50 gives lower expected gains to the seller than selling for \$80 with probability .833, when the object is worth \$0 to him). So this trading plan, which maximizes the probability of trade, can only be implemented with some kind of mediation. Thus, this analysis shows the importance of using mediation in bargaining. Mediation maximizes the set of feasible trading plans, because it allows each individual to reveal his or her information without reducing other individuals' incentive to reveal their information as well.

Selecting an Efficient Plan

Some incentive-compatible trading plans (such as in matrix 6) may be clearly inefficient, in the sense that there are other incentive-compatible plans (such as in matrix 7) that all the parties would obviously prefer. We say that an incentive-compatible trading plan is *incentive-efficient* (or simply, *efficient*) if there does not exist any other incentive-compatible plan that gives a higher expected payoff to every possible type of every individual involved. Incentive-efficiency is the basic welfare criterion that a mediator should try to satisfy if the mediator is to serve his or her clients well. To appreciate the importance of this remark, compare it to the more naive criterion of guaranteeing that the parties should never miss an opportunity to make a mutually advantageous trade; such a criterion would be impossible to satisfy in situations like this example. As we have seen, trying to pretend that these incentive constraints do not exist or can be costlessly satisfied, as in matrix 1, may just create a worse situation, as in matrices 4–6. It is generally better to accept the incentive constraints and analyze them with a goal of finding the least costly way of satisfying them.

At the time of bargaining (in this example), each party already knows how much the object is worth to him or her, so he or she really cares only about the expected payoffs for his or her true type. However, a mediator does not know these private informational types when the mediator is formulating a mediation plan, so the mediator can be sure that a change in the plan would be considered an improvement by all parties only if every possible type of every individual would consider it an improvement. This is why the incentive-efficiency criterion should involve checking to make sure that there is no way to make all possible types of all individuals better off.

To characterize the incentive-efficient plans for this example, where there are two individuals, each of whom has two possible types, we have four expected payoffs to compute. Let S_0 denote the expected payoff to the seller if he is weak, let S_{80} denote the expected payoff to the strong seller, let B_{100} denote the expected payoff to the weak buyer, and let B_{20} denote the expected payoff to the strong buyer. For the plans shown in matrix 2, these expected payoffs are

$$S_0 = B_{100} = (.5 \times 50) + (.5 \times q \times y);$$

$$S_{80} = B_{20} = (.5 \times q) \times (20 - y).$$

Using a mathematical technique called linear programming, one can analyze the incentive constraints for this example and prove that any incentive-compatible plan that satisfies the equation

$$(3 \times S_0) + (5 \times S_{80}) + (3 \times B_{100}) + (5 \times B_{20}) = 200 \tag{4}$$

must be incentive-efficient. It is straightforward to check that the plans in matrix 2 are incentive-efficient if and only if

$$q = 25/(50 - y), \tag{5}$$

which is the maximum feasible value for the probability q, given the price y. It can be shown that the plans in matrices 4, 5, and 6 are not incentive-efficient, and they generate expected payoffs that do not satisfy equation (4).

The question still remains: Which incentive-efficient plan should a mediator actually implement? A further criterion that an impartial mediator may want to satisfy is *interpersonal equity* (that is, giving each party gains from an agreement that are, in some sense, commensurate with what that party is contributing to the other parties in the agreement). Game theorists have sought to develop a general mathematical theory of equity in games for many years with some incomplete success. (Concepts like the Shapley value and Nash bargaining-solution have been defined for this purpose. For an introduction to these ideas, see for example Shubik 1983; Young 1988; or Myerson 1991.) In this special example, however, we can define interpersonal equity without such general theories. The buyer and seller have a clear symmetry of position (notice how their possible values are symmetrically arrayed around $50), so an interpersonally equitable mediation plan should have the symmetric structure shown in matrix 2. Thus, the plans shown in matrices 4 and 5 can be ruled out as inequitable. The efficient and equitable mediation plans are those shown in matrix 2, when $q = 25/(50 - y)$ and y is not greater than 20.

This still leaves an infinite number of efficient and equitable mediation plans. One of them, corresponding to $y = 20$, is shown in matrix 3. Two more, corresponding to $y = 10$ and $y = 0$ respectively, are shown in matrices 7 and 8.

		Buyer's Value	
		$20	$100
Seller's Value	$80	(0, *)	(.625, $90)
	$0	(.625, $10)	(1, $50)

Matrix 7. An incentive-efficient trading plan

		Buyer's Value	
		$20	$100
Seller's Value	$80	(0, *)	(.5, $100)
	$0	(.5, $0)	(1, $50)

Matrix 8. An incentive-efficient trading plan

These equitable and efficient plans differ in what they offer the weak and strong types of each individual. Among these plans, the strong types of both individuals get the highest expected payoffs in matrix 8 (where $S_{80} = B_{20} = 5$), and they get the lowest expected payoffs in matrix 3 (where $S_{80} = B_{20} = 0$). On the other hand, the weak types get the highest expected payoffs in matrix 3 (where $S_0 = B_{100} = 33.3$), and they get the lowest expected payoffs in matrix 8 (where $S_0 = B_{100} = 25$). Thus, choosing among these efficient and equitable plans involves making some kind of *intertype compromise*, as well as interpersonal compromise. We must ask: Is it better to use a mediation plan that favors the strong types (matrix 8), one that favors the weak types (matrix 3), or something in between (matrix 7)?

One way to resolve this question is to assume that the individuals themselves should be able to negotiate its answer. There is good reason to believe that such negotiations would result in a plan that favors the strong types, as in matrix 8. To see why, notice that all of these efficient and equitable plans lose their incentive compatibility if one party learns that the other party is weak. (As we have seen, it is generally better to claim that you are strong if you know that your opponent is weak.) Thus, to avoid giving the impression of weakness, each individual would be likely to take the negotiating position favored by his or her strong type in these premediation negotiations, even if he or she was actually weak. That is, arguing for the plan in matrix 8 is the best *inscrutable* tactic for an individual who does not want to seem weak.

To see this another way, notice that, if the buyer were convinced that the seller were strong (having value $80), then the buyer probably would be willing to accept a plan like that in matrix 5, where a take-it-or-leave-it price is set at $90. That is, a price of $90 would seem quite equitable to the buyer if she were sure that the object was worth $80 to the seller. Thus, before the mediation plan is fixed, the seller would like to do everything possible to convince the buyer that he (the seller) is strong. One way that the seller might do this is to reject the services of a mediator who proposed to implement one of the plans from matrix 3 or matrix 7. That is, a mediator who proposed to use a plan that was relatively more favorable to the weak types might be exposed to a "mediator-bashing" tactic, where one party attacks the currently proposed mediation plan as a way to signal something about his or her type to the other party. Plans like matrix 8, which favor the strong types, are most likely to be stable against such mediator-bashing tactics.

Cost of Time in Bargaining

In the preceding analysis, we have not explicitly considered the role of time in bargaining. One might argue that, when the object has not been traded, the seller and buyer should continue bargaining as long as they believe that there is any positive probability of finding a mutually beneficial trade, so that the buyer should eventually get the object if it is worth more to her than to the seller. This argument, if true, would seem to contradict our conclusion that there must be a positive probability of the seller keeping the object when it is actually worth more to the buyer. In fact, there is no contradiction, because the preceding analysis has been based on an assumption that a trade must occur now or never. If we drop this assumption, *delay of trade* may replace *failure to trade* as the costly signal or threat that provides the incentive to make concessions. However, the total expected signaling costs are not generally decreased by the possibility of future bargaining. In fact, as I now show, total signaling costs may actually be increased because the possibility of future bargaining may reduce individuals' incentives to make serious offers at any given time.

To properly analyze this issue, we need to make some specific assumptions about how the two individuals (seller and buyer) in our example would feel about the prospect of trading at a later time. Let us assume that the object of bargaining is durable, so that it can be sold at any time. To account for the cost of time, let us suppose (to be specific) that there is a 10 percent annual interest rate. Thus, a dollar to be earned at some time t years in the future is worth less than a dollar earned now, because a dollar now could be invested at the 10 percent interest rate to generate more than a dollar in t years. In fact, when the possibility or reinvesting interest income is taken into account, a dollar earned t years in the future is worth only $e^{-.1t}$ dollars now, where $e \approx 2.718$ is a mathematical constant.

With such discounting of future gains, consider a simple open-ended bargaining game. In this game, the seller and buyer each make some initial price demand, which will remain constant over time until one of the two individuals concedes, at which time trade will occur at the price demanded by the other individual. To be specific, suppose that initially the seller demands a price of \$90 and the buyer demands \$10. The strong type of each individual (with value \$80 or \$20) would be unwilling to accept the other's demand, but the weak type (with value \$0 or \$100) would prefer to accept it than to never trade at all. A weak individual might not want to immediately accept the other's demand, however, in hopes that his or her own demand might be accepted instead, if the other individual is weak. In the equilibrium of this bargaining game, neither individual ever knows when the other might concede, because the weak type of each individual would decide randomly how long to wait for the other's concession.

To see how the probabilities of concession are determined in this equilibrium, let s be a very small number (like $1/365$). Suppose that, at some time, a weak seller is trying to decide whether to concede immediately or to go on waiting another fraction s of a year. Let $p(s)$ denote the probability that the buyer will concede to the seller's demand during such a time interval of length s. If the seller decides to wait for this interval of length s before conceding, then at the end of the interval he may get either \$90 (from the buyer's concession) with probability $p(s)$, or \$10 (from the seller's own concession at the end) with probability $1 - p(s)$. On the other hand, if the seller concedes to the buyer's demand now, then the seller gets \$10 immediately, which will be worth approximately $\$10 \times (1 + .1 \times s)$ at the end of the same time interval of length s, when it is invested at the 10 percent rate. (This formula is only approximate because I am ignoring, for now, the small compound interest correction.) Thus, to make the weak seller indifferent between conceding immediately and waiting s years more (so that the seller is willing to let his own concession time be randomly determined), we need

$$[90 \times p(s)] + \{10 \times [1 - p(s)]\} \approx 10 \times (1 + .1 \times s),$$

and so

$$p(s) \approx s/80. \tag{6}$$

(Here, $\approx$ means "is approximately equal to," and this approximation is very good as long as s is small.) A similar argument from the buyer's viewpoint shows that the probability of a concession by the seller during this period must also satisfy (6). According to this formula, at any time in this bargaining game, the probability of a concession by either individual must be only $(1/365)/80 \approx 0.000034$ during the next day of bargaining, or only about $1/80$ during the next year. If the probability of one individual conceding were higher than this, then the other individual would surely choose to wait. If the probability of one individual conceding were lower than this, then the weak type of the other individual would surely choose to concede immediately.

Exact formulas for the probability of a concession during any given period can be computed from relation (6), using integral calculus. These formulas show that it may take up to 55.4 years for a weak individual to finally make a concession in this open-ended bargaining game. The expected time until someone finally concedes to the other's demand is 24.5 years if only one individual is weak, and is 15.4 years if both individuals are weak. (To be exact, if an individual is weak, then the probability in equilibrium that he or she would choose to concede sometime during the first t years of this bargaining process is $2 \times (1 - e^{-t/80})$, for any number t between 0 and 55.4. When $t = 55.4$, this probability equals one.)

Because of the discounted value of future profits, each individual would be indifferent between the following two alternatives: (i) trading at some given price after waiting t years; and (ii) trading at this given price now, with probability $e^{-.1t}$, or never trading, with probability $1 - e^{-.1t}$. If we let T denote the random time when someone finally accepts the other's demand in this equilibrium of our bargaining game, then the expected value of the quantity $e^{-.1T}$ is .222 if only one individual is weak, and it is .356 if both

		Buyer's Value $20	Buyer's Value $100
Seller's Value	$80	(0, *)	(.222, $90)
	$0	(.222, $10)	(.356, $50)

Matrix 9. A trading plan that is equivalent to an equilibrium of an open-ended bargaining game

individuals are weak. (These numbers can be computed from the probability formula in the preceding paragraph, using integral calculus.) Thus, this equilibrium of our open-ended bargaining game is equivalent to a trading plan, shown in matrix 9, in which trade is supposed to occur either now or never. (Here, as before, the probability of trade and the expected price if trade occurs are shown in each cell. The expected price of $50 in the cell where both are weak is the average of the two possible prices $10 and $90, which are then equally likely, depending on who concedes first.) Each type of each individual gets the same expected present-discounted value of gains from trade in this now-or-never trading plan as in the equilibrium of our open-ended bargaining game with initial demands of $90 and $10.

The trading plan in matrix 9 satisfies all the incentive constraints implied by the revelation principle. Furthermore, it is clearly worse for both individuals than the incentive-compatible trading plan shown in matrix 6, for example. More generally, as long as both individuals use the same interest rate for discounting future profits, any equilibrium of any open-ended bargaining game must be equivalent to some now-or-never trading plan that satisfies all the incentive constraints that I have discussed. In this sense, open-ended bargaining cannot help the individuals to avoid the cost of satisfying the incentive constraints, and may actually make matters worse.

If we generalize our open-ended bargaining game to allow each individual to choose his or her initial demand (instead of arbitrarily assuming that these demands are $10 and $90), then there are many more equilibria to be considered. In particular, for any y between $0 and $20, there is an equilibrium in which the buyer's initial demand is y and the seller's initial demand is $100 − y. If y is very close to zero, then the expected time before the first concession in this equilibrium is very large, and the expected present-discounted values of gains from trade are very close to zero; so we may refer to this as a *standoff equilibrium.* In such a standoff equilibrium neither individual has much incentive to concede at any time, because the other's demand is so extreme. Furthermore, in such an equilibrium, neither individual wants to switch to a less extreme demand, because he or she fears that he or she

would then be perceived as surely being weak and would then be expected (by the other player) to fully concede soon, so that a less extreme demand would not elicit any more rapid concession from the other individual.

Such standoff equilibria may actually be very useful to a mediator, because they offer a way to nullify the effect of future bargaining opportunities and, thus, to achieve trading plans like the one in matrix 8, rather than the one in matrix 9. If a mediator can persuade the buyer and seller to focus on playing a standoff equilibrium in all bargaining that might follow *after* the current mediation efforts, then trade may essentially be "now or never" for the two individuals, as we previously assumed. That is, a mediator may be able to effectively prevent postmediation bargaining by suggesting that, if a trade is not recommended, then any effort to reopen bargaining with further price offers should be taken as a sign of weakness, and that an individual who is thus revealed to be weak should also be expected to quickly concede to all further demands. Thus, the analysis in the earlier sections of this chapter, where we assumed that trade must occur now or never and will not occur at all if the mediator recommends against trading, may also accurately characterize what can be accomplished by a mediator in situations where the buyer and seller actually have open-ended opportunities to continue bargaining.

Conclusions

The analysis of this example has led us through many of the basic issues that arise in practical mediation. This analysis may be extended to more complicated examples, provided that the structure of all individuals' preferences and information is given in some well-quantified form (see, for example, Myerson 1984; Hurwicz, Schmeidler, and Sonnenschein 1985; Roth 1985). A more difficult task is applying this kind of analysis to real situations where people's preferences and information are not so easily given any simple quantitative representation. In such situations, it may be impossible to come up with a simple quantitative "solution" like matrix 8, but the insights generated by our analysis of quantitative examples may still be applied. The general insights discussed in this chapter may be summarized as follows.

When the various parties in bargaining have private information, communicating separately with a mediator in confidential caucuses may help them to achieve better outcomes, because each party can reveal information to a mediator without reducing other parties' incentives to reveal their information. This statement is based on the assumption that the mediator will keep the parties' revelations confidential until final recommendations are made.

After an initial session in which the parties introduce the mediator to the structure of the problem, the mediator should think analytically about the "mediation plan" that describes how his or her final recommendations will depend on the information that the parties will confidentially reveal to him or her. In particular, the mediator should think about mediation plans that are "incentive compatible" in the sense that they would not give any party an incentive to lie about information that the mediator solicits. If a mediator uses a mediation plan that seems equitable and efficient but is not incentive compatible, then rational and intelligent parties may pervert the plan by lying in a way that may lead, with high probability, to unfair agreements or even disagreement. Furthermore, any equilibrium of any bargaining game can be simulated by an equivalent incentive-compatible mediation plan, so a mediator loses no power by restricting himself or herself to incentive-compatible plans.

To avoid mediator-bashing tactics by bargainers who want to convince each other of the strength of their positions, a mediator may have to choose a mediation plan that is better for a party whose private information actually puts that party in a stronger position than other parties might suspect.

There may be some situations in which no incentive-compatible mediation plan, and no rational equilibrium of any bargaining game, could guarantee that the parties would reach a mutually beneficial agreement whenever such an agreement exists. That is, costly disagreement is not necessarily a result of irrational behavior by bargainers. On the contrary, a positive probability of costly disagreement may be necessary to give a rational bargainer some incentive to admit the weakness of his or her position when he or she actually *is* weak. Thus, the appropriate question for an analytical mediator is not how to guarantee that the best possible agreement will always be reached, but, rather, it is how to formulate a mediation plan that will maximize the parties' expected benefits from agreement without creating incentives for them to lie in confidential caucuses or to engage in mediator bashing.

A choice among mediation plans is a decision under uncertainty that involves compromise between different parties, as well as compromise between the interests of the different possible types of any one party. There generally is not an incentive-compatible mediation plan that would be best for all types of all parties.

When two parties bargain without any deadlines, they may get trapped in a standoff equilibrium, in which bargaining drags on so long as to virtually nullify the value of the agreements that might eventually be achieved. In such a standoff equilibrium, each party insists on an extreme demand and is afraid to moderate its demand, lest such moderation be taken as a sign of weakness and of imminent acceptance of the other party's extreme demand. Because these demands are so extreme, each party is willing to wait a very long time, in hopes of getting his or her own demand accepted, rather than immediately conceding to reach an agreement at the other party's demand.

A mediator may try to exploit the possibility of standoff equilibria in postmediation bargaining to induce the various parties to accept whatever

recommendation may be suggested, provided that this recommendation is not worse for any party than bargaining indefinitely without agreement. That is, a mediator may encourage the parties to accept his or her recommendations by suggesting that any party who attempted to negotiate an agreement different from the mediator's recommendation would be perceived by the other parties as actually being in a weak position and prepared to give away almost everything in subsequent bargaining.

BIBLIOGRAPHY

Hurwicz, L., D. Schmeidler, and H. Sonnenschein, eds. 1985. *Social Goals and Social Organization.* Cambridge: Cambridge University Press.

Myerson, R. B. 1984. "Two-Person Bargaining Problems with Incomplete Information." *Econometrica* 52:461–87.

Myerson, R. B. 1991. *Game Theory: Analysis of Conflict.* Cambridge, Mass.: Harvard University Press.

Roth, A. E., ed. 1985. *Game-Theoretic Models of Bargaining.* Cambridge: Cambridge University Press.

Shubik, M. 1983. *Game Theory in the Social Sciences: Concepts and Solutions.* Cambridge, Mass.: MIT Press.

Young, H. P. 1988. "Fair Division." Working paper. Brookings Institution. Photocopy.