Econometrica, Vol. 50, No. 3 (May, 1982)

A THEORY OF DISAGREEMENT IN BARGAINING

BY VINCENT P. CRAWFORD[1]

This paper proposes a simple theory to explain bargaining impasses, which is based on Schelling's view of the bargaining process as a struggle between bargainers to commit themselves to favorable bargaining positions. Because bargaining impasses are generally Pareto-inefficient, anything involving a positive probability of impasse is Pareto-inefficient as well. It is demonstrated that in spite of this avoidable inefficiency, when successful commitment is uncertain and irreversible it can still be rational for individuals to attempt commitment and thereby risk an impasse; in a leading special case, the model reduces to a Prisoner's Dilemma game, in which only strategic-dominance arguments are needed to establish this conclusion. Further, making commitment more difficult, or changing the costs of disagreement in a way that makes available a wider range of settlements that are better for both bargainers than disagreement, need not always lower the probability of impasse, in spite of the conventional wisdom to the contrary.

It is also a good rule not to put overmuch confidence in the observational results
that are put forward *until they are confirmed by theory*.
–Sir Arthur Stanley Eddington

1. INTRODUCTION

BARGAINING, BROADLY CONSTRUED, is a pervasive phenomenon in modern economies, ranging from labor negotiations to trade agreements to strategic arms limitation talks. One need only consider these examples in the light of past experience to realize that the potential welfare gains from improving the efficiency of bargaining outcomes are enormous, perhaps even greater than those that would result from a better understanding of the effects of macroeconomic policy. Yet the problem of designing environments to yield improved bargaining outcomes has been all but ignored by economists.

A major part of this design problem is ensuring that impasses are avoided as often as possible. Because such disagreements, whether they take the form of strikes, trade restrictions, or arms races, tend to be very costly, reducing their likelihood is of great welfare importance. But, before this aspect of the problem can even be approached, a theory that relates the likelihood of disagreement to the bargaining environment is needed. Such a theory would serve an important purpose in guiding attempts to determine this relationship empirically or experimentally, even if it did not yield strong theoretical conclusions.

Almost all microeconomic and game-theoretic models of bargaining beg the question of what determines the probability of disagreement by *assuming* that an

[1]This paper incorporates material from University of California, San Diego Discussion Papers 79-3, written in November, 1978 and bearing the same title, and 80-18, "A Model of the Commitment Process in Bargaining," written in August, 1980. Support from NSF Grant SES 79-05550 is gratefully acknowledged. I also owe thanks for helpful suggestions to Peter Berck, Norman Clifford, Clifford Donn, David Lilien, Mark Machina, William Samuelson, Joel Sobel, Hugo Sonnenschein, Ichiro Takahashi, William Thomson, and Allan Young. Participants in seminar presentations at Caltech, UCSD, Michigan, Cornell, MIT, VPI, Harvard, and Berkeley also made useful comments.

efficient settlement is always reached.[2] This is probably due to the simple and elegant theoretical results often available under the efficiency assumption and to the common belief that inefficient outcomes are inconsistent with rational behavior by well-informed bargainers. But plainly, any theory of bargaining that assumes away the possibility of disagreement must fail to capture an aspect of bargaining that is of central importance in the design problem mentioned above.

This paper proposes a simple theory that explains the probability of disagreement in bargaining, and, it is hoped, will therefore prove more useful in studies of the design problem than existing theories. The theory develops Schelling's [18] view of the bargaining process as a struggle between bargainers to commit themselves to—that is, to convince their opponents that they will not retreat from—advantageous bargaining positions. The potential benefits of commitment are clear, since once one's opponent is convinced, his best strategy is to yield if he can.

It is shown that if the outcome of the commitment process is both uncertain and irreversible, it can be rational for bargainers to take actions that imply a positive probability of disagreement, an outcome ex ante inferior for both to outcomes feasible through negotiation. The theory, which determines the probability or frequency of impasse endogenously, permits an evaluation of the assumption, common in the industrial relations and law and economics literature, that enlarging the set of feasible settlements that are at least as good for both bargainers as disagreement—commonly called the *contract zone* in this literature—makes a negotiated settlement more likely. It turns out that this need not be true: in quite "well-behaved" bargaining situations, enlarging the contract zone by changing the disagreement outcome may actually increase the probability of an impasse.[3] It is also shown that attempts (like the common requirement in labor law to bargain "in good faith") to make commitment more difficult, with the goal of reducing the probability of impasse, may have perverse effects.

[2] See, for example, Harsanyi [9] and the references therein, Kalai and Smorodinsky [11], Nash [14, 15], and Roth [17] and the references therein. Notable exceptions are Cross [4] and Ashenfelter and Johnson [1]. But the models developed there are somewhat ad hoc, in that bargainers' motivations for behaving as they are assumed to do are weak. Chatterjee and W. Samuelson [2] have developed an interesting model, discussed in footnote 4 below, that explains the occurrence of disagreement by focusing on bargainers' uncertainty about each other's preferences. And Harsanyi [8, pp. 329–334] presents an example to show that under uncertainty, the fact that bargainers cannot make binding agreements before they have all relevant information about preferences and feasible outcomes may prevent them from reaching an agreement that is fully Pareto-efficient relative to the information that is collectively available. But Harsanyi makes no attempt to explain the occurrence of impasses, and the theory of bargaining outlined there and in Harsanyi and Selten [10], unlike the one developed here, would predict fully efficient outcomes in the absence of uncertainty about preferences and feasible outcomes. An interesting recent development, which explains the occurrence of impasses by exploring bargainers' incentives to maintain reputations for "toughness," is the work of Rosenthal and Landau [16].

[3] While the rationale for this assumption is rarely made explicit, it often appears to stem from an analogy between bargaining and individual behavior, where large costs (taking uncertainty and costs of decision-making into account) are more likely to be avoided than small costs. Although this analogy is superficially plausible, adopting it as an *assumption* is quite risky. There is little reason to suppose that bargaining, one of the most interactive of economic situations, is behaviorally analogous to individual decision-making in all respects.

The paper is organized as follows. Section 2 discusses modeling issues and presents the bargaining model, describing the commitment process and the rules that determine bargaining outcomes. Section 3 analyzes the model under the assumption of full noncooperative game-theoretic rationality; the solution concept employed is Harsanyi's [**8**] Bayesian Nash equilibrium, with an additional requirement of perfectness (see Selten [**21**]). Section 4 analyzes the model under the alternative assumption that bargainers use simple heuristics (rather than the assumptions of perfectness and full rationality) to evaluate the uncertain future consequences of current attempted commitments. It is shown that for a leading special case of these heuristics, the bargaining game can be reduced to a Prisoner's Dilemma, providing a simple "textbook" explanation of the fact that bargainers do not always manage to avoid disagreement and showing that the implied inefficiency can arise even if one is willing to accept only strategic-dominance arguments about bargainers' rational strategy choices. For a more general class of heuristics, a simple condition which guarantees that both bargainers will attempt commitment to incompatible positions, thereby risking impasse, is provided. The analysis of Sections 3 and 4 confirms Schelling's [**20**, Chapter 3] suggestion that uncertainty may enhance the strategic usefulness of attempting commitment, shows that Harsanyi's [**7**, p. 182; **9**, p. 187] claim that attempting commitment is irrational because it creates a risk of impasse is not valid if commitment is uncertain, and yields new insight into the properties of the "demand game" proposed by Nash [**15**] as a noncooperative model of bargaining to provide an alternative justification for the cooperative bargaining solution he axiomatized in [**14**].

Section 5 studies some of the comparative statics properties of the model, showing that commonly held beliefs about the effects of enlarging the contract zone and of making commitment more difficult may be invalid and are, at any rate, not justified on *a priori* grounds. Section 6 concludes by discussing some possible directions for future research in this area.

2. THE BARGAINING MODEL

This section outlines a model of the bargaining process that is simple, but rich enough to explain the occurrence of disagreements under reasonable behavioral assumptions. The theory follows Schelling's classic paper [**18**] in focusing on the commitment aspects of bargaining. Schelling defines commitment impressionistically and by way of examples, but the essential idea seems to involve making a demand and "burning one's bridges," or taking actions during the negotiation process that increase the future cost of backing down from one's demand. The potential benefits from such a strategy arise from the possibility that one's opponent will thereby become convinced that one will in fact not retreat, and that he will therefore decide to yield to one's demand. Schelling [**18**, pp. 295–96] suggests that the possibility of mutually incompatible commitments may explain the occurrence of impasses:

> In threat situations, as in ordinary bargaining, commitments are not altogether clear; each party cannot exactly estimate the costs and values to the other side of the two related actions involved in the threats; the process of commitment may be a progressive one, the commitments acquiring their firmness by a sequence of actions. Communication is often neither entirely impossible nor entirely reliable; while certain evidence of one's commitment can be communicated directly, other evidence must travel by newspaper or hearsay, or be demonstrated by actions. In these cases the unhappy possibility of both acts occurring, as a result of simultaneous commitment, is increased. Furthermore, the recognition of this possibility of simultaneous commitment becomes itself a deterrent to the taking of commitments.

(In the present context, Schelling's "threats" correspond to relying on the disagreement outcome rather than agreeing on a settlement.)

But Schelling [**18**] does not address further the question of why bargainers might attempt commitment. The puzzle is simple: attempting commitment creates a risk of impasse, which is generally Pareto-inefficient *ex post*.[4] And any distribution of outcomes that puts positive probability on an inefficient outcome is not *ex ante* Pareto-efficient either. This paper proposes a plausible explanation of why bargainers attempt commitment when there are feasible negotiated settlements that are better for both *ex ante* than attempting commitment.

Schelling [**18**] argues convincingly that commitment is an important component of real bargaining and that it typically involves significant elements of uncertainty and irreversibility. The uncertainty is intrinsic to the process, which is primarily psychological (and uncertain even to psychologists). The irreversibility arises primarily because attempting commitment involves making statements about one's relative evaluations of disagreement and agreement on one's position, and linking one's reputation to the maintenance of one's position. A union leader who, in the hope that management is listening, has told his members they should replace him if he fails to get them a given wage increase cannot comfortably back down from that position. If it later turns out that he cannot obtain that wage increase in negotiations, he may actually prefer a strike. (Of course his original preferences, which presumably better reflected those of his membership, are still the relevant ones for judging outcomes.) A management representative who has stated publicly that his company cannot grant the wage increase sought by the union without going out of business is in much the same position.

Schelling [**20**, Chapter 3] suggests that if attempting commitment is not *certain* to cause an impasse, it might be a viable bargaining strategy even when disagreement is extremely costly. The rest of this section discusses the issues that arise in building a model to confirm that uncertainty can provide a resolution of the puzzle posed above, and outlines the model.

First, it is important to note that an element of irreversibility is an essential component of a sensible model of the commitment process. As in all models with uncertainty, unless actions whose effects cannot be undone are taken before the

[4] Here and in what follows, "*ex ante*" and "*ex post*" refer to before and after the uncertainty inherent in the bargaining process is resolved.

uncertainty is resolved, uncertainty can have no lasting effect. In the present context, bargainers could reconsider their decisions whenever an impasse seemed imminent, and it would generally be irrational for them to do otherwise.

Uncertainty is equally essential. With irreversibility alone, questions of timing take on primary importance, as Schelling [**19**, Chapter 2 and Appendix B] pointed out. A bargainer who knew he could be the first to communicate an irrevocable demand to his opponent would find it to his advantage to do so, ending negotiations on the spot. And if bargainers must communicate their demands simultaneously, there is great *strategic* uncertainty, in the form of multiple Nash equilibria: under certainty, the best strategy is to demand a lot if one's opponent demands a little, and vice versa. Aside from the fact that in games with multiple equilibria and payoff functions that allow inefficient outcomes it is difficult to justify the assumption that players will coordinate their strategy choices so that a Nash equilibrium arises (see, however, Nash [**15**, pp. 131–136]); the resulting predictions are sensitive to the theory used to predict which, if any, equilibrium will arise. Since no such theory has yet been widely accepted (see, however, Harsanyi [**9**, Chapter 7] and Schelling [**19**, Chapter 4]), it seems a better research strategy to avoid uncertainty about which theory is appropriate by incorporating uncertainty into the game.

At the most general level, commitment, and the entire bargaining process, is a complex multistage game with incomplete information, in which bargainers make demands, take actions to increase the difficulty of retreating from their demands, learn from their opponents' actions, and periodically reconsider their strategy choices. In a detailed model of the process, bargainers' freedom to undo the effects of their decisions would be eroded only gradually over time, and uncertainty would be resolved more or less continually. While much of interest might be learned by building such a detailed model, I shall adopt the alternative strategy of building the simplest possible model in which attempting and succeeding at commitment are the outcomes of bargainers' rational decisions, and in which uncertainty and irreversibility can have an effect on the outcome. This model, which has two stages and abstracts from all uncertainty except that in the commitment process, is likely to share many of the important features of more realistic models. I believe that the approach taken here will best serve to elucidate the possibilities inherent in more general models, and in real bargaining.

A useful starting point for expositing the model is the "demand game" presented by Nash [**15**] as an alternative justification for his "fixed-threats" solution, developed in [**14**]. (In Nash [**15**], the demand game is later combined with a "threat game" to provide a model of what is now commonly called "variable-threats" bargaining. But the demand game also stands on its own as a model of fixed-threats bargaining.) In Nash's demand game, there is only one stage, in which bargainers simultaneously make demands, in utility terms. If the demands are compatible, in the sense of being collectively feasible, each bargainer gets the utility level he demanded; if the demands are incompatible, disagreement is the outcome.

Nash [15] observes that in his demand game, any Pareto-efficient pair of demands in the contract zone is in Nash (noncooperative) equilibrium; this is the basic source of the strategic uncertainty referred to above. He then characterizes his fixed-threats cooperative solution as the only Nash equilibrium of the demand game that is the limit of equilibria of "smoothed" games as the amount of smoothing goes to zero. (In the smoothed games, bargainers assume they are certain to get their demands if they are compatible, but that the probability of getting them, or, equivalently, of their compatibility, falls off rapidly toward zero as their distance from the set of feasible settlements increases. Nash suggests that these probabilities can be thought of as reflecting uncertainty about preferences and the information structure of the game.)

The model used here differs from Nash's in three respects. First, bargainers are not *required* to make demands (that is, to attempt commitment); one of their options is to bargain cooperatively, which entails no risk of impasse and leads, if both bargainers adopt that strategy, to a known Pareto-efficient compromise settlement. Second, if both bargainers make demands that are more than compatible and allow them to stand, they share the surplus according to a rule that generalizes Nash's but need not preclude an efficient outcome. Finally, the most important difference lies in a richer specification of how demands are made and backed up. In Nash's demand game, demands are simply irrevocable; thus, one might view his model (although he apparently did not) as a model of commitment in which commitments are completely certain. I shall adopt the alternative assumption that whenever a bargainer makes a demand, he also sets in motion a process that will make it costly, to an uncertain extent, for him to later accept less than his demand, but that he may freely choose to accept less, as long as he pays the cost. It is this cost-generating process by which bargainers give meaning to their demands.[5]

More precisely, bargaining is viewed as a two-stage process, in which bargainers are perfectly informed about everything except their costs of backing down. In the first stage, bargainers simultaneously decide whether or not to attempt commitment. An attempt, if one is made, consists of the announcement of a demand (in utility terms) and a draw from a probability distribution, whose realization is the cost (again in utility terms) that must be borne if the bargainer in question later decides to accept anything less than his demand. In the second stage, each bargainer learns his own, but not his opponent's, cost of backing

[5] A related model of bargaining is developed by Chatterjee and Samuelson [2]. Although they do not point out the connection, their model is essentially a generalization of Nash's demand game, in which the irrevocability of demands, or what I have called commitment, is certain, but bargainers are uncertain about each other's disagreement utilities. (Minor differences between their model and Nash's are that they make assumptions that imply a linear utility-possibility frontier, and employ a different rule for compromising when more-than-compatible demands have been made.) Chatterjee and Samuelson show that in their model, Nash equilibrium demands (which are, equivalently, "Bayesian" equilibrium demands in the game with incomplete information; see Harsanyi [8]) generally involve "shading,"or demanding more than one's disagreement utility. Given the uncertainty, shading makes it possible that with positive probability no bargain will be struck even if there is a nonempty contract zone. Thus, Chatterjee and Samuelson's model provides an alternative explanation of the occurrence of impasses.

down (the outcome of his draw); whether or not his opponent attempted commitment; and what demand, if any, he made. He then decides, taking into account this information, whether or not to retreat from his demand; if not, he is said to have "achieved commitment" to the position given by his first-stage demand. The second-stage part of his strategy takes the form of a rule that relates his action to the situation in which he finds himself. These second-stage decisions then determine the final outcome as in Nash's demand game, with the qualification noted above: incompatible demands from which neither bargainer retreats lead to the disagreement outcome; and compatible demands lead to a compromise that yields each bargainer at least what he demanded. In other situations, which cannot arise in Nash's game, the outcome is as follows: if exactly one bargainer has made a demand and not backed down from it, the Pareto-efficient settlement in which he gets his demand is the final outcome; and if neither bargainer has made a demand and stuck to it, the outcome is a compromise settlement. This compromise settlement is assumed to be Pareto-efficient, to avoid begging the question that lies at the heart of this paper: whether rational behavior in bargaining must lead to efficient outcomes.

In the above specification of the bargaining game, it is assumed that bargainers do not learn their costs of backing down until after their first-period decisions. In general multi-stage games with incomplete information, players can learn about their opponents by observing their actions in early stages and drawing inferences based on game-theoretic rationality (or other behavioral) assumptions. Players must, therefore, evaluate early-stage actions taking into account the later-stage effects of the information those actions reveal to opponents. (See Kreps and Wilson [**12**] for an especially clear discussion of these and other problems that arise in modeling behavior in multi-stage games with incomplete information.) These issues do not arise in my specification of the bargaining game, because a bargainer's choice of first-stage strategy is based on precisely the same information that is assumed to be available to his opponent, who can, therefore, draw no additional inferences from it. (Information is transmitted by second-stage actions, but by then, given my assumptions, it is too late for this information to change the outcome.) This fact allows complex issues of strategic information transmission, which are conceptually separate from the issues considered here (although certainly important in real bargaining, mainly in connection with information about preferences, however), to be avoided.

The bargaining environment can be formally described as follows. It includes two bargainers, indexed $i = 1, 2$. The index j, when it appears, refers to the bargainer other than i; and to avoid needless repetition of "$i = 1, 2$", i alone will be understood to refer to each bargainer. Each bargainer i is assumed to have preferences over agreements and probability distributions of agreements that can be represented by a von Neumann-Morgenstern utility function, denoted u^i. These preferences may reflect the anticipated effects of the current agreement on future negotiations, but bargainers are assumed not to contemplate coordinating their current bargaining strategies with strategies in future negotiations. Following Nash [**14**] and many others, I shall assume that the set of utility pairs, denoted U, that can be attained by negotiating an agreement is a closed, bounded, convex, and nonempty subset of two-dimensional Euclidean space.[6] And I shall consider only the "fixed-threats" case, where the *disagreement outcome*—denoted $(\underline{u}^1, \underline{u}^2)$ and defined as the pair of expected utilities bargainers associate with the (possibly uncertain) consequences of not negotiating an agreement—is independent of bargainers' actions. The disagreement outcome is assumed not to be Pareto-efficient in U. The *contract zone*, denoted V, is defined as the set of utility pairs $(u^1, u^2) \in U$ such that $u^i \geq \underline{u}^i, i = 1, 2$. For simplicity, it is assumed that the part of the boundary of the utility-possibility set that lies in V is strictly downward-sloping and differentiable. And a bargainer faced with a choice between outcomes between which he is indifferent will be assumed always to choose as his opponent prefers.

Let $\bar{u}^i$ denote the largest utility for bargainer i that is compatible with $(u^1, u^2) \in V$; given the above assumptions, $\bar{u}^i > \underline{u}^i$. Write the equation of the portion of the utility-possibility frontier that lies in V, which is downward-sloping by assumption, as $u^j \equiv \phi(u^i)$, and let $\psi \equiv \phi^{-1}$, so that $u^i \equiv \psi(u^j)$ also represents the utility-possibility frontier in V. (Even though the superscript i is normally understood to refer to either bargainer, this notation will be used to indicate whether the utility-possibility frontier is viewed as parameterized by u^i or u^j as they appear elsewhere in the argument; no confusion should result.) A *compatible* pair of utilities is defined as one such that, if u^1 and u^2 are the utilities, $(u^1, u^2) \in U$; *more-than-compatible* utilities lie, as a pair, in the interior of U; and *just-compatible* utilities lie on the utility-possibility frontier.

The outcome function is formally specified as follows. Without loss of generality, we may restrict bargainer i's possible commitment positions to the interval $[\underline{u}^i, \bar{u}^i]$. If neither bargainer achieves commitment, the outcome is a Pareto-efficient compromise, denoted $(\tilde{u}^1, \tilde{u}^2)$; it is assumed that $\underline{u}^i < \tilde{u}^i < \bar{u}^i$, so that $(\tilde{u}^1, \tilde{u}^2) \in V$. If both bargainers achieve commitment, to incompatible positions, the outcome is $(\underline{u}^1, \underline{u}^2)$. If bargainers achieve commitment to compatible positions $(\hat{u}^1, \hat{u}^2)$, the outcome $(u^1, u^2) \in U$ satisfies $u^1 \geq \hat{u}^1$ and $u^2 \geq \hat{u}^2$; it is not required to be Pareto-efficient. However, u^i is assumed to be a strictly increasing function of $\hat{u}^i$ in this case, as long as the positions $(\hat{u}^1, \hat{u}^2)$ remain compatible. Nash's [**15**] rule, where $u^i = \hat{u}^i$, is a simple example of one satisfying this assumption. Finally, if bargainer i alone achieves commitment, to a position $\hat{u}^i \leq \bar{u}^i$, the outcome is $(\hat{u}^i, \phi(\hat{u}^i))$.

The justification of this outcome function is clear, given the above definition of achieving commitment. Figure 1, which illustrates a typical bargaining environment, may make it easier to remember the notation; the contract zone is the shaded portion of the utility-possibility set in the figure.

To complete the specification of the model, consider the determinants of the costs of backing down for bargainers who have attempted commitment. In principle, a bargainer's cost might be systematically related to the position he

[6] Convexity could be dispensed with, but since bargainers are allowed to choose strategies that lead to probabilistic outcomes, it seems unnatural to prohibit them from negotiating random settlements. Without such a prohibition, U is well known to be convex.

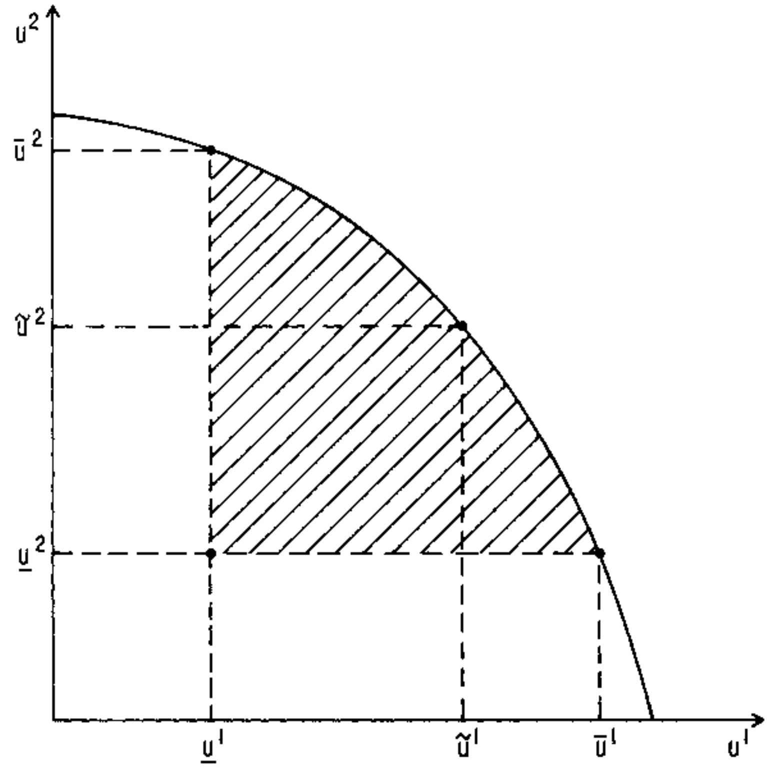

FIGURE 1

attempts commitment to; and he might, by his choice of negotiating tactics, also be able to exert an influence on the distribution of these costs that is separate from the influence of his commitment position. Introspection and casual empiricism have not made clear to me whether it should tend to be harder to back down from an extreme position or a moderate one, and I have even less intuition about how this effect should interact with bargaining tactics. For these reasons, and because these complications seem to add little to what light the model sheds on bargaining, I shall study explicitly only the leading special case where the distribution of costs, given that a bargainer has attempted commitment to some position, is independent of that position and of his other actions. The only influence he can exert on these costs is by exercising his option not to attempt commitment. This assumption does not, of course, imply that the probability of achieving commitment is necessarily independent of the commitment position as well. Additional simplicity is gained at little expense by assuming that bargainers' cost distributions are independent.

More formally, let $F^i(c^i)$ denote the distribution function of bargainer i's cost, in utility terms. F^i and F^j are assumed to represent independent and, for most of the analysis, continuous distributions, and to be common knowledge. The supports of the F^i contain only nonnegative values, so there is never a "subsidy" for backing down.

3. PERFECT BAYESIAN NASH EQUILIBRIUM

This section analyzes the model developed in Section 2 when the solution concept is Harsanyi's [**8**] Bayesian Nash equilibrium, with an additional requirement of perfectness (see Selten [**21**]). These assumptions embody what has come to be known as full game-theoretic rationality in multi-stage noncooperative games with incomplete information; they therefore provide the most stringent possible test of the model's ability to rationalize the occurrence of impasses in bargaining. The result obtained here, that impasses can occur with positive probability in equilibrium, does not stem from bargainers having irrational expectations or making suboptimal decisions.

A Bayesian Nash equilibrium, in the context of the model, is simply a Nash equilibrium in first-stage actions and in the rules that relate bargainers' second-stage actions to the situation created by first-stage actions and to their observed costs of backing down. Each bargainer responds optimally to his opponent's strategy choice, taking into account its implications in view of his probabilistic beliefs about his opponent's cost. An additional requirement of perfectness will be imposed, which has the effect of ruling out equilibria in which a bargainer makes implausible bluffs about what he will do in the second stage that are not called in equilibrium. To put it another way, perfectness requires bargainers always to respond optimally in the second stage to any situation that might be created by their actions in the first stage. The resulting equilibrium concept is equivalent to Kreps and Wilson's [**12**] notion of "sequential rationality."

The Bayesian Nash equilibrium is both the natural generalization of the ordinary Nash equilibrium to games with incomplete information and a natural extension of the familiar concept of rational-expectations equilibrium to situations where strategic interactions are important. It has the further, closely related advantage that rational players who have chosen strategies that are in Bayesian Nash equilibrium, and are given an opportunity to revise them before the uncertainty in the game is resolved, will never do so.

When a Bayesian Nash equilibrium exists in the model (the existence question is discussed further below), it can be characterized as follows. First, note that no equilibrium can have neither bargainer attempting commitment: in such a situation, either bargainer could attempt commitment to a position better than the compromise settlement for himself and not back down in the second stage (which is consistent with perfectness), obtaining his demand. For similar reasons equilibrium cannot, with one possible exception, involve only one bargainer attempting commitment: in such cases, that bargainer would always have an incentive to increase his demand. The exception occurs when bargainer i is already demanding $\bar{u}^i$, in which case it might be an equilibrium strategy for bargainer j to acquiesce. But in that case, j might just as well attempt commitment to $\underline{u}^j$, a convention that allows the simple classification of equilibria presented below.

It follows from the above observations that equilibrium can always be taken to involve each bargainer attempting commitment to some position; but these positions need not be incompatible. In fact, there are two significantly different types of equilibria in this model: those with compatible commitments and those with incompatible commitments. Given my assumption that the utility a bargainer gets if both bargainers achieve commitment, to compatible positions, increases strictly with his position as long as compatibility is preserved, it is immediately clear that compatible commitments can be in equilibrium only if they are just compatible. It is also clear that a bargainer can never gain by unilaterally backing down from a commitment position that is just compatible with that of his opponent, even if the cost of doing so turns out to be zero. Thus, equilibrium involves either just-compatible commitments (in which it can be assumed without loss of generality that neither bargainer backs down in the second stage, no matter what his cost) or incompatible commitments. The just-compatible commitment equilibria are the ones identified by Nash **[15]**, although in general, not all of those equilibria are equilibria here as well. Relaxing Nash's assumption that demands are certain to be irrevocable both destroys some compatible-commitment equilibria and allows the existence of some incompatible-commitment equilibria, as we shall see.

Suppose that bargainers have attempted commitment to the incompatible positions $(\hat{u}^i, \hat{u}^j)$ in the first stage. How do they decide when to back down in the second stage? Under my assumption of perfectness, the rules that answer this question must be in Nash equilibrium in the second-stage game created by bargainers' choices of demand in the first stage. It is clear that best-response decision rules will involve cutoff levels of costs, below which bargainers will back down, and above which they will stand firm. Suppose the F^i represent continuous distributions, so that how bargainers break ties is unimportant. The equilibrium cutoff levels are determined as follows.

Let d^i denote bargainer i's cutoff level of costs. When $c^i \leq d^i$, he backs down in the second stage; otherwise, he stands firm. Given the definitions and assumptions, it is easy to verify that bargainer i's expected payoff in this game, when demands are incompatible, is given by

$$(3.1) \qquad w^i(d^i, d^j) \equiv F^i(d^i)\left[F^j(d^j)\tilde{u}^i + \left(1 - F^j(d^j)\right)\psi(\hat{u}^j)\right] + \left[1 - F^i(d^i)\right]\left[F^j(d^j)\hat{u}^i + \left(1 - F^j(d^j)\right)\underline{u}^i\right] - \int_0^{d^i} c^i f^i(c^i)\, dc^i,$$

where f^i denotes the density associated with F^i. This expression for $w^i(d^i, d^j)$ is derived by considering the four possible combinations of bargainers backing down and standing firm, weighted by their probabilities, and subtracting the expected costs incurred by backing down. An easy computation reveals that

$$(3.2) \qquad w_1^i(d^i, d^j) \equiv f^i(d^i)\left[F^j(d^j)(\tilde{u}^i - \hat{u}^i) + \left(1 - F^j(d^j)\right)\left(\psi(\hat{u}^j) - \underline{u}^i\right) - d^i\right].$$

Suppose, for example, that the support of F^i is the interval $[0, e^i]$, where $e^i > \bar{u}^i - \underline{u}^i$. Then it is not hard to see that i's best choice of d^i must be interior, and must, therefore, satisfy $w_1^i(d^i, d^j) = 0$, which holds, given that $f^i(d^i) > 0$ whenever $d^i \in [0, e^i]$, if and only if the term in brackets on the right-hand side of (3.2) equals zero. Another easy computation reveals that whenever $w_1^i(d^i, d^j) = 0$,

$$(3.3) \qquad w_{11}^i(d^i, d^j) = -f^i(d^i) < 0.$$

Thus, it follows that the w^i are strictly quasi-concave. Since they are also clearly continuous, and the strategy spaces can be taken to be compact and convex without ruling out any "good" strategies, Debreu's Social Equilibrium Existence Theorem (restated as Theorem 1 in Dasgupta and Maskin [5]) implies the existence of a pure-strategy Nash equilibrium in the second-stage game, given any incompatible choices of $\hat{u}^i$ and $\hat{u}^j$ in the first stage. In general, it is not possible to prove uniqueness without considerably stronger restrictions on the payoff functions, but I shall assume uniqueness to facilitate the rest of this discussion.

Given this determination of equilibrium strategies in the second stage, a perfect Bayesian Nash equilibrium in the entire game can be constructed, in dynamic programming fashion, by finding Nash equilibrium demands in the first stage, evaluating the second-stage consequences of first-stage actions using the already determined second-stage equilibrium strategy rules. Of course, that there is always an equilibrium in pure strategies in the second-stage game does not imply that there is always an equilibrium, in pure *or* mixed strategies, in the game taken as a whole. In fact, it does not appear possible to prove a general existence result under the maintained assumptions, due to the discontinuities that may arise in the payoff functions when bargainers' demands are just compatible.[7] I shall provide a partial remedy for this by exhibiting a class of examples in which equilibrium generally exists and, later, by providing restrictive, but not unreason-

[7] The discontinuities prevent the use of Debreu's Social Equilibrium Existence Theorem (restated as Theorem 1 in Dasgupta and Maskin [5]). Dasgupta and Maskin have shown [5, Theorems 3 and 4 and n. 16] that at least a mixed-strategy Nash equilibrium will exist, given the compactness and convexity of the strategy spaces, provided only that payoff functions are upper semi-continuous and graph continuous. (Graph continuity requires, roughly, that the graph of a player's payoff, viewed as a function of his own actions, vary continuously with changes in others players' actions.) If, in addition, the payoff functions are quasi-concave, pure-strategy existence is obtained. These more general results do not imply existence in the present model, because it is not generally true that both payoff functions are upper semi-continuous when bargainers' demands are just compatible. The intuitive reason for this is that by moving from a just-compatible to a just-incompatible demand, a bargainer may induce both his opponent and himself to back down frequently enough that the compromise settlement occurs with significant probability. (Allowing for such changes in second-stage actions is the proper way to evaluate the consequences of such a change, since it is the rule that determines the bargainer's second-stage actions, rather than the action itself, that must be held constant in viewing the perfect Nash equilibrium as a Nash equilibrium in first-stage demands, when the payoff functions are dynamic-programming value functions.) If a bargainer does this starting from a configuration of demands that is worse for himself than the compromise settlement, it can make him better off, even though there is also a risk of impasse that was not present before. This failure of upper semi-continuity appears to be intrinsic to the model rather than to the present formulation, although of course it does not *imply* nonexistence of equilibrium.

able, assumptions about behavior and the bargaining environment under which existence is guaranteed.

The remainder of this section analyzes in detail a class of simple examples in which pure-strategy existence can be guaranteed under easily interpretable and reasonable assumptions, and in which for some parameter configurations there are only incompatible-commitment equilibria. It is hoped that the analysis of these examples will also serve to illustrate better the workings of the model.

Normalize $\underline{u}^i = \underline{u}^j = 0$ and $\bar{u}^i = \bar{u}^j = 1$, and depart from the earlier assumptions by letting F^i and F^j be Bernoulli distributions, with probabilities q^i and q^j respectively of yielding a cost greater than unity, and probabilities $1 - q^i$ and $1 - q^j$ of yielding zero costs. I shall assume to avoid trivialities that q^i and q^j lie strictly between zero and one. When costs are high in this case, bargainers always stand firm in the second stage; when costs are low, they are zero. Thus, costs are never paid; they serve only to make it effectively "impossible" to back down, with given probability. As will become clear, however, the game is far from trivial even in this simple case.

The reader can easily verify that my analysis of the necessary conditions that must be satisfied at compatible-commitment equilibria remains valid here: with inessential exceptions, any such equilibria must involve just-compatible demands and must not be vulnerable to defections involving incompatible demands. In the remainder of this section, I shall argue first that the examples are capable of supporting compatible-commitment equilibria for some parameter configurations; second, that there are configurations where incompatible-commitment equilibria always exist; and finally, that there are configurations where only incompatible-commitment equilibria exist.

While it is difficult to characterize compatible-commitment equilibria in general, it is not hard to show that when $\tilde{u}^i \geq 1 - q^j$ and $\tilde{u}^j \geq 1 - q^i$, $(\hat{u}^i, \hat{u}^j) = (\tilde{u}^i, \tilde{u}^j)$ is such an equilibrium. Consider possible defections from the hypothesized equilibrium configuration by bargainer i, and recall the perfectness requirement, which implies that the consequences of a defection must be evaluated under the assumption that bargainer j responds optimally (in the sense of choosing an equilibrium strategy) to the second-stage situation created by the defection. To evaluate these consequences, we must first consider bargainers' choices of strategy in the second stage when they have attempted commitment to incompatible positions in the first stage. Recall that at perfect equilibria, bargainers always stand firm when costs are high, because the costs outweigh any possible gains. Bargainers' unconditional expected payoffs, as a function of the actions that they take when costs are low and taking into account their optimal actions when costs are high, are as follows, with i's expected payoff given first in each case. If bargainer i stands firm, if j also stands firm the outcome is $(\underline{u}^i, \underline{u}^j)$, and if j backs down the outcome is $(q^j\underline{u}^i + (1 - q^j)\hat{u}^i, q^j\underline{u}^j + (1 - q^j)\phi(\hat{u}^i))$. If bargainer i backs down, if j stands firm the outcome is $(q^i\underline{u}^i + (1 - q^i)\psi(\hat{u}^j), q^i\underline{u}^j + (1 - q^i)\hat{u}^j)$, and if j also backs down the outcome is $(q^iq^j\underline{u}^i + q^i(1 - q^j)\hat{u}^i + (1 - q^i)q^j\psi(\hat{u}^j) + (1 - q^i)(1 - q^j)\tilde{u}^i, q^iq^j\underline{u}^j + q^i(1 - q^j)\phi(\hat{u}^i) + (1 - q^i)q^j\hat{u}^j + (1 - q^i)(1 - q^j)\tilde{u}^j)$.

Given the definition of the outcome function when commitments are compatible, from $(\hat{u}^i, \hat{u}^j) = (\tilde{u}^i, \tilde{u}^j)$ only a defection to some incompatible commitment $\mathring{u}^i > \tilde{u}^i$ could possibly yield a better outcome for bargainer i. Inspection of the conditional payoff function given above reveals that in the case under consideration, any such defection makes backing down when costs are low a dominant strategy for bargainer j (that is, optimal for either of i's possible second-stage actions) no matter what commitment position in $(\tilde{u}^i, \bar{u}^i)$ bargainer i defects to. Thus, setting $\mathring{u}^i$ at (or near) unity is the optimal defection of this type for bargainer i, and given that $\tilde{u}^i \geq 1 - q^j$, backing down when costs are low is his optimal second-stage policy. This defection yields bargainer i an expected payoff of $q^i(1 - q^j) + (1 - q^i)\tilde{u}^i$, which is unprofitable when compared to his original payoff if and only if $\tilde{u}^i \geq 1 - q^j$. Given the symmetry of the situation aross bargainers, imposing the analogous condition $\tilde{u}^j \geq 1 - q^i$ for bargainer j guarantees that $(\hat{u}^i, \hat{u}^j) = (\tilde{u}^i, \tilde{u}^j)$ is, in fact, a compatible-commitment equilibrium. Generally, it is not unique; I have singled this one out because it is easy to identify.

I shall now argue that when q^i and q^j are near enough to zero that $\tilde{u}^j < 1 - q^i$ and $\tilde{u}^j < 1 - q^i$, there always exist pure-strategy incompatible-commitment equilibria. This confirms Schelling's [**20**, Chapter 3] intuition that low probabilities of success (because they imply a low probability that one's opponent will succeed) tend to favor attempting commitment to incompatible positions. Conceivably, for some values of the parameters there are equilibria where one, or even both, of the $\hat{u}^i \leq 1(=\bar{u}^i)$ constraints on bargainers' demands are binding. Since in these cases the analysis is unduly complicated by possible multiplicity of equilibria in the second-stage game, I shall confine my discussion to interior equilibria, where this multiplicity is easily dealt with. The analysis yields a simple characterization of incompatible-commitment equilibria.

The analysis of interior incompatible-commitment equilibria is greatly simplified by the observation that such equilibria must have both bargainers backing down in the second stage if (and only if) costs are low. Any other combination of second-stage strategies yields either bargainer i less than $\psi(\hat{u}^j)$ or bargainer j less than $\phi(\hat{u}^i)$; since bargainers can unilaterally guarantee themselves these amounts by defecting to just-compatible commitments in the first stage and standing firm in the second stage (which is consistent with perfectness), such strategies are incompatible with equilibrium. That bargainers are in equilibrium in the second-stage game backing down when costs are low places the following restrictions on interior incompatible-commitment equilibria, recalling the $\underline{u}^i = 0$ normalization:

$$q^i(1 - q^j)\hat{u}^i + (1 - q^i)q^j\psi(\hat{u}^j) + (1 - q^i)(1 - q^j)\tilde{u}^i \geq (1 - q^j)\hat{u}^i,$$

which reduces to

$$\text{(3.4)} \qquad \hat{u}^i \leq \frac{q^j}{1 - q^j}\psi(\hat{u}^j) + \tilde{u}^i;$$

and

$$q^i(1-q^j)\psi(\hat{u}^i)+(1-q^i)q^j\hat{u}^j+(1-q^i)(1-q^j)\tilde{u}^j \geq (1-q^i)\hat{u}^j,$$

which reduces to

$$\hat{u}^j \leq \frac{q^i}{1-q^i}\phi(\hat{u}^i)+\tilde{u}^j. \tag{3.5}$$

An interior incompatible-commitment equilibrium must have (3.4) and (3.5) satisfied with equality, because in these examples, given the anticipated equilibrium in the second-stage game, the probabilities of backing down are independent of $\hat{u}^i$ and $\hat{u}^j$ as long as (3.4) and (3.5) are satisfied. Thus, setting $\hat{u}^i$ or $\hat{u}^j$ lower than these inequalities allow, but still incompatible, gives up something without gaining any advantage in return. In addition, as noted above, an equilibrium with incompatible commitments must yield bargainer i at least $\psi(\hat{u}^j)$ and bargainer j at least $\phi(\hat{u}^i)$; thus,

$$q^i(1-q^j)\hat{u}^i+(1-q^i)q^j\psi(\hat{u}^j)+(1-q^i)(1-q^j)\tilde{u}^i \geq \psi(\hat{u}^j) \tag{3.6}$$

and

$$q^i(1-q^j)\phi(\hat{u}^i)+(1-q^i)q^j\hat{u}^j+(1-q^i)(1-q^j)\tilde{u}^j \geq \phi(\hat{u}^i). \tag{3.7}$$

Substituting (3.4), with equality, into (3.6) and simplifying yields $\tilde{u}^i \geq \psi(\hat{u}^j)$, which is equivalent, given the Pareto efficiency of $(\tilde{u}^i, \tilde{u}^j)$, to $\hat{u}^j \geq \tilde{u}^j$. Similar use of (3.5), with equality, reduces (3.7) to $\hat{u}^i \geq \tilde{u}^i$. Given the $\hat{u}^i \leq 1$ and $\hat{u}^j \leq 1$ constraints, these conditions are automatically satisfied when (3.4) and (3.5) are satisfied with equality. So the question of existence of interior incompatible-commitment equilibria reduces to the question: Can (3.4) and (3.5) both be satisfied with equality for some $(\hat{u}^i, \hat{u}^j)$ in $(\tilde{u}^i, 1)\times(\tilde{u}^j, 1)$?

To see that they can, substitute the equality form of (3.5) into the equality form of (3.4) to obtain

$$\hat{u}^i = \frac{q^j}{1-q^j}\psi\left[\frac{q^i}{1-q^i}\phi(\hat{u}^i)+\tilde{u}^j\right]+\tilde{u}^i. \tag{3.8}$$

When $\hat{u}^i = \tilde{u}^i$, the right-hand side of (3.8) becomes $q^j\psi[\tilde{u}^j/(1-q^i)]/(1-q^j)+\tilde{u}^i$, so the left-hand side is less than the right-hand side if and only if $\tilde{u}^j < 1-q^i$. When $\hat{u}^i = 1$, the right-hand side simplifies to $\tilde{u}^i/(1-q^j)$, so the left-hand side exceeds the right-hand side if and only if $\tilde{u}^i < 1-q^j$. It therefore follows by the Intermediate Value Theorem from the continuity of ψ and ϕ that, when $\tilde{u}^j < 1-q^i$ and $\tilde{u}^i < 1-q^j$, there exists a value of $\hat{u}^i$ strictly between $\tilde{u}^i$ and unity that satisfies (3.8). Now consider the corresponding value of $\hat{u}^j$, as given by the equality form of (3.5). When $\hat{u}^i < 1$, $\hat{u}^j > \tilde{u}^j$. And when $\hat{u}^i > \tilde{u}^i$, that ϕ is strictly decreasing implies that $\hat{u}^j < \tilde{u}^j/(1-q^i)$. Thus, if $\tilde{u}^j < 1-q^i$ as assumed above, the corresponding value of $\hat{u}^j$ also satisfies the desired restrictions, and we have an interior incompatible-commitment equilibrium, which is completely characterized by the equality forms of (3.4) and (3.5). It does not seem possible to establish uniqueness in general, however.

This section closes with an example to show that there are parameter configurations in the region where incompatible-commitment equilibria always exist for which there are no compatible-commitment equilibria. The example will be constructed by deriving a necessary condition that compatible-commitment equilibria must satisfy, and then constructing a parameter configuration where this necessary condition cannot be satisfied; it is hoped that this will be more informative than simply pulling the parameter values in question out of a hat.

Since any equilibrium pair of compatible demands must be just compatible, either $\hat{u}^i \leq \tilde{u}^i$ or $\hat{u}^j \leq \tilde{u}^j$. Suppose the first for definiteness; since the example will be symmetric across bargainers, this involves no loss of generality. Given that $\hat{u}^i \leq \tilde{u}^i$, (3.4) holds with strict inequality for any $\hat{u}^j < 1$. Thus, by simple first-order stochastic dominance arguments, if bargainer j defects to such an incompatible commitment position, backing down (when costs are low) is a strictly dominant strategy for bargainer i in the second-stage game. This is true for *any* $\hat{u}^j < 1$, so bargainer j can guarantee himself a payoff as close as desired to $1 - q^i$ by defecting to a position near unity and standing firm in the second stage. Imposing the perfectness requirement can only raise this expected payoff further, since bargainer i has a dominant second-stage strategy in this case. It follows that a necessary condition for the demands $(\hat{u}^i, \hat{u}^j)$ to be a compatible-commitment equilibrium is that $\hat{u}^j \geq 1 - q^i$ if $\hat{u}^j \geq \tilde{u}^j$.

Consider a compatible pair of demands $(\hat{u}^i, \hat{u}^j)$ with $\hat{u}^j \geq 1 - q^i > \tilde{u}^j$ and $\hat{u}^i = \psi(\hat{u}^j)$. Bargainer i has the option of defecting to $\hat{u}^i + \epsilon$, where ϵ is near zero. Since $\hat{u}^i < \tilde{u}^i$, this strategy is potentially advantageous in that it may induce bargainer j to back down when costs are low in order to reduce the risk of impasse. Of course, this risk of impasse is costly for i as well, but the costs may be outweighed by the increased probability of getting the compromise settlement. For such a defection to help, two things must be true: the defection must cause bargainer j to back down in the second stage, and the resulting expected payoff for i must exceed $\psi(\hat{u}^j)$. The first condition is just inequality (3.5), which simple algebra reveals to be satisfied with strict inequality for all $\hat{u}^j \in [1 - q^i, 1]$ and for $\hat{u}^i$ near $\psi(\hat{u}^j)$, provided that $\tilde{u}^j > (1 - 2q^i)/(1 - q^i)$. Impose this and the symmetric condition $\tilde{u}^i > (1 - 2q^j)/(1 - q^j)$; these are easily seen to be compatible with the $1 - q^i > \tilde{u}^j$ and $1 - q^j > \tilde{u}^i$ restrictions. They are also compatible with the fact that $\tilde{u}^i = \psi(\tilde{u}^j)$ when ψ is linear and $q^i = q^j > 1/3$, for example. The second condition, for small ϵ, requires that

$$q^i(1 - q^j)\hat{u}^i + (1 - q^i)q^j\psi(\hat{u}^j) + (1 - q^i)(1 - q^j)\tilde{u}^i > \psi(\hat{u}^j). \tag{3.9}$$

Given that $\hat{u}^i = \psi(\hat{u}^j)$, (3.9) reduces to

$$(1 - q^i)(1 - q^j)(\tilde{u}^i - \hat{u}^i) > q^i q^j \hat{u}^i, \tag{3.10}$$

which, recalling that $\hat{u}^i = \hat{u}^i - \underline{u}^i$ given the normalization, is easily interpretable in terms of weighing the benefits of the greater probability of compromise against the costs of the greater risk of impasse. For defections of the kind being analyzed to rule out all possible compatible-commitment equilibria, (3.10) must hold for all $\hat{u}^i \in [0, \psi(1 - q^i)]$, given that $\hat{u}^j \geq 1 - q^i$ at any candidate for an equilibrium; preserving symmetry guarantees that the analogous condition for bargainer j will be satisfied.

Now suppose that ψ is linear, the $\bar{u}^i = \bar{u}^j = 1/2$, and that $q^i = q^j = 7/20$. This preserves symmetry, and it is easy to check that for these parameter values, the conditions for a successful defection are always met. In particular, (3.10) is satisfied for all $\hat{u}^i < 169/436$, but $\hat{u}^i \leq \psi(1 - q^i) = q^i = 7/20$, which is less than 169/436. Thus, there can exist *only* incompatible-commitment equilibria for these parameter values.

4. EQUILIBRIUM WITH EXPECTATIONAL HEURISTICS

Assumptions of full game-theoretic rationality such as those maintained in Section 3 provide a useful discipline, whose value is obvious when the goal is to show that the occurrence of inefficient bargaining outcomes is compatible with rational behavior by bargainers. And it is tempting simply to view perfect Bayesian Nash equilibrium as the "right" equilibrium concept for noncooperative multi-stage games with incomplete information, and to reject all others as being irrational or arbitrary. But in my opinion, it would be unfortunate if the search for truly descriptive behavioral assumptions in such games were to end there.

My reasons for this opinion are as follows. Suppose we give the "fully rational" theory the greatest possible benefit of the doubt by accepting the usual iterative story about how players come to choose equilibrium strategies, in which they are free to keep revising their actions until both are satisfied. Suppose further, for the sake of argument, that the implied dynamic process converges. Even under these circumstances, in the model of this paper, players are not relieved of the need to formulate expectations about their opponents' actions, because irreversible actions whose payoffs depend on those expectations must be taken in the first stage. Most will agree that it is reasonable to assume that what a player does in the first stage can be viewed as maximizing expected von Neumann-Morgenstern utility, where the expectation is taken over his probability distributions of his cost and of his opponent's future actions. The question is: Where does the latter distribution come from?

The standard story among game theorists is that players are (except as limited by incomplete information) fully informed about the game, assume that their opponents are fully rational, and actually compute self-confirming strategy rules for all players under these assumptions. Bayesian Nash equilibrium strategies (and, if best responses are unique, only such strategies) are self-confirming. Thus, unless bargainers play strategies that are in Bayesian Nash equilibrium, at least one will eventually have to revise his probabilistic expectations about opponents' actions. (Compare the concept of "stable conformistic expectations" that characterizes Harsanyi and Selten's [**10**, pp. P92–P93] "strict" equilibrium points.) This is why the Bayesian Nash equilibrium is a natural generalization of the rational-expectations equilibrium to environments where strategic interactions are important.

Given the great complexity of the simplified bargaining game analyzed in Section 3, it seems a reasonable conclusion that the search for a sensible solution concept to describe real bargainers should not end here. But the story just outlined, it can be argued, is a straw man. What is really being assumed is not this kind of unlimited computational skill, but rather that players who live in a population of rational players will come to learn the correct probability distributions of what their opponents will do in various strategic situations. While this view may have some justification, its theoretical underpinnings are extremely weak. And even if the learning process could be counted on to converge to the correct distributions, this view's reliance on the assumption that players have spent a long time in a stable environment (but are not repeatedly matched with the same players, for then the situation would eventually cease to be one of incomplete information) greatly limits its applicability, in a way that does not also limit the applicability of the approach about to be proposed.

In this section, I shall study the implications of an intermediate position. It will be assumed that bargainers reach a Nash equilibrium in first-stage actions, but that their expectations about what will happen in the second stage are formed more simply than fully "rational" expectations would be. This approach has two main advantages: it is uncomplicated enough to be a reasonable candidate for a descriptive model, and it does not ignore the most immediate (first-stage) strategic aspects of bargainers' choices of demand. In evaluating this approach, the reader may find it useful to recall to what extent he used game-theoretic rationality assumptions in deciding on the best strategy to pursue in the negotiations accompanying his last purchase of a house. The analysis that follows attributes, as I hope the reader's recollections will show, a reasonable level of strategic sophistication to bargainers, and leads to conclusions that differ in some interesting respects from, but basically confirm, the results in Section 3. The assumptions made here also provide a framework in which it is convenient to study the relationship of the probability of impasse to the bargaining environment, as is done in Section 5.

I shall begin with the simplest possible assumption about expectations: that the probability bargainer i assigns to his deciding not to back down in the second stage is a constant, denoted p_i, and assumed to lie strictly between zero and unity. It is also assumed that bargainers view their decisions whether or not to back down as probabilistically independent events (a natural extension from Section 3, where independence of the cost distributions implies conditional independence of bargainers' decisions whether or not to back down), and, for simplicity, that both share the same perception of the probabilities of these events. This case is closest in spirit to the examples of Section 3, in which the equilibrium probabilities of backing down are constant, as long as bargainers' commitment positions leave them enough incentive to back down when costs are low. The present assumption fails to be fully rational because it ignores these constraints on the commitment positions. Finally, I shall maintain the assumption made in the examples of Section 3, that costs are either zero or prohibitively high. This seems to allow the points of this and the next section to be made most simply, and the reader can easily check how things would change with more general assumptions about the cost distributions.

When the probabilities are constant, a striking result is obtained. A strategy is said to *dominate* another strategy if it yields the player who employs it an outcome at least as good, no matter what strategy his opponent employs. I shall now argue that in this case, successive deletion of dominated strategies reduces the bargaining game to a Prisoner's Dilemma. In it, each bargainer has two strategies: not attempting commitment, and attempting commitment to the position in the contract zone most favorable to him. This is so in spite of the fact that neither of these strategies *alone* dominates all other strategies. After the deletion of dominated strategies, the latter strategy dominates the former for each player, so that if bargainers do not play dominated strategies, bargainers attempt commitment to $(\hat{u}^i, \hat{u}^j) = (\overline{u}^i, \overline{u}^j)$ and there is a positive probability $p_1 p_2$ of impasse.[8] Since the impasse outcome $(\underline{u}^1, \underline{u}^2)$ is not Pareto-efficient, the distribution of outcomes that results is not ex ante Pareto-efficient, even though bargainers could have avoided it only by playing dominated strategies.

The argument proceeds in three steps. First, I shall show that attempting commitment to a position $\hat{u}^i \leq \tilde{u}^i$ is dominated by not attempting commitment. Then, I shall observe that attempting commitment to a position $\hat{u}^i > \overline{u}^i$ is dominated by attempting commitment to $\overline{u}^i$. And finally, I shall argue that after deletion of the above dominated strategies, attempting commitment to a position $\hat{u}^i$ such that $\tilde{u}^i < \hat{u}^i < \overline{u}^i$ is also dominated by attempting commitment to $\overline{u}^i$.

To see that attempting commitment to $\hat{u}^i \leq \tilde{u}^i$ is dominated by not attempting commitment, note that there is no difference between these strategies unless the attempt succeeds. Assuming it does, bargainer i obtains utility $\hat{u}^i$ if bargainer j does not achieve commitment; u^i, where $\hat{u}^i \leq u^i \leq \psi(\hat{u}^j)$, if j achieves commitment to a compatible position $\hat{u}^j$ (where the latter inequality follows from the requirements that $u^j \geq \hat{u}^j$ and $(u^i, u^j) \in U$); and $\underline{u}^i$ if j achieves commitment to an incompatible position $\hat{u}^j$ or to a position $\hat{u}^j > \overline{u}^j$. If, on the other hand, bargainer i does not attempt commitment, he obtains, respectively, utilities $\tilde{u}^i, \psi(\hat{u}^j)$, and $\psi(\hat{u}^j)$ or $\underline{u}^i$. In each case, i does at least as well by not attempting commitment.

Attempting commitment to $\hat{u}^i > \overline{u}^i$ is dominated by attempting commitment to $\overline{u}^i$ simply because if commitment is achieved, the former strategy always yields utility $\underline{u}^i$, while the latter sometimes yields $\underline{u}^i$, but also yields $\tilde{u}^i$ and $\overline{u}^i$ with positive probabilities.

[8] Since there is no guarantee that bargainers' expectations about second-stage events are correct, $p_1 p_2$ need not be the "true" probability of impasse, which depends on how bargainers actually behave in the second stage. To keep the discussion as simple as possible, I shall maintain throughout the practice of making *ex ante* judgments using bargainers' own expectations.

Finally, attempting commitment to $\hat{u}^i$, where $\tilde{u}^i < \hat{u}^i < \bar{u}^i$, is dominated by attempting commitment to $\bar{u}^i$. To see this, note first that p_i is the same for all $\hat{u}^i$. By the above arguments, only cases where bargainer j does not attempt commitment or attempts commitment to $\hat{u}^j$, where $\tilde{u}^j < \hat{u}^j \leq \bar{u}^j$, need be considered. All strategies under consideration are equivalent unless the attempted commitment succeeds. If bargainer i succeeds in committing himself to position $\hat{u}^i$, he obtains utility $\hat{u}^i$ if bargainer j does not achieve commitment and $\underline{u}^i$ if j does achieve commitment, since in this case the commitments must be incompatible because $\hat{u}^i > \tilde{u}^i$, $\hat{u}^j > \tilde{u}^j$, and $(\tilde{u}^i, \tilde{u}^j)$ is Pareto-efficient. Thus, since p_i is constant, it is clear that attempting commitment to $\bar{u}^i$ dominates all other strategies in this class for bargainer i.

Consider the bargaining game, reduced by deletion of dominated strategies to a game in which each bargainer i has only two strategies: not attempting commitment, and attempting commitment to the position $\bar{u}^i$. I shall now argue that the latter strategy in fact strictly dominates the former for both bargainers.

There is clearly no difference between the two strategies unless commitment is successful. If bargainer i achieves commitment to $\bar{u}^i$ and bargainer j does not achieve commitment, i obtains utility $\bar{u}^i$, whereas he could have obtained only $\tilde{u}^i < \bar{u}^i$ by not attempting commitment. If, on the other hand, bargainer j achieves commitment to $\bar{u}^j$, which is necessarily incompatible with $\bar{u}^i$, bargainer i obtains utility $\underline{u}^i$ whether he attempts commitment or not, regardless of whether his attempted commitment is successful. Thus, attempting commitment to $\bar{u}^i$ stochastically dominates not attempting commitment for bargainer i.

To complete the interpretation of the game, note that it is natural to assume that $(\tilde{u}^1, \tilde{u}^2)$, the outcome when neither bargainer attempts commitment, is *ex ante* Pareto-superior to the outcome when both bargainers attempt commitment. It is unlikely, given the convexity of the utility-possibility set, that a bargainer would negotiate an agreement inferior to what he could obtain *ex ante* by attempting commitment; and the prospect of attempting commitment is at least as good as the outcome when both bargainers attempt commitment, because having an opponent attempt commitment is the worst case for a bargainer who has attempted commitment.

To summarize, in the constant-probabilities case the bargaining game can be reduced by successive deletion of dominated strategies to a classical Prisoner's Dilemma, in which attempting commitment to the most favorable position in the contract zone dominates not attempting commitment for both bargainers. There is, therefore, a unique Nash equilibrium, which results provided only that bargainers do not employ dominated strategies. Intuitively, this is so because, given the deletion of dominated strategies carried out above, attempting commitment to $\bar{u}^i$ increases bargainer i's chance of getting a favorable outcome at no cost, since the probabilities of success are constant and an impasse is no worse for i than letting bargainer j achieve commitment to $\bar{u}^j$. The Nash equilibrium of the commitment game has the usual property of the noncooperative equilibria of Prisoner's Dilemma games—in spite of its clear individual rationality, it leads to an outcome that is collectively "irrational" because the positive probability of impasse that results implies that its distribution of outcomes is not *ex ante* Pareto-efficient. While I would not wish to argue that the extreme demands that occur in equilibrium in this case are realistic, the result provides a nice "textbook" example and a striking demonstration that, under these expectational assumptions, inefficiency can occur even if one is willing to accept only strategic-dominance arguments about bargainers' rational actions.

At this point, it is natural to consider more sophisticated heuristics relating bargainers' second-stage expectations to the positions to which they attempt commitment in the first stage. Consider what determines these expectations in the fully rational model of Section 3. There, bargainers' first-stage estimated probabilities that they will not back down in the second stage depend on both bargainers' commitment positions and, in general, the entire bargaining environment. For example, bargainer i's probability in the continuous model of Section 3 is $1 - F^i(d^i)$, where d^i is the equilibrium cutoff level (assumed unique for the purposes of this discussion), which depends on both commitment positions and the entire bargaining environment. In Section 5's comparative statics analysis, $(\underline{u}^i, \underline{u}^j)$ is the only aspect of the bargaining environment that is allowed to vary; therefore, for my purposes, $\hat{u}^i$, $\underline{u}^i$, $\hat{u}^j$, and $\underline{u}^j$ are the variables on which bargainers' estimated probabilities might be allowed to depend.

What is a reasonable form for this dependence? Since we are studying the implications of bounded rationality, there is little but intuition and common sense to go on here (although the same sort of intuition and common sense come into play in deciding what factors to take into consideration in a fully rational model). I have therefore chosen plausible hypotheses that lead to a rich, but analytically tractable model. Bargainer i's probability of successful commitment —of not backing down in the second stage—is assumed to be a twice continuously differentiable function of $\hat{u}^i$ and $\underline{u}^i$, denoted $p^i(\hat{u}^i, \underline{u}^i)$. In the continuous model of Section 3, comparative statics calculations not reproduced here reveal that a fully rational $p^i(\cdot)$ should depend positively on $\underline{u}^i$, negatively on $\underline{u}^j$, and ambiguously on $\hat{u}^i$ and $\hat{u}^j$. I shall assume however that, denoting partial derivatives by subscripts in the usual way, $p_1^i(\cdot) < 0$, $p_2^i(\cdot) > 0$, and $p_{11}^i(\cdot) \leq 0$, except where the natural zero/one probability boundaries come into play. In words, the likelihood of successful commitment decreases at an increasing (algebraically decreasing) rate with increases in the position to which commitment is attempted, and increases when the bargainer's disagreement utility rises. To avoid trivialities, I shall also assume that $p^i(\tilde{u}^i, \underline{u}^i) > 0$ for all $\underline{u}^i$. (Note that $\tilde{u}^i$ may depend on $\underline{u}^i$.) I shall continue to assume that bargainers are well informed about everything but the outcomes of attempted commitments, and that they agree on the forms of the p^i functions.

In this model, as in the model of Section 3, it does not seem possible to demonstrate without further assumptions that pure-strategy, or even mixed-strategy, Nash equilibria always exist. To see why this is so, note that the proof given in the analysis of the constant-probabilities case that not attempting commitment dominates attempting commitment to a position $\hat{u}^i \leq \tilde{u}^i$ is still valid in the variable-probabilities case, because the proof involved only the value of the commitment probability at a single position. Thus, such overly conservative commitment strategies can be ruled out a priori, and the notation can be simplified by assigning to $\hat{u}^i = \tilde{u}^i$ the meaning of not attempting commitment (rather than Section 3's meaning of attempting commitment to $\tilde{u}^i$, which is dominated here by not attempting commitment). Given this notational convention, which will be maintained from now on, each bargainer has, in effect, a compact strategy space—the set of $\hat{u}^i$ such that $\tilde{u}^i \leq \hat{u}^i \leq \bar{u}^i$—and the question of existence can be resolved by examining the behavior of the payoff functions.

Let $v^i(\hat{u}^i, \hat{u}^j)$ denote bargainer i's expected payoff when bargainers' strategies are $(\hat{u}^i, \hat{u}^j)$. Given my assumptions, the payoff function can be written:

$$(4.1)\qquad v^i(\hat{u}^i, \tilde{u}^j) \equiv p^i(\hat{u}^i, \underline{u}^i)\hat{u}^i + \left[1 - p^i(\hat{u}^i, \underline{u}^i)\right]\tilde{u}^i,$$

$$v^i(\tilde{u}^i, \hat{u}^j) \equiv p^j(\hat{u}^j, \underline{u}^j)\psi(\hat{u}^j) + \left[1 - p^j(\hat{u}^j, \underline{u}^j)\right]\tilde{u}^i,$$

and

$$\begin{aligned} v^i(\hat{u}^i, \hat{u}^j) \equiv{}& p^i(\hat{u}^i, \underline{u}^i)p^j(\hat{u}^j, \underline{u}^j)\underline{u}^i + p^i(\hat{u}^i, \underline{u}^i)\left[1 - p^j(\hat{u}^j, \underline{u}^j)\right]\hat{u}^i \\ &+ \left[1 - p^i(\hat{u}^i, \underline{u}^i)\right]p^j(\hat{u}^j, \underline{u}^j)\psi(\hat{u}^j) \\ &+ \left[1 - p^i(\hat{u}^i, \underline{u}^i)\right]\left[1 - p^j(\hat{u}^j, \underline{u}^j)\right]\tilde{u}^i \text{ if } \hat{u}^i > \tilde{u}^i \text{ and } \hat{u}^j > \tilde{u}^j. \end{aligned}$$

The v^i functions are clearly continuous in $(\hat{u}^i, \hat{u}^j)$ whenever $\hat{u}^i > \tilde{u}^i$ and $\hat{u}^j > \tilde{u}^j$, but discontinuous when either $\hat{u}^i = \tilde{u}^i$ or $\hat{u}^j = \tilde{u}^j$. Thus, the standard results used to guarantee the existence of Nash equilibrium are not applicable.[9]

In spite of these technical problems, the results in the constant-probabilities case indicate that pure-strategy Nash equilibria will exist much of the time, in particular when $p^i(\hat{u}^i, \underline{u}^i)$ does not fall too rapidly with increases in $\hat{u}^i$ for either bargainer i. In fact, if (but not necessarily only if)

$$(4.2)\qquad p^j(\tilde{u}^j, \underline{u}^j)\underline{u}^i + \left[1 - p^j(\tilde{u}^j, \underline{u}^j)\right]\bar{u}^i > \tilde{u}^i$$

holds for $i = 1, 2$, there always exists a pure-strategy Nash equilibrium, as I shall now argue. The reason is that condition (4.2), which depends only on the parameters of the bargaining environment, guarantees that bargainer i will wish to attempt commitment to some position even if bargainer j is attempting

[9]See footnote 7. In the present context, it is not hard to show that, because a bargainer is better off if his opponent does not attempt commitment than if he attempts commitment to a position that differs only slightly from the compromise settlement, and because the bargainer himself is better off not attempting commitment than attempting commitment to such a position, the v^i are everywhere upper semi-continuous. The problem is that Dasgupta and Maskin's [5] condition of graph continuity is not satisfied here, because there is a radical change in the graph of $v^i(\hat{u}^i, \hat{u}^j)$, viewed as a function of $\hat{u}^i$, when $\hat{u}^j$ rises above $\tilde{u}^j$. This change occurs because such an increase in $\hat{u}^j$ creates a risk of impasse that was not present before. As a result, it is easy to imagine situations where bargainers' reaction correspondences, even allowing mixed strategies, are not upper hemi-continuous. In such situations, there may not exist any pure-strategy, or even mixed-strategy, Nash equilibria.

commitment to a position near $\tilde{u}^j$. When $\hat{u}^j$ is near $\tilde{u}^j$, the relative attractiveness of attempting commitment is at its lowest, both because the associated risk of impasse is highest there and because the cost of yielding to j's demand is lowest. Thus, it is intuitively clear that if bargainer i attempts commitment in these unfavorable circumstances, he will, *a fortiori*, always attempt commitment. Note that $p^i(\tilde{u}^i, \underline{u}^i)$ and $p^j(\tilde{u}^j, \underline{u}^j)$, the probabilities of bargainers' successful commitment to their compromise settlements, can always be made low enough so that (4.2) is satisfied. This can be viewed as a confirmation of Schelling's [**20**, Chapter 3] suggestion that low success probabilities favor attempting commitment.

More formally, straightforward computations reveal that when (4.2) is satisfied, there is always some value of $\hat{u}^i$ (in fact, $\hat{u}^i = \bar{u}^i$ will do, although it is not generally a best response) such that $v^i(\hat{u}^i, \hat{u}^j) > v^i(\tilde{u}^i, \hat{u}^j)$ for *all* $\hat{u}^j$ such that $\tilde{u}^j \leq \hat{u}^j \leq \bar{u}^j$. Further, it is easy to verify that for all such $\hat{u}^j$,

$$\lim_{\hat{u}^i \to \tilde{u}^i} v_1^i(\hat{u}^i, \hat{u}^j) > 0. \tag{4.3}$$

It follows that $\hat{u}^i$ can be restricted to lie in the compact and convex interval $[\tilde{u}^i + \epsilon, \bar{u}^i]$, for some $\epsilon > 0$, without ruling out any "good" strategies. Thus, if condition (4.2) holds, each bargainer i will always choose to attempt commitment to some position $\hat{u}^i > \tilde{u}^i$. His strategy space can therefore be taken to be compact and convex; and given (4.1), his payoff function will be continuous in the relevant range. If, in addition, the payoff function $v^i(\hat{u}^i, \hat{u}^j)$ is quasiconcave in $\hat{u}^i$, $i = 1, 2$, Debreu's Social Equilibrium Existence Theorem (see Dasgupta and Maskin [**5**, Theorem 1]) yields the existence of a Nash equilibrium in pure strategies.

To see that $v^i(\hat{u}^i, \hat{u}^j)$ is quasiconcave in $\hat{u}^i$, note first that when $\hat{u}^j > \tilde{u}^j$,

$$v_1^i(\hat{u}^i, \hat{u}^j) \equiv p_1^i(\hat{u}^i, \underline{u}^i)\Big[p^j(\hat{u}^j, \underline{u}^j)\big[\underline{u}^i - \psi(\hat{u}^j)\big] + \big[1 - p^j(\hat{u}^j, \underline{u}^j)\big] \times (\hat{u}^i - \tilde{u}^i)\Big] + p^i(\hat{u}^i, \underline{u}^i)\big[1 - p^j(\hat{u}^j, \underline{u}^j)\big] \tag{4.4}$$

and

$$v_{11}^i(\hat{u}^i, \hat{u}^j) \equiv p_{11}^i(\hat{u}^i, \underline{u}^i)\Big[p^j(\hat{u}^j, \underline{u}^j)\big[\underline{u}^i - \psi(\hat{u}^j)\big] + \big[1 - p^j(\hat{u}^j, \underline{u}^j)\big] \times (\hat{u}^i - \tilde{u}^i)\Big] + 2p_1^i(\hat{u}^i, \underline{u}^i)\big[1 - p^j(\hat{u}^j, \underline{u}^j)\big]. \tag{4.5}$$

It is easy to verify from (4.4) and (4.5) that if $v_1^i(\hat{u}^i, \hat{u}^j) = 0$, $v_{11}^i(\hat{u}^i, \hat{u}^j) < 0$; thus, $v^i(\hat{u}^i, \hat{u}^j)$, viewed as a function of the single variable $\hat{u}^i$, is strictly quasiconcave. It follows that whenever (4.2) holds for $i = 1, 2$, there is a pure-strategy Nash equilibrium, at which bargainers both attempt commitment, to incompatible positions.

In spite of this result, it is not true that in the general case where (4.2) does not hold for $i = 1, 2$, a Nash equilibrium, if one exists, necessarily involves both bargainers attempting commitment. To see this, consider the case where, for all

$\underline{u}^j$, $p^j(\hat{u}^j,\underline{u}^j)$ is unity (or nearly unity) for all $\hat{u}^j$ up to and including a certain level between $\tilde{u}^j$ and $\bar{u}^j$ and falls rapidly beyond that point, so that in equilibrium $\hat{u}^j$ will generally be set at (or near) that level. This might be the case, for example, if there is some feature of the bargaining situation (for example, a Council on Wage and Price Stability guideline) that makes it especially credible that bargainer j will not accept less than some particular settlement. If bargainer i attempts commitment to some position $\hat{u}^i$ where $p^i(\hat{u}^i,\underline{u}^i) > 0$, (4.1) and the fact that $p^j(\hat{u}^j,\underline{u}^j) \approx 1$ imply that

$$(4.6) \qquad v^i(\hat{u}^i,\hat{u}^j) \approx p^i(\hat{u}^i,\underline{u}^i)\underline{u}^i + \left[1 - p^i(\hat{u}^i,\underline{u}^i)\right]\psi(\hat{u}^j) < \psi(\hat{u}^j).$$

On the other hand, if bargainer i does not attempt commitment (or, equivalently, attempts commitment to a position where $p^i(\cdot) = 0$),

$$(4.7) \qquad v^i(\hat{u}^i,\hat{u}^j) \approx \psi(\hat{u}^j).$$

Thus, in this case there is a Nash equilibrium where one bargainer attempts commitment while the other does not.

It does, however, follow from my assumption that $p^i(\tilde{u}^i,\underline{u}^i) > 0$ and (4.1) that $(\hat{u}^i,\hat{u}^j) = (\tilde{u}^i,\tilde{u}^j)$—that is, neither bargainer attempting commitment—can never be a Nash equilibrium. For, since p^i is a continuous function, there will always exist some $\hat{u}^i > \tilde{u}^j$ such that $p^i(\hat{u}^i,\underline{u}^i) > 0$ as well; if $\hat{u}^j = \tilde{u}^j$, this $\hat{u}^i$ yields a higher value of $v^i(\hat{u}^i,\hat{u}^j)$ than $\tilde{u}^i$, by (4.1).

5. COMPARATIVE STATICS

The main interest of making the probability of impasse variable is that it provides a framework in which the relationship between the bargaining environment and the probability of impasse can be investigated. This section presents comparative statics results that indicate how the probability of impasse responds to changes in the size of the contract zone (caused by changes in the costs of disagreement) and to changes in the difficulty of commitment.

The conclusions are easy to summarize: two assumptions that are almost invariably maintained in the industrial labor relations and law and economics literature about bargaining—that the frequency of impasse is reduced by increases in the size of the contract zone (brought about by increases in the costs of disagreement) and by increases in the difficulty of commitment—cannot be supported on theoretical grounds if commitment to incompatible demands is part of the explanation for the occurrence of impasses. The comparative statics results that test these assumptions are ambiguous in the models studied here; and this ambiguity is highly robust to changes in behavioral assumptions and to special assumptions about the bargaining environment. A nonpathological example shows that strong and implausible assumptions, at the very least, would be required to resolve the ambiguities and justify the conventional wisdom.

To demonstrate the extent of this robustness, I shall adopt rather special assumptions for the remainder of this section. First, only equilibria where both bargainers attempt commitment, to incompatible positions, will be considered.

Second, since the ambiguous sensitivities of the p^i functions implied by the analysis of Section 3 would result immediately in ambiguous conclusions here, I shall instead maintain the assumptions about the p^i functions used in Section 4; attention is further restricted to the leading special case where $p^i(\hat{u}^i, \underline{u}^i)$ can be written in the form $p^i(\hat{u}^i - \underline{u}^i)$. (No confusion should result from this abuse of notation.) Finally, it will be assumed that the $\tilde{u}^i$, viewed as differentiable functions $\tilde{u}^i(\underline{u}^i, \underline{u}^j)$ of the disagreement outcome, satisfy $0 \le \tilde{u}^i_1(\cdot) \le 1$ and $\tilde{u}^i_2(\cdot) \le 0$ for all $(\underline{u}^i, \underline{u}^j)$; these assumptions are satisfied in most descriptive bargaining theories (see, for example, Kalai and Smorodinsky [**11**] and Nash [**14**]), and are quite plausible.[10]

In what follows, I shall suppress the arguments of functions for notational clarity whenever this can be done without causing confusion. First, consider the effects of varying the costs of disagreement. When the p^i can be written in the form $p^i(\hat{u}^i - \underline{u}^i)$,

$$\text{(5.1)} \qquad \frac{dp^ip^j}{d\underline{u}^i} \equiv p^ip^j_1 \frac{d\hat{u}^j}{d\underline{u}^i} + p^jp^i_1\left[\frac{d\hat{u}^i}{d\underline{u}^i} - 1\right],$$

where $\hat{u}^i$ and $\hat{u}^j$ now denote the equilibrium attempted commitments for given values of $\underline{u}^i$ and $\underline{u}^j$. If $dp^ip^j/d\underline{u}^i > 0$, shrinking the contract zone by raising $\underline{u}^i$ also raises the probability of impasse; this is the conventional wisdom. Given that $p^j_1(\cdot) < 0$ and $p^i_1(\cdot) < 0$, this will be true in general only if $d\hat{u}^j/d\underline{u}^i \le 0$ and $d\hat{u}^i/d\underline{u}^i \le 1$, with strict inequality holding at least once.

Under the maintained assumptions, straightforward but extremely tedious calculations, not reproduced here, reveal that it is not true in general that $d\hat{u}^j/d\underline{u}^i \le 0$ or $d\hat{u}^i/d\underline{u}^i \le 1$. Further, this ambiguity is not of the type that can be resolved by naive application of Samuelson's Correspondence Principle: the denominators of the expressions for $d\hat{u}^j/d\underline{u}^i$ and $d\hat{u}^i/d\underline{u}^i$ can be signed by postulating the local stability of a simple gradient adjustment process, but the numerators remain sufficiently indeterminate that the above questions cannot be resolved a priori.

In the $p^i(\hat{u}^i - \underline{u}^i)$ case, the indeterminacy can be shown to depend on interactions between bargainers' strategy choices, in the following sense. Suppose the problem is simplified (as was done in Section 4, in the example used to show that there could be Nash equilibria at which one bargainer did not attempt commitment) by specifying p^j so that bargainer j's optimal commitment strategy leads to a constant probability of success for all $\hat{u}^i$. Then only bargainer i's equilibrium conditions need be considered.

The first-order condition for the problem that determines i's optimal commitment strategy requires that $v^i_1(\hat{u}^i, \hat{u}^j) = 0$, and the second-order sufficient condition requires that $v^i_{11}(\hat{u}^i, \hat{u}^j) < 0$; these derivatives are given by (4.4) and (4.5). As

[10] One might reasonably argue that the cooperative outcome ought to be determined instead by the actual strategic possibilities—in this case, the utilities bargainers can guarantee themselves by attempting or not attempting commitment—rather than the disagreement outcome. From this point of view, that the Nash and Raiffa-Kalai-Smorodinsky solutions satisfy my assumptions that $0 \le \tilde{u}^i_1(\cdot) \le 1$ and $\tilde{u}^i_2(\cdot) \le 0$ is less compelling. But the assumptions still appear quite plausible.

noted in Section 4, the second-order condition is satisfied whenever the first-order condition is; it follows that $v^i(\hat{u}^i, \hat{u}^j)$ is strictly quasi-concave in $\hat{u}^i$ and has a unique maximum, which is characterized by the first-order condition, for any given value of $\hat{u}^j$.

Total differentiation of the first-order condition with respect to $\underline{u}^i$ reveals that

$$\frac{d\hat{u}^i}{d\underline{u}^i} \equiv \frac{p^i_{11}\left[p^j\left(\underline{u}^i - \psi(\hat{u}^j)\right) + (1-p^j)(\hat{u}^i - \tilde{u}^i)\right] - p^i_1\left[2p^j - 1 - (1-p^j)\tilde{u}^i_1\right]}{2p^i_1(1-p^j) + p^i_{11}\left[p^j\left(\underline{u}^i - \psi(\hat{u}^j)\right) + (1-p^j)(\hat{u}^i - \tilde{u}^i)\right]}$$

$$(5.2) \qquad < 1,$$

where the inequality follows, after a simple computation, from the facts that $\tilde{u}^i_1 \leq 1$ and the denominator, which is $v^i_{11}(\hat{u}^i, \hat{u}^j)$, is strictly negative. Similarly, total differentiation of the first-order condition with respect to $\underline{u}^j$ yields

$$(5.3) \qquad \frac{d\hat{u}^i}{d\underline{u}^j} \equiv \frac{p^i_1\left[p^j\psi_1 + (1-p^j)\tilde{u}^i_2\right]}{2p^i_1(1-p^j) + p^i_{11}\left[p^j\left(\underline{u}^i - \psi(\hat{u}^j)\right) + (1-p^j)(\hat{u}^i - \tilde{u}^i)\right]} < 0,$$

where the inequality follows immediately from my assumptions and the fact that the denominator is, again, negative by the second-order condition. These results, in conjunction with (5.1), imply that when bargainers' strategy choices do not interact, the conventional wisdom relating the probability of impasse to the size of the contract zone is confirmed in the $p^i(\hat{u}^i - \underline{u}^i)$ case. Of course, these are extremely stringent requirements.

In deriving (5.3), I made use of the assumption that, from bargainer i's standpoint, $d\hat{u}^j/d\underline{u}^j \equiv 1$. This assumption follows from the assumptions in the $p^i(\hat{u}^i - \underline{u}^i)$ case, given that bargainer i knows the form of the function p^j and really believes that $p^j(\cdot)$ will remain constant when $\underline{u}^j$ changes. But it can be argued that this is too sophisticated to be really plausible. If one instead adopts the position that bargainer i, in addition to expecting $p^j(\cdot)$ to remain constant, expects no change in $\hat{u}^j$ when $\underline{u}^j$ changes, the $p^j\psi_1$ term in the numerator of the right-hand side of (5.3) disappears, altering the magnitude, but not the sign, of $d\hat{u}^i/d\underline{u}^j$.

To make things more concrete, one would like to supplement the formal ambiguity referred to above with a simple example in which bargainers' strategy choices interact and shrinking the contract zone actually decreases the probability of an impasse. Unfortunately, I have been unable to find an analytically tractable example with this property in the $p^i(\hat{u}^i - \underline{u}^i)$ case, which I consider the most interesting; interactions between strategy choices, while necessary for the perverse result in this case, also seem to prevent explicit solutions. The following simple example, in which $p^i(\cdot)$ depends only on $\hat{u}^i$ and $p^j(\cdot)$ is constant, may help to dispel some of the reader's dissatisfaction with this state of affairs and illustrate why the perverse result can occur.

Suppose that the relevant part of the utility-possibility set lies in the nonnegative quadrant and that its frontier is given by $u^i \equiv \psi(u^j) \equiv 1 - u^j$. Let $p^i(\hat{u}^i, \underline{u}^i)$

$\equiv 1 - \hat{u}^i$ and $p^j(\hat{u}^j, \underline{u}^j) \equiv p^j$, a constant. By continuity, small amounts of dependence on $\underline{u}^i$ and $\underline{u}^j$ could be introduced without significantly altering the result. The compromise agreement that is reached if neither bargainer achieves commitment is taken to be the Nash [14] cooperative solution (or equivalently, the Raiffa-Kalai-Smorodinsky [11] solution, which is the same in this case), with $(\underline{u}^i, \underline{u}^j)$ as the threat point. Thus, $\tilde{u}^i \equiv (\underline{u}^i - \underline{u}^j + 1)/2$ and $\tilde{u}^j \equiv (\underline{u}^j - \underline{u}^i + 1)/2$. (One might argue as in footnote 10 that the strategic possibilities, rather than the disagreement outcome alone, ought to determine the $\tilde{u}^i$. But since the utilities bargainers can guarantee themselves by attempting or not attempting commitment are influenced by the $\tilde{u}^i$, this would involve a recursive definition of the $\tilde{u}^i$, adding much complexity but little insight.)

Easy calculations show that condition (4.2) guarantees that bargainer i will always attempt commitment at a Nash equilibrium provided that $p^j < 1/2$; similarly, bargainer j will always attempt commitment when $\underline{u}^j < \underline{u}^i$.[11] When these conditions are satisfied, it is clear that $\hat{u}^j = 1 - \underline{u}^i$ is bargainer j's optimal commitment strategy. Given this and the above assumptions, bargainer i's expected utility when he attempts commitment to $\hat{u}^i$ is given by

$$v^i(\hat{u}^i, 1 - \underline{u}^i) \equiv p^j \underline{u}^i + (1 - \hat{u}^i)(1 - p^j)\hat{u}^i + \hat{u}^i(1 - p^j)(\underline{u}^i - \underline{u}^j + 1)/2. \tag{5.4}$$

Maximizing the expression on the right-hand side of (5.4) with respect to $\hat{u}^i$ is clearly equivalent to maximizing $(1 - \hat{u}^i)\hat{u}^i + \hat{u}^i(\underline{u}^i - \underline{u}^j + 1)/2$. The second-order sufficient condition is satisfied everywhere, so solving the first-order condition yields the unique optimal strategy:

$$\hat{u}^i = (\underline{u}^i - \underline{u}^j + 3)/4. \tag{5.5}$$

Now we are ready to examine the effects on the probability of impasse of changing $\underline{u}^i$ and $\underline{u}^j$. In Nash equilibrium, this probability is given by

$$p^i(\hat{u}^i, \underline{u}^i)p^j \equiv p^j(1 - \underline{u}^i + \underline{u}^j)/4. \tag{5.6}$$

Thus, $dp^ip^j/d\underline{u}^i \equiv -p^j/4 < 0$, so increasing $\underline{u}^i$, which shrinks the contract zone, actually makes an impasse *less* likely. On the other hand, $dp^ip^j/d\underline{u}^j \equiv p^j/4 > 0$, so shrinking the contract zone by increasing $\underline{u}^j$ makes impasse more likely. (If $\underline{u}^i$ is near zero, but not exactly zero, it is possible to raise $\underline{u}^j$ while continuing to satisfy the $\underline{u}^j < \underline{u}^i$ constraint.) Thus, the conventional wisdom is valid in this example when $\underline{u}^j$ changes, but not when $\underline{u}^i$ changes. These results follow simply from the fact that changing the disagreement outcome also changes the relative costs and benefits of attempting commitment to different positions, possibly, as the example shows, in a way that induces bargainers to take more extreme positions when the contract zone shrinks, making impasse less likely.

[11] Readers of footnote 10 may be curious about whether $\tilde{u}^i$ and $\tilde{u}^j$ in this example are in fact at least as good for bargainers i and j as what they could guarantee themselves by attempting commitment. When a bargainer attempts commitment, the worst case is when the other bargainer also does. It is not always true that $\tilde{u}^i$ and $\tilde{u}^j$ are better for i and j than these worst cases, but it can be shown to be true near $(\underline{u}^i, \underline{u}^j) = (0, 0)$, for example, whenever, roughly, $1/9 < p^j < 1/3$.

To summarize the results obtained so far, it is impossible to establish on a priori grounds that the conventional wisdom about the effect of changing the size of the contract zone on the probability of impasse is always valid. But there is some weak theoretical evidence to support it in the leading special case where $p^i(\hat{u}^i, \underline{u}^i)$ can be written as $p^i(\hat{u}^i - \underline{u}^i)$. The expected relationship holds in this case when interactions between bargainers' strategy choices are relatively unimportant.

Certain provisions in the customs and laws governing bargaining behavior—for example, the common requirement in labor law to bargain "in good faith"—can be viewed as attempts to make commitment more difficult, presumably with the purpose of reducing the probability of impasse. This section concludes by asking whether such attempts that stop short of making commitment completely impossible, viewed in the present framework, appear likely to succeed.

More formally, rewrite the function p^i as $p^i(\hat{u}^i, \underline{u}^i, \alpha)$, where α is a shift parameter that shifts only $p^i(\cdot)$; without essential loss of generality, it is assumed that $p^i_3(\cdot) > 0$ everywhere. Lowering α makes commitment more difficult, in the sense of lowering the probability of successful commitment to any position. Now,

$$\text{(5.7)} \qquad \frac{dp^ip^j}{d\alpha} \equiv p^ip^j_1\frac{d\hat{u}^j}{d\alpha} + p^j\left[p^i_1\frac{d\hat{u}^i}{d\alpha} + p^i_3\right],$$

where $\hat{u}^i$ and $\hat{u}^j$ denote the equilibrium attempted commitments for the given value of α. If $dp^ip^j/d\alpha > 0$, making commitment more difficult by lowering α lowers the probability of impasse, which is the conventional wisdom. Given that $p^j_1(\cdot)$ and $p^i_1(\cdot)$ are everywhere negative, this is *guaranteed* in general only if $d\hat{u}^i/d\alpha < 0$ and $d\hat{u}^j/d\alpha < 0$.

As with the above analysis of the effects of changes in the costs of disagreement on the $\hat{u}^i$, the comparative statics of Nash equilibria when α changes are complex and unilluminating, leading to formal ambiguities very similar to those encountered above. But as before, studying the response when strategy choices do not interact gives a feel for the likely effects of changing α, and can be viewed as an approximate analysis of the response of Nash equilibria when interactions between bargainers' strategy choices are not expected to be important.

Straightforward computations verify that in this case,

$$\text{(5.8)} \qquad \frac{d\hat{u}^i}{d\alpha} \equiv -\frac{p^i_{13}\left[p^j\left(\underline{u}^i - \psi(\hat{u}^j)\right) + (1-p^j)(\hat{u}^i - \bar{u}^i)\right] + p^i_3(1-p^j)}{2p^i_1(1-p^j) + p^i_{11}\left[p^j\left(\underline{u}^i - \psi(\hat{u}^j)\right) + (1-p^j)(\hat{u}^i - \bar{u}^i)\right]}.$$

The term in brackets in the numerator of the right-hand side of (5.8) must be strictly positive by the first-order condition; the denominator must be negative by the second-order condition. Because $p^i_3(\cdot) > 0$ by assumption, it is plain that $d\hat{u}^i/d\alpha$ is strictly positive if $p^i_{13}(\cdot) > 0$ and ambiguous in sign otherwise. (A similar computation shows that $d\hat{u}^j/d\alpha$ is always ambiguous in sign.) In any event, $d\hat{u}^i/d\alpha$ can never be unambiguously negative, so that it is certainly not possible to validate the conventional wisdom theoretically in this case. The common reasoning errs in considering only the partial effect of changing α, which is obviously in the "right" direction. But the total effect, taking into account the responses of $\hat{u}^i$ and $\hat{u}^j$ (which tend to go in the "wrong" direction), may easily be counterintuitive.

6. CONCLUSION

This paper develops a theory that explains why rational bargainers might take actions that could lead to an impasse.[12] By formalizing Schelling's [**18, 19, 20**] view of the bargaining process as a struggle between bargainers to commit themselves to favorable bargaining positions, the theory determines the probability of impasse endogenously and permits an investigation of its relationship to the bargaining environment. Its crucial elements, which allow an explanation of the occurrence of inefficient outcomes without assuming irrationality, are the uncertainty and irreversibility of commitment. To the extent that commitment is part of the explanation for the impasses frequently observed in real bargaining, doubt is cast on two widely held beliefs about the influence of the bargaining environment on the frequency of impasse. Contrary to what is assumed almost universally in those parts of the law and economics and industrial relations literature that consider the question, increasing the costs of disagreement and thereby enlarging the contract zone need not decrease the probability of impasse. Malouf and Roth [**13**] report experimental evidence that casts further doubt on this assumption. Neither is there support for the common belief that making commitment more difficult (in the sense of lowering the probability of success) makes impasse less likely. The reason is that the "common-sense" argument used to reach this conclusion considers only the partial effect of lowering the success probability; it ignores the resulting changes in attempted commitment positions, whose effects generally go the other way and could easily swamp the partial effect.

Two lines of research appear especially promising at this point. As mentioned above, Chatterjee and W. Samuelson [**2**] have developed an alternative explanation of the occurrence of impasses, based on a stylized complete irreversibility of demands (certainty about commitment, in the language of this paper) and on

[12] It might be argued that the rationalization of the occurrence of bargaining impasses presented here is incomplete because my assumption that bargainers can communicate only by their demands is unduly restrictive. When outcomes are always efficient, bargaining is essentially a zero-sum game, and direct communication cannot play a significant role. But when impasses are possible, bargainers have a mutual interest in avoiding them, and it is no longer clear that communication would not occur in equilibrium if it were allowed. (In the present model, bargainers might wish to communicate about their costs of backing down between the first and second stages.) In a recent paper [3], Joel Sobel and I have considered the general question of strategic communication in an abstract setting, obtaining results that seem to shed some light on whether the above criticism is valid. In [3], it is shown under reasonable assumptions (which have not, however, been verified for the present model) that perfect communication is not compatible with noncooperative rationality unless agents' interests completely coincide, and that once their interests differ by a given, "finite" amount, only no communication is compatible with rationality. Thus, it seems likely that direct communication would not alter the qualitative results obtained above.

bargainers' uncertainty about each others' preferences. Since in their model, bargainers generally find it advantageous to demand more than their disagreement utilities, the possibility of impasse due to uncertainty about preferences is present. As it happens, their model also casts doubt on the conventional wisdom about the effects of expanding the contract zone. In another recent paper, Sobel and Takahashi [22] have combined uncertainty about preferences with uncertainty about the number of stages which shares some of the effects of uncertainty about commitment, as it is conceptualized here. They obtain interesting results about how the frequency of impasse depends on the bargaining environment. Natural next steps in this direction would be to build models with endogenous numbers of stages, and to deal with the issues of strategic information transmission that are avoided in Chatterjee and Samuelson [2], Sobel and Takahashi [22], and the present paper.

The second line of research involves a deeper investigation of the process by which commitment is achieved. In this paper, bargainers communicate only by making demands and making it costly to back down from them. And the actual process of commitment is not modeled explicitly; rather, its influence on bargainers' strategy choices is summarized by the probability distributions of their costs of backing down. To explain the occurrence of impasses and to evaluate the effects of changes in the bargaining environment on their likelihood, it suffices to place certain mild, plausible restrictions on these distributions. The situation is somewhat similar to that encountered in consumer and producer theory, where many of the conclusions do not depend on detailed knowledge of technology and tastes. But to make sharper predictions, or to test the theory empirically, it would be highly desirable to know more about the nature of communication and commitment in bargaining. This would require a detailed model of the process by which bargainers back up their demands; Schelling [18; 19, Appendix B] and Ellsberg [6] contain fascinating discussions of the issues that such a model would have to resolve.

University of California, San Diego

Manuscript received December, 1979; last revision received May, 1981.

REFERENCES

[1] Ashenfelter, Orley, and George Johnson: "Bargaining Theory, Trade Unions, and Industrial Strike Activity," *American Economic Review*, 59(1969), 35–49.

[2] Chatterjee, Kalyan, and William Samuelson: "The Simple Economics of Bargaining," Boston University School of Management Working Paper 2/81, 1979; revised 1981.

[3] Crawford, Vincent P., and Joel Sobel: "Strategic Information Transmission," *Econometrica*, forthcoming.

[4] Cross, John: *The Economics of Bargaining*. New York: Basic Books, 1969.

[5] Dasgupta, Partha, and Eric Maskin: "The Existence of Economic Equilibrium: Continuity and Mixed Strategies," Stanford IMSSS Technical Report 252, 1977.

[6] Ellsberg, Daniel: "The Theory and Practice of Blackmail," in *Bargaining: Formal Theories of Negotiation*, ed. by Oran R. Young. Urbana, Illinois: University of Illinois Press, 1975.

[7] Harsanyi, John C.: "On the Rationality Postulates Underlying the Theory of Cooperative Games." *Journal of Conflict Resolution*, 5(1961), 179–196.

[8] ———: "Games With Incomplete Information Played by 'Bayesian' Players II: Bayesian Equilibrium Points," *Management Science*, 14(1968), 320–334.

[9] ———: *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. London and New York: Cambridge University Press, 1977.

[10] Harsanyi, John C., and Reinhard Selten: "A Generalized Nash Solution for Two-Person Bargaining Games With Incomplete Information," *Management Science*, 18(1972), P80–P106.

[11] Kalai, Ehud, and Meir Smorodinsky: "Other Solutions to Nash's Bargaining Problem," *Econometrica*, 43(1975), 513–518.

[12] Kreps, David M., and Robert Wilson: "Sequential Equilibria," *Econometrica*, forthcoming.

[13] Malouf, Michael, and Alvin Roth: "Disagreement in Bargaining: An Experimental Study," *Journal of Conflict Resolution*, 25(1981), 329–348.

[14] Nash, John: "The Bargaining Problem," *Econometrica*, 18(1950), 155–162.

[15] ———: "Two-Person Cooperative Games," *Econometrica*, 21(1953), 128–140.

[16] Rosenthal, Robert, and Henry Landau: "A Game-Theoretic Analysis of Bargaining with Reputations," *Journal of Mathematical Psychology*, 20(1979), 1353–1366.

[17] Roth, Alvin E.: *Axiomatic Models of Bargaining*. Berlin and New York: Springer-Verlag, 1979.

[18] Schelling, Thomas C.: "An Essay on Bargaining," *American Economic Review*, 46(1956), 281–306; reprinted as Chapter 2 of Schelling [19].

[19] ———: *The Strategy of Conflict*. London and New York: Oxford University Press, 1963.

[20] ———: *Arms and Influence*. New Haven and London: Yale University Press, 1966.

[21] Selten, Reinhard: "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4(1975), 25–55.

[22] Sobel, Joel, and Ichiro Takahashi: "A Multi-Stage Model of Bargaining," University of California, San Diego, Discussion Paper 80 – 25, 1980.