Equilibrium Behavior in Crisis Bargaining Games*

Jeffrey S. Banks, *University of Rochester*

This paper analyzes a general model of two-player bargaining in the shadow of war, where one player possesses private information concerning the expected benefits of war. I derive conclusions about equilibrium behavior by examining incentive compatibility constraints, where these constraints hold regardless of the game form; hence, the qualitative results are "game-free." I show that the higher the informed player's payoff from war, the higher is his or her equilibrium payoff from settling the dispute short of war, and the higher is the equilibrium probability of war. The latter result rationalizes the monotonicity assumption prevalent in numerous expected utility models of war. I then provide a general result concerning the equilibrium relationship between settlement payoffs and the probability of war.

1. Introduction

A common perception among analysts studying crisis bargaining situations is that the presence of informational asymmetries plays a key role in determining the behavior of the participants (cf. Powell 1987 and the citations therein). For quite some time, however, the tools necessary to explore such private information environments rigorously did not exist, thereby restricting the analyst to a class of models—namely, complete information models—which were clearly inappropriate for the task at hand. Beginning with the seminal work of Harsanyi (1967–68), game theory has advanced to a stage where it is now capable of dealing with issues of incomplete information, leading to numerous applications in economics and, to a lesser extent, political science. On the crisis bargaining front, various authors have incorporated these advances to reformulate earlier theories and to generate predictions concerning the role of information transmission, acquisition, and misperception in determining crisis bargaining outcomes (e.g., Powell 1987; Morrow 1989; Bueno de Mesquita and Lalman 1989).

One of the benefits of formulating a game-theoretic model is the necessity of explicitly modeling all of the relevant decisions by the participants, the timing of such decisions, and so forth. Yet such precision can also be seen as a drawback in that it may be unclear whether the conclusions deduced from a particular model are robust to other specifications of the game. Such a limitation is particularly acute in models of bargaining: should one party be able to make a "take-it-or-leave it" offer to the other? Does one player make all the offers, while the other simply accepts or rejects? Is the appropriate model one of alternating of-

fers, and if so, how long can the bargaining persist? Such indeterminacy in the selection of the "right" model potentially undermines the applicability of results derived from any particular model.

However, it turns out that there exists a class of results that concern equilibrium behavior in games with incomplete information which are robust to the specifics of the game the players actually play. That is, these results have the feature that they hold for *any* equilibrium in *any* game in which private information is present; in this sense then the results are "game-free." The results are derived from a set of constraints known as *incentive compatibility* conditions, where these conditions are a necessary feature of any optimal strategy adopted by a player with private information. In the current paper these conditions are examined in the context of a simple crisis bargaining situation in which one of the participants possesses private information concerning the benefits and costs of war. Examples of such information include a country's military capabilities (Morrow 1989) and the political fallout from war (Bueno de Mesquita and Lalman 1989). Rather than specify a particular process through which the participants interact (i.e., the game they play), we simply assume that through *some* bargaining process the participants either settle the dispute or do not. If they fail to settle, a war ensues; otherwise, they agree on some resolution of the dispute.

We are able to show that in any equilibrium of any game with the above format, the probability of war is an increasing function of the expected benefits from war of the informed player. Thus, whereas decision-theoretic models at times *assume* that stronger countries are more likely to engage in war (cf. Bueno de Mesquita and Lalman 1986; Lalman 1988), we are able to *derive* such a condition as a necessary consequence of optimal behavior. Further, the expected benefits from successfully concluding the bargaining short of war are also increasing in the informed player's expected benefits of war. Therefore, in any equilibrium of a crisis bargaining game, the following trade-off occurs: "stronger" countries (i.e., those with greater expected benefits from war) are more likely to end up in a war; yet if the bargaining negotiations are successful and war is averted, stronger countries receive a better settlement as well. Further, these conclusions hold regardless of the specifics of the bargaining game or the selection of a particular equilibrium from the set of equilibria in such a game.

Following the derivation of these monotonicity results, we proceed to characterize the "equilibrium" relationship between the probability of war and the expected benefits conditional on no war. That is, given a probability of war, where this is a function of the informed player's information, we can derive the "settlement" function that together with the former, constitutes equilibrium behavior. In this way then we can identify the subset of (pairs of) functions "rationalizable" as equilibrium behavior and derive further inferences about such behavior in crisis bargaining games.

2. The Model

The model concerns the behavior of two players, labeled 1 and 2, who attempt to resolve a dispute through some bargaining process; failure to resolve the dispute leads to war. Let $X = [0, 1]$ denote the set of all possible outcomes from the bargaining process other than war, where X contains any notion of a status quo ante, x_0, and let w denote the war outcome. We assume that both players 1 and 2 are risk neutral with respect to outcomes in X and that their preferences are diametrically opposed on X; thus, let the utility of player 1 from an outcome $x' \in X$ be simply x', while the utility for player 2 is $1 - x'$.[1] The utility for players 1 and 2 from the war outcome is denoted u and v, respectively, where we think of (u, v) as reduced form expressions that summarize the expected benefits of war. That is, during the bargaining process, the players will have expectations concerning the likelihood of winning a war should one occur, the gains from winning the war, the losses from losing the war, and the costs involved; these expressions are aggregated into the players' expected benefits of war. Further, player 1 is assumed to possess private information concerning the values of (u, v), while player 2 does not. For example, player 1 may know more about his own military capabilities than does player 2; therefore, since the expected benefits of war will be a function of 1's military capability, 1 will possess an informational advantage vis-à-vis 2 about the values (u, v).

I model this in the usual Harsanyi (1967–68) framework as a Bayesian environment where player 1's private information is described by a set of "types" T, where for each type $t \in T$ there exists a unique pair of values (u, v). Thus, we can write u and v as functions of the parameter t. Player 1 knows the actual value of $t \in T$ prior to making any decisions, while player 2 possesses a common knowledge prior probability $f(\cdot)$ over the set T, where $f(t) > 0$ for all $t \in T$. Let $T = [\underline{t}, t] \subset \mathbb{R}_+$, and assume $u(\cdot)$ is differentiable and strictly increasing in t, so that higher types receive greater expected benefits from engaging in a war.

At this point the common game-theoretic approach is to posit a particular game form for players 1 and 2 and then to analyze the resulting Bayesian game, where a game form characterizes (1) the set of decisions D_i available to player i, $i = 1, 2$, and (2) a (probabilistic) outcome function G describing the likelihood of any one outcome in $X \cup \{w\}$ occurring as a function of the chosen decisions (d_1, d_2).[2] Thus, in any game a decision profile generates a probability

---

[1] Relaxing risk neutrality and assuming instead that player 1 (2) has a strictly increasing (decreasing) utility function over X would not alter the monotonicity results in section 3 (see note 7) and would simply make the characterization result in section 4 more cumbersome.

[2] If the game is one with sequential moves, then a player's decision describes the action he or she would take in every contingency.

p of war occurring and a probability distribution over the set X of settlement outcomes conditional on no war; by risk neutrality we can associate with the latter the expected settlement $x \in [0, 1]$ conditional on no war. Therefore, we can view the outcome function G as a pair of mappings

$$g_s : D_1 \times D_2 \rightarrow [0, 1], \qquad g_w : D_1 \times D_2 \rightarrow [0,1],$$

where $g_s(d_1, d_2)$ is the expected settlement given the decisions (d_1, d_2), and $g_w(d_1, d_2)$ is the associated probability of war. Since player 1 knows the value of $t \in T$ prior to any decision making, he is able to condition his choice of $d_1 \in D_1$ on the realized value of t. Thus, a (pure) strategy for player 1 in the Bayesian game is a function $\sigma_1 : T \rightarrow D_1$, where $\sigma_1(t) \in D_1$ is the decision of player 1 when his type is $t \in T$. Player 2 does not possess any private information; thus, a (pure) strategy for player 2 is simply a selection $\sigma_2 \in D_2$.

A strategy profile (σ_1, σ_2) and a type $t \in T$ thus generate, through the outcome function G, a probability of war $g_w(\sigma_1(t), \sigma_2)$ and an expected settlement $g_s(\sigma_1(t), \sigma_2)$ conditional on no war. Since the players' preferences over such outcomes are well defined, we can discuss the optimality of a player's decision given the opponent's decision and, hence, describe a notion of equilibrium in a game form (D_1, D_2, G). For Bayesian games the appropriate generalization of the Nash equilibrium concept is known as Bayesian equilibrium (cf. Myerson 1985), where a strategy profile (σ_1, σ_2) constitutes a Bayesian equilibrium if (1) for all $t \in T$, $\sigma_1(t)$ is a best response to σ_2 and (2) σ_2 is a best response to σ_1 based on player 2's beliefs $f(\cdot)$ concerning player 1's type (and hence, through σ_1, player 1's decision).

As noted in the Introduction, however, the motivation for the current paper concerns not the qualitative properties of equilibrium behavior in a particular Bayesian game but rather properties of any equilibrium in any Bayesian game. Therefore, the analytical trade-off chosen here is toward general results that are not a function of the particulars of the game structure (i.e., D_1, D_2, G) or the selection of a single equilibrium within a Bayesian game, at the expense of a precise prediction concerning the behavior of the participants and the subsequent ability to carry out comparative statics exercises. However, as we shall see, the general results do have the flavor of comparative statics results in that they describe changes in outcomes as a function of a variable—namely, player 1's type, upon which player 1 can condition his behavior but player 2 cannot. In particular, all of the results will specify the relative likelihood of any outcome as a function of player 1's private information concerning the expected benefits from going to war.

From the above discussion, we see that any strategy profile (σ_1, σ_2) in a Bayesian game generates an outcome (x, p) as a function of player 1's type, by $x(t) = g_s(\sigma_1(t), \sigma_2)$, $p(t) = g_w(\sigma_1(t), \sigma_2)$. Let

$$\Omega = \{(x, p) : x : T \to [0, 1], p : T \to [0, 1]\}$$

denote the set of all possible outcomes from all possible Bayesian games. Clearly not every element of Ω is necessarily derived from an equilibrium of some game; thus, what we would like is a criterion for selecting those elements of Ω that are rationalizable in the sense that they are generated as equilibrium behavior of some Bayesian game. Let

$$U(t; x, p) = p(t)u(t) + [1 - p(t)]x(t)$$

denote player 1's expected utility from the outcome (x, p) given type $t \in T$, where we assume that there exists a strategy profile (σ_1, σ_2) generating (x, p). To determine whether or not σ_1 and σ_2 constitute equilibrium strategies in some game would obviously require knowledge of all available strategies and (through G) outcomes, since, for example, $\sigma_1(t)$ must be the best action from the set D_1 for player 1 if type t. Yet even without such knowledge, we can identify a class of alternative strategies and outcomes that exist for player 1. Since player 1's type only affects the war utilities (u, v) and not the available decisions D_1, one alternative for player 1 to any strategy σ_1 is to have some type t "mimic" the behavior suggested for some other type t', that is, play according to $\sigma_1(t')$ rather than $\sigma_1(t)$. Since player 2's strategy is independent of player 1's (by the Nash assumption), this then generates the outcome $(x(t'), p(t'))$ rather than $(x(t), p(t))$.[3] Define

$$U(t', t; x, p) = p(t')u(t) + [1 - p(t')]x(t')$$

as the expected utility for player 1 from acting as if his type were t' when his type is actually t. If there exists types $t, t' \in T$ such that $U(t) < U(t', t)$, then player 1 can choose strategy σ_1', defined as $\sigma_1'(\hat{t}) = \sigma_1(\hat{t})$ for all $\hat{t} \neq t$ and $\sigma_1'(t) = \sigma_1(t')$, receive the same expected utility for all $\hat{t} \neq t$ and receive a strictly higher expected utility for t. Since the definition of Bayesian equilibrium assumes optimal behavior for player 1 "type-by-type," this then contradicts the assumption of (σ_1, σ_2) being an equilibrium or, in particular, the assumption of σ_1 being a best response to σ_2. But since this holds for all games where there exists a strategy profile generating (x, p), this implies that if $U(t) < U(t', t)$ for some $t, t' \in T$ then the outcome (x, p) is not associated with equilibrium behavior in *any* game. Thus, a necessary condition for an outcome (x, p) to be generated by equilibrium behavior is that it be incentive compatible (d'Aspremont and Gerard-Varet 1979).[4]

---

[3] Recall that in a sequential move game the *actions* chosen by one player may be a function of the *actions* of the opponent, yet a player's *strategy*, which assigns a (possibly different) action at each of the player's information sets, is chosen independently of the opponent's strategy.

[4] Incentive compatibility is also sufficient: let $D_1 = T$, D_2 be any set, and for all $d_2 \in D_2$ let $g_s(t, d_2) = x(t)$, $g_w(t, d_2) = p(t)$ (i.e., player 2's role is suppressed). Then since (x, p) is incentive

DEFINITION: An outcome $(x, p) \in \Omega$ is *incentive compatible* if and only if for all $t, t' \in T$, $U(t; x, p) \geq U(t', t; x, p)$.

In particular, for any $t, t' \in T$, incentive compatibility implies the following inequalities hold:

$$p(t)u(t) + [1 - p(t)]x(t) \geq p(t')u(t) + [1 - p(t')]x(t'), \tag{1}$$

$$p(t')u(t') + [1 - p(t')]x(t') \geq p(t)u(t') + [1 - p(t)]x(t). \tag{2}$$

Equation (1) says that type t receives at least as high an expected utility from the outcome $(x(t), p(t))$ as he would from $(x(t'), p(t'))$, while equation (2) says that for t' the opposite is true. Thus, our principle criterion for identifying the set of equilibrium outcomes is to examine only those outcomes that are incentive compatible.[5]

An additional restriction I place on outcomes has more to do with the nature of the games I wish to examine, in the following sense: suppose (x, p) is derived from some equilibrium profile, and $t \in T$ is such that $p(t) < 1$, that is, with some probability player 1 does not go to war if his type is t. Then we would expect that in any reasonable game $x(t)$, the equilibrium payoff from resolving the dispute, would be at least as large as $u(t)$, the expected payoff from war. Otherwise, so long as there exists some bargaining strategy (e.g., always demanding everything) which generates a payoff of at least $u(t)$, player 1 would never accept a settlement less than $u(t)$. Thus, the additional constraint is that the outcome (x, p) be "individually rational," in the sense of generating a payoff to player 1 that is at least as high as he could get from simply fighting, where this holds for each type (i.e., "interim" individual rationality). Given $(x, p) \in \Omega$ let $T_b(x, p) = \{t \in T : p(t) < 1\}$ denote those types who with positive probability resolve the dispute in the bargaining process. Individual rationality then implies that for all $t \in T_b$, $x(t) \geq u(t)$ or, equivalently, that for all $t \in T$, $U(t) \geq u(t)$. Let $\Omega^* \subseteq \Omega$ denote the set of outcomes (x, p) that are incentive compatible and individually rational.

3. Monotonicity Results

In this section we derive some qualitative features of elements of the set Ω^*, with the conclusion being that such features hold in any equilibrium of any Bayesian game where the set of outcomes and the preferences (i.e., the environ-

---

compatible, the strategy $\sigma_1(t) = t$ is optimal for player 1, thereby generating (x, p). The result that incentive compatibility is necessary and sufficient for equilibrium behavior is known in the economics literature as the *revelation principle* (cf. Dasgupta, Hammond, and Maskin 1979; Myerson 1979; and Rosenthal 1978).

[5]Myerson and Satterthwaite (1983) use this approach to characterize equilibrium outcomes in a bilateral bargaining environment. Incentive compatibility conditions and the revelation principle

ment) are as described above. Our first result concerns the likelihood of war as a function of player 1's type.

LEMMA 1: If $(x, p) \in \Omega^*$, then $p(t)$ is weakly increasing on T.

PROOF: Let $t' > t$. Subtracting the right-hand side (RHS) of equation (2) from the left-hand side (LHS) of equation (1), and the LHS of (2) from the RHS of (1), we get

$$x(t) - p(t)[x(t) - u(t)] - \{x(t) - p(t)[x(t) - u(t')]\} \geq \tag{3}$$
$$x(t') - p(t')[x(t') - u(t)] - \{x(t') - p(t')[x(t') - u(t')]\}.$$

Canceling terms, we get

$$p(t')[u(t') - u(t)] \geq p(t)[u(t') - u(t)]. \tag{4}$$

Since $t' > t$ and $u(\cdot)$ is strictly increasing, $p(t') \geq p(t)$.   QED

Thus, given the environment outlined in section 2, for any game form (D_1, D_2, G) and any equilibrium (σ_1, σ_2) of the resulting Bayesian game, the probability of war $g_w(\sigma_1(t), \sigma_2)$ is weakly increasing in t (i.e., in equilibrium the probability of war is an increasing function of player 1's expected benefits from war). This justifies the assumption in the expected utility models of Bueno de Mesquita and Lalman (1986) and Lalman (1988) that a decision maker with a higher expected benefit from war will be more likely to go to war; indeed Lemma 1 shows this to be the *only* assumption consistent with rational behavior in an incomplete information environment.[6] It also shows how the presence of such monotonicity in the equilibria analyzed by Morrow (1989) is not an artifact of the particular game form assumed nor an artifact of any selection from among the set of Bayesian equilibria in the game.

With regard to the expected settlement $x(t)$ conditional on not fighting, it is clear that for $t \in T_b$ such a value is not relevant, since these types always go to war. For the remaining types, however, the next result shows the monotonicity implied by incentive compatibility and individual rationality.

LEMMA 2: If $(x, p) \in \Omega^*$ then $x(t)$ is weakly increasing on T_b.

PROOF: Let $t', t \in T_b$ and $t' > t$, so (by Lemma 1) $1 > p(t') \geq p(t)$. Since $x(t) \geq u(t) \, \forall \, t \in T_b$ (by individual rationality),

$$p(t')u(t') + [1 - p(t')]x(t') \leq p(t)u(t') + [1 - p(t)]x(t'). \tag{5}$$

---

are also useful for deriving optimal allocation schemes (Harris and Raviv 1981), optimal contracts in principle-agent settings (Holmstrom 1979), and even equilibrium strategies in particular Bayesian games (Banks 1989).

[6]See Lalman (1988) for a discussion of the monotonicity assumption in expected utility models.

The LHS of equation (5) is equal to the LHS of equation (2); thus, combining (5) and (2) yields

$$p(t)u(t') + [1 - p(t)]x(t) \leq p(t)u(t') + [1 - p(t)]x(t'). \tag{6}$$

Canceling terms and then dividing both sides by $[1 - p(t)]$ (which is nonzero, since $p(t) < 1$) implies $x(t') \geq x(t)$.   QED

Thus, while higher types go to war at least as often as lower types, they also receive at least as high expected benefits if no war is fought.[7] The next result shows that, if one of these relations is strict the other must be as well.

LEMMA 3: If $(x, p) \in \Omega^*$, $t' > t$, and $t, t' \in T_b$, then $x(t') > x(t)$ if and only if $p(t') > p(t)$.

PROOF: Suppose not; by Lemmas 1 and 2 there are only two cases to consider:

(i) $x(t') > x(t)$ and $p(t') = p(t)$; but this contradicts equation (1), since $p(t) < 1$.

(ii) $x(t') = x(t)$ and $p(t') > p(t)$; but this contradicts equation (2).
   QED

Therefore, in crisis bargaining situations, equilibrium analysis predicts the following trade-off between the gains from settling the dispute and the probability of war: as the expected benefits of war increase, the informed player receives a better negotiated settlement but in addition runs a greater risk of war. Furthermore, this prediction is derived from the general properties of optimizing behavior of the participants and hence will hold in any crisis bargaining model with the incomplete information environment detailed in section 2.

Incentive compatibility of course also implies such trade-offs are beneficial for all types; indeed, the next result shows that the equilibrium expected utility of player 1 is increasing in t. Let $T_w = \{t \in T : p(t) > 0\}$ denote those types that with positive probability go to war.

LEMMA 4: If $(x, p) \in \Omega^*$, then $U(t; x, p)$ is continuous, weakly increasing on T, and strictly increasing on T_w.

PROOF: Suppose $t' > t$, $t, t' \in T_w$, and $U(t) \geq U(t')$, implying

$$p(t)u(t) + [1 - p(t)]x(t) \geq p(t')u(t') + [1 - p(t')]x(t'). \tag{7}$$

---

[7]Suppose we drop risk neutrality and assume player 1 has a strictly increasing utility function z over X, so $z(x(t))$ denotes 1's utility from the settlement $x(t)$. Let $4(t)$ denote the expected *utility* conditional on no war for type t from the outcome (x, p). Then it is easily seen that Lemma 1 continues to hold, while Lemma 2 holds with 4 replacing x, i.e. the expected utility, rather than the expected settlement.

Then since $u(t') > u(t)$,

$$p(t)u(t') + [1 - p(t)]x(t) \geq p(t')u(t') + [1 - p(t')]x(t'), \tag{8}$$

implying if $t, t' \in T_w$ (i.e., $p(t), p(t') > 0$) then $U(t, t'; x, p) > U(t')$, contradicting incentive compatibility. If $t, t' \notin T_w$ then clearly incentive compatibility implies $x(t) = x(t')$, yielding $U(t) = U(t')$. To see $U(\cdot)$ is continuous, note that $U(\cdot)$ monotone implies that any discontinuities are jump discontinuities, so for all $t \in T$ the left- and right-hand limits of $U(\cdot)$ at t, $\lim_{t_-} U(\cdot)$ and $\lim_{t_+} U(\cdot)$, exist. If $U(\cdot)$ is discontinuous at t, then $\lim_{t_+} U(\cdot) - \lim_{t_-} U(\cdot) \geq \epsilon > 0$. Choose types $t - \delta$ and $t + \delta$; then since $u(\cdot)$ is assumed to be differentiable and hence continuous, for δ sufficiently small $U(t - \delta; x, p) < U(t + \delta, t - \delta; x, p)$, contradicting incentive compatibility.   QED

From Lemma 4 we know the equilibrium utility of player 1 is increasing in his type. However, a different result comes about when we consider the expected gain in utility for player 1 above that generated by war. For any $(x, p) \in \Omega$ let $\Delta(t; x, p) = U(t; x, p) - u(t)$ denote this difference.

LEMMA 5: If $(x, p) \in \Omega^*$, then $\Delta(t; x, p)$ is weakly decreasing on T and strictly decreasing on T_b.

PROOF: If $t \in T \backslash T_b$ then $U(t) = u(t)$, so the result follows. For $t \in T_b$, incentive compatibility implies that for all t',

$$\Delta(t) = [x(t) - u(t)] \cdot [1 - p(t)] \geq [x(t') - u(t)] \cdot [1 - p(t')]. \tag{9}$$

Let $t' > t$ and $t' \in T_b$; then since $u(\cdot)$ is strictly increasing,

$$[x(t') - u(t)] \cdot [1 - p(t')] > [x(t') - u(t')] \cdot [1 - p(t')] = \Delta(t'). \tag{10}$$

Combining equations (9) and (10), we get $\Delta(t) > \Delta(t')$.   QED

Thus, the gain from participating in the bargaining process and potentially resolving the dispute over simply going to war is decreasing in player 1's expected benefits from such a war. In addition, Lemma 5 implies that if (x, p) is incentive compatible, then we need only check the individual rationality constraint $U(t; x, p) - u(t) \geq 0$ at $t_b \equiv \sup \{t \in T_b\}$, since if it is satisfied at t_b by Lemma 5 it will be satisfied for all $t < t_b$ as well.

It is easily seen that none of the above results are sensitive to player 2's prior belief $f(\cdot)$ concerning 1's type, the functional form of $u(\cdot)$, the assumption that T is not finite, or (for that matter) the preferences or actions of player 2. Rather, these monotonicity results are derived simply through the optimizing behavior of player 1 and the willingness and ability of player 1 to differentiate his bargaining behavior as a function of his information concerning the expected benefits of war. Hence, what drives the results is not the competition among the players per

se but the ability of player 1 to make his decisions contingent upon payoff-relevant and private information.

Suppose that we add a little bit more structure to the bargaining process we envision. In particular, let player 1 be the "initiator" of the crisis, in that the first move of the process has player 1 selecting whether to stay with the status quo ante, namely, the outcome $(x_o, 0)$, or begin the bargaining. This structure then places an additional "individual rationality" restriction on the equilibrium set of outcomes in that for all $t \in T$ the following condition must hold: $U(t; x, p) = p(t)u(t) + [1 - p(t)]x(t) \geq x_o$. This follows, since now player 1 can guarantee himself a payoff of x_o by simply failing to initiate a crisis. For any $(x, p) \in \Omega$ let $T_s = \{t \in T : x(t) = x_o, p(t) = 0\}$ denote those types $t \in T$ that unilaterally select the $(x_o, 0)$ outcome.[8] Now if $(x, p) \in \Omega^*$ is such that $T_s \neq \phi$, then for all $t \notin T_s$, $p(t) > 0$. This follows, since if not, then for some $t \notin T_s$ $p(t) = 0$ but $x(t) > x_o$, which contradicts Lemma 3. Therefore, $T_s = T \backslash T_w$, and all types that do not receive the status quo outcome face a positive probability of going to war. This conforms to the "selection bias" noted in Bueno de Mesquita (1981) and Morrow (1989), in that, conditional on a crisis occurring (i.e., player 1 not selecting $(x_o, 0)$), the posterior probability distribution of player 1's type should not be the same as the prior belief but rather should place positive weight only on those types not in T_s. In addition, Lemma 1 tells us that this posterior distribution should place greater weight (relative to the prior) on higher types. The types of initiators that begin a crisis are thus not "typical" in the sense of being the expected type according to the prior $f(\cdot)$ and neither are those that engage in war. Thus, for example, there will always exist a selection bias in the observed military capabilities of those countries that initiate crises and fight wars.

4. A Characterization Theorem

As noted above, all of the monotonicity results in section 3 go through if the set of types T is finite. With continuous types, however, we are able succinctly to characterize the set Ω^* by using calculus-based techniques. Since $x(\cdot)$, $p(\cdot)$, and $U(\cdot)$ are monotone increasing and T is a closed interval, $x(\cdot)$, $p(\cdot)$, and $U(\cdot)$ are differentiable almost everywhere (i.e., except on a set of measure zero) (Royden 1968). In particular, for almost all $t \in T$, either

(i) $p(t) = 1$ and $\partial p/\partial t = 0$,
(ii) $p(t) < 1$ and $\partial x/\partial t = \partial p/\partial t = 0$, or
(iii) $p(t) < 1$ and $\partial x/\partial t > 0$, $\partial p/\partial t > 0$.

---

[8] If we imagine an arbitrarily small cost to initiating the bargaining process, then the outcome $(x_o, 0)$ cannot occur as an equilibrium outcome subsequent to initiating the process; hence, $(x_o, 0)$ will only occur when player 1 selects this at the outset.

Types where case (i) holds are in $T \backslash T_b$ in that these types always go to war. We can think of those types where case (ii) holds as (locally) *pooling*, in that they either adopt the same behavioral strategies in the underlying game, or they adopt different strategies wherein these differences are irrelevant to the behavior by player 2 and the subsequent outcome. Those in case (iii) are *separating*, in that they are adopting distinctly different behavior in the underlying game.[9] Fix an outcome $(x, p) \in \Omega^*$, and consider the increase in the equilibrium utility $U(t)$ for player 1 as t increases:

$$\frac{\partial U}{\partial t} = \frac{\partial x}{\partial t} \cdot [1 - p(t)] - \frac{\partial p}{\partial t} \cdot [x(t) - u(t)] + \frac{\partial u}{\partial t} \cdot p(t). \tag{11}$$

For the pooling types and $t \in T \backslash T_b$, $\partial U / \partial t = \partial u / \partial t \cdot p(t)$. For the separating types, note that if $(x, p) \in \Omega^*$, the incentive compatibility condition $U(t) \geq U(t', t)$ holds with equality at $t' = t$. This along with the differentiability of $x(\cdot)$ and $p(\cdot)$ implies the following "local" incentive compatibility condition:

$$\left. \frac{\partial U(t', t)}{\partial t'} \right|_{t'=t} = \frac{\partial x}{\partial t} \cdot [1 - p(t)] - \frac{\partial p}{\partial t} \cdot [x(t) - u(t)] = 0. \tag{12}$$

Plugging equation (12) into equation (11), we get that for separating types $\partial U / \partial t = \partial u / \partial t \cdot p(t)$ as well. Thus, we have proven the following result.

LEMMA 6: If $(x, p) \in \Omega^*$, then for almost all $t \in T$,

$$\frac{\partial U}{\partial t} = \frac{\partial u}{\partial t} \cdot p(t). \tag{13}$$

Lemma 6 is analogous to the "envelope theorem" for single-person optimization problems (cf. Takayama 1986). Increasing player 1's type has a "direct" effect on $U(t)$ through the increase in players 1's utility from war and an "indirect" effect through changes (if any) in the functions $x(\cdot)$ and $p(\cdot)$. Now given the behavior suggested by (x, p), we can think of each type as solving an optimization program with regard to which type to act like, with the implication of incentive compatibility being that in equilibrium each type optimally selects his true type. But then local incentive compatibility implies that these indirect effects vanish as we vary the "parameter" t along the derived solutions to player 1's optimization program, which is simply the envelope theorem. Thus, in any equilibrium of any Bayesian game, the increase in player 1's equilibrium utility as a function of an increase in his type can be expressed as a simple function of the probability of war, $p(t)$, and the marginal gain in expected benefits from war, $\partial u / \partial t$.

[9] Here I do not mean necessarily to imply that player 1 signals his information to player 2, since player 1's behavior may only differ at some "final" move prior to war, where player 2 would not have any subsequent moves.

Note that local incentive compatibility actually holds for all types, not just separating types, since if $t \notin T_b$ then $[1 - p(t)] = \partial p/\partial t = 0$, while for $t \in T_b$ that are locally pooling $\partial x/\partial t = \partial p/\partial t = 0$. Thus, if $(x, p) \in \Omega^*$ then equation (12) is satisfied almost everywhere. In addition, it turns out that local incentive compatibility, along with $p(\cdot)$ increasing and $U(\cdot)$ continuous, implies "global" incentive compatibility.

LEMMA 7: If $(x, p) \in \Omega$ is such that $p(t)$ is increasing on T, $U(t; x, p)$ is continuous on T, and equation (12) holds, then (x, p) is incentive compatible.

PROOF: Rewrite $U(t', t)$ as

$$U(t', t) = U(t') + p(t')[u(t) - u(t')]. \tag{14}$$

Since equation (14) holds identically (i.e., for all $t' \in T$), the derivatives of both sides are equal. Thus, for almost all $t, t' \in T$,

$$\frac{\partial U(t', t)}{\partial t'} = \frac{\partial U}{\partial t'} + \frac{\partial p}{\partial t'}[u(t) - u(t')] - \frac{\partial u}{\partial t'} \cdot p(t'). \tag{15}$$

From Lemma 6, the first and last terms cancel. Thus,

$$\frac{\partial U(t', t)}{\partial t'} = \frac{\partial p}{\partial t'}[u(t) - u(t')]. \tag{16}$$

Since $u(\cdot)$ is increasing and $\partial p/\partial t \geq 0$, $\partial U(t', t)/\partial t' \geq 0$ if $t > t'$ and $\partial U(t', t)/\partial t' \leq 0$ if $t < t'$, so that $U(t', t)$ is weakly increasing on $[\underline{t}, t)$ and weakly decreasing on $(t, \bar{t}]$. This plus the continuity of $U(t)$, which implies the continuity of $U(t', t)$ at t, implies for all $t \in T$, $t \in \mathrm{argmax}_{t'} U(t', t)$, so that (x, p) is incentive compatible.   QED

Thus (by Lemmas 1, 4, and 7), $p(t)$ increasing, $U(t; x, p)$ continuous, and local incentive compatibility (i.e., equation (12)) are necessary and sufficient conditions for incentive compatibility. Integrating both sides of equation (13), we see that for almost all t,

$$U(t) = x(t)[1 - p(t)] + p(t)u(t) = U(\underline{t}) + \int_{\underline{t}}^{t} \frac{\partial u(\hat{t})}{\partial \hat{t}} \cdot p(\hat{t})d\hat{t}. \tag{17}$$

Using integration by parts, we can rewrite the integral in equation (17) as

$$\int_{\underline{t}}^{t} \frac{\partial u(\hat{t})}{\partial \hat{t}} \cdot p(\hat{t})d\hat{t} = -\int_{\underline{t}}^{t} \frac{\partial p(\hat{t})}{\partial \hat{t}} \cdot u(\hat{t})d\hat{t} + p(t)u(t) - p(\underline{t})u(\underline{t}). \tag{18}$$

Plugging this into equation (17) and rearranging terms, we get

$$x(t) = \left\{ x(\underline{t})[1 - p(\underline{t})] - \int_{\underline{t}}^{t} \frac{\partial p(\hat{t})}{\partial \hat{t}} \cdot u(\hat{t})d\hat{t} \right\} / [1 - p(t)]. \tag{19}$$

Hence, given a weakly increasing function $p : T \rightarrow [0, 1]$ and a value $x(\underline{t})$, equation (19) can be used to derive the expected settlement function $x : T \rightarrow [0, 1]$ necessary for (x, p) to be incentive compatible; for values of t where $p(\cdot)$ is nondifferentiable, $x(\cdot)$ is derived by the requirement that $U(t; x, p)$ be continuous (such a set of types is countable, since $p(\cdot)$ is monotonic on a closed interval). Finally, since incentive compatibility and $x(t_b) \geq u(t_b)$ are sufficient conditions for individual rationality (by Lemma 5), we have the following characterization of outcomes in the set Ω^*.

> THEOREM: Let $(x, p) \in \Omega$; then $(x, p) \in \Omega^*$ if and only if $p(t)$ is weakly increasing on T, $U(t; p, x)$ is continuous on T, $x(t_b) \geq u(t_b)$, and for almost all $t \in T_b$ $x(t)$ is as in equation (19).

Hence, given an increasing probability of war function $p(t)$, we can solve for the settlement conditional on no war $x(t)$ that will "rationalize" $p(t)$, in that the pair (x, p) constitute equilibrium behavior of some Bayesian game; if no such function $x(t)$ exists then $p(t)$ could not have been derived from equilibrium behavior.[10] Alternatively, any outcome (x, p) derived from equilibrium behavior in a Bayesian game must satisfy equation (19) and the individual rationality condition $x(t_b) \geq u(t_b)$.

> EXAMPLE: Let $T = [0, 1]$, $u(t) = t/2$, and $p(t) = 1 - e^{-t}$, so that $p(0) = 0$ and $p(1) = 1 - 1/e$; then $x(t) = (t + 1)/2 - [1/2 - x(0)]e^t$. Individual rationality implies $x(1) \geq 1/2$, so that if $1/2 \geq x(0) \geq (e - 1)/2e$, the pair (x, p) is feasible (i.e., $(x, p) \in \Omega$), incentive compatible, and individually rational.

From equation (12), we can also say something about whether $x(t)$ is increasing faster or slower than $p(t)$ at any separating type:

$$\left. \frac{\partial x}{\partial t} \right|_{t'} \gtreqless \left. \frac{\partial p}{\partial t} \right|_{t'} \quad \text{as} \quad [x(t') - u(t')] \gtreqless [1 - p(t')]. \tag{20}$$

Therefore, if the difference in expected utility from resolving the dispute versus war is large relative to the probability of resolving the dispute, then the expected utility from resolution is increasing faster than the probability of war. Further, since equation (19) holds for amost all types, we see that if $(x, p) \in \Omega^*$ then the function $x(t)[1 - p(t)]$ is decreasing in t. Thus,

$$\frac{\partial x}{\partial t} \frac{1}{x(t)} \leq \frac{\partial p}{\partial t} \frac{1}{1 - p(t)}. \tag{21}$$

Multiplying both sides by t (recall $T \subset \mathbb{R}_+$) results in expressions known in economics as *elasticities* (cf. Takayama 1986), where the elasticity $\epsilon_x \equiv |[\partial x/$

---

[10]Of course, the reverse analysis works as well (i.e., given an increasing function $x : T \rightarrow [0, 1]$, we can solve for the required $p : T \rightarrow [0, 1]$).

$\partial t]\cdot[t/x(t)]|$ measures the percentage change in x due to a percentage change in t; similarly for $\epsilon_p \equiv |[\partial(1 - p(t))/\partial t]\cdot[t/(1 - p(t))]| = [\partial p/\partial t]\cdot[t/(1 - p(t))]$. Elasticities are useful in that they give a dimension-free measure of the responsiveness of a function, in contrast to a derivative. From equation (21), we have the following result.

> COROLLARY: If $(x, p) \in \Omega^*$, then $\epsilon_p \geq \epsilon_x$, that is, the probability of no war, $1 - p(t)$, is more elastic than $x(t)$, the settlement conditional on no war.

Therefore, a 1% increase in player 1's expected benefits from war leads to a greater percentage decrease in the probability of no war than the percentage increase in the expected benefits from resolving the dispute short of war.

5. Conclusion

This paper has analyzed a simple model of crisis bargaining with incomplete information where, rather than specify the actual game the participants play, we derived results which hold for any equilibrium of any such game. In this fashion we can unambiguously determine the effect on crisis bargaining outcomes (i.e., the probability of war and the benefits from resolving the dispute short of war) of the expected benefits from war. We see that the higher the informed player's benefits from war, the more likely the dispute will end in war; conversely, if the dispute is settled short of war, the better is the negotiated settlement.

From a methodological perspective, it is important to point out that the approach taken in the current paper should not be viewed as a substitute for the more common approach of explicitly modeling the game. This immediately follows by noting what the incentive compatibility approach *cannot* do. Most important, this approach cannot address the issue of the informed player's *perceived* level of expected benefits from war, where such a perception is summarized ex ante by player 2's prior belief $f(\cdot)$ concerning 1's type. Since the function $f(\cdot)$ is an actual parameter of the model, meaningful results on the effect of changes in $f(\cdot)$ on equilibrium outcomes requires the explicit modeling approach. Rather, these two approaches should be seen as complimentary, in that incentive compatibility can generate certain types of results, while the specifics of the game form hypothesized can generate others. In particular, a "two-step" approach to incomplete information games might be useful, where the first step would be to generate as many results as possible from simply the specification of the environment and the resulting incentive compatibility constraints, and then move on to the specification of a particular game form and the determination of a particular behavioral prediction.

In terms of generalizing the current model, the most obvious extension

would be to have player 2 possess private information as well. In such a situation, then, both players would face incentive compatibility constraints, and the results would then pertain to the behavior of both players. If we think of player 2's information concerning her own expected benefits from war, then it is not too difficult to foresee how the monotonicity results of section 3 will generalize. However, the constraints on each player's behavior will also include the prior beliefs that concern the opponent's type, so that generalizing the characterization theorem of section 4 will prove to be a little trickier. Other possible extensions, such as expanding the outcome space to include the temporal length of the bargaining prior to either compromise or war, should be explored in further research.

Manuscript submitted 31 August 1989
Final manuscript received 18 December 1989

REFERENCES

Banks, Jeffrey. N.d. "A Model of Electoral Competition with Incomplete Information." *Journal of Economic Theory,* forthcoming.

Bueno de Mesquita, Bruce. 1981. *The War Trap.* New Haven: Yale University Press.

Bueno de Mesquita, Bruce, and David Lalman. 1986. "Reason and War." *American Political Science Review* 80:1113–30.

————. N.d. "Domestic Opposition and Foreign War." *American Political Science Review,* forthcoming.

Dasgupta, Partha, Peter Hammond, and Eric Maskin. 1979. "The Implementation of Social Choice Rules: Some Results on Incentive Compatibility." *Review of Economic Studies* 46:185–216.

d'Aspremont, Claude, and Louis-Andre Gerard-Varet. 1979. "Incentives and Incomplete Information." *Journal of Public Economics* 11:25–45.

Harris, Milton, and Arthur Raviv. 1981. "Allocation Mechanisms and the Design of Auctions." *Econometrica* 49:1477–99.

Harsanyi, John. 1967–68. "Games with Incomplete Information Played by Bayesian Players." *Management Science* 14:159–82, 320–34, 486–502.

Holmstrom, Bengt. 1979. "Moral Hazard and Observability." *Bell Journal of Economics* 10:74–91.

Lalman, David. 1988. "Conflict Resolution and Peace." *American Journal of Political Science* 32:590–615.

Morrow, James. 1989. "Capabilities, Uncertainty, and Resolve: A Limited Information Model of Crisis Bargaining." *American Journal of Political Science* 33:941–72.

Myerson, Roger. 1979. "Incentive Compatibility and the Bargaining Problem." *Econometrica* 47:61–74.

————. 1985. "Bayesian Equilibrium and Incentive Compatibility: An Introduction." In *Social Goals and Social Organization: Essays in Memory of Elisha Pazner,* ed. Leonid Hurwicz, David Schmeidler, and Hugo Sonnenschein. Cambridge: Cambridge University Press.

Myerson, Roger, and Mark Satterthwaite. 1983. "Efficient Mechanisms for Bilateral Trading." *Journal of Economic Theory* 29:265–81.

Powell, Robert. 1987 "Crisis Bargaining, Escalation, and MAD." *American Political Science Review* 81:717–35.

Rosenthal, Robert. 1978. "Arbitration of Two-Party Disputes under Uncertainty." *Review of Economic Studies* 45:595–604.

Royden, H. 1968. *Real Analysis,* 2d ed. New York: Macmillan.

Takayama, Akira. 1986. *Mathematical Economics.* 2d ed. Cambridge: Cambridge University Press.