Center for Mathematical Studies in Economics and Management Science

Northwestern University, Evanston, Illinois 60201

Discussion Paper No. 691

CREDIBLE NEGOTIATION STATEMENTS AND COHERENT PLANS

by

Roger B. Myerson

August 1986
revised, August 1987

Abstract. Criteria are developed to determine which negotiation statements are credible with respect to a reference payoff allocation for the negotiator's possible types. An attractive reference allocation is any limit of reference allocations that admit no credible statements. A coherent plan is a negotiation statement that all types can make with likelihood one that is credible with respect to an attractive reference allocation. Coherent plans are shown to exist. Semicoherent plans are also defined, without reference allocations. Sequentially coherent plans are defined for multistage games with no simultaneous moves. Other negotiation structures are considered, including mechanism design by an informed principal.

1. Introduction.

Nash [18] argued that bargaining and cooperation in games should be studied using the methodology of noncooperative game theory, by modelling the bargaining process itself as a noncooperative game and analyzing its equilibria. Although this research program has generated a number of remarkable results and insights, there has been a growing sense that its power is limited by the fact that many natural bargaining-game models have very large sets of Nash equilibria.

The advantage of Nash's program is that the derivation of behavior from rational decision-making by individuals has traditionally been much clearer in noncooperative game theory than in cooperative game theory. However, what seems to be missing from noncooperative models is the essential assumption that people listen and understand each other when they communicate to coordinate their decisions.

The goal of this paper is to try to provide new foundations for cooperative game theory by developing a theory of coherent decision-making by individual negotiators. To clarify the individualistic foundations of the analysis, most of the paper focuses on the determination of a single negotiation statement by one individual. We assume that this individual negotiator has an ability to make himself understood by other individuals which goes beyond what is assumed in noncooperative game theory and which gives the model its "cooperative" nature.

In general, negotiation may be defined as any communication process in which individuals try to determine or influence the effective equilibrium that they will play thereafter in some game. In many economic situations, it seems reasonable to believe that individuals will be able to coordinate effectively

with each other and implement an equilibrium that is (at least)
Pareto-undominated within the set of equilibria (given some appropriate
refinement of the equilibrium concept). Such a belief may be derived from an
assumption that they can negotiate, in this sense.

However, the theory of noncooperative games with signalling and
communication (based largely Aumann's [1] concept of correlated equilibrium
and Kreps and Wilson's [10] concept of sequential equilibrium) derives the
meaning of all statements and signals from the equilibrium in which they are
used. Thus, noncooperative game theory does not permit us to suppose that
communication can determine the effective equilibrium that is actually played,
because the meanings of all statements are supposed to be determined by the
effective equilibrium itself. In fact, if the mere act of saying something
does not directly affect any payoffs then there is always a "babbling"
equilibrium, in which every player randomizes over the set of his possible
statements independently of his information and his payoff-relevant actions,
and in which all other players ignore his meaningless statements. Such
analysis suggests that communication can only increase the set of equilibria
(as, indeed, the correlated equilibria include all Nash equilibria), and cannot
provide a way to select among equilibria.

To escape from this conclusion, we must drop the assumption that
statements have no absolute meanings beyond what is endogenously determined
by the equilibrium in which they are used. Instead, we must introduce an
assumption that negotiation statements have literal meanings that are
exogenously defined, at least under certain circumstances, as statements in
a rich language like English which every player understands. For example,
the message "let's meet at the train station tomorrow at noon" has a given

literal meaning in any conversation between two English-speaking individuals who are residents of a town with one train station and who both want to meet again soon. We may suppose that negotiations to determine equilibrium behavior in the game are conducted in such a language in which statements have exogenous literal meanings. Then, using the assumption that such literal meanings are understood by the players in negotiations, a negotiation structure may indeed determine a set of "coherent" equilibria that is narrower than the set of equilibria of the corresponding game without negotiations, instead of enlarging the set of equilibria as a general communication system does.

Rigorously describing the role of literal meanings in such negotiations can be a subtle problem, however, because rational individuals cannot be expected to believe every possible statement. That is, some statements have literal meanings that simply are not credible. For example, if a stranger approached you on the street and said "let me hold your wallet for a minute and I will give it back to you with twice as much money," you would understand what he is trying to say, but you would probably not believe him or do as he asks. In a society where many people would accept and believe such a statement, others would have an incentive to abuse their trust by using this statement with a different effective meaning, so that it would become effectively synonymous with "give me your wallet and watch how fast I can run away with it."

Thus, we need to develop general criteria to determine which statements cannot be credibly used according to their literal meanings. To be true to our assumption that literal meanings of statements are understood by all players in negotiations, we should make such criteria as narrow as possible and assume that individuals will accept the literal meaning of each others'

statements unless there is a strong logical reason to distrust them.

The literal meaning of any negotiator's statement may be analyzed into three components: an allegation that describes some private information which may be known by the negotiator; a promise that describes how the negotiator may plan to choose his own future actions and messages; and a request (or suggestion) that describes strategies for the other players which the negotiator may want or expect them to use hereafter. After hearing a statement by some negotiator, the other players may ask themselves whether the following three conditions are satisfied.

(1.1)   If all other players believe the negotiator's allegations and promises, then it should be rational for them all to obey his requests.

(1.2)   If the negotiator expects all other players to obey his requests, then it should be rational for him to fulfill his promises.

(1.3)   The information that the negotiator alleges should be consistent with the information that could be inferred about him from the fact that he wants other players to use the strategies that he has requested, rather than some other strategies that they might have otherwise been willing to use.

If (1.1) is satisfied, then we say that the statement (or the request in it) is tenable. If (1.2) is satisfied, then we say that the statement (or the promise in it) is reliable. If (1.3) is satisfied, then we say that the statement (or the allegation in it) is plausible. If a statement is tenable, reliable, and plausible, then we say that is credible.

Our key assumption is that, if a statement is credible in this sense, so that conditions (1.1) - (1.3) are all satisfied, then the other players will believe what the negotiator alleges and promises about his own information and strategy, and they will all obey his requests. On the other hand, if any of these conditions are not satisfied, then the literal meaning of the statement is not credible, and so the inferences and responses that the other players make to this statement (that is, the effective meaning of the statement) can be arbitrarily determined in any way permitted by Kreps and Wilson's [10] concept of sequential equilibrium, without further regard for the literal meaning of the statement.

Of these three conditions, plausibility (1.3) has been the hardest to formulate in a rigorous mathematical model, because it implicitly relies on some notion of what the negotiator could have achieved by making some other negotiation statement. Following common game-theoretic terminology, we may refer to the private information that a player has at the time that he makes a negotiation statement as his type. Then the formal definition of plausibility requires us to determine reference payoffs that represent a conjecture about what each type of the negotiator could have gotten "otherwise." Once such reference payoffs are specified, one might infer from a request that the negotiator's actual type is unlikely to be in the set of types that would get less than their reference payoffs if this request were obeyed.

As we seek to develop a solution concept to predict the outcome of negotiations, we can assume without loss of generality that all statements that are actually used with positive probability in our solution are used consistently with their literal meanings. If this assumption were violated then we could make it true by simply redefining the way that literal meanings

are assigned to statements in the language that negotiators use.

The assumption that literal meanings are understood by all players has analytical force only when it is applied to a rich language of statements that are supposed to get zero probability in our solutions. Farrell [5] calls such statements that are not used in the predicted solution neologisms. The assumption that the literal meaning of these neologisms would be understood by all players may constrain the response that such neologisms would elicit if they were used, so that they might become profitable alternatives that tempt negotiators away from the predicted solution. Thus, as Farrell has argued, our solutions must be defined so that there are no neologisms that a negotiator could more profitably use. That is, we want our analysis ultimately to predict the negotiation statements and equilibrium strategies that would be used by the each player in a given game with a negotiation structure; but to make such predictions, we must be able to show that no player would be able to find any other credible negotiation statements that would be more profitable for him than the predicted outcome. To guarantee this, most neologisms should not be credible in our solutions.

This observation is the key to determining the reference payoffs that are required to formalize the concept of credibility. A negotiator's reference payoffs must be determined in a way that narrows his range of credible statements down sufficiently so that a specific coherent plan can be predicted. This idea is used in Section 5 to derive our fundamental definition of coherent plans. That is, a coherent plan is defined to be a plan that is credible with respect to a standard of credibility that admits essentially no other credible statements.

This paper may be viewed as a part of the growing literature on ways of

selecting among equilibria of games with signalling or communication. (See also Kalai and Samet [7], Kohlberg and Mertens [8], McLennan [13], Kreps [9], Banks and Sobel [2], and Cho and Kreps [3], for other recent contributions in this area.) Kumar's [11] investigation of sequential selection of mechanisms is also closely related to the subject of this paper. This paper builds most directly on the papers of Farrell [5], Grossman and Perry [6], and Myerson [16]. In each of these three papers, credible statements and signals can play a role in selecting among equilibria, because of the constraint that they should be interpreted according to their literal or natural meanings. Farrell [5] has most clearly articulated the focal role of credible literal meanings in such analysis. The dynamic models studied in this paper are adapted from the work of Grossman and Perry [6]. (See also Okuno-Fujiwara and Postlewaite [19] for other interesting modifications of Grossman and Perry's solution concept.) The fundamental solution concept used in this paper is a generalization of the solution concept developed by Myerson [16].

The plan of this paper is as follows. A basic model of the negotiation problem faced by one player in one stage of a game is developed in Sections 2 and 3. In Section 4, we define a generalization of this negotiation model which contains all the mathematical structures needed in Section 5. Section 5 develops the formal definition of a coherent plan and the general existence theorem, and also introduces a related but weaker concept of semicoherent plans. Section 6 contains two simple examples that illustrate the properties of coherent plans. In Section 7, we show how the model of Sections 2 and 3 may be embedded in a multistage game in which players move one at a time, to define the sequentially coherent plans of such a game. Section 8 discusses

some other negotiation structures, and, using the general negotiation model

of Section 4, shows that the concept of coherent plans generalizes the concept

of neutral optima that was defined by Myerson [16] for the problem of mechanism

design by an informed principal. Section 9 contains the proofs of all

theorems.

Readers in need of illustrative examples may find it helpful to skip ahead

and look at Section 6 before reading the more technical Sections 2 - 5 in

detail.

2. A basic model of negotiation statements.

In this and the next two sections, we develop a series of models of

negotiation statements and of the environment confronting a negotiator. Before

developing our first model of a negotiation move, however, a brief digression

on mathematical notation is necessary. In general, for any finite set X, we

will let $\Delta(X)$ denote the set of all probability distributions over X, so that

$$\Delta(X) = \{p:X \to \mathbb{R} | \sum_{x \in X} p(x) = 1, \text{ and } p(y) \geq 0 \quad \forall y \in X\}.$$

For any finite set X, we let $\Lambda(X)$ denote the set of subprobability

distributions on X, where a subprobability distribution differs from a

probability distribution in that the sum of the weights may be less than or

equal to one, instead of only equal to one. That is,

$$\Lambda(X) = \{p:X \to \mathbb{R} | \sum_{x \in X} p(x) \leq 1, \text{ and } p(y) \geq 0 \quad \forall y \in X\}.$$

As usual, [0,1] denotes the interval from 0 to 1, including both endpoints;

and (0,1] denotes the half-open interval from 0 to 1, including 1 but

excluding 0. For any set Y and any finite set Z, we let Y^Z denote the set

of all functions from Z into Y.

Let us now consider the problem faced by one player, whom we may refer to as **the negotiator**, when he has an opportunity to make a negotiation statement to the other players in a game. When the negotiator makes his statement, he may have some private information, which we refer to as his type, and he may have a range of payoff-relevant actions available to him. We let T denote the set of possible types for the negotiator, and we let C denote the set of possible actions that the negotiator must choose among. We let S denote the set of all possible combinations of pure strategies which will be available to the other players jointly in the game after the negotiator makes his negotiation statement. For any (c,s,t) in $C \times S \times T$, we let $U(c,s,t)$ denote the expected utility payoff to the negotiator if his type is t, he chooses action c, and the other players subsequently choose s. We assume that T, C, and S are nonempty finite sets.

Let us now assume that the negotiator is selecting a statement that will be his final statement to the other players. (In the next section, we will drop this assumption and allow the negotiator to follow his negotiation statement by subsequent messages.) In this statement, the negotiator may offer information about his type and the way that his action will depend on his type, and he may make a request as to how the other players should choose their strategies (possibly with randomization)..

The information that the negotiator alleges about his type can be characterized by a vector of likelihoods in $[0,1]^T$, such that every other player should update his beliefs (in $\Delta(T)$) about the negotiator's type using Bayes' formula with this vector of likelihoods. That is, if the negotiator announces some likelihood vector $\lambda = (\lambda(t))_{t \in T}$ in $[0,1]^T$, then he is alleging

that, for any t in T, any player who had prior beliefs p in $\Delta(T)$ before the statement should, after the statement, assign posterior probability $p(t)\lambda(t)/\sum_{r\in T} p(r)\lambda(r)$ to the event that t is the negotiator's actual type. The likelihood $\lambda(t)$ is interpreted as the conditional probability that this statement would be made if t were the negotiator's actual type.

The way that the negotiator promises to choose his actions can be described by a randomized strategy in $\Delta(C)^T$. That is, if the negotiator announces some strategy $\gamma = (\gamma(c|t))_{c\in C, t\in T}$ in $\Delta(C)^T$, then he is promising to use action c with probability $\gamma(c|t)$ if his type is t.

The request that the negotiator makes on the other players can always be described by some jointly randomized strategy $\sigma = (\sigma(s))_{s\in S}$ in $\Delta(S)$, in which $\sigma(s)$ denotes the probability that the others should use their joint pure strategy s.

Thus, the set of possible final statements that the negotiator could make may be identified with the set $[0,1]^T \times \Delta(C)^T \times \Delta(S)$. When we represent a statement this way, the likelihood vector in $[0,1]^T$ is the negotiator's allegation, the strategy in $\Delta(C)^T$ is the negotiator's promise, and the strategy in $\Delta(S)$ is the negotiator's request.

To complete our basic model of the negotiator's problem, we need some way to specify or determine what strategies in $\Delta(S)$ may be rational for the players who will move after the negotiator in the game. (The utility function $U(\cdot)$ is the basic determinant of rational behavior for the negotiator in this model.) So let us assume that we are given some correspondence $F:[0,1]^T \times \Delta(C)^T \twoheadrightarrow \Delta(S)$ that characterizes the tenable requests that could be made on the players other than i, for any allegation and promise that the negotiator might make. That is, for any (λ,γ) in $[0,1]^T \times \Delta(C)^T$, $F(\lambda,\gamma)$

represents the set of all correlated strategies in $\Delta(S)$ that could be rationally implemented by the players in the subgame that will follow the negotiator's current move, if everyone believed the negotiator's allegation λ and promise γ. Thus, $\sigma \in F(\lambda,\gamma)$ if and only if the statement (λ,γ,σ) is tenable in the sense of (1.1). We may refer to this correspondence F as the tenability correspondence. The actual construction of this correspondence $F(\bullet)$ must ultimately depend on our developing a theory of rational behavior in such subgames, but we defer such questions until Section 7 and assume for now that this correspondence is given.

We assume that this correspondence F satisfies three basic properties:

(2.1) $\quad F(\lambda,\gamma) \neq \varnothing, \quad \forall \lambda \in [0,1]^T, \quad \forall \gamma \in \Delta(C)^T$;

(2.2) $\quad F(\alpha\lambda,\gamma) = F(\lambda,\gamma), \quad \forall \alpha \in (0,1], \quad \forall \lambda \in [0,1]^T, \quad \forall \gamma \in \Delta(C)^T$;

(2.3) $\quad F(\bullet)$ is upper-semicontinuous.

Condition (2.1) asserts that, whatever information is given about the negotiator's type and strategy, there must be some strategy combination for the other players thereafter that constitutes rational behavior for them. Condition (2.2) is a homogeneity condition that follows from Bayes' formula. Multiplying a vector of likelihoods by a positive scalar α does not affect the posterior probabilities in $\Delta(T)$ that any observer would calculate, and so the set of tenable requests should remain the same. Condition (2.3) is a topological regularity condition, asserting that the graph of F is a closed subset of $[0,1]^T \times \Delta(C)^T \times \Delta(S)$.

3. Negotiation statements with subsequent messages.

A negotiation statement may include a promise to communicate additional information, by transmitting further messages through some communication channel. That is, we may think of communication by a player as a two-stage process, consisting of an introductory negotiation statement and subsequent messages. The player's negotiation statement advocates some equilibrium of the subsequent game with communication, and his subsequent messages form a part of this communication equilibrium. Since the intended interpretation of the subsequent messages is defined by the negotiation statement that precedes them, our analysis should focus on the negotiation statement itself. However, the possibility of such postnegotiation communication has important implications for our analysis of negotiation statements.

With such possibilities for further communication after his negotiation statement, a negotiator can make a statement of the form "I am about to transmit to you, through some suitable communication channel, either the statement μ_1 or the statement μ_2." We may refer to such a statement as the introductory sum of statements μ_1 and μ_2. For any two possible negotiation statements μ_1 and μ_2, if the sum of their alleged likelihoods is never greater than one then the introductory sum of μ_1 and μ_2 should also be a possible negotiation statement. The model developed in the preceding section ignores the possibility of postnegotiation messages, and the representation of negotiation statements as vectors in $[0,1]^T \times \Delta(C)^T \times \Delta(S))$ does not include all such introductory sums. To include all such introductory sums in our set of possible negotiation statements, we must redefine the set of negotiation statements to be some larger set Ω, into which the statements in $[0,1]^T \times \Delta(C)^T \times \Delta(S)$ can be naturally embedded. We now show how this

can be accomplished by letting $\Omega = \Lambda(C \times S)^T$.

In our notation, $\Lambda(C \times S)^T$ denotes the set of all functions from the set T into the set of all subprobability distributions on $C \times S$, so

$$\Lambda(C \times S)^T = \left\{ \mu = (\mu(c,s|t))_{c \in C, s \in S, t \in T} \;\middle|\; \begin{array}{l} \mu(c,s|t) \geq 0 \quad \forall c \in C, \; \forall s \in S, \; \forall t \in T; \\[1em] \text{and} \;\; \sum_{c \in C} \sum_{s \in S} \mu(c,s|t) \leq 1, \quad \forall t \in T \end{array} \right\}.$$

For any μ in $\Lambda(C \times S)^T$ and any t in T, we define

$$L(\mu|t) = \sum_{c \in C} \sum_{s \in S} \mu(c,s|t).$$

That is, $L(\mu|t)$ is the sum of the subprobabilities in the distribution on $C \times S$ that is designated for t by μ.

For any triple $(\lambda, \gamma, \sigma)$ in $[0,1]^T \times \Delta(C)^T \times \Delta(S)$, let $\lambda * \gamma * \sigma$ be defined so that $\mu = \lambda * \gamma * \sigma$ iff

$$\mu(c,s|t) = \lambda(t) \, \gamma(c|t) \, \sigma(s), \quad \forall c \in C, \; \forall s \in S, \; \forall t \in T.$$

Notice that the vector μ determined by such a star product will be in the set $\Lambda(C \times S)^T$. Furthermore, given $\mu = \lambda * \gamma * \sigma$, we can reconstruct λ from μ by the formula

$$\lambda(t) = L(\mu|t),$$

and, whenever $\lambda(t) \neq 0$, we can reconstruct γ and σ from μ by the formulas

$$\gamma(c|t) = \Sigma_{s \in S} \, \mu(c,s|t)/\lambda(t),$$

$$\sigma(s) = \mu(c,s|t)/(\gamma(c|t) \, \lambda(t)).$$

Thus, the negotiator's allegation, promise, and request can all be essentially reconstructed from their star product in $\Lambda(C \times S)^T$. The only terms that are lost in this representation are the strategic promises and requests that would be made by types that are alleged to have zero probability of making this statement. So any negotiation statement that can be represented by a triple in $[0,1]^T \times \Delta(C)^T \times \Delta(S)$ can also be represented by a vector in $\Lambda(C \times S)^T$,

by taking this star product of the three vectors in $[0,1]^T$, $\Delta(C)^T$, and $\Delta(S)$;
and this representation in $\Lambda(C \times S)^T$ does not suppress any relevant
information about the negotiator's allegations, promises, and requests. This
representation may seem unintuitive at first, but it will greatly improve both
the simplicity and power of our notation to henceforth use this
representation. The literal interpretation of a negotiation statement μ in
$\Lambda(C \times S)^T$ is that, for each (c,s,t) in $C \times S \times T$, if the negotiator's type
were t then $\mu(c,s|t)$ would be the probability that the negotiator would make
this negotiation statement, would choose his action c, and would have the other
players after him use their pure joint strategy s.

The tenability correspondence F from Section 2 can be represented in
$\Lambda(C \times S)^T$ by the set $\hat{G}$ defined as follows:

$\hat{G} = \{\mu| \exists \lambda \in [0,1]^T, \exists \gamma \in \Delta(C)^T,$ and $\exists \sigma \in F(\lambda,\gamma)$ such that $\mu = \lambda * \gamma * \sigma\}$.
That is, if $\mu \in \hat{G}$ then the request in μ is tenable, so that the other players
would be willing to obey it if they believed the negotiator's allegations and
promises.

Representing negotiation statements as vectors in $\Lambda(C \times S)^T$ has an
important advantage over the representation in $[0,1]^T \times \Delta(C)^T \times \Delta(S)$, because
sums of vectors in $\Lambda(C \times S)^T$ can be interpreted as representing introductory
sums of statements. That is, if $\mu = \mu_1 + \mu_2$ in $\Lambda(C \times S)^T$, then μ can be
interpreted as the statement "I am about to announce either the statement μ_1
or μ_2." If the negotiator would make this introductory statement before
announcing either μ_1 or μ_2, then indeed $\mu_1(c,s|t) + \mu_2(c,s|t)$ would be the
conditional probability that he would make this statement and subsequently
use action c and have the other players use their joint strategy s if his type
were t. Not every vector in $\Lambda(C \times S)^T$ can be expressed as a star product

of three vectors in $[0,1]^T$, $\Delta(C)^T$, and $\Delta(S)$; but every vector in $\Lambda(C \times S)^T$ can be expressed as a sum of such star products. Thus, we can identify $\Lambda(C \times S)^T$ with the set of all possible introductory sums of statements that the negotiator could make in $[0,1]^T \times \Delta(C)^T \times \Delta(S)$. In this sense, $\Lambda(C \times S)^T$ represents the set of all possible negotiation statements, when subsequent messages are allowed.

We let G denote the set of all vectors in $\Lambda(C \times S)^T$ that can be expressed as sums of vectors in $\hat{G}$. That is, $\mu \in G$ iff $\mu \in \Lambda(C \times S)^T$ and there exists some finite set $\{\mu_1, \ldots, \mu_k\}$ such that $\{\mu_1, \ldots, \mu_k\} \subseteq \hat{G}$ and $\mu_1 + \ldots + \mu_k = \mu$. Thus, any vector in G represents the introductory sum of a set of final statements, each of which expresses a tenable request. We may therefore refer to G as the set of <u>tenable</u> statements in $\Lambda(C \times S)^T$. Condition (2.2) implies that G is convex. Convexity, condition (2.3), and Caratheodory's Theorem (see [20, section 17]) imply that G is also a closed subset of $\Lambda(C \times S)^T$.

A negotiator may choose to transmit messages through some mediator or communication channel. We may think of such a mediator as being an agent of the negotiator who filters and transforms messages from the negotiator according to some random rule that is has been selected by the negotiator himself. We now show how such a mediator could help the negotiator to implement the terms of any statement in G.

For any μ in G, suppose that $\mu = \sum_{j=1}^{k} \mu_j$, where each $\mu_j = \lambda_j * \gamma_j * \sigma_j \in \hat{G}$ and $\sigma_j \in F(\lambda_j, \gamma_j)$. A mediator could help the negotiator to implement the terms of μ by the following scheme. After the negotiation statement μ is announced, the negotiator should report his type to the mediator. (Since the mediator is going to filter the information anyway, there

is no loss of generality in assuming that the negotiator is asked to report all of his information to the mediator.) Then the mediator should select (but not yet reveal) his final message from the set $\{\mu_1, \ldots, \mu_k\}$. If the negotiator reported type t to the mediator then the probability of the mediator selecting μ_j should be $\lambda_j(t)/L(\mu|t)$. Next, the mediator should confidentially recommend an action to the negotiator, recommending action c with probability $\gamma_j(c|t)$ if t was reported and final announcement μ_j was selected. Finally, after the negotiator has chosen his action in C, the mediator should publicly announce the μ_j that he has selected and should request that the other players select their strategies according to the corresponding σ_j, so that each pure strategy s is recommended with probability $\sigma_j(s)$. Suppose that, for each type t, $L(\mu|t)$ is the probability that this particular mediation scheme would have been selected by the negotiator when his type is t. Then the statement that that this mediation scheme is to be used can be represented by the given vector μ.

To verify this representation, notice that, if the negotiator's type is t, then the conditional probability that this mediation scheme would be selected and that c and s would ultimately be implemented, assuming that everyone would be honest and obedient to the mediator, is

$$L(\mu|t) \ \textstyle\sum_{j=1}^{k} (\lambda_j(t)/L(\mu|t)) \ \gamma_j(c|t) \ \sigma_j(s) = \sum_{j=1}^{k} \mu_j(c,s|t) = \mu(c,s|t).$$

Furthermore, the mediator's final announcement of μ_j is also being used consistently with its literal meaning, because, when the negotiator's type is t, then the probability that the mediator will select μ_j using this scheme and c and s will be implemented is

$$L(\mu|t) \ (\lambda_j(t)/L(\mu|t)) \ \gamma_j(c|t) \ \sigma_j(t) = \mu_j(c,s|t),$$

assuming that everyone will be honest and obedient to the mediator. Notice that, since $\mu_j = \lambda_j * \gamma_j * \sigma_j \in \hat{G}$, the request σ_j in μ_j is tenable.

In this context, we can now begin to formulate the question of reliability of such a statement μ. If the negotiator is honest and all players are obedient to the mediator under the mediation scheme described above, then the expected payoff to type t of the negotiator when he selects this scheme is

$$\sum_{c\in C} \sum_{s\in S} (\mu(c,s|t)/L(\mu|t)) \ U(c,s,t),$$

because, under this scheme, given the reported type t, the probability of any (c,s) being ultimately recommended by the mediator is

$$\sum_{j=1}^{k} (\lambda_j(t)/L(\mu|t)) \ \gamma_j(c|t) \ \sigma_j(s) = \mu(c,s|t)/L(\mu|t).$$

On the other hand, suppose that the negotiator plans to lie to the mediator, claiming that his type is r when it is really t, and suppose that he plans to disobey the mediator according to some function $\delta:C \rightarrow C$, so that he will choose action $\delta(c)$ if the mediator recommends c. (We need to consider manipulative strategies for disobedience that are functions from C into C, because the negotiator's optimal disobedience may be a function of the action recommended to him, when this recommended action is correlated with the mediator's subsequent requests to the other players. We are assuming here that the negotiator will choose his action in C before the mediator's final announcement.) Then the expected payoff to type t of the negotiator from manipulating this mediation scheme in this way is

$$\sum_{c\in C} \sum_{s\in S} (\mu(c,s|r)/L(\mu|r)) \ U(\delta(c),s,t).$$

Let D denote the set of all functions from C into C, that is

$$D = C^C.$$

For any types t and r in T, and any function δ in D, we let

$$V(\mu|t) = \sum_{c\in C} \sum_{s\in S} \mu(c,s|t) \ U(c,s,t),$$

and

$$\hat{V}(\mu,\delta,r|t) = \sum_{c \in C} \sum_{s \in S} \mu(c,s|r) \ U(\delta(c),s,t).$$

Then to guarantee that the negotiator would not want to disobey and lie to his mediator, as is required for reliability, we need that

(3.1) $V(\mu|t)/L(\mu|t) \geq \hat{V}(\mu,\delta,r|t)/L(\mu|r)$, $\forall t \in T$, $\forall r \in T$, $\forall \delta \in D$.

Thus, the statement μ may be considered **reliable**, in the sense of (1.2), if condition (3.1) is satisfied.

These functions $V(\mu|t)$ and $\hat{V}(\mu,\delta,r|t)$ may be called <u>discounted-payoff</u> <u>functions</u>, because they are equal to the product of the negotiator's expected payoff (under honesty and manipulation, respectively) from μ multiplied by the likelihood of the statement μ for the reported type ($L(\mu|t)$ or $L(\mu|r)$). Notice that $V(\mu|t)$, $\hat{V}(\mu,\delta,r|t)$, and $L(\mu|t)$ are all linear functions of μ, for any given t, r, and δ. This linearity makes the discounted-payoff functions more convenient to work with than the corresponding expected payoffs.

Condition (3.1) does not address the question of whether the initial information supposedly conveyed by the statement μ is plausible in the sense of (1.3). That is, (3.1) cannot confirm that the probability of the negotiator announcing μ would, as claimed, be $L(\mu|t)$ if his type is t. Criteria to determine whether these likelihoods are plausible are developed in Section 5. Before that, in Section 4, we develop a more abstract model of negotiation statements, to generalize further the model just developed and to bring its essential structures into better focus.

4. Abstract negotiation problems.

To summarize the mathematical properties of the model derived above, we define one further mathematical model of a negotiation problem.

An abstract <u>negotiation problem</u> is any $(\Omega, G, T, D, L, V, \hat{V})$ such that Ω is a closed convex subset of some topological vector space, G is a compact subset of Ω, T and D are nonempty finite sets, L is a function mapping $G \times T$ into the interval $[0,1]$, V is a function mapping $G \times T$ into the real numbers $\mathbb{R}$, $\hat{V}$ is a function mapping $G \times D \times T \times T$ into $\mathbb{R}$, and the following properties are satisfied:

(4.1) $L(\mu|t)$, $V(\mu|t)$, and $\hat{V}(\mu,\delta,r|t)$ are continuous linear functions

of μ in Ω, $\forall t \in T$, $\forall r \in T$, $\forall \delta \in D$;

(4.2) $\alpha\mu \in G$, $\forall \mu \in G$, $\forall \alpha \in (0,1]$;

(4.3)   if $L(\mu_1|t) + L(\mu_2|t) \leq 1$ $\forall t \in T$,   then $\mu_1 + \mu_2 \in G$, $\forall \mu_1 \in G$, $\forall \mu_2 \in G$;

(4.4) $V(\mu|t)/L(\mu|t)$ and $\hat{V}(\mu,\delta,t|r)/L(\mu|t)$ are bounded functions

on $\{\mu \in G | L(\mu|t) > 0\}$, $\forall t \in T$, $\forall r \in T$, $\forall \delta \in D$;

(4.5) $\forall \lambda \in (0,1]^T$, $\exists \mu \in G$ such that, $\forall t \in T$, $\forall r \in T$, $\forall \delta \in D$,

$L(\mu|t) = \lambda(t)$ and $V(\mu|t)/L(\mu|t) \geq \hat{V}(\mu,\delta,r|t)/L(\mu|r)$;

(4.6) $\forall \lambda \in [0,1]^T$, $\forall \mu \in \Omega$, $\exists(\lambda * \mu) \in \Omega$ such that, $\forall t \in T$, $\forall r \in T$, $\forall \delta \in D$,

$V(\lambda * \mu|t) = \lambda(t) V(\mu|t)$,

$\hat{V}(\lambda * \mu,\delta,r|t) = \lambda(r) \hat{V}(\mu,\delta,r|t)$,   and

$L(\lambda * \mu|t) = \lambda(t) L(\mu|t)$.

In the interpretation of this abstract model, Ω is the set of possible negotiation statements. G is the set of tenable statements in Ω; that is, G is the set of statements that include requests that the other players would be willing to obey if they believed the negotiator's allegations and promises. T is the set of possible types for the negotiator, and D is the set of possible

manipulative strategies that the negotiator could use if he chose to disobey his own strategic promise.

$L(\mu|t)$ is the likelihood of type t making statement μ, according to the literal meaning of μ. For any vector μ in G and any scalar α between 0 and 1, the vector $\alpha\mu$ can be interpreted as the statement "after observing an extraneous event that had probability α, I have decided to make the statement μ." Condition (4.2) asserts that such a statement is tenable if μ is tenable. The sum of two vectors μ_1 and μ_2 in G can be interpreted as the statement, "I am about to make either the statement μ_1 or μ_2." Condition (4.3) asserts that this introductory sum is a tenable statement if each of the individual statements is tenable and the sum of their likelihoods never exceeds one.

The functions V and $\hat{V}$ are discounted-payoff functions. They are defined so that, if the requests in μ are obeyed by all other players, then

$$V(\mu|t)/L(\mu|t)$$

is the negotiator's expected payoff from μ when his type is t and he is honest and obedient to the terms of μ, and

$$\hat{V}(\mu,\delta,r|t)/L(\mu|r)$$

is his expected payoff when his type is t but he pretends that his type is r and then disobeys the terms of μ in choosing his actions according to the manipulative strategy δ. The functions V and $\hat{V}$ are linear in μ, by condition (4.1), and they must be divided by the corresponding likelihoods to compute his expected utility payoffs. Condition (4.4) asserts that expected utility payoffs are bounded. Condition (4.5) asserts that, given any vector of positive likelihoods λ, there exists some tenable statement that is consistent with the allegation λ and that gives the negotiator no incentive to lie or

disobey his own promise.

Condition (4.6) asserts that we can define a star product between allegations in $[0,1]^T$ and statements in Ω. The star product $\lambda * \mu$ represents the statement that is derived from μ by appending the further allegation that all players should reupdate their posterior beliefs after μ using the likelihoods in λ. In the model of Section 3, where $\Omega = \Lambda(C \times S)^T$, this star product is defined by

$$(\lambda * \mu)(c,s|t) = \lambda(t) \mu(c,s|t), \quad \forall c \in C, \quad \forall s \in S, \quad \forall t \in T.$$

This statement $\lambda * \mu$ is not necessarily tenable, even if μ is, since the additional information conveyed by λ may disturb the other players' incentives to obey the requests in μ.

Lemma 1. The model constructed in Section 3 above satisfies all the properties of an abstract negotiation problem defined in this section.

Proof. See Section 9.

5. Reference payoffs and coherent plans.

In our model, a statement is tenable (in the sense of (1.1)) if it is in the set G. Condition (3.1) is a formalization of reliability (1.2). However, we have not yet considered the third component of credibility, the plausibility (1.3) of a statement. That is, we still need to develop criteria to answer the question of whether the types that would want to make this statement are the types that the negotiator is alleging that his actual type is among. The answer to this question depends on what each type of the negotiator could have gotten if he had made some other statement instead.

Thus, to assess the plausibility of a statement, we need some <u>reference payoff</u> <u>allocation</u> $w = (w(t))_{t \in T}$ in $\mathbb{R}^T$, where the reference payoff $w(t)$ is interpreted as the expected payoff that type t of the negotiator could have safely gotten by making some other statement. Once such a reference allocation w has been specified, one might suppose that a statement μ is plausible if, for every type t in T,

(5.1)   if $L(\mu|t) > 0$ then $V(\mu|t)/L(\mu|t) \geq w(t)$,

and

(5.2)   if $V(\mu|t)/L(\mu|t) > w(t)$ then $L(\mu|t) = 1$.

Condition (5.1) asserts that any type that could possibly make the statement μ should not get less from μ than the reference payoff. Condition (5.2) asserts that any type that would get strictly more from μ than the reference payoff should surely make the statement μ.

Conditions (3.1) and (5.2) are difficult to work with. Notice first that, if $L(\mu|t) = 0$ or $L(\mu|r) = 0$ then the ratios in these inequalities are not well defined. Conditions (4.1) and (4.4) guarantee that the numerator on either side of (3.1) must equal zero whenever the denominator does, but this still leaves us the problem of dividing zero by zero. Furthermore, even when the denominators are nonzero, the nonlinearity of these conditions (3.1) and (5.2) makes them difficult to work with. (Actually, there are ways to try to resolve the problem of dividing zero by zero here. One way is to arbitrarily pick some sequence of statements $\{\mu_k\}_{k=1}^{\infty}$ such that $\mu_k \rightarrow \mu$ as $k \rightarrow \infty$, and all ratios $V(\mu_k|t)/L(\mu_k|t)$ and $\hat{V}(\mu_k,\delta,r|t)/L(\mu_k|r)$ are well-defined and convergent as $k \rightarrow \infty$. Then we may define $V(\mu|t)/L(\mu|t)$ and $\hat{V}(\mu,\delta,r|t)/L(\mu|r)$ to be the limits of these ratios in the sequences. Such sequences can always be found because of the boundedness condition (4.4).)

So let us consider a weaker credibility criterion, which allows more statements to be accepted as credible and which will be easier to work with. We say that μ is <u>credible with respect to w</u> iff

(5.3) $\mu \in G$,

(5.4) $L(\mu|t) > 0$ for at least one t in T,

(5.5) $V(\mu|t) + (1 - L(\mu|t))\, w(t) \geq w(t)$, $\forall t \in T$,

(5.6) $V(\mu|t) + (1 - L(\mu|t))\, w(t) \geq \hat{V}(\mu,\delta,r|t) + (1 - L(\mu|r))\, w(t)$,

   $\forall t \in T$, $\forall r \in T$, $\forall \delta \in D$.

Condition (5.3) asserts that μ is tenable, so that all the players moving after the negotiator would be willing to obey the requests in μ if they believed the negotiator's promise and allegation in μ.

We say that a statement μ in G is <u>null</u> iff $L(\mu|t) = 0$ for every t in T. Thus, a null statement could be interpreted as an assertion of the form "I would never make this statement ..." Condition (5.4) asserts that a credible statement cannot be null.

Conditions (5.5) and (5.6) can be readily interpreted if we allow that the negotiator may direct a mediator or agent to negotiate on his behalf. Such a mediator might announce the statement μ on behalf of the negotiator as a part of to the following scheme. First, the negotiator confidentially reports his type to the mediator. For each type t, if the negotiator reports that his type is t then, the mediator will announce μ with probability $L(\mu|t)$, otherwise he do something else (make some other statement or perhaps just be silent). We may let the reference payoff w(t) represent the expected payoff to type t of the negotiator when the statement μ is not made. If the negotiator is honest and obedient and all other players are obedient then the expected payoff to type t from this scheme is

$L(\mu|t) \ (V(\mu|t)/L(\mu|t)) + (1 - L(\mu|t)) \ w(t)$

$\qquad = V(\mu|t) + (1 - L(\mu|t)) \ w(t)$.

(Recall that $V(\mu|t)/L(\mu|t)$ is the expected payoff to type t from the terms of μ after it is announced.) On the other hand, if the negotiator's actual type is t but he pretends that his type is r and then, if μ is announced, disobeys the terms of his promise according to the manipulative strategy δ, then his expected payoff is

$L(\mu|r) \ (\hat{V}(\mu,\delta,r|t)/L(\mu|r)) + (1 - L(\mu|r)) \ w(t)$

$\qquad = \hat{V}(\mu,\delta,r|t) + (1 - L(\mu|r)) \ w(t)$.

Thus, condition (5.5) asserts that, for any type t, the negotiator's expected payoff from this scheme should not be less than the reference payoff that he could have gotten by some other negotiation statement. In fact, (5.5) is equivalent to (5.1), and so (5.5) may be interpreted as a weak plausibility condition.

Condition (5.6) asserts that the negotiator's expected payoff under this scheme, implemented honestly and obediently, should not be less than what he could expect to get by lying and disobeying. Notice that, if $L(\mu|t) = 1$ for all t, then (5.6) is equivalent to (3.1). More generally, if μ satisfies (5.2) (letting the 0/0 ratios be defined as discussed parenthetically above), then (3.1) implies (5.6). Thus, (5.6) can be interpreted as a reliability condition that is weaker than (3.1) in many important cases. The difference between (3.1) and (5.6) is that (5.6) assumes that the negotiator can report his type to the mediator before any negotiation statement is announced and can allow the mediator to control the decision to announce μ according to the terms of μ itself. The reference payoffs are needed in (5.6) to account for the negotiator's payoff when μ is not announced under such a scheme.

Thus, (5.3)-(5.6) constitute a natural definition of credibility for mediated negotiation statements, once a reference allocation w has been specified. Given any w in $\mathbb{R}^T$, the set of all statements that satisfy (5.5) and (5.6) is a compact convex set, and so the set of credible statements with respect to w is convex. The central theoretical question that remains is how to determine the appropriate reference allocation w for our analysis.

With no loss of generality, we may assume that there is only one negotiation statement that is actually made by the negotiator with positive probability, so that this statement has likelihood one for all types. Any given theory that predicts that all the statements in some set $\{\mu_1, \ldots, \mu_k\}$ will be made with positive probability is equivalent to a theory that asserts that the negotiator's initial statement will be μ, where $\mu = \mu_1 + \ldots + \mu_k$ (so μ can be interpreted as "I or my agent will soon announce the one of the messages in the set $\{\mu_1, \ldots, \mu_k\}$, which should be interpreted according to its literal meaning"). In effect, we can assume without loss of generality that the fundamental negotiation statement made by the negotiator is actually uninformative, because any further informative communication could be subsumed by messages in a communication equilibrium that is established by the negotiation statement. This argument is called the inscrutability principle and is discussed further in [16]. Since, under the given theory, the statements in $\{\mu_1, \ldots, \mu_k\}$ were supposed to be the only statements used with positive probability by any type, their likelihoods must sum to one, so

(5.7) $L(\mu|t) = 1, \quad \forall t \in T$.

We define a plan to be any statement μ that satisfies this condition (5.7).

Thus, a theory of negotiations should predict one credible plan that the negotiator should be expected to negotiate, no matter what his type is. In

general, however, each type of the negotiator will have different preferences over the set of tenable statements. To compel all types to announce the same plan, the reference allocation must define a standard of credibility that accepts one plan but rejects all other statements that any types might prefer over it. This property is the key to determining the reference allocation.

We say that an allocation w in $\mathbb{R}^T$ is <u>strongly attractive</u> iff there are no credible statements with respect to w. That is, for any strongly attractive w, if we altered the game by giving the negotiator a new option, called "the easy way out," that would pay w(t) to each type t, then the only credible negotiation statement for the negotiator would be "I am taking the easy out" (which would correspond to a null statement in the original game). Obviously, if the components w(t) are all higher than the upper bounds on the utility functions (given in (4.4)), then w must be strongly attractive, by (5.4) and (5.5). It is straightforward to check that the set of strongly attractive allocations is an open subset of $\mathbb{R}^T$.

We say that w is <u>attractive</u> iff it is the limit of a sequence of strongly attractive vectors in $\mathbb{R}^T$. Thus, if w is attractive then any statement is not credible with respect to reference allocations that are arbitrarily close to w. Furthermore, as the following theorem asserts, for any attractive reference allocation and any reliable plan that any type might be tempted to advocate, there is a plausible inference that the other players could make about the negotiator's type after this plan is announced, such that this plan with this inference would not be tenable.

<u>Theorem 1.</u> Let w be an attractive reference allocation, and let μ be any plan that satisfies (5.7) and the reliability conditions (3.1) or (5.6). Suppose that $V(\mu|\hat{t}) > w(\hat{t})$ for at least one $\hat{t}$ in T (so that at least one

type would find it profitable to announce μ, relative to w). Then there exists

some likelihood vector λ in $[0,1]^T$ such that $\lambda * \mu \notin G$ and, for every t in T,

$$\lambda(t) \begin{cases} = 1, & \text{if } V(\mu|t) > w(t), \\ = 0, & \text{if } V(\mu|t) < w(t), \\ \in \{0,1\}, & \text{if } V(\mu|t) = w(t). \end{cases}$$

Proof. See Section 9.

We say that a statement μ is a coherent plan iff $L(\mu|t) = 1$ for every

t in T and there exists some attractive reference allocation w such that μ is

credible with respect to w. Thus, although μ may not be the unique credible

statement with respect to w, if we allow that μ is judged with respect to w

but every other statement is judged with respect to some strongly attractive

vector arbitrarily close to w, then μ is the unique credible statement. (This

may seem like a double standard, but it is only infinitesimally so.)

Equivalently, we may suppose that the other players would accept the

announcement of μ as inscrutable or uninformative about the negotiator's type;

but, after the announcement of any other reliable plan that might tempt some

of the negotiator's types, the other players would update their beliefs

according to the plausible likelihood vector that is given by Theorem 1, which

would destroy the tenability of the plan.

We can now state our general existence theorem.

Theorem 2. For any negotiation problem, as defined in Section 4, there

exists at least one coherent plan in G.

Proof. See Section 9.

We say that a plan μ is incentive compatible iff it is tenable and

reliable, in the senses of (5.3) and (3.1). Notice that a plan μ is coherent iff it is incentive compatible and there exists some attractive reference allocation w such that $V(\mu|t) \geq w(t)$, for every t in T.

A plan μ is <u>strongly dominated</u> for the negotiator iff there is an incentive-compatible plan $\hat{\mu}$ such that $V(\hat{\mu}|t) > V(\mu|t)$, for every t in T. A coherent plan cannot be strongly dominated. (If $\hat{\mu}$ were an incentive-compatible plan such that $V(\hat{\mu}|t) > V(\mu|t) \geq w(t)$ for every t, then $\hat{\mu}$ would be credible with respect to any reference allocation sufficiently close to w.) Thus, the negotiator's expected payoff allocations that are generated by coherent plans must be in the undominated frontier of the incentive-feasible set, which is generally a $(|T|-1)$-dimensional surface in $\mathbb{R}^T$.

In fact, there is reason to believe that the set of coherent plans is usually a small, zero-dimensional set (finite or countable). Notice that the set of attractive reference allocations is a $(|T|-1)$-dimensional surface, since it is the boundary of an open set in $\mathbb{R}^T$. By definition, an attractive reference allocation is one that admits some credible statements but almost fails to admit any. So it is reasonable to expect that the set of statements that are credible with respect to an attractive reference allocation is minimal in size. Such a "minimal" set could not be a single point (since, for any α in (0,1], $\alpha\mu$ is credible with respect to w if μ is credible with respect to w), but a minimal credible set could be a segment of a one-dimensional ray. Then the condition that such a one-dimensional set should include statements that are plans, in the sense of having $L(\mu|t) = 1$ for all t, gives us $|T|-1$ independent equations to satisfy. Thus, the set of attractive reference allocations that support coherent plans should usually be zero-dimensional. Obviously, this is not a rigorous argument, but it strongly

suggests that the set of coherent plans is small for most negotiation problems.

To guarantee that coherent plans exist, we must allow that the attractive reference allocation w that supports a coherent plan μ may be different from the negotiator's expected payoff allocation from μ. (See the second example in Section 6.) In such cases, the attractive reference allocation may not coincide with the negotiator's expected payoffs from any incentive-compatible plan.

Strict inequality between expected payoffs from the coherent plan μ and reference payoffs from the attractive allocation w that supports μ can only occur for types that are, in a sense, free-riding on the other types in the negotiation of μ. To make this idea precise, let $\hat{T}$ denote the set of all types that get the same expected payoff from the coherent plan μ as from the attractive reference allocation w that supports it. Then $\hat{T}$ is nonempty and there is no other incentive-compatible plan that is strictly better than μ for all the types in $\hat{T}$. (If some incentive-compatible plan $\hat{\mu}$ were strictly better than μ for all the types in $\hat{T}$, then some convex combination of μ and $\hat{\mu}$ would be an incentive-compatible plan that was strictly better than w for all types, which would violate Theorem 1.) Thus, the desirability of μ for the negotiator would be evident even if we ignored the preferences of his types outside of $\hat{T}$.

The attractive reference allocation w that supports a coherent plan μ may be quite difficult to interpret when $w(t) < V(\mu|t)$ for some type t of the negotiator. The best general interpretation that we can offer is that the reference allocation represents a hypothetical conjecture that the followers make about what the negotiator would have expected to get, as a function of his type, if he had made some negotiation statement other than

the one that he actually made. Such a conjecture is necessary to define a
criterion for evaluating the plausibility of the negotiator's allegations.

A weaker concept of coherence may be defined that avoids the use of
reference allocations, but it requires us to restrict the negotiator's
statements to plans, which involve no nontrivial allegations. We may say that
μ is a <u>semicoherent plan</u> (or that μ is <u>coherent in the weak sense</u>) iff μ is
an incentive-compatible plan and, for any other incentive-compatible plan ν
such that $V(\nu|\hat{t}) > V(\mu|\hat{t})$ for some $\hat{t}$ in T, there exists some likelihood
vector λ in $[0,1]^T$ such that $\lambda * \nu \notin G$ and $\lambda(t) = \max_{r \in T} \lambda(r) > 0$ for every
t in T such that $V(\nu|t) > V(\mu|t)$. That is, μ is semicoherent iff, for any
other incentive-compatible plan ν that the negotiator might prefer, there is
an inference that the followers might make about the negotiator, if he
negotiated for ν, such that they do not take his negotiation for ν as evidence
against any type that prefers ν over μ, but they would not be willing to obey
his requests in ν after making this inference.

It is easy to see that a semicoherent plan cannot be strongly dominated
for the negotiator by any other incentive-compatible plan. Furthermore,
Theorem 1 implies that any coherent plan is semicoherent, so the set of
semicoherent plans is nonempty. In Section 6, we show an example in which
the set of semicoherent plans is much larger than the set of coherent plans,
and we show another example in which the two sets coincide.

6. Sender-receiver examples.

Two-player sender-receiver games have been studied to gain insights into
the problems of communication in games, since Crawford and Sobel [4]. In these

games, one player, the sender, has all the private information and the the

other player, the receiver, has all the payoff-relevant actions. The sender

moves first, sending some message or signal to the receiver, who then chooses

an action. In analyzing sender-receiver games here, we assume that the sender

is also the sole negotiator. A statement by the sender is tenable iff he is

requesting the receiver to use only actions that maximize the receiver's

expected payoff given the sender's allegation. In this section, we consider

two sender-receiver games, to illustrate some of the key properties of coherent

plans.

Throughout this paper, we have assumed that communication between a

negotiator and the other players could be facilitated by a mediator. In fact,

this assumption is crucial to our general existence theorem. Our first example

illustrates the importance of this mediation assumption.

In this game (suggested by R. Aumann, based on a similar example studied

by Moulin and Vial [14]), player 1, the sender, has a set of three possible

types {1a,1b,1c} which are initially considered by player 2 to be equally

likely. Player 2, the receiver, has a set of three possible actions {x,y,z}.

The payoffs to players 1 and 2 respectively depend on player 1's type and

player 2's action according to Table 1.

		Player 2's actions		
		x	y	z
	1a	0,0	5,4	4,5
Player 1's types	1b	4,5	0,0	5,4
	1c	5,4	4,5	0,0

TABLE 1.

Notice that player 2 is initially indifferent between his three actions, since each type has probability 1/3. Type 1a of player 1 would most prefer that 2 choose y, but player 2 would choose z if he were convinced that 1's type was 1a. On the other hand, y is also good for type 1c of player 1 (although x would be slightly better for 1c), and player 2 would be willing to obey a request to choose y if he believed that the request could have come from 1a or 1c with approximately equal likelihood. Thus, type 1a would like to be pooled with 1c, but distinguished from 1b, to be able to tenably request y. Similarly, 1c would like to be pooled with 1b but distinguished from 1a, to be able to tenably request action x; and 1b would like to be pooled with 1a but distinguished from 1c, to be able to tenably request z.

Suppose, for now, that player 1 can only communicate with player 2 by direct face-to-face communication, so that any signal that player 1 sends must be received unaltered by player 2. With such communication, there are only fully pooling equilibria in this game. That is, no matter what may be the set of signals available to player 1, in any equilibrium of the signalling game, player 2 must always have the same (uniform) posterior distribution over player 1's types after receiving any signal that has positive probability, and player 2's strategy must be independent of the signal sent by player 1. Thus, for any equilibrium, there must be some pair of 1's types which could both do strictly better (getting payoffs of 5 and 4) if, with equal likelihood, they would announce that the third type has likelihood zero and request that 2 should use his best action given this information. Furthermore, the third type would lose relative to the equilibrium from this request. Thus, for any reference allocation that is close to a feasible allocation (generated for player 1's types by a fully pooling equilibrium), there is a credible statement

that at least two of player 1's types could more profitably use. In fact, there can be no coherent plans for this example without mediation.

Now suppose that player 1 can also send messages to player 2 through a mediator or noisy channel, as we have assumed throughout Sections 3 through 6. Then, as Theorem 2 guarantees, there does exist a coherent plan that would be implemented by a mediator as follows. First player 1 confidentially reports his type to the mediator. If 1 reports 1a then, with probability 5/6, the mediator asks 2 to choose y, and, with probability 1/6, the mediator asks 2 to choose z. If 1 reports 1b then, with probability 5/6, the mediator asks 2 to choose z, and, with probability 1/6, the mediator asks 2 to choose x. If 1 reports 1c then, with probability 5/6, the mediator asks 2 to choose x, and, with probability 1/6, the mediator asks 2 to choose y. In this coherent plan, each type of player 1 gets the payoff 5 with probability 5/6 and gets the payoff 4 with probability 1/6. The attractive reference payoffs that support this coherent plan are $4\frac{5}{6}$ for each of the three types of player 1. No two types could simultaneously get more than $4\frac{5}{6}$ by an allegation that assigns likelihood zero to the third type, since at least one of these types would get 4 from player 2's best response.

The mediator is needed in this plan because player 2 is not willing to obey any request to choose some action unless there is at least a 1/6 probability that it is the second-most preferred action of player 1. So a mediator is needed to filter player 1's request and guarantee that there is always at least a 1/6 probability that the request that 2 hears is for 1's second-most preferred action.

Mathematically, introducing mediation into the communication process helps to guarantee the existence of coherent plans, because mediation convexifies

the set of credible statements. For example, without mediation, the reference
allocation that gives 3.1 to each of player 1's types would allow a credible
statement in which any pair of 1's types are given likelihood 1, but there
would be no credible statement in which all three types have positive
likelihood. Another useful mathematical analogy may be with the theory of
the core. Coherent plans are similar to core allocations in traditional
cooperative game theory, in that no coalition of types can block a coherent
plan with a credible statement. Under this analogy, allowing mediated
statements is mathematically analogous to taking the balanced cover of a game,
which guarantees nonemptiness of the core.

The coherent plan described above is the unique coherent plan for this
game. There are infinitely many semicoherent plans, however, including all
incentive-compatible plans that are not strongly dominated for the sender and
give an expected payoff greater than 4 to each type of the sender.

For a second example, consider the following sender-receiver game,
proposed by Farrell [5]. Player 1, the sender, has two possible types, 1a
and 1b, which player 2 initially considers to be equally likely. Player 2
has three possible actions, x, y, and z. The payoffs to players 1 and 2
respectively depend on 1's type and 2's action as in Table 2.

		Player 2's actions		
		x	y	z
	1a	2,3	-1,0	0,2
Player 1's types				
	1b	1,0	0,3	2,2

TABLE 2.

In this game, type 1a would like to reveal his type to player 2, but type

1b would prefer to not be distinguished from type 1a. Farrell [5] shows that this game has no neologism-proof equilibria, in the sense that he defines. However, there is a unique coherent plan for player 1 in this game, and it is also the unique semicoherent plan. In this plan, the probability distribution over player 2's actions depends on 1's type according to Table 3.

		Player 2's actions		
		x	y	z
	1a	.8	0	.2
Player 1's types				
	1b	.4	.2	.4

TABLE 3.

This plan cannot be implemented without a mediator. (Indeed, Farrell asserts that direct face-to-face communication would only allow one pooling equilibrium in which player 2 chooses z for sure.) The plan in Table 3 is supported as a coherent plan by the attractive reference allocation that gives 1.6 to type 1a and gives 0 to type 1b.

Notice that this reference allocation of 1.6 for 1a and 0 for 1b gives strictly less to type 1b than the expected payoff of 1.2 that 1b gets in the coherent plan itself. In fact, the coherent plan's actual payoff allocation of 1.6 for 1a and 1.2 for 1b is not attractive, because any reference allocation close to this would permit player 1 to credibly allege that he is type 1a and request that player 2 choose action x. The reference allocation (1.6, 0) can be attractive even though it is weakly dominated by (1.6, 1.2), because decreasing the reference allocation for type 1b makes it harder for player 1 to plausibly allege that he is type 1a. This example shows that, for a plan to be coherent, it may be necessary to use a reference allocation

that is different from (and weakly dominated by) the actual expected payoff allocation from the plan.

Farrell [5] and Grossman and Perry [6] have defined solution concepts that are closely related to the coherent plans of this paper. Both of these papers assume that mediation is not used, and both implicitly equate reference payoffs with expected equilibrium payoffs. General existence theorems are not possible for the solution concepts developed in these papers, as the examples in this section would suggest.

7. Sequentially coherent plans.

When a player makes a negotiation move in a dynamic multistage game, the tenability correspondence F should be derived from an analysis of the game after his move. In Section 2, we simply assumed that this correspondence was exogenously given, but we can now show how this correspondence may be derived.

Let us consider a dynamic multistage game with n players, numbered 1 to n, who get to move in order of their numbers (player 1 first, n last). During each player's move, he can make public statements and choose some payoff-relevant action. To keep our notation from getting too complex, we assume in this section that each player moves only once, there are no simultaneous moves, and all statements are observed by all players. However, we do not assume that a player's payoff-relevant actions are necessarily observed by the other players.

Such a game may be formally described by a model of the form

$$\Gamma = (p_0, \ (\Theta_i, \ T_i, \ \tau_i, \ C_i, \ p_i, \ u_i)_{i=1}^n, \ \Theta_{n+1}),$$

where, each T_i, C_i, and Θ_i is a nonempty finite set, τ_i is a function from

Θ_i into T_i, p_0 is a probability distribution over Θ_1, p_i is a function from $C_i \times \Theta_i$ into $\Delta(\Theta_{i+1})$, and u_i is a function from Θ_{n+1} into $\mathbb{R}$. We assume that

$$p_0(\theta_1) > 0, \quad \forall \theta_1 \in \Theta_1,$$

and, for every i in $\{1, 2, \ldots, n\}$,

$$\forall t_i \in T_i, \quad \exists \theta_i \in \Theta_i \quad \text{such that} \quad t_i = \tau_i(\theta_i),$$

and

$$\forall \theta_{i+1} \in \Theta_{i+1}, \quad \exists(c_i, \theta_i) \in C_i \times \Theta_i \quad \text{such that} \quad p_i(\theta_{i+1}|c_i, \theta_i) > 0.$$

In the interpretation of this model, Θ_i is the set of all possible <u>states</u> of the game at the beginning of player i's move. Θ_{n+1} is the set of all possible final states or <u>outcomes</u> of the game. T_i is the set of possible <u>types</u> (or private-information states) for player i when he makes his move, and $\tau_i(\theta_i)$ is i's type if the state of the game is θ_i when he moves. C_i is the set of possible payoff-relevant <u>actions</u> that i can choose among. The probability that the initial state of the game will be θ_1 is $p_0(\theta_1)$. If θ_i is the state when i moves and i chooses action c_i then $p_i(\theta_{i+1}|c_i, \theta_i)$ is the probability that θ_{i+1} will be the state when i+1 moves. If the final outcome of the game is θ_{n+1} then the utility payoff for player i is $u_i(\theta_{n+1})$. The assumptions listed above guarantee that any state and type could occur with positive probability, if the players choose their actions appropriately.

For any finite set X, we let $\Delta^0(X)$ denote the set of all probability distributions on X that assign positive probability to every element in X, so

$$\Delta^0(X) = \{q \in \Delta(X) | \quad q(x) > 0 \quad \forall x \in X\}.$$

A <u>conditional probability system</u> on X is defined by Myerson [17] to be function $\bar{q}$ that maps nonempty subsets of X into probability distributions on X such that, for any sets Y and Z such that $\emptyset \neq Y \subseteq Z \subseteq X$,

$$\sum_{y \in Y} \bar{q}(y|Y) = 1, \quad \text{and}$$

$$\bar{q}(x|Z) = \bar{q}(x|Y) \sum_{y \in Y} \bar{q}(y|Z), \quad \forall x \in Y.$$

We let $\Delta^*(X)$ denote the set of all conditional probability systems on X. Any probability distribution q in $\Delta^0(X)$ generates a conditional probability system $\bar{q}$ in $\Delta^*(X)$ by the formula

$$\bar{q}(x|Y) = \begin{cases} q(x)/\sum_{y \in Y} q(y), & \text{if } x \in Y, \\ 0, & \text{if } x \notin Y. \end{cases}$$

It can be shown (see Myerson [17]) that any conditional probability system in $\Delta^*(X)$ is the limit of a sequence of conditional probability systems that can be generated from distributions in $\Delta^0(X)$ in this way.

At the beginning of player i's moves, the beliefs of all players about the state of the game are derived from some conditional probability system in $\Delta^*(\Theta_i)$. When player i announces an allegation and a promise, he is in effect announcing how to compute the new conditional probability system in $\Delta^*(\Theta_{i+1})$. For any π_i in $\Delta^*(\Theta_i)$, any allegation λ_i in $(0,1]^{T_i}$, and any strategy γ_i in $\Delta^0(C_i)^{T_i}$, a conditional probability distribution $b_i(\pi_i,\lambda_i,\gamma_i)$ in $\Delta^*(\Theta_{i+1})$ is determined by Bayes' formula.

This function b_i can be defined precisely as follows. Suppose that π_i, λ_i, and γ_i are given as above. For any X that is a nonempty subset of Θ_{i+1}, let

$$Y = \{\theta_i \in \Theta_i | \exists c_i \in C_i \text{ and } \exists \theta_{i+1} \in X \text{ such that } p_i(\theta_{i+1}|c_i,\theta_i) > 0\}.$$

Then, for any x in X, let

$$\beta(x,X) = \sum_{y \in Y} \sum_{c_i \in C_i} \pi_i(y|Y) \lambda_i(\tau_i(y)) \gamma_i(c_i|\tau_i(y)) p_i(x|c_i,y).$$

Then, for any θ_i in Θ_i,

$$(b_i(\pi_i,\lambda_i,\gamma_i))(\theta_i|X) = \begin{cases} \beta(\theta_i,X)/\sum_{x \in X} \beta(x,X), & \text{if } \theta_i \in X, \\ 0, & \text{if } \theta_i \notin X. \end{cases}$$

This Bayesian-updating function $b_i : \Delta^*(\Theta_i) \times (0,1]^{T_i} \times \Delta^0(C_i)^{T_i} \to \Delta^*(\Theta_{i+1})$ is continuous on its domain. However, λ_i must be in $(0,1]^{T_i}$ and γ_i must be in $\Delta^0(C_i)^{T_i}$ (that is, both λ_i and γ_i must have all strictly positive components) to guarantee that Bayes' formula never involves dividing by zero in the definition of $b_i(\pi_i, \lambda_i, \gamma_i)$. To define Bayesian updating when some components of i's allegation and promise may be zero, we must define

$$B_i : \Delta^*(\Theta_i) \times [0,1]^{T_i} \times \Delta(C_i)^{T_i} \twoheadrightarrow \Delta^*(\Theta_{i+1})$$

to be the minimal upper-semicontinuous correspondence containing the graph of b_i. Then $B_i(\pi_i, \lambda_i, \gamma_i)$ is always a nonempty set of conditional probability systems on Θ_{i+1}, and

$$B_i(\pi_i, \lambda_i, \gamma_i) = \{b_i(\pi_i, \lambda_i, \gamma_i)\} \quad \forall \lambda_i \in (0,1]^{T_i}, \quad \forall \gamma_i \in \Delta^0(C_i)^{T_i}, \quad \forall \pi_i \in \Delta^*(\Theta_i).$$

To derive a model as in Section 2 to represent the negotiation problem faced by any player i in this game, we need to define the set of pure joint strategies for players after i. Player n has no followers, so let

$$S_n = \{\emptyset\}.$$

For any other player i, the set of pure joint strategies for players after i can be defined recursively by the formula

$$(7.1) \qquad S_i = (C_{i+1} \times S_{i+1})^{T_{i+1}}.$$

That is, a joint strategy for the players after i is any rule for determining the action of player i+1 and the joint strategies for the players after i+1 as a function of the type of player i+1. (The choice of a joint strategy for the players after i+1 can depend on i+1's type because his statement may convey information to them.)

We can recursively define utility functions for actions and strategies at earlier stages in the game by the following formulas:

$$U_i^n(c_n, \varnothing, \theta_n) = \sum_{\theta_{n+1} \in \Theta_{n+1}} p_n(\theta_{n+1}|c_n, \theta_n) \, u_i(\theta_{n+1}),$$

and, for any $j < n$,

$$U_i^j(c_j, s_j, \theta_j) = \sum_{\theta_{j+1} \in \Theta_{j+1}} p_j(\theta_{j+1}|c_j, \theta_j) \, U_i^{j+1}(s_j(\tau_{j+1}(\theta_{j+1})), \theta_{j+1}).$$

(Notice that $s_j(\tau_{j+1}(\theta_{j+1})) \in C_{j+1} \times S_{j+1}$ if $s_j \in S_j$.) Then $U_i^j(c_j, s_j, \theta_j)$ is the expected utility payoff for player i if, when j moves, the state is θ_j, j's action is c_j, and s_j is the joint strategy to be used by the players after j.

We now define $U_i : C_i \times S_i \times T_i \times \Delta^*(\Theta_i) \to \mathbb{R}$ by the formula

$$U_i(c_i, s_i, t_i|\pi_i) = \sum_{\theta_i \in \Theta_i} \pi_i(\theta_i|t_i) \, U_i^i(c_i, s_i, \theta_i),$$

where

$$\pi_i(\theta_i|t_i) = \pi_i(\theta_i| \{x \in \Theta_i | \tau_i(x) = t_i\}).$$

That is, $U_i(c_i, s_i, t_i|\pi_i)$ is the expected utility payoff for player i at his move if his type is t_i, his action is c_i, the joint strategy for the players following him is s_i, and conditional probability system π_i characterizes current beliefs about the state of the game.

We can now formulate the negotiation problem that player i faces in this game in the terms of the basic model that we developed in Section 2. Obviously, when the negotiator is player i, we should let $C = C_i$, $T = T_i$, and $S = S_i$ in the model of Section 2. Player i's negotiation problem depends on the current beliefs about the state of the game, which are characterized by some conditional probability system π_i in $\Delta^*(\Theta_i)$. Given such a π_i, the utility function $U(\bullet)$ in Section 2 should be identified with the function $U_i(\bullet|\pi_i)$. To complete the model, it now remains only to specify a tenability

correspondence F.

When $i = n$, the set of follower's joint strategies S_n is a trivial one-point set (since there are no followers), and we may define F_n by the formula $F_n(\lambda_n, \gamma_n | \pi_n) = \Delta(S_n)$. Then, letting $F(\cdot) = F_n(\cdot | \pi_n)$ completes the formulation of the last player n's negotiation problem in the model of Section 2. Following the construction of Sections 2 through 5, let $\phi_n(\pi_n)$ be the set of coherent plans for player n, for any given conditional probability system π_n in $\Delta^*(\Theta_n)$. By Theorem 2,

$$\varnothing \neq \phi_n(\pi_n) \subseteq \Delta(C_n \times S_n)^{T_n}, \quad \forall \pi_n \in \Delta^*(\Theta_n).$$

Now, for any player i, suppose inductively that a correspondence $\phi_{i+1}(\cdot)$ such that

$$(7.2) \qquad \varnothing \neq \phi_{i+1}(\pi_{i+1}) \subseteq \Delta(C_{i+1} \times S_{i+1})^{T_{i+1}}, \quad \forall \pi_{i+1} \in \Delta^*(\Theta_{i+1}),$$

has been defined to represent the set of coherent plans for player i+1 when he moves in this game, as a function of the public information at his move. Notice that $\Delta(C_{i+1} \times S_{i+1})^{T_{i+1}}$ can be naturally embedded in $\Delta(S_i)$, by the recursive definition (7.1) of S_i. Using this natural embedding, we define the correspondence $\psi_{i+1} : \Delta^*(\Theta_{i+1}) \twoheadrightarrow \Delta(S_i)$ so that $\sigma_i \in \psi_{i+1}(\pi_{i+1})$ iff there exists some μ_{i+1} in $\phi_{i+1}(\pi_{i+1})$ such that

$$\sigma_i(s_i) = \prod_{t_{i+1} \in T_{i+1}} \mu_{i+1}(s_i(t_{i+1}) | t_{i+1}), \quad \forall s_i \in S_i.$$

The sets $\phi_{i+1}(\pi_{i+1})$ and $\psi_{i+1}(\pi_{i+1})$ are obviously isomorphic, but they have different interpretations: $\phi_{i+1}(\pi_{i+1})$ is the set possible negotiation statements that player i+1 could coherently make; whereas $\psi_{i+1}(\pi_{i+1})$ is a set of requests that player i could make as a part of his negotiation statement. Since tenable requests of one negotiator must correspond to plans that would be confirmed by the next negotiator, we should identify $\psi_{i+1}(\pi_{i+1})$ with the

set of tenable requests that player i can make when π_{i+1} represents the updated
beliefs generated by i's allegation and promise.

So let the correspondence $f_i : [0,1]^{T_i} \times \Delta(C_i)^{T_i} \times \Delta^*(\Theta_i) \twoheadrightarrow \Delta(S_i)$ be
defined by the equation

$$f_i(\lambda_i, \gamma_i | \pi_i) = \{\sigma_i | \exists \pi_{i+1} \in B_i(\pi_i, \lambda_i, \gamma_i) \text{ such that } \sigma_i \in \psi_{i+1}(\pi_{i+1})\}.$$

That is, $f_i(\lambda_i, \gamma_i | \pi_i)$ is the set of all tenable requests by player i that would
be confirmed by player i+1, with some Bayesian updated beliefs, after player
i announced allegation λ_i and promise γ_i, if π_i represented the beliefs at the
beginning of i's move. Our inductive assumption (7.2) implies that
$f_i(\lambda_i, \gamma_i | \pi_i) \neq \emptyset$, as required by (2.1); and the definitions of b_i and B_i
guarantee that $f_i(\alpha\lambda_i, \gamma_i | \pi_i) = f_i(\lambda_i, \gamma_i | \pi_i)$ for any scalar α in $(0,1]$, as
required by (2.2). But we cannot guarantee that f_i is upper-semicontinuous,
as required by (2.3). So, for any π_i in $\Delta^*(\Theta_i)$, we define $F_i(\cdot | \pi_i)$ to be the
minimal upper-semicontinuous correspondence containing $f_i(\cdot | \pi_i)$. Then $F_i(\cdot | \pi_i)$
is nonempty-valued, homogeneous, and upper-semicontinuous, as conditions
(2.1)-(2.3) require. Letting $F(\cdot) = F_i(\cdot | \pi_i)$ completes the formulation of
player i's negotiation problem in the model of Section 2. Following the
construction of Sections 2 through 5, let $\phi_i(\pi_i)$ be the set of coherent plans
for player i, given the conditional probability system π_i in $\Delta^*(\Theta_i)$. By
Theorem 2,

(7.3) $\emptyset \neq \phi_i(\pi_i) \subseteq \Delta(C_i \times S_i)^{T_i}, \quad \forall \pi_i \in \Delta^*(\Theta_i),$

which justifies the inductive assumption (7.2). Thus, we have recursively
defined $\phi_i(\pi_i)$ to be the set of plans that player i could negotiate coherently,
given beliefs π_i, when he and each of his followers take full account of the
fact that a tenable request must be one that would not be renegotiated by
subsequent players.

Let $\bar{p}_0$ be the initial conditional probability system on Θ_1 generated by p_0. Then $\psi_1(\bar{p}_0)$ may be called the set of <u>sequentially coherent plans</u> of the game Γ, and it represents our fundamental cooperative solution concept for this game. These are the randomized joint strategies for all players that could be coherently announced in the first stage of the game and would not be contradicted by any coherent negotiation statement by any other player. Notice that $\psi_1(\bar{p}_0) \subseteq \Delta(S_0)$, where (unwinding the recursive definition (7.1))

$$S_0 = \underset{i=1}{\overset{n}{\times}} C_i^{(T_1 \times \ldots \times T_i)}.$$

So any sequentially coherent plan describes how every player's action may depend on the types of all players before him. Because a coherent plan is incentive compatible for the negotiating player at every stage, any sequentially coherent plan is a sequentially rational communication equilibrium of Γ in the sense of Myerson [17].

Plans that are <u>sequentially semicoherent</u> or (<u>sequentially coherent in the weak sense</u>) may be similarly defined, by letting $\phi_i(\pi)$ be recursively defined as the set of semicoherent plans for player i, above formula (7.3).

In this construction, we twice took an upper-semicontinuous extension of a function or correspondence: in going from b_i to B_i, and from f_i to F_i. The extension from b_i to B_i is an application of the basic idea of Kreps and Wilson [10]: that Bayesian updated posteriors after events of probability zero should be computed by taking the limits of the posteriors that would be computed in a slightly perturbed system in which everything has positive probability.

The extension from f_i to F_i resolves a more novel technical problem, however. To see why this extension is needed, consider the following

three-person game. First, player 1 chooses Heads or Tails. Then player 2

has an opportunity to negotiate, but he has only one trivial ("inaction")

option in C_2. Then player 3 chooses Heads or Tails. Suppose that neither 2

nor 3 observes 1's action. The payoffs are as follows: player 1 gets \$1 if

he matches 3 and gets \$0 otherwise; 2 gets \$1 if 3 chooses Tails and gets

\$0 otherwise; and 3 gets \$1 if he does not match 1 and gets \$0 otherwise.

Let ρ denote the probability that 1 chooses Heads in this game. The

coherent plans for player 3 are just his own best responses: Heads if

$\rho < .5$; Tails if $\rho > .5$; and any randomized strategy if $\rho = .5$. These are

also the tenable requests for player 2, so that his coherent plans must be

to request Heads if $\rho < .5$, and to request Tails (which he prefers) if

$\rho \geq .5$. This discontinuity at $\rho = .5$ creates a serious dilemma for

player 1. If 1 promises to use Heads with probability ρ such that $\rho \geq .5$,

then 2 should tell 3 to choose Tails, in which case 1's promise to use Heads

with positive probability is not reliable. On the other hand, if he promises

to use Heads with probability ρ such that $\rho < .5$, then 3 must choose Heads,

in which case 1's promise to use Tails with positive probability is not

reliable. Obviously, the unique equilibrium of this game is when 1 and 3

independently randomize with equal probability between Heads and Tails (so

$\rho = .5$), but 2's negotiation move seems to interfere with this plan (since

2 wants to steer 3 to Tails when 3 is indifferent). The technical resolution

of this paradox is in our extension from f_1 to F_1. While $f_1(.5)$ includes only

the plan in which 3 obeys a request from 2 to choose Tails, $F_1(.5)$ includes

a plan in which 3 does Heads as well as a plan in which 3 does Tails, because

for any positive ϵ, there are plans in $f_1(.5-\epsilon)$ in which 3 chooses Heads.

Thus, 1 can promise to use Heads with probability .5 and tenably "request"

that 3 should also randomize between Heads and Tails. To interpret this technical resolution, we might suppose that 3 has an inclination to obey 2's request as long as the expected cost to himself from doing so is less than some very small or infinitesimal number; and so 1 should choose Heads with a probability that is infinitesimally less than .5, to exactly counteract 3's inclination to obey player 2.

An alternative solution concept could be proposed in which the two upper-semicontinuous extensions are replaced by one. Given any π_i in $\Delta^*(\Theta_i)$, we could define $\hat{F}_i(\cdot|\pi_i)$ to be the minimal upper-semicontinuous correspondence on $[0,1]^{T_i} \times \Delta(C_i)^{T_i}$ that extends $\psi_{i+1}(b_i(\pi_i,\cdot))$, which is defined only on $(0,1]^{T_i} \times \Delta^o(C_i)^{T_i}$. Then, we could generate a somewhat different formulation of i's negotiation problem by using $\hat{F}_i(\cdot|\pi_i)$ as i's tenability correspondence instead of $F_i(\cdot|\pi_i)$. Using $\hat{F}_i$ in this way would be more in the spirit of Selten's [23] concept of trembling-hand perfect equilibrium, whereas our definition of sequential coherence using F_i is more in the spirit of Kreps and Wilson's concept of sequential equilibrium.

The assumption that players move one at a time, in a given order, is essential to the analysis of this section. For example, the analysis of the "Battle of the Sexes" game of Luce and Raiffa [12] depends on which player moves first, even if the player who moves second does not directly observe the action of the player who moves first. We only assume that the first player can commit to an action and make a negotiation statement before the second player can make any commitments. With this assumption, if the man is moves first in the "Battle of the Sexes" then the unique sequentially coherent plan is for both players to choose the actions that he most prefers ("going to the prize fight"), because he will announce that he is using his action from this

equilibrium and this statement will be credible. On the other hand, if the woman moves first, then the unique sequentially coherent plan is for both to use the actions that she most prefers ("going to the ballet").

The above example suggests that there is often an advantage to moving first. On the other hand, there are games in which moving later may be an advantage. For example, consider a three-person game in which players 1 and 2 have no payoff-relevant alternatives, and player 3's alternative actions are Heads and Tails. Suppose that 3 is indifferent between Heads and Tails, but 1 prefers that 3 should choose Heads, while 2 prefers that 3 should choose Tails. Under the assumption that 2 gets to make the last negotiation statement before 1 moves, the unique sequentially coherent plan is for 3 to choose Tails. In the initial negotiation statement, player 1 would like to request that 3 should choose Heads, but this request would not be tenable because it would be immediately contradicted by player 2. Thus, when two players are both trying to influence a third player by their negotiation statements, the last person to speak before his move may have the advantage.

8. Other negotiation structures.

In Section 7, we assumed that a player has an opportunity to negotiate whenever he moves in a game. It is straightforward to relax this assumption, but we must then add to the structure of the game a specification of which moves are "negotiation moves" and which are not.

For example, we might assume that a particular player i can make mediated public announcements at his move, but that these announcements are not necessarily treated as negotiation statements. Substantively, this means

either that player i cannot make his announcements in a language with given literal meanings (i.e., he can wave his arms or put lanterns in a church steeple but cannot say anything in English), or that he cannot be confident (for some exogenous reason) that the following players would attentively listen, understand, and respect literal meanings of his statements when they passed appropriate credibility tests. To take this revised assumption into account, the only change that is needed in the definition of sequentially coherent plans is, for such a nonnegotiating player i, to let $\Phi_i(\pi_i)$ be the set of incentive-compatible plans (rather than the set of coherent plans) when this set is defined just above formula (7.3) in Section 7.

If it is assumed that some given players can neither negotiate nor communicate when they move, then other changes are needed in the definition of sequentially coherent plans for multistage games. For example, suppose that between the negotiation moves of players i and j, where i moves before j, there is a sequence of moves by other players who can not communicate when they move. Then the strategies for the players between i and j that are specified in a tenable request by player i must form a sequential equilibrium for these players, given the requested behavior for player j and his followers and given i's allegation and promise. Furthermore, the strategies for player j and his followers that are specified in a tenable request by player i must form a coherent plan for the game starting at j's move, given the allegation and promise by i and given the requested strategies for the players between i and j. Using such a notion of tenability, it should be possible to generalize the model in Section 7 to allow for such an exogenously given set of noncommunicative players.

One extension that is straightforward to make in the model of Section 7

is to allow that a player may have more than one move in a game. To reduce
such games to the model of the Section 7, we can simply use the temporary-agent
device suggested by Selten [23]. That is, each time a particular player moves
again, we could suppose that he is represented by a different agent, who has
the same utility function and the same type-information as the player,
including a recollection of his past types and actions. Then treating these
agents as the "players" gives us an equivalent game in which no one moves
twice. (The only subtle point which needs to be checked, to verify this
equivalence, is that there is no advantage for a player's later agent to know
whether an earlier agent might have disobeyed or lied to the mediator who
transmitted his public messages. Since the messages that the mediator
transmitted for the earlier agent are known by everyone, and since we assume
that the later agent's type includes all information about the earlier agent's
type and action, there is no advantage for the later agent to learn anything
else about the earlier agent's private communications with his mediator.)

Although many of the most interesting models of bargaining (as, for
example, the model of Rubinstein [21, 22]) do assume that players move one
at a time, the assumption that there are no simultaneous moves is restrictive
in general. It is not clear how our model could be extended to allow for two
players to make simultaneous negotiation statements that are heard by all the
other players. The problem is that each player's set of tenable requests must
depend on what the other player is announcing. If they simultaneously make
different contradictory requests on a later player who is willing to obey
either one, which does he obey? (When the negotiators spoke one at a time,
we assumed in Section 7 that the later request would be obeyed.)

One way to avoid this dilemma is to suppose that players who negotiate

simultaneously must be speaking to different disjoint sets of players. A
simple model with this kind of negotiation structure was considered by Myerson
[15], in an analysis of equilibria among several principals who head separate
corporations. In this model, the set of incentive-compatible (that is, tenable
and reliable) plans for each corporation depends upper-semicontinuously on
the plans of other corporations. Consequently, each principal's optimal
incentive-compatible plan (that is, his coherent plan) for his corporation
may depend discontinuously on the plans chosen by other principals for their
corporations. This discontinuity may prevent the existence of any equilibrium
in which each principal is negotiating his optimal incentive-compatible plan
for his own corporation, given the plans that the other principals are
simultaneously negotiating. However, a natural quasi-equilibrium concept can
be defined, for which a general existence theorem can be proven (see [15]).
The essential idea behind this quasi-equilibrium concept is that our definition
of "optimality" or "coherence" must be weakened, so that the set of coherent
plans for any negotiator is convex and depends upper-semicontinuously on the
plans that are chosen simultaneously by other negotiators. This technical
point is likely to be relevant in any model with simultaneous negotiation.

In the model of Section 7, the only communication between players was
assumed to be in public statements and messages that all players observe. We
may want to drop this assumption and allow a mediator to transmit different
confidential messages to each player in the game. In [16], a model with such
full communication potential is analyzed under the assumption that one player,
called the principal, has all of the negotiating ability. That is, in the
model of [16], confidential messages can be transmitted to and from any player,
but only the credible public statements of the principal are necessarily

interpreted according to their exogenous literal meanings.

To formulate this model in the current context, let $\{1,2,\ldots,n\}$ be the set of players, and let player 1 be the principal. Let $\bar{C}_i$ be the set of actions available to player i, and let $\bar{T}_i$ be the set of his possible types. Let $\bar{C} = \times_{j=1}^{n} \bar{C}_j$, $\bar{C}_{-i} = \times_{j \neq i} \bar{C}_j$, $\bar{T} = \times_{j=1}^{n} \bar{T}_j$, $\bar{T}_{-i} = \times_{j \neq i} \bar{T}_j$. For any c in $\bar{C}$ and t in $\bar{T}$, let $u_i(c,t)$ denote the utility payoff to player i when c is the list of actions chosen by the players and t is the list of their actual types. Let $p_i(t_{-i}|t_i)$ denote the probability that player i would assign to the event that t_{-i} in $\bar{T}_{-i}$ is the list of types of the other players if t_i were i's own type.

A coordination plan or <u>mechanism</u> for this game is any μ in $\Delta(\bar{C})^{\bar{T}}$, specifying a probability distribution over the combinations of players' actions for any possible combination of players' types. For any such μ, any player i, any t_i and r_i in $\bar{T}_i$, and any function $\delta:\bar{C}_i \to \bar{C}_i$, let

$$\hat{V}_i(\mu,\delta,r_i|t_i) = \sum_{t_{-i}\in\bar{T}_{-i}} \sum_{c\in\bar{C}} p_i(t_{-i}|t_i)\, \mu(c|t_{-i},r_i)\, u_i((c_{-i},\delta(c_i)),t).$$

(Here c_i is the i-component of the vector c in $\bar{C}$, $(c_{-i},\delta(c_i))$ is the vector in $\bar{C}$ differing from c in that the i-component is changed from c_i to $\delta(c_i)$, (t_{-i},r_i) is the vector in $\bar{T}$ with i-component equal to r_i and all other components as in t_{-i}, and $t = (t_{-i},t_i)$.) Similarly, let

$$V_i(\mu|t_i) = \sum_{t_{-i}\in\bar{T}_{-i}} \sum_{c\in\bar{C}} p_i(t_{-i}|t_i)\, \mu(c|t)\, u_i(c,t).$$

Then a plan μ is <u>incentive compatible</u> iff, for every player i,

(8.1) $V_i(\mu|t_i) \geq \hat{V}_i(\mu,\delta,r_i|t_i),\quad \forall t_i\in\bar{T}_i,\quad \forall r_i\in\bar{T}_i,\quad \forall\delta:\bar{C}_i \to \bar{C}_i.$

The wide communication possibilities in this model cannot be subsumed in a model of the form considered in Sections 2 and 3, but they can be subsumed in the more general model considered in Section 4. To translate this model

into the formulation of Section 4, when the negotiator is player 1, we equate

the negotiator's set of types T and manipulative strategies D in Section 4

with the sets $\bar{T}_1$ and $\bar{C}_1^{\bar{C}_1}$ here respectively. The set of possible negotiation

statements (which is necessarily more general than the set of plans) for

player 1 is

$$\Omega = \{\mu \in \Lambda(\bar{C})^{\bar{T}} | \quad \exists \lambda \in [0,1]^{\bar{T}_1} \text{ such that } \forall t \in \bar{T}, \sum_{c \in \bar{C}} \mu(c|t) = \lambda(t_1)\}.$$

(Here t_1 is the 1-component of $t = (t_i)_{i=1}^{n}$.) That is, a statement is any

result of a star product between an allegation in $[0,1]^{\bar{T}_1}$ and a plan in $\Delta(\bar{C})^{\bar{T}}$,

where such a star product is defined by the formula

$$(\lambda * \mu)(c|t) = \lambda(t_1) \mu(c|t).$$

Notice that the formulas for V_i and $\hat{V}_i$ above can be applied to any statement

μ in Ω. Thus, the negotiator's discounted-payoff functions V and $\hat{V}$ in

Section 4 can be equated with the functions V_1 and $\hat{V}_1$ here respectively. We

may define the likelihood function L by the equations

$$L(\mu|t_1) = \sum_{c \in \bar{C}} \mu(c|t), \quad \forall t_{-1} \in \bar{T}_{-1}.$$

Finally, the set of tenable statements G is the set of statements that satisfy

the incentive constraints (8.1) for every player i other than player 1.

When we translate this principal's mechanism-design problem into the

formulation of Section 4 in this way, then the set of coherent plans is exactly

the set of neutral optima for the principal, as defined by in [16]. Similarly,

the set of semicoherent plans coincides with the principal's core defined in

[16]. To prove that the principal's neutral optima are coherent plans, notice

that w is a strongly attractive reference allocation in this negotiation

problem iff there are no nonnegative vectors λ and μ, other than the zero

vectors, that solve the following system of constraints:

$$\sum_{c\in C} \mu(c|t) = \lambda(t_1), \quad \forall t\in \bar{T};$$

$$V_i(\mu|t_i) \geq \hat{V}_i(\mu,\delta_i,r_i|t_i), \quad \forall i\neq 1, \quad \forall r_i\in\bar{T}_i, \quad \forall t_i\in\bar{T}_i, \quad \forall \delta_i:\bar{C}_i \rightarrow \bar{C}_i;$$

$$V_1(\mu|t_1) + (1 - \lambda(t_1))\ w(t_1) \geq \hat{V}_1(\mu,\delta_1,r_1|t_1) + (1 - \lambda(r_1))\ w(t_1),$$

$$\forall r_1\in\bar{T}_1, \quad \forall t_1\in\bar{T}_1, \quad \forall \delta_1:\bar{C}_1 \rightarrow \bar{C}_1;$$

$$V_1(\mu|t_1) + (1 - \lambda(t_1))\ w(t_1) \geq w(t_1), \quad \forall t_1\in\bar{T}_1.$$

Using theorems of the alternative (see [20, section 22]), or the duality theorem of linear programming, one can show that there are no nonzero solutions in the nonnegative orthants to these linear constraints iff there exist nonnegative vectors (ℓ,a) such that, for every type t_1 in $\bar{T}_1$,

$$(\ell(t_1) + \sum_{r_1} \sum_{\delta_1} a(\delta_1,r_1|t_1))\ w(t_1) - \sum_{r_1} \sum_{\delta_1} a(\delta_1,t_1|r_1)\ w(r_1) > \sum_{t_{-1}} \nu(t),$$

where, for every t in $\bar{T}$,

$$\nu(t) =$$

$$\max_{c\in\bar{C}}\ \ell(t_1)\ p_1(t_{-1}|t_1)\ u_1(c,t) + \sum_{i=1}^{n} \sum_{\delta_i} \sum_{r_i} a_i(\delta_i,r_i|t_i)\ p_i(t_{-i}|t_i)\ u_i(c,t)$$

$$- \sum_{i=1}^{n} \sum_{\delta_i} \sum_{r_i} a_i(\delta_i,t_i|r_i)\ p_i(t_{-i}|r_i)\ u_i((c_{-i},\delta_i(c_i)),(t_{-i},r_i)).$$

These dual constraints, derived from the credibility conditions, are almost the same as the conditions in Theorem 7 of [16] that characterize the principal's neutral optima. Notice that, since all the inequalities constraining (ℓ,a) are strict, we can always perturb ℓ slightly to make its components strictly positive, instead of merely nonnegative. Then, by Lemma 1 of [16], for any incentive-compatible plan μ, a sequence of strongly attractive allocations can converge to an attractive reference allocation that supports μ as a coherent plan iff there exists a sequence of warranted claims that satisfy the conditions of Theorem 7 of [16] for the same plan μ.

9. Proofs.

Proof of Lemma 1. Conditions (4.1), (4.4), and (4.6) follow easily from the definitions of L, V, and $\hat{V}$. (Notice that the V/L and $\hat{V}$/L ratios in (4.4) are expected values of bounded payoffs in the finite range of U(•).) Conditions (4.2) and (4.3) follow straightforwardly from (2.2) and the way that $\hat{G}$ and G are derived from F. Compactness of G follows from upper-semicontinuity of F (2.3) and Caratheodory's theorem.

To show that (4.5) is satisfied, we use a fixed-point argument. Let λ be given as in (4.5). For any (γ,σ) in $\Delta(C)^T \times \Delta(S)$, define $K(\gamma,\sigma)$ so that $(\hat{\gamma},\hat{\sigma}) \in K(\gamma,\sigma)$ iff $\hat{\sigma}$ is in the convex hull of $F(\lambda,\gamma)$, $\hat{\gamma} \in \Delta(C)^T$, and

$$\sum_{c \in C} \sum_{s \in S} \hat{\gamma}(c|t) \, \sigma(s) \, U(c,s,t) = \max_{c \in C} \sum_{s \in S} \sigma(s) \, U(c,s,t), \quad \forall t \in T.$$

Using upper-semicontinuity of F, it is straightforward to show that K satisfies all the conditions of the Kakutani fixed-point theorem, so there exists some $(\bar{\gamma},\bar{\sigma})$ such that $(\bar{\gamma},\bar{\sigma}) \in K(\bar{\gamma},\bar{\sigma})$. Let $\mu = \lambda * \bar{\gamma} * \bar{\sigma}$. Since $\bar{\sigma}$ is in the convex hull of $F(\lambda,\bar{\gamma})$, μ is in G. Furthermore, $L(\mu|t) = \lambda(t)$ for every t, and

$$\sum_{c \in C} \sum_{s \in S} \mu(c,s|t) \, U(c,s,t)/\lambda(t) = \sum_{c \in C} \sum_{s \in S} \bar{\gamma}(c|t) \, \bar{\sigma}(s) \, U(c,s,t)$$

$$\geq \sum_{c \in C} \sum_{s \in S} \bar{\gamma}(c|r) \, \bar{\sigma}(s) \, U(\delta(c),s,t) = \hat{V}(\mu,\delta,r|t)/\lambda(r)$$

for every t and r in T and every δ in D. Thus μ satisfies condition (4.5).

Q.E.D.

Proof of Theorem 1. Let $\hat{w}$ be any strongly attractive reference allocation such that $|\hat{w}(t) - w(t)| < |V(\mu|t) - w(t)|$ for every t such that $V(\mu|t) \neq w(t)$. Let $\lambda(t) = 1$ if $V(\mu|t) \geq \hat{w}(t)$, and let $\lambda(t) = 0$ if $V(\mu|t) < \hat{w}(t)$. Condition (3.1) and the equations in (4.6) than imply that the statement $\lambda * \mu$ and the reference allocation $\hat{w}$ would satisfy

conditions (5.5) and (5.6). Clearly $\hat{\lambda}(t) = 1$, so $\lambda * \mu$ is not null. If $\lambda * \mu$ were in G, then $\lambda * \mu$ would be credible with respect to the strongly attractive vector $\hat{w}$, which is impossible. Q.E.D.

Given a negotiation problem as defined in Section 4, let M be a number such that

$$M \geq |V(\mu|t)/L(\mu|t)| \quad \forall t \in T, \quad \forall \mu \in G \text{ such that } L(\mu|t) > 0.$$

Such a bound M exists by condition (4.4).

Lemma 2. Given any w in $\mathbb{R}^T$ and any $\hat{t}$ in T, if $w(\hat{t}) < -M$ then there exists some μ in G such that μ is credible with respect to w and $L(\mu|\hat{t}) = 1$.

Proof of Lemma 2. For any small positive number ε, define the correspondence $J_\varepsilon : [-M,M]^T \times [\varepsilon,1]^T \twoheadrightarrow [-M,M]^T \times [\varepsilon,1]^T$ so that $(z,m) \in J_\varepsilon(y,\lambda)$ iff

(1) there exists some μ in G such that $L(\mu|r) = \lambda(r) \quad \forall t \in T$, and

$$z(t) = V(\mu|t)/\lambda(t) \geq \hat{V}(\mu,\delta,r|t)/\lambda(r) \quad \forall \delta \in D, \quad \forall t \in T, \quad \forall r \in T,$$

and

(2) $m(t) = 1$ if $y(t) > w(t)$,

$m(t) = \varepsilon$ if $y(t) < w(t)$

$m(t) \in [\varepsilon,1]$ if $y(t) = w(t)$.

Condition (4.5) guarantees that $J_\varepsilon(y,\lambda)$ is a nonempty set. Condition (4.1) and the assumption that G is compact and convex guarantee that $J_\varepsilon(y,\lambda)$ is convex and that J_ε is upper-semicontinuous. Thus, by the Kakutani fixed-point theorem, there exists some $(y_\varepsilon,\lambda_\varepsilon)$ such that $(y_\varepsilon,\lambda_\varepsilon) \in J_\varepsilon(y_\varepsilon,\lambda_\varepsilon)$. Clearly, $\lambda_\varepsilon(\hat{t}) = 1$, since $y_\varepsilon(\hat{t}) \geq -M > w(\hat{t})$. Notice that $\lambda_\varepsilon(t) = \varepsilon$ if $y_\varepsilon(t) < w(t)$, and $\lambda_\varepsilon(t) = 1$ if $y_\varepsilon(t) > w(t)$. Let μ_ε be chosen to satisfy (1) when $z = y = y_\varepsilon$ and $m = \lambda = \lambda_\varepsilon$.

By compactness, we can assume that the sequences $\{y_\varepsilon\}$, $\{\lambda_\varepsilon\}$, and $\{\mu_\varepsilon\}$

are all convergent as ϵ goes to zero. We let $\bar{y}$, $\bar{\lambda}$, and $\bar{\mu}$ denote the limits

of these sequences. Notice $\bar{\mu} \in G$, because G is compact. For any t in T,

if $\bar{y}(t) < w(t)$ then $\bar{\lambda}(t) = 0$, and if $\bar{y}(t) > w(t)$ then $\bar{\lambda}(t) = 1$. Thus,

for any t in T,

$$V(\bar{\mu}|t) + (1 - L(\bar{\mu}|t)) \, w(t) = (\bar{y}(t) - w(t)) \, \bar{\lambda}(t) + w(t)$$

$$= \max\{\bar{y}(t), w(t)\}.$$

Thus, (5.5) is satisfied by $\bar{\mu}$. Furthermore,

$$\bar{\lambda}(r) \, \bar{y}(t) = \bar{\lambda}(r) \lim_{\epsilon \to 0} y_\epsilon(t) \geq \bar{\lambda}(r) \lim_{\epsilon \to 0} \hat{V}(\mu_\epsilon, \delta, r|t) / \lambda_\epsilon(r)$$

$$= \hat{V}(\bar{\mu}, \delta, r/t)$$

for any t, r, and δ. (Notice that, if $\bar{\lambda}(r) = L(\bar{\mu}|r) = 0$ then

$\hat{V}(\bar{\mu}, \delta, r|t) = 0$, by continuity (4.1) and boundedness (4.4).) Thus,

$$\bar{V}(\bar{\mu}|t) + (1 - L(\bar{\mu}|t)) \, w(t) = \max\{\bar{y}(t), w(t)\} \geq \bar{\lambda}(r) \, \bar{y}(t) + (1 - \bar{\lambda}(r)) \, w(t)$$

$$\geq \hat{V}(\bar{\mu}, \delta, r|t) + (1 - L(\bar{\mu}|t)) \, w(t),$$

so $\bar{\mu}$ satisfies (5.6). Finally, $L(\bar{\mu}|\hat{t}) = \bar{\lambda}(\hat{t}) = 1$, so $\bar{\mu}$ is not null. Thus,

$\bar{\mu}$ is credible with respect to w.   Q.E.D.

Proof of Theorem 2. Notice first that, if a statement μ is credible with

respect to w, then there exists a statement $\bar{\mu}$ that is credible with respect

to w such that $\Sigma_{t \in T} \, L(\bar{\mu}|t) \geq 1$. To prove this, let

$$\bar{\mu} = (1/(\max\{L(\mu|t)| \; t \in T\}))\mu.$$

This $\bar{\mu}$ satisfies (5.5) and (5.6) by linearity. Condition (4.2) implies that

$(1/(k \, \max\{L(\mu|t)| \; t \in T\}))\mu$ is in G, for some sufficiently large integer k,

and then (4.3) implies that this $\bar{\mu}$ is in G.

Now, for any w in $[-(M+2), M+2]^T$, let $H_1(w) \subset \mathbb{R}^T$ be defined so that

$y \in H_1(w)$ iff there exists some statement μ that is credible with respect

to w such that $\Sigma_{r \in T} \, L(\mu|r) \geq 1$ and $y(t) = w(t) + L(\mu|t)$ for every t in T.

It is straightforward to check that $H_1(\bullet)$ is an upper-semicontinuous and convex-valued correspondence.

$H_1(w) = \emptyset$ iff w is strongly attractive. Let $H_2(w)$ be defined so that $H_2(w) = \{w - \underset{\sim}{1}_T\}$ if w is strongly attractive, $H_2(w) = H_1(w)$ if w is not attractive, and $H_2(w)$ is the convex hull of $H_1(w)$ and $\{w - \underset{\sim}{1}_T\}$ if w is attractive but not strongly attractive. It is straightforward to check that H_2 is nonempty-convex-valued and upper-semicontinuous. (Here, $\underset{\sim}{1}_T$ is the vector in $\mathbb{R}^T$ in which all components are equal to one.)

For any w in $[-(M+2),M+2]^T$, $H_2(w) \subseteq [-(M+2),M+2]^T$. To verify this, notice first that, if $y \in H_2(w)$ then $|y(t) - w(t)| \leq 1$ for every t, so that $y(t)$ can fail to be in $[-(M+2),M+2]$ only when $|w(t)| \geq M + 1$. If $w(t) > M$, then $L(\mu|t) = 0$ for any μ that is credible with respect to w (by (5.5) and the definition of the bound M), so that $y \in H_2(w)$ implies that $y(t) \leq w(t) \leq M + 2$. On the other hand, if $w(t) < -M$ for some t, then w is not attractive, by Lemma 2, and so $H_2(w) = H_1(w)$, and so $y \in H_2(w)$ implies that $y(t) \geq w(t) \geq -(M + 2)$. Thus, $y \in H_2(w)$ implies that $y \in [-(M+2),M+2]^T$.

By the Kakutani fixed-point theorem, there exists some $\bar{w}$ in $[-(M+2),M+2]^T$ such that $\bar{w} \in H_2(\bar{w})$. Notice that $w \notin H_1(w)$ for every w in $[-(M+2),M+2]^T$, so $H_2(\bar{w}) \neq H_1(\bar{w})$, which implies that $\bar{w}$ is attractive. So $\bar{w}$ is a convex combination of the vector $\bar{w} - \underset{\sim}{1}_T$ and a vector in $H_1(\bar{w})$. Therefore, by definition of H_1, there is some number α and some statement μ that is credible with respect to $\bar{w}$ and such that $L(\mu|t) = \alpha$ for every t in T. Let $\bar{\mu} = (1/\alpha)\,\mu$. Then $\bar{\mu} \in G$ (by conditions (4.2) and (4.3)), and $\bar{\mu}$ is credible with respect to $\bar{w}$. Furthermore, $L(\bar{\mu}|t) = 1$ for every t in T, so $\bar{\mu}$ is a coherent plan.   Q.E.D.

REFERENCES

1. R. Aumann, "Subjectivity and Correlation in Randomized Strategies," Journal of Mathematical Economics 1 (1974), 67-96.

2. J. Banks and J. Sobel, "Equilibrium Selection in Signalling Games," Cal Tech and U.C.S.D. working paper, 1985.

3. I. Cho and D. Kreps, "More Signalling Games and Stable Equilibria," Stanford University working paper, 1985.

4. V. Crawford and J. Sobel, "Strategic Information Transmission," Econometrica 50 (1982), 579-594.

5. J. Farrell, "Meaning and Credibility in Cheap-Talk Games," GTE Laboratories working paper, 1986.

6. S. Grossman and M. Perry, "Perfect Sequential Equilibrium," Journal of Economic Theory 39 (1986), 97-119.

7. E. Kalai and D. Samet, "Persistent Equilibria in Strategic Games," International Journal of Game Theory 13 (1984), 129-144.

8. E. Kohlberg and J.-F. Mertens, "On the Strategic Stability of Equilibria," Econometrica 54 (1986), 1003-1037.

9. D. Kreps, "Signalling Games and Stable Equilibria," Stanford University working paper, 1984.

10. D. Kreps and R. Wilson, "Sequential Equilibria," Econometrica 50 (1982), 863-894.

11. P. Kumar, "Sequential Incentive Mechanism Design and the Incomplete Revelation Principle," Carnegie-Mellon University working paper, 1986.

12. R. Luce and H. Raiffa, Games and Decisions, Wiley, New York, 1957.

13. A. McLennan, "Justifiable Beliefs in Sequential Equilibrium," Econometrica 53 (1985), 889-904.

14. H. Moulin and J.-P. Vial, "Strategically Zero-Sum Games: the class whose completely mixed equilibria cannot be improved upon," International Journal of Game Theory 7 (1978), 201-221.

15. R. Myerson, "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems," Journal of Mathematical Economics 10 (1982), 67-81.

16. R. Myerson, "Mechanism Design by an Informed Principal," Econometrica 51 (1983), 1767-1797.

17. R. Myerson, "Multistage Games with Communication," Econometrica 54 (1986), 323-358.

18. J. Nash, "Noncooperative Games," Annals of Mathematics 54 (1951), 289-295.

19. M. Okuno-Fujiwara and A. Postlewaite, "Forward Induction and Equilibrium Refinement," University of Pennsylvania working paper, 1987.

20. R. Rockafellar, Convex Analysis, Princeton University Press, Princeton, 1970.

21. A. Rubinstein, "Perfect Equilibria in a Bargaining Model," Econometrica 50 (1982), 97-109.

22. A. Rubinstein, "A Bargaining Model with Incomplete Information about Time Preferences," Econometrica 53 (1985), 1151-1172.

23. R. Selten, "Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games," International Journal of Game Theory 4 (1975), 25-55.