Discussion Paper No. 527

TWO-PERSON BARGAINING PROBLEMS WITH INCOMPLETE INFORMATION

by

Roger B. Myerson

June 1982

J.L. Kellogg Graduate School of Management
Northwestern University
Evanston, Illinois 60201

Abstract. A generalization of the Nash bargaining solution is defined for two-person bargaining problems with incomplete information. These solutions form the smallest set satisfying three axioms: a probability-invariance axiom, an extension (or independence of irrelevant alternatives) axiom, and a random-dictatorship axiom. A bargaining solution can also be characterized as an incentive-compatible mechanism that is both equitable and efficient in terms of some virtual utility scales for the two players.

Acknowledgements. Research for this paper was supported by the Kellogg Center for Advanced Study in Managerial Economics and Decision Sciences, and by a research fellowship from I.B.M.

TWO-PERSON BARGAINING PROBLEMS WITH INCOMPLETE INFORMATION

by Roger B. Myerson

1. Introduction

Consider first the simplest of bargaining games, in which two risk-neutral players can divide \$100 in any way that they agree on, or else they each get \$0 if they fail to agree. In this example, there is a natural common scale (dollars) for making interpersonal comparisons of utility, and both players have equal power to prevent an agreement, so \$50 for each individual is the obvious bargaining solution. This 50-50 split is fair, in the sense that each player gains as much from the agreement as he is contributing to the other player, as measured in the natural utility scale. One goal of cooperative game theory is to provide a formal definition of fair equitable agreements for the widest possible class of bargaining games. Such a theory of fair bargaining solutions can be useful both for prescriptive purposes, providing guidelines for arbitrators, and for descriptive purposes, if we assume that individuals tend to reach agreements in which each gains as much as he contributes to the other.

The bargaining solution of Nash [1950, 1953] is the best-known solution concept for two-person bargaining problems. It selects a unique Pareto-efficient utility allocation for any bargaining problem with complete information, and it coincides with the 50-50 split for the simple example above.

A game with incomplete information is a game in which each player may have private information (about the payoff structure of the game) which the others do not know, at the time when the game is played. Harsanyi and Selten [1972] proposed an extension of the Nash bargaining solution for two-person games with incomplete information, and a modified version of this solution concept was used in Myerson [1979a]. However, this solution concept uses probabilities in a way which cannot be based on the essential decision-theoretic structure of the bargaining game. In this paper, we will develop a new generalization of the Nash bargaining solution for games with incomplete information. (See Myerson [1982] for an analogous generalization of the NTU value.)

In a bargaining game with incomplete information, the players may be uncertain about each other's preferences or endowments. To describe such situations, we shall use the concept of Bayesian bargaining problem, based on ideas from Harsanyi [1967-68]. Formally, a two-person Bayesian bargaining problem Γ is an object of the form

$$(1.1) \qquad \Gamma = (D, d^*, T_1, T_2, u_1, u_2, p_1, p_2)$$

whose components are interpreted as follows. D is the set of collective decisions or feasible outcomes available to the two players if they cooperate, and $d^* \varepsilon D$ is the conflict outcome which the players must get by default if they fail to cooperative. For each player i (i=1,2), T_i is the set of possible types for player i. That is, each t_i in T_i represents a complete description of player i's relevant characteristics: his preferences, beliefs, and endowments. Each u_i is a function from $D \times T_1 \times T_2$ into the real numbers $\mathbb{R}$, such that $u_i(d, t_1, t_2)$ is the payoff which player i would get if d in D were chosen and if (t_1, t_2) were the vector of players' types. These payoff numbers are measured in a vonNeumann-Morgenstern utility scale for each player.

Without loss of generality, we shall assume that utilities are normalized so that $u_i(d^*,t_1,t_2) = 0$ for all i, t_1, t_2. Each p_i is a function that specifies the conditional probability distribution that each type of player i would assess over the other player's possible type. That is, $p_1(t_2|t_1)$ is the conditional probability of player 2 being of type t_2, as would be assessed by player 1 if he were of type t_1. Similarly, $p_2(t_1|t_2)$ is the conditional probability of player 1 being of type t_1, as would be assessed by player 2 if he were of type t_2. To simplify our notation, we let T denote the set of all possible type-pairs $t=(t_1,t_2)$; that is,

$$T = T_1 \times T_2.$$

Any pair $t=(t_1,t_2)$ in T represents a possible <u>state</u> of the players' information. For mathematical simplicity, we shall assume that D and T are finite sets throughout this paper.

The players' conditional probability distributions p_1 and p_2 are said to be <u>consistent</u> iff there exists some <u>common prior</u> probability distribution p on T such that p_1 and p_2 can be derived from p by Bayes theorem, so that

$$p_i(t_{-i}|t_i) = p(t)/p^i(t_i), \qquad \forall i\varepsilon\{1,2\}, \quad \forall(t_1,t_2)\varepsilon T,$$

where

$$p^i(s_i) = \sum_{t_{-i}\varepsilon T_{-i}} p(t_{-i},s_i), \quad \forall i\varepsilon\{1,2\}, \quad \forall s_i\varepsilon T_i.$$

(We use the notation $T_{-1} = T_2$, $t_{-1} = t_2$, $T_{-2} = T_1$, $t_{-2} = t_1$, and $t = (t_1,t_2)$ throughout this paper. We use (t_{-i},s_i) to stand for (t_1,s_2) or (s_1,t_2), depending on whether i=2 or i=1.) We will not need to assume consistency for any results in this paper, but it will be helpful when we interpret these results in Section 9.

The players in a bargaining problem do not have to agree on a specific outcome in D; instead they may agree on some <u>decision rule</u> or <u>mechanism</u>, which is a contract specifying how the choice should depend on the players' types. We will allow randomized strategies, so a decision mechanism is here defined to be any real-valued function μ on the domain D × T such that

$$(1.2) \qquad \sum_{c \in D} \mu(c|t) = 1 \quad \text{and} \quad \mu(d|t) \geq 0, \qquad \forall d \in D,\ \forall t \in T$$

That is, $\mu(d|t)$ is the probability of choosing outcome d in the mechanism μ, if t_1 and t_2 are the players' types.

Given any mechanism μ satisfying (1.2), we let $U_i(\mu|t_i)$ denote conditionally expected utility for player i, given that he is of type t_i, if the mechanism μ is implemented. That is, for any i in $\{1,2\}$ and any t_i in T_i,

$$(1.3) \qquad U_i(\mu|t_i) = \sum_{t_{-i} \in T_{-i}} \sum_{d \in D} p_i(t_{-i}|t_i)\, \mu(d|t)\, u_i(d,t)$$

Since the players can agree on a mechanism, they do not need to reveal anything about their actual types in the negotiating process. That is, instead of player 1 saying "I demand decision d" if he is type t_1 and saying "I demand decision $\hat{d}$" if he is type $\hat{t}_1$, he can say "I demand a mechanism with $\mu(d|t_1,t_2) = 1$ and $\mu(\hat{d}|\hat{t}_1,t_2) = 1$" in both types, and thus make the same effective demands without revealing whether t_1 or $\hat{t}_1$ is true. Throughout this paper, we shall assume that neither player will ever deliberately reveal any information about his true type until the mechanism is agreed upon. This assumption, that players will bargain inscrutably, has been discussed and justified in the context of principal-subordinate cooperation in Myerson [1981].

In order to conceal his type, each player must phrase his bargaining offers and demands in a way which is independent of his type. However, this need for inscrutability can create a new kind of dilemma for a player. It can easily happen that player 1, when he knows only his own type, would be indifferent between two mechanisms μ and $\hat{\mu}$; but player 2 might prefer μ over $\hat{\mu}$ if t_2 is his type, and $\hat{\mu}$ over μ if $\hat{t}_2$ is his type. So player 2 would prefer to argue for μ if t_2 is true and for $\hat{\mu}$ if $\hat{t}_2$ is true. However, such a policy by player 2 would reveal information to player 1, which could destroy player 1's indifference between μ and $\hat{\mu}$. For example, 1 might be indifferent between betting that 2 can or cannot speak French, until 1 learns that 2 wants to bet he can.

Thus, each player must be careful to use a bargaining strategy which maintains a balance between the conflicting goals that he would have if he were of different types, (maximizing $U_i(\cdot|t_i)$ or $U_i(\cdot|\hat{t}_i)$) , even though he already knows his actual type. That is, in bargaining games with incomplete information, we need to understand not only how fair compromises between players 1 and 2 should be defined, but also how fair compromises between alternative types of the same player should be defined.

2. Feasible and Efficient Mechanisms

We must now clarify one additional question of interpretation relating to our Bayesian bargaining problems: are the players' types verifiable or unverifiable? If the types are verifiable, it means that players can costlessly prove their types to each other. One may think of a verifiable type as consisting of information written on a government-certified identification card, which each player keeps hidden during the bargaining but can pull out to prove his type afterwards. If the types are unverifiable, it means that players cannot prove their types to each other, and so each player would lie about his type whenever such a lie might be profitable. For example, an unobservable subjective preference would be unverifiable in this sense. When types are unverifiable, players will not reveal their types honestly unless they are given incentives to do so.

Actually, by appropriately redefining the set of choices D, one can describe any situation with verifiable types by a more elaborate model with unverifiable types (by building the verification procedure into the definition of a "chosen outcome"); so the unverifiable-types assumption is more general. Thus, in this paper, we shall restrict our attention to the case of unverifiable types.

Recall that $U_i(\mu|t_i)$ denotes the expected utility for player i in mechanism μ if his type is t_i and both players report their types honestly. We let $U_i^*(\mu,s_i|t_i)$ denote the expected utility for player i in mechanism μ if his type were t_i but he pretended that his type were s_i in implementing the mechanism, while the other player remained honest. That is

$$U_i^*(\mu,s_i|t_i) = \sum_{t_{-i}\varepsilon T_{-i}} \sum_{d\varepsilon D} p_i(t_{-i}|t_i)\, \mu(d|t_{-i},s_i)\, u_i(d,t).$$

A mechanism μ is (Bayesian) incentive compatible iff

$$(2.1) \qquad U_i(\mu|t_i) \geq U_i^*(\mu,s_i|t_i), \qquad \forall i\varepsilon\{1, 2\},\ \forall t_i\varepsilon T_i,\ \forall s_i\varepsilon T_i.$$

Condition (2.1) asserts that if player i is of type t_i, then his expected utility $U_i(\mu|t_i)$ from participating honestly in mechanism μ cannot be less than his expected utility from pretending to be of any other type s_i. That is, (2.1) asserts that honest participation in the choice mechanism μ is a

Nash equilibrium for the two players. If (2.1) were violated, then at least one type of one player would be tempted to lie about his type and so, since types are unverifiable, the mechanism μ could not be implemented. It can be shown that even dishonest equilibrium behavior in more general mechanisms cannot achieve any expected utility allocations which are not also achieved by incentive-compatible mechanisms satisfying (2.1); see Myerson [1979a], for example. Thus there is no loss of generality in restricting our attention to such incentive-compatible direct revelation mechanisms.

Besides lying about his type, a player can also force the conflict outcome if his expected utility is less than zero. A mechanism μ is said to be individually rational iff

(2.2) $$U_i(\mu|t_i) \geq 0 \qquad \forall i\varepsilon\{1,2\}, \quad \forall t_i \varepsilon T_i.$$

A mechanism μ is feasible iff it satisfies the probability conditions (1.2) and is incentive compatible and individually rational.

A mechanism μ is incentive-efficient iff μ is feasible and there does not exist any other incentive-compatible mechanism $\hat{\mu}$ such that

(2.3) $$U_i(\hat{\mu}|t_i) \geq U_i(\mu|t_i) \qquad \forall i\varepsilon\{1,2\}, \quad \forall t_i \varepsilon T_i,$$

with strict inequality for at least one t_i. That is, if μ is incentive-efficient, then there is no other incentive-compatible mechanism that gives higher expected utility to at least one type of one player without giving lower expected utility to any type of either player. Notice that, if there were another feasible mechanism $\hat{\mu}$ satisfying (2.3), then an outside arbitrator could be sure that both players would be willing to accept a change from μ to $\hat{\mu}$, no matter what their types are. See Holmström and Myerson [1981] for a wider discussion of efficiency with incomplete information.

Given any mechanism μ, we let $U(\mu)$ denote the vector of all $U_i(\mu|t_i)$ conditionally-expected utility levels for each player, given each of his possible types. That is,

$$U(\mu) = \left(\left(U_i(\mu|t_i)\right)_{t_i \varepsilon T_i}\right)_{i \varepsilon \{1,2\}},$$

so $U(\mu)$ is a vector with $|T_1|+|T_2|$ components.

Our definition of incentive-efficiency implicitly uses $U(\mu)$ as the relevant utility allocation vector for welfare analysis. It would not be appropriate to average player i's expected utility over his various types, because we are assuming that he already knows his true type at the time of bargaining. On the other hand, an arbitrator (or an external social theorist) does not know which t_i is true, so welfare analysis must be based on consideration of all of the $U_i(\mu|t_i)$ numbers, for all possible t_i. Even if the players bargain without the help of an arbitrator, all of the components of $U(\mu)$ may be significant in determining whether mechanism μ is chosen (not just the components corresponding to the two actual types), because each player must express a compromise among the preferences of all of his possible types in bargaining, in order to not reveal his true type during the bargaining process.

3. The probability-invariance axiom

Let BP denote the set of all two-person Bayesian bargaining problems of the form (1.1). Then a solution concept for bargaining problems is a mapping $S(\cdot)$ such that, for any Bayesian bargaining problem Γ in BP, $S(\Gamma)$ is a set of feasible mechanisms for Γ. That is, if $\mu \varepsilon S(\Gamma)$ then μ should be considered a fair bargaining solution for the two players in Γ. Our theoretical problem is to find a reasonable definition of such a solution correspondence $S(\cdot)$.

Following Nash [1950], we will approach this problem axiomatically. We will present some basic properties that a fair solution correspondence should satisfy and derive a generalization of Nash's bargaining solution from these properties.

Harsanyi and Selten [1972] proposed that the solution to a Bayesian bargaining problem should be the mechanism which maximizes

$$(3.1) \qquad \left(\prod_{t_1 \varepsilon T_1} (U_1(\mu|t_1))^{p^1(t_1)}\right)\left(\prod_{t_2 \varepsilon T_2} (U_2(\mu|t_2))^{p^2(t_2)}\right)$$

over the set of all feasible mechanisms (although they defined the set of feasible mechanisms somewhat differently from in this paper). They assumed that the players' probability distributions are consistent, so that the marginal probabilities $p^i(t_i)$ are well-defined. Formula (3.1) is a natural generalization of the product-maximization formula characterizing the Nash [1950] bargaining solution, and Harsanyi and Selten have derived it from a very convincing set of axioms.

A fundamental property of the Nash bargaining solution is that it depends only on the decision-theoretically significant structures of the problem. (Nash's scale-invariance axiom follows from this property.) For a solution defined on general Bayesian bargaining problems, this property implies the following axiom:

Axiom 1. (Probability-invariance) Consider any two Bayesian bargaining problems $\Gamma = (D, d^*, T_1, T_2, u_1, u_2, p_1, p_2)$ and $\hat{\Gamma} = (D, d^*, T_1, T_2, \hat{u}_1, \hat{u}_2, \hat{p}_1, \hat{p}_2)$, having the same decision sets, type sets, and conflict outcome. Suppose that

$$p_i(t_{-i}|t_i)\, u_i(d,t) = \hat{p}_i(t_{-i}|t_i)\, \hat{u}_i(d,t), \quad \forall i \in \{1,2\}, \quad \forall d \in D, \quad \forall t \in T.$$

Then these two bargaining problems must have the same solutions; that is,

$S(\Gamma) = S(\hat{\Gamma})$.

To see why the probability-invariance axiom must hold, notice that whenever we compute an expected utility, we always multiply probabilities by utilities, as in the axiom. Thus, both bargaining problems in the axiom have the same sets of feasible mechanisms, and each mechanism μ generates the same vector $U(\mu)$ of conditionally expected utilities in both problems.

In effect, the probability-invariance axiom states that probabilities cannot be meaningfully defined separability from utilities, when state-dependent utility functions are allowed (see Myerson [1979b] for a basic development of this idea). This axiom was first observed by Aumann and Maschler [1967]. It implies that there is no loss of generality in considering only bargaining problems in which the two players' types are stochastically independent, provided that $u_i(d,t)$ is allowed to depend on both components of t in any arbitrary way. For example, Γ is equivalent to $\hat{\Gamma}$ where

(3.2) $\quad \hat{p}_i(t_{-i}|t_i) = 1/|T_{-i}|, \quad \hat{u}_i(d,t) = |T_{-i}|\; p_i(t_{-i}|t_i)\, u_i(d,t), \; \forall i, \; \forall d \in D, \; \forall t \in T.$

(In Myerson [1976], this axiom was appplied to n-person dynamic games, in which the probabilities of some players' types may depend on the decisions of earlier players, to reduce dynamic problems to equivalent static problems in which all players' types are determined simultaneously and independently.)

For our present purposes, the most important application of the probability-invariance axiom is to rule out Harsanyi and Selten's solution, because the probability exponents in (3.1) depend on the probabilities separately from the utility functions. Thus, we are presented with a dilemma: Harsanyi and Selten have derived (3.1) uniquely from a convincing set of axioms, and yet this criterion violates the probability-invariance axiom. To resolve this dilemma, we must relax one of Harsanyi and Selten's assumptions. The assumption to be weakened will be their axiom of irrelevant alternatives.

4. The Extension Axiom

Given two bargaining problems Γ and $\hat{\Gamma}$ in BP, we say that $\hat{\Gamma}$ extends Γ iff Γ and $\hat{\Gamma}$ can be written in the form

$$\Gamma = (D, d^*, T_1, T_2, u_1, u_2, p_1, p_2)$$

$$\hat{\Gamma} = (\hat{D}, d^*, T_1, T_2, \hat{u}_1, \hat{u}_2, p_1, p_2),$$

where $\hat{D} \supseteq D$ and $\hat{u}_i(d,t) = u_i(d,t)$ for every (d,t) in $D \times T$. That is, $\hat{\Gamma}$ extends Γ iff the two bargaining problems differ only in that more decisions are available in $\hat{\Gamma}$.

For bargaining problems with complete information, Nash's axiom of independence of irrelevant alternatives (IIA) defines a relationship between the solutions of one problem and the solutions of its extensions. In this paper, we shall consider the following condition, which generalizes a stronger version of Nash's IIA axiom.

Axiom 2 (Extension) Suppose that μ is an incentive-efficient mechanism for a bargaining problem Γ. Suppose also that there exist bargaining problems $\{\Gamma^k\}_{k=1}^{\infty}$ and mechanisms $\{\mu^k\}_{k=1}^{\infty}$ such that each Γ^k is an extension of Γ, each μ^k is in the solution set $S(\Gamma^k)$, and

$$\limsup_{k\to\infty} U_i^k(\mu^k|t_i) \leq U_i(\mu|t_i), \qquad \forall i \in \{1,2\}, \ \forall t_i \in T_i.$$

Then μ must be in the solution set $S(\Gamma)$.

When the hypothesis of this axiom applies, it means that there are ways to increase the set of decision options available to the players (without changing the conflict outcome) such that both players would be willing to settle for expected utility payoffs that are arbitrarily close to what they can get from the feasible mechanism μ. Thus, following the argument of Nash's IIA axiom, the players ought to be willing to settle for the mechanism μ even when these extra decision options are not available.

Our definition of an "extension" of a bargaining problem is more restrictive than that used by Harsanyi and Selten [1972] in their IIA axiom, so our extension axiom is weaker than their IIA axiom. (Actually, their IIA axiom allows no sequential approximation, so the two axioms are not quite comparable. But their solution would also satisfy a stronger version of their IIA axiom that would be strictly stronger than our extension axiom.) Essentially, Harsanyi and Selten considered that a bargaining problem $\hat{\Gamma}$ "extends" Γ iff Γ and $\hat{\Gamma}$ have the same type sets and the same probability distributions over types, and

$$\{\hat{U}(\mu) \mid \mu \text{ is feasible in } \hat{\Gamma}\} \supseteq \{U(\mu) \mid \mu \text{ is feasible in } \Gamma\}.$$

Any extension $\hat{\Gamma}$ in our sense would satisfy this definition, but there are many other bargaining problems that give a larger set of feasible utility allocations than Γ, but which cannot be constructed from Γ by adding new decision-options to the set D.

5. The Random-Dictatorship Axiom

Let us consider again the "Divide the Dollars" game, in which two risk-neutral players can divide \$100 among themselves, provided that they can agree on the division. If player 1 were a "dictator" with all of the bargaining ability, then he could insist on essentially all of the money (or a 99% share). After all, player 2 would be better off with any infinitesimal share than with the zero that he gets in the conflict outcome. On the other hand, if player 2 had all of the bargaining ability, then he could insist on essentially all of the money.

Now, when the two players have equal bargaining ability, one equitable solution would be to randomize equally between these two dictatorial outcomes. Indeed, this plan of giving each player an equal chance of getting the entire $100 is a Nash bargaining solution for this game. Since the players are both risk-neutral, this random-dictatorship plan is equivalent to splitting the money equally, in that each player gets the same expected payoff of $50.

More generally, consider any two-person bargaining problem with complete information (no uncertainty yet). Let x_1 be the highest expected utility that player 1 can achieve, subject to the constraint that player 2 gets his conflict payoff of zero. Similarly, let x_2 be the highest expected utility that player 2 can achieve, subject to the constraint that player 1 gets his conflict payoff of zero. A random dictatorship that gives each player an equal chance of enforcing his best outcome, subject to the constraint that the other gets zero, would give expected utility $\frac{1}{2}x_i$ to each player i. If this utility allocation $(\frac{1}{2}x_1, \frac{1}{2}x_2)$ is Pareto-efficient then it is the Nash solution of the bargaining problem.

This last observation is very important. Its proof is that the random dictatorship is Pareto-efficient iff the Pareto frontier is a straight line from $(x_1,0)$ to $(0,x_2)$ in utility space, as in the "Divide the Dollars" game. In such games, the Nash bargaining solution always selects the midpoint of this line.

Now let us return to Bayesian bargaining problems with incomplete information. In this case, it may not be obvious how to even define a "random dictatorship" mechanism, because it may not be obvious what decision or mechanism a player would demand if he had all of the bargaining ability. For example it may be that each type of player 1 has a different incentive-compatible mechanism that would be optimal for it; and it may happen that none of these "potentially optimal" mechanisms for player 1 would be incentive compatible if player 2 could infer player 1's type from the fact that he demanded this mechanism. In such cases player 1 would have to demand a mechanism that seemed to be a fair compromise between the alternative goals of his possible types, so as to prevent player 2 from learning 1's type, even if player 1 had all of the bargaining ability. What such a "fair compromise" should be is a subtle question, and has been analyzed in detail in Myerson [1981].

However, there are some Bayesian bargaining problems in which there is a clear decision that each player should demand, if he could have all of the bargaining ability. We will now focus our attention on a class of such problems.

We may say that a mechanism μ always implements a decision d iff $\mu(d|t) = 1$ for every t in T. We say that a decision d is incentive-efficient iff the mechanism that always implements d is incentive-efficient.

A decision e_1 in D is a strongly optimal decision for player 1 iff e_1 is incentive-efficient and $u_2(e_1,t) = 0$ for every t in T. Similarly, a strongly optimal decision for player 2 is any decision in D that is incentive-efficient and gives zero utility to player 1 in all states.

Two important facts about these "strong optima" will justify the name.

First, there can be at most one strongly optimal decision for a player, up to equivalence in utility.

<u>Proposition 1</u>. If e_1 and $\hat{e}_1$ are both strongly optimal decisions for player 1 in the bargaining problem Γ, then $u_1(e_1,t) = u_1(\hat{e}_1,t)$ for all t.

<u>Proof</u>. If not, let μ be the mechanism that systematically selects e_1 or $\hat{e}_1$, depending on which one is better for player 1. Then this mechanism is incentive compatible, since player 1 gets the decision that he prefers and player 2 always gets zero utility no matter what he reports. But μ dominates e_1 and $\hat{e}_1$ as mechanisms, unless e_1 and $\hat{e}_1$ are utility-equivalent. Since strong optima are incentive-efficient, the equivalence asserted in Proposition 1 must hold. Q.E.D.

Our second proposition shows why a player would have to demand his strongly optimal decision, if there were such a decision and he had all of the bargaining ability. When S_1 is any nonempty subset of T_1, we say that a mechanism μ is <u>feasible given S_1</u> iff satisfies the probability constraints (1.2), is incentive compatible and individually rational for player 1 in the usual sense (that is, μ satisfies (2.1) and (2.2) for i=1), and would be incentive compatible and individually rational for player 2 <u>after</u> he inferred that 1's type was in S_1. That is, to be feasible given S_1, μ must satisfy:

$$(4.1) \qquad \sum_{t_1 \varepsilon S_1} \sum_{d \varepsilon D} p_2(t_1|t_2)\, \mu(d|t)\, u_2(d,t)$$

$$\geq \sum_{t_1 \varepsilon S_1} \sum_{d \varepsilon D} p_2(t_1|t_2)\, \mu(d|t_1,r_2)\, u_2(d,t), \qquad \forall t_2 \varepsilon T_2,\ \forall r_2 \varepsilon T_2;$$

and

chance at being allowed to demand any feasible mechanism, constrained only by the requirements of incentive-compatibility and individual-rationality. If each player has a strongly optimal decision e_i, then a random dictatorship will be equivalent to a .50-.50 randomization between e_1 and e_2. A random dictatorship is certainly equitable (since each player gets an equal chance to control the decision), but it is not necessarily incentive-efficient. However, if the randomization between e_1 and e_2 is an incentive-efficient mechanism, then it is a fair and efficient bargaining solution, and should be included in our solution set $S(\Gamma)$. The following axiom summarizes this conclusion.

<u>Axiom 3</u>. (<u>Random dictatorship</u>) In a bargaining problem Γ, suppose that there exist strongly optimal decisions e_1 and e_2 for players 1 and 2 respectively. Suppose also that the mechanism $\bar{\mu}$ defined by

$$\bar{\mu}(e_1|t) = \bar{\mu}(e_2|t) = \frac{1}{2}, \qquad \forall t \in T$$

is incentive-efficient. Then $\bar{\mu} \in S(\Gamma)$.

It is important to recognize that the hypotheses of this axiom are quite restrictive, which makes this a fairly weak axiom. In many bargaining problems, there may be no strongly optimal decisions (so that the outcomes of a random dictatorship may be hard to predict), or else the randomization between the strongly optimal decisions may not be efficient. This axiom only states that, in cases where the random dictatorship is well-understood and incentive-efficient, then it should be considered a fair bargaining solution.

$$(4.2) \qquad \sum_{t_1 \varepsilon S_1} \sum_{d \varepsilon D} p_2(t_1|t_2)\, \mu(d|t)\, u_2(d,t) \geq 0, \qquad \forall t_2 \varepsilon T_2.$$

<u>Proposition 2</u> Suppose that e_1 is a strongly optimal decision for player 1, and μ is any mechanism. Let S_1 be the set of all types of player 1 that prefer μ over e_1; that is,

$$S_1 = \{t_1 | U_1(\mu|t_1) > \sum_{t_2 \varepsilon T_2} p_1(t_2|t_1)\, u_1(e_1,t)\}.$$

If S_1 is nonempty then μ is <u>not</u> feasible given S_1.

<u>Proof</u>. This is essentially a special case of Theorem 1 in Myerson [1981]. The basic idea is that, if μ were feasible given S_1, then we could construct another feasible mechanism that would dominate e_1, by implementing e_1 if $t_1 \notin S_1$, and implementing μ otherwise. But e_1 is incentive-efficient, and so cannot be dominated. Q.E.D.

Suppose e_1 is strongly optimal for player 1. Then, if player 1 tried to demand some other incentive-compatible mechanism μ, and if player 2 inferred from this demand that player 1 must be in the set of types that prefer μ over e_1, then μ could not be feasible given this information: some type of player 2 would have an incentive either to lie or to insist on the conflict outcome. Thus, if a player has a strongly optimal decision, and he is given an opportunity to demand any feasible mechanism, then he cannot do better than to demand his strong optimum, no matter what his type is. Any other demand would be self-defeating, because of the information that it would reveal. (For more on this issue, see Myerson [1981]. Our strongly optimal decisions here are just a special subclass of the "strong solutions" discussed in that paper.)

Now consider the random dictatorship, in which each player gets a 50%

6. Definition of neutral bargaining solutions.

There are many solution correspondences which satisfy the probability-invariance, extension, and random-dictatorship axioms. For example, letting $S(\Gamma)$ be the set of all incentive-efficient mechanisms would satisfy all three axioms. Our goal as theorists is to find a theory of bargaining that determines the smallest possible set of solutions, so as to get strongest possible predictions. So we would like to know how small a solution set we can define, for each bargaining problem, and still satisfy these two axioms. We would like to know which mechanisms must be contained in any solution correspondence that satisfies these axioms.

We formally define a neutral bargaining solution of a two-person Bayesian bargaining problem Γ to be any mechanism μ with the property that, for every solution correspondence $S(\cdot)$ that satisfies the probability-invariance, extension, and random-dictatorship axioms, μ is in $S(\Gamma)$. We let $\bar{S}(\Gamma)$ be the set of all neutral bargaining solutions of Γ. That is, if we let $\underset{\sim}{H}$ denote the set of all solution correspondences that satisfy our three axioms then

$$(6.1) \qquad \bar{S}(\Gamma) = \bigcap_{S \in \underset{\sim}{H}} S(\Gamma).$$

We can now state three of our main results.

Theorem 1. As a solution correspondence, $\bar{S}(\cdot)$ itself satisfies the three axioms: probability-equivalence, extension, and random-dictatorship.

Theorem 2. For any two-person Bayesian bargaining problem Γ, $\bar{S}(\Gamma) \neq \emptyset$.

<u>Theorem 3.</u> If Γ is a bargaining problem with complete information (in that T_1 and T_2 contain only one type each) then $\mu \varepsilon \bar{S}(\Gamma)$ if and only if μ is a Nash bargaining solution of Γ.

Theorem 1 is easy to check, because any intersection of correspondences that satisfy the three axioms will satisfy these axioms as well. To prove Theorem 2 and 3, we will need a more useful characterization of the neutral bargaining solutions in $\bar{S}(\Gamma)$, as will be developed in the next two sections. The proofs are thus deferred to Section 11.

7. The Primal and Dual Problems

Let us now consider a fixed Bayesian bargaining problem Γ, as in (1.1) To simplify our notation we let

$$\Omega = \mathbb{R}^{T_1} \times \mathbb{R}^{T_2},$$

and we let Ω_+ and Ω_{++} denote the nonnegative and strictly positive orthants of Ω respectively.

Given any vector $\lambda = \left((\lambda_i(t_i))_{t_i \varepsilon T_i}\right)_{i \varepsilon \{1,2\}}$ in Ω_+, we define the <u>primal problem</u> for λ to be the optimization problem

$$(7.1) \qquad \underset{\mu}{\text{maximize}} \sum_{i=1}^{2} \sum_{t_i \varepsilon T_i} \lambda_i(t_i)\, U_i(\mu|t_i)$$

subject to (1.2) and (2.1).

That is, the primal problem for λ is to find an incentive-compatible mechanism that maximizes the λ-weighted sum of the expected utilities of all types of both players. With D and T assumed to be finite sets, the primal problem is just a linear programming problem.

These primal problems are important because every incentive-efficient mechanism must be an optimal solution of the primal problem for some λ in Ω_{++}. This fact follows from the supporting-hyperplane theorem, and the fact that the set of all utility allocations $U(\mu)$ generated by incentive-compatible mechanisms is a closed convex polyhedron in Ω. (We have omitted the individual-rationality constraints (2.2) in the primal problem for λ, because any incentive-efficient mechanism will still not be dominated, in the sense of (2.3), when these constraints are dropped.)

Let us now formulate the dual of problem (7.1). We let $\alpha_i(s_i|t_i)$ denote the dual variable (or shadow price) corresponding to the incentive constraint (2.1), which asserts that player i should not expect to gain from reporting his type as s_i, if t_i is his true type. We let $\underset{\sim}{A}$ denote the set of all vectors $\alpha = \left(\alpha_i(s_i|t_i)\right)_{i\varepsilon\{1,2\},s_i\varepsilon T_i,t_i\varepsilon T_i}$ such that

$$\alpha_i(s_i|t_i) \geq 0 \quad \forall i\varepsilon\{1,2\},\ \forall s_i\varepsilon T_i,\ \forall t_i\varepsilon T_i,$$

$$\text{and}\quad \alpha_i(t_i|t_i) = 0 \quad \forall i\varepsilon\{1,2\},\ \forall t_i\varepsilon T_i.$$

If we multiply the incentive constraints (2.1) by their dual variables and add them into the objective function of (7.1), we get the following Lagrangian function

$$(7.2)\qquad \sum_{i=1}^{2} \sum_{t_i\varepsilon T_i} \lambda_i(t_i)\, U_i(\mu|t_i) + \sum_i \sum_{t_i} \sum_{s_i} \alpha_i(s_i|t_i)\left(U_i(\mu|t_i) - U_i^*(\mu,s_i|t_i)\right)$$

Substituting in the formulas for $U_i(\mu|t_i)$ and $U_i^*(\mu,s_i|t_i)$ it is straightforward to show that (7.2) is equal to

$$(7.3) \qquad \sum_{t\in T} \sum_{d\in D} \sum_{i=1}^{2} \mu(d|t)\, V_i(d,t,\lambda,\alpha)$$

where

$$(7.4) \qquad V_i(d,t,\lambda,\alpha) = \Big(\big(\lambda_i(t_i) + \sum_{s_i\in T_i} \alpha_i(s_i|t_i)\big)\, p_i(t_{-i}|t_i)\, u_i(d,t)$$

$$- \sum_{s_i\in T_i} \alpha_i(t_i|s_i)\, p_i(t_{-i}|s_i)\, u_i(d,(t_{-i},s_i))\Big).$$

Formula (7.4) will be very important. We will refer to $V_i(d,t,\lambda,\alpha)$ as player i's virtual evaluation of decision d in state t, with respect to λ and α.

By standard Lagrangian analysis, we know that an incentive-compatible mechanism μ will be an optimal solution of the primal for λ iff there exists some α in $\underset{\sim}{A}$ such that

if $U_i(\mu|t_i) > U_i^*(\mu,s_i|t_i)$ then $\alpha_i(s_i|t_i) = 0$, $\forall i\varepsilon\{1,2\}$, $\forall t_i\varepsilon T_i$, $\forall s_i\varepsilon T_i$,

and μ maximizes the Lagrangian (7.3) subject only to the probability constraints (1.2). Clearly, the Lagrangian is maximized by putting all probability weight, in each $\mu(\cdot|t)$ distribution, on the decisions that maximize the sum of the two players' virtual evaluations.

The appropriate vector α for use in this Lagrangian analysis is the vector that solves the dual of (7.1). This dual problem for λ can be written as follows:

$$(7.5) \qquad \underset{\alpha\varepsilon \underset{\sim}{A}}{\text{minimize}} \sum_{t\varepsilon T} \underset{d\varepsilon D}{\text{maximum}} \sum_{i=1}^{2} V_i(d,t,\lambda,\alpha).$$

Each $V_i(d,t,\lambda,\alpha)$ is linear in α, so this dual problem is a linear programming problem.

8. Characterization theorems

We now have the machinery to state our first main characterization theorem.

Theorem 4 μ is a neutral bargaining solution in $\bar{S}(\Gamma)$ if and only if μ is an incentive-efficient mechanism and there exist sequences $\{\lambda^k\}_{k=1}^{\infty}$, $\{\alpha^k\}_{k=1}^{\infty}$, and $\{\omega^k\}_{k=1}^{\infty}$ such that:

(8.1) $\lambda^k \varepsilon\ \Omega_{++}$, $\alpha^k \varepsilon \underset{\sim}{A}$, and $\omega^k \varepsilon \Omega_+$, $\forall k$;

(8.2) $$\Big(\lambda_i^k(t_i) + \sum_{s_i \varepsilon T_i} \alpha_i^k(s_i|t_i)\Big)\ \omega_i^k(t_i) - \sum_{s_i \varepsilon T_i} \alpha_i^k(t_i|s_i)\ \omega_i^k(s_i)$$
$$= \sum_{t_{-i} \varepsilon T_{-i}} \max_{d \varepsilon D} \sum_{j=1}^{2} V_j(d,t,\lambda^k,\alpha^k)/2\ ,\quad \forall i \varepsilon \{1,2\},\ \forall t_i \varepsilon T_i,\ \forall k;$$

and

(8.3) $$\limsup_{k \to \infty} \omega_i^k(t_i) \leq U_i(\mu|t_i), \qquad \forall i \varepsilon \{1,2\},\ \forall t_i \varepsilon T_i.$$

We defer the proof of Theorem 4 to Section 11. Instead, we shall devote this and the next section to analysis of these conditions, to give a better understanding of their significance.

Since (8.1) and (8.2) are linearly homogenous in λ^k and α^k, and since each λ^k has strictly positive components, we can assume without loss of generality that the (λ^k, α^k) are normalized so that $\|\lambda^k\| + \|\alpha^k\| = 1$. Then these sequences lie in a simplex and have a convergent subsequence. By (8.3), the ω^k are also bounded and have a convergent subsequence. So we can assume without loss of generality that the $(\lambda^k,\alpha^k,\omega^k)$ sequences are convergent to some limits (λ,α,ω), where the limiting vectors λ and α are not both zero.

Now (8.2) and (8.3) imply that

(8.4) $$\sum_i \sum_{t_i} \lambda_i(t_i)\ U_i(\mu|t_i) \geq \sum_i \sum_{t_i} \lambda_i(t_i)\ \omega_i(t_i) = \sum_t \max_{d \varepsilon D} \sum_{j=1}^{2} V_j(d,t,\lambda,\alpha).$$

Since μ is feasible in the primal for λ, and α is feasible in the dual for λ, (8.4) implies that μ and λ are also optimal solutions of the primal and dual respectively. Furthermore, by duality theory, we must have equality in (8.4), so the following complementarity condition holds for every type t_i of each player:

$$(8.5) \qquad \omega_i(t_i) = U_i(\mu|t_i) \quad \text{or} \quad \lambda_i(t_i) = 0.$$

To summarize, we have proven the following theorem, which gives the most tractable conditions for computing neutral bargaining solutions.

Theorem 5. If μ is a neutral bargaining solution in $\bar{S}(\Gamma)$ then μ is incentive-efficient and there exist λ in Ω_+, α in $\underset{\sim}{A}$, and ω in Ω_+ such that

(8.6) $\qquad (\lambda,\alpha) \neq (\underset{\sim}{0},\underset{\sim}{0})$;

(8.7) $\qquad \mu$ is an optimal solution of the primal problem for λ;

(8.8) $\qquad \alpha$ is an optimal solution of the dual problem for λ;

$$(8.9) \qquad \Big(\lambda_i(t_i) + \sum_{s_i \varepsilon T_i} \alpha_i(s_i|t_i)\Big)\, \omega_i(t_i) - \sum_{s_i \varepsilon T_i} \alpha_i(t_i|s_i)\, \omega_i(s_i)$$

$$= \sum_{t_{-i} \varepsilon T_{-i}} \max_{d \varepsilon D} \sum_{j=1}^{2} V_j(d,t,\lambda,\alpha)/2\,, \qquad \forall i \varepsilon \{1,2\}, \quad \forall t_i \varepsilon T_i; \quad \text{and}$$

$$(8.10) \qquad \omega_i(t_i) \leq U_i(\mu|t_i), \qquad \forall i \varepsilon \{1,2\}, \qquad \forall t_i \varepsilon T_i.$$

Futhermore, λ and ω must satisfy the complementarity condition (8.5) for every t_i.

If these conditions are satisfied with λ in Ω_{++} then we can satisfy the conditions of Theorem 4 with the constant sequences $(\lambda^k,\alpha^k,\omega^k) = (\lambda,\alpha,\omega)$. Thus, we get the following theorem.

Theorem 6. If λ, α, ω, and μ together satisfy conditions (8.7)-(8.10), and $\lambda \varepsilon \Omega_{++}$ (that is, all $\lambda_i(t_i)$ are strictly positive) then μ is a neutral bargaining solution in $\bar{S}(\Gamma)$.

Notice that the conditions (8.7)-(8.10) form a well-determined system, in the sense that there are as many equations as variables. Given any λ, the primal problem (7.1) determines μ, the dual problem (7.5) determines α, and equation (8.9) determines ω. (In fact, (8.9) is a nonsingular system of equations for ω if $\lambda \varepsilon \Omega_{++}$.) Then (8.10) and the resulting complementarity condition (8.5), give us $|T_1|+|T_2|$ equations, which are enough equations to determine λ. This suggests a conjecture that the set of neutral bargaining solutions may be generically finite.

The interpretation of equation (8.9) still needs to be developed. Summing the equations of (8.9) over t_i in T_i and applying (8.5) we get

$$\sum_{t_i \varepsilon T_i} \lambda_i(t_i)\, U_i(\mu|t_i) = \sum_{t_i \varepsilon T_i} \lambda_i(t_i)\, \omega_i(t_i)$$

$$= \sum_{t \varepsilon T} \max_{d \varepsilon D} \sum_{j=1}^{2} V_j(d,t,\lambda,\alpha)/2 .$$

Since the last expression is independent of i, we get

$$(8.11) \qquad \sum_{t_1 \varepsilon T_1} \lambda_1(t_1)\, U_1(\mu|t_1) = \sum_{t_2 \varepsilon T_2} \lambda_2(t_2)\, U_2(\mu|t_2).$$

Thus, (8.9) gives us an equity condition between the two players, namely, that the weighted sums of their possible conditionally expected utilities should be equal. However, (8.9) is a stronger condition than (8.11). (For example, one can show that any Harsanyi-Selten solution would satisfy (8.7) and (8.11) for some nonzero vector λ, even though it would not generally satisfy (8.9).) As was discussed at the end of Section 1, a bargaining solution concept must specify criteria for equitable compromise between the different possible types of each player, as well as between two players. The equations of (8.9) define these intertype-equity conditions in our solution theory.

Extending the terminology of Myerson [1981], we say that the vector ω is warranted by λ and α, and $\omega_i(t_i)$ is the warranted claim of type t_i, iff ω satisfies (8.9) for λ and α. Then (8.10) says that μ gives to each type at least its warranted claim in expected utility. In the next section we will try to show in what sense it may be appropriate to consider such $\omega_i(t_i)$ as equitable or "warranted" claims.

9. The Virtual Bargaining Problem

In this section, we will assume that the players' probability distributions p_1 and p_2 in the bargaining problem Γ are consistent with a common prior distribution p, and that $p(t) > 0$ for every t in T.

Let μ be an incentive-efficient mechanism for the bargaining problem Γ. Let λ be a vector in Ω_{++} such that μ is an optimal solution of the primal for λ, and let α be an optimal solution of the dual for λ.

The virtual bargaining problem with respect to λ and α is defined as follows. The sets of types T_1 and T_2 and the probability distributions p_1 and p_2 are the same as in Γ, but the utility functions are v_1 and v_2, defined by the formula

(9.1) $$v_i(d,t) = V_i(d,t,\lambda,\alpha)/p(t), \quad \forall d \in D, \quad \forall t \in T,$$

where V_i is as in (7.4). We may refer to v_i as the virtual utility function of player i. In the analysis of this virtual bargaining problem, we will assume that it has two other properties different from the original bargaining problem Γ. The players' types are assumed to be verifiable in the virtual bargaining problem, so that there are no incentive constraints. And the virtual utility is assumed to be transferable between the two players in the virtual bargaining problem, as if the virtual payoff were in money.

One reason for considering this virtual bargaining problem is that our original mechanism μ is the obvious agreement for the players to reach in this game. This is because, with verifiable types and sidepayments, the players should simply agree to choose their decision, in each state, so as to maximize the sum of their payoffs, which can later be redistributed according to any other standard of equity. Since μ is an optimal solution of the primal for λ, for any state t, $\mu(d|t)$ only puts positive probability on those decisions d that maximize

$$V_1(d,t,\lambda,\alpha) + V_2(d,t,\lambda,\alpha) = p(t)\left(v_1(d,t) + v_2(d,t)\right)$$

So μ does systematically maximize the sum of the players' virtual-utility payoffs in every state.

Now consider what would be an equitable agreement in the virtual bargaining problem, if the players' virtual-utility payoffs were transferable? The virtual bargaining problem is very simple: there are no incentive constraints, the players are paid in transferable units like money, and the virtual payoffs for the conflict outcome are always zero. So an obviously equitable agreement is that each of the players should get half of the total virtual payoff available in each state. The expected virtual utility for type t_i of player i in this equitable allocation would be

$$(9.2)\qquad x_i(t_i) = \sum_{t_{-i}\varepsilon T_{-i}} p_i(t_{-i}|t_i) \max_{d\varepsilon D} (v_1(d,t) + v_2(d,t))/2$$

$$= \sum_{t_{-i}\varepsilon T_i} \max_{d\varepsilon D} \sum_{j=1}^{2} V_j(d,t,\lambda,\alpha)/(2p^i(t_i)),$$

where $p^i(t_i)$ is the marginal probability of type t_i of player i in the common prior. We may say that a mechanism is virtually equitable if it gives an expected virtual utility to each type t_i equal to its equitable allocation $x_i(t_i)$.

We may now ask, what allocations of real utility could correspond to this equitable allocation of virtual utility? Suppose that $\hat{\mu}$ is a mechanism that is virtually equitable and incentive compatible. Then

$$(9.3)\qquad x_i(t_i) = \sum_{t_{-i}\varepsilon T_{-i}} \sum_{d\varepsilon D} \hat{\mu}(d|t)\; p_i(t_{-i}|t_i)\; v_i(d,t)$$

$$= \sum_{t_{-i}\varepsilon T_{-i}} \sum_{d\varepsilon D} \hat{\mu}(d|t)\; V_i(d,t,\lambda,\alpha)/p^i(t_i)$$

$$= \Big(\big(\lambda_i(t_i) + \sum_{s_i} \alpha_i(s_i|t_i)\big)\, U_i(\hat{\mu}|t_i) - \sum_{s_i} \alpha_i(t_i|s_i)\, U_i^*(\hat{\mu},t_i|s_i)\Big)/p^i(t_i)$$

$$\geq \Big(\big(\lambda_i(t_i) + \sum_{s_i} \alpha_i(s_i|t_i)\big)\, U_i(\hat{\mu}|t_i) - \sum_{s_i} \alpha_i(t_i|s_i)\, U_i(\hat{\mu}|s_i)\Big)/p^i(t_i)$$

(The third equality here just follows from the definition of V_i (7.4). The final inequality uses incentive-compatibility of $\hat{\mu}$.) Notice that the inequality in (9.3) would be an equality if $\hat{\mu}$ were an optimal solution of the primal for λ, by compementary slackness of dual optima. Now suppose that ω is a vector of warranted claims satisfying (8.9) for λ and α. Then (9.2) and (9.3) imply

$$(9.4)\qquad \Big(\lambda_i(t_i) + \sum_{s_i \in T_i} \alpha_i(s_i|t_i)\Big)\, U_i(\hat{\mu}|t_i) - \sum_{s_i \in T_i} \alpha_i(t_i|s_i)\, U_i(\hat{\mu}|s_i)$$

$$\leq \sum_{t_{-i} \in T_{-i}} \max_{d \in D} \sum_{j=1}^{2} V_i(d,t,\lambda,\alpha)/2$$

$$= \Big(\lambda_i(t_i) + \sum_{s_i \in T_i} \alpha_i(s_i|t_i)\Big)\, \omega_i(t_i) - \sum_{s_i \in T_i} \alpha_i(t_i|s_i)\, \omega_i(s_i).$$

Since this inequality holds for every t_i, straightforward algebra implies (see Lemma 1 in Myerson [1981]) that

$$(9.5)\qquad U_i(\hat{\mu}|t_i) \leq \omega_i(t_i), \quad \forall i \in \{1,2\}, \quad \forall t_i \in T_i,$$

with equality if $\hat{\mu}$ is an optimal solution of the primal for λ.

Thus we can at last interpret the warrant equations (8.9) in the characterization theorems. Equation (8.9) implicitly defines ω to be the highest vector of expected utilities that the various types of the two players could get in any virtually-equitable mechanism that is incentive-compatible in Γ (or in any extension of Γ for which μ is still an optimal solution of the primal for λ). The condition (8.10) asserts that, to be a neutral bargaining solution, the mechanism μ must give every type of each player a real expected utility that is at least as large as what it could get in a virtually-equitable mechanism.

One of the basic difficulties we face in understanding cooperation under uncertainty is that a mechanism which is efficient within the set of all incentive-compatible mechanisms might not be efficient ex post, after the players learn each others' types. In order to satisfy the constraints of incentive compatibility, it may be necessary to accept a positive probability of an outcome that is bad for both players. For example, in union-management negotiations, if the management is of the "type" that cannot pay high wages then, to prove its type, it might have to accept a positive probability of a costly strike that neither side wants. Otherwise, if they simply used some strikeless mechanism in which the wage agreement was increasing in the managements' ability to pay, the management type with high ability would always pretend to be the type with low ability. But it may be difficult to understand how the players can commit themselves to implement a mechanism with a strike of any duration, since managements' type (low ability to pay) is revealed as soon as the strike begins, and then both sides would prefer to settle at a low wage.

Virtual utility gives us one way to explain how the players might implement an incentive-efficient mechanism μ that may be inefficient ex post. In the heat of bargaining, as each player feels the pressure of his incentive-compatibility constraints (in that he has difficulty getting the other player to trust him), he might begin to act as if he wants to maximize his virtual utility, rather than his real utility. That is, instead of saying that the incentive constraints (2.1) force the players to accept ex post inefficiency, we may say that the incentive constraints force the players to transform their effective preferences from the real to the virtual utility functions. We may refer to this idea as the <u>virtual utility hypothesis.</u> The incentive-efficient mechanism μ maximizes the sum of virtual utilities in every state, so it will be efficient ex post in terms of virtual utilities.

We may say that one type s_i <u>jeopardizes</u> another type t_i of player i, in the efficient mechanism μ, iff the constraint that says s_i should not gain by claiming to be t_i (i.e., $U_i(\mu|s_i) \geq U_i^*(\mu,t_i|s_i)$) is binding in the primal for λ and its shadow price $\alpha_i(t_i|s_i)$ is positive. Then the virtual utility of type t_i differs from the actual utility of type t_i in that the virtual utility exaggerates the difference from the types that jeopardize t_i. That is, equations (7.4) and (9.1) construct the virtual utility of type t_i as a positive multiple of the actual utility of type t_i minus a multiple of the "false" utilities of types that jeopardize t_i. So the virtual-utility hypothesis of the preceding paragraph may be restated as follows: when incentive constraints are binding in a bargaining process, then a player in one type may begin to act according to virtual preferences which exaggerate the difference from the other types that he needs to distinguish himself from.

The argument leading up to (9.5) showed that this virtual-utility hypothesis is more than just a convenient way to describe how the players can come to agree on a mechanism that is ex post inefficient. It is also embodied in the warrant equations (8.2) and (8.9) that characterize the neutral bargaining solutions. If the players made interpersonal-equity comparisons in terms of virtual utility, then each type t_i could justly demand half of the expected virtual utility that t_i contributes by cooperating (that is $x_i(t_i)$ from (9.2)). But (8.9) and (8.10) guarantee that a neutral bargaining solution either satisfies these fair virtual demands (if Theorem 6 applies), or gives no less expected (real) utility to each type of each player than he could get from any mechanism that does satisfy these fair virtual demands.

10. Example

Let us consider an example which was first discussed in Myerson [1979a]. In this example, the two players can jointly build a road which both would use, and which would cost \$100 to build. The road is commonly known to be worth \$90 to player 2; but its value to player 1 on depends on his type, which is unknown to player 2. If 1's type is 1h ("high") then the road is also worth \$90 to him, and player 2 assigns subjective probability 0.9 to this event. However, if 1's type is 1ℓ ("low") then the road is only worth \$30 to him, and player 2 assigns probability 0.1 to this event. The problem is to decide whether the road should be built, and if so, how much each player should pay.

To formally model this problem, we let $T_1 = \{1h, 1\ell\}$, $T_2 = \{2\}$, $D = \{d_0, d_1, d_2\}$, $p(1h) = 0.9$, $p(1\ell) = 0.1$, with the utility functions as follows:

(u_1,u_2)	d_0:	d_1:	d_2:
t_1=1h:	(0,0)	(-10,90)	(90,-10)
$t_1=1\ell$:	(0,0)	(-70,90)	(30,-10)

Since player 2 has only one possible type, we may ignore the t_2 variable throughout this analysis. The decision options in D are interpreted as follows:

d_0 is the decision not to build the road;

d_1 is the decision to build the road at 1's expense; and

d_2 is the decision to build the road at 2's expense.

The conflict outcome is $d^* = d_0$. There is no need to consider intermediate financing options, because they can be represented by "randomized" strategies (assuming that both players are risk-neutral). For example, letting $\mu(d_1|t_1) = 0.4$ and $\mu(d_2|t_1) = 0.6$ is equivalent to building the road and charging \$40 to player 1 and \$60 to player 2 in state t_1.

It can be shown (see Myerson [1979a]) that the set of incentive-efficient utility allocations satisfying individual rationality is a triangle in Ω_+, with extreme points $(U_1(\mu|1h), U_1(\mu|1\ell); U_2(\mu))$ as follows:

$$(80,20;0), \quad (60,0;20), \quad (0,0;72).$$

The first of these allocations is implemented by having player 1 pay \$10 and player 2 pay \$90 for the road independently of the state, or by using the mechanism μ_1 where

$$\mu_1(d_1|t_1) = 0.1, \quad \mu_1(d_2|t_1) = 0.9, \quad \forall t_1.$$

The second of these allocations is implemented by having player 1 pay \$30 and player 2 pay \$70 independently of the state, or by using μ_2 where

$$\mu_2(d_1|t_1) = 0.3, \quad \mu_2(d_2|t_1) = 0.7, \quad \forall t_1.$$

The third of these allocations is implemented by having player 1 pay \$90 and player 2 pay \$10 for the road if 1's type is 1h, and by not building the road if 1's type is 1ℓ; or by using the mechanism μ_3 where

$$\mu_3(d_1|1h) = 0.9, \quad \mu_3(d_2|1h) = 0.1, \quad \mu_3(d_0|1\ell) = 1.0 .$$

Notice that μ_1 above is the best feasible mechanism for both types of player 1, and μ_3 is the best feasible mechanism for player 2. (Mechanism μ_2 is best for player 2 among feasible mechanisms that guarantee that the road will be built, but player 2 gets a higher expected utility from μ_3 because his subjective probability of type 1ℓ is so small.) Thus, a random dictatorship would implement the mechanism $\mu_4 = 0.5\mu_1 + 0.5\mu_3$, that is

$$\mu_4(d_1|1h) = 0.5, \quad \mu_4(d_2|1h) = 0.5,$$

$$\mu_4(d_1|1\ell) = 0.05, \quad \mu_4(d_2|1\ell) = 0.45, \quad \mu_4(d_0|1\ell) = 0.5 .$$

If 1's type is 1h then μ_4 is equivalent to building the road and having both players pay \$50 each; if 1's type is 1ℓ then μ_4 is equivalent to having a 50% probability of not building the road, and a 50% probability of building the road with 1 paying \$10 and 2 paying \$90. Although there is a 5% probability ex ante that the road will not be built in μ_4 $(p(1\ell)\ \mu_4(d_0|1\ell) = 0.05)$, this mechanism is incentive-efficient, with

$$U_1(\mu_4|1h) = 40, \quad U_1(\mu_4|1\ell) = 10, \quad U_2(\mu_4) = 36.$$

Thus, the argument justifying the random-dictatorship axiom also suggests that

μ_4 should be a bargaining solution for this game.

In fact μ_4 is the neutral bargaining solution for this game. To check this, notice first that all the incentive-efficient mechanisms satisfy

$$\frac{13}{15}\, U_1(\mu|1h) + \frac{2}{15}\, U_1(\mu|1\ell) + U_2(\mu) = 72,$$

so that all incentive-efficient mechanisms must be optimal solutions of the primal for λ, where

$$\lambda_1(1h) = \frac{13}{15}, \quad \lambda_1(1\ell) = \frac{2}{15}, \quad \lambda_2 = 1.$$

The unique optimal solution of the dual for λ (7.5) is

$$\alpha_1(1\ell|1h) = \frac{1}{30}, \quad \alpha_1(1h|1\ell) = 0.$$

That is, type 1h jeopardizes 1ℓ, because the only problematical incentive constraint is that player 1 should not have any incentive to claim that his valuation for the road is low (\$30) if it is actually high (\$90). With λ and α as above, the virtual utility functions are (by (9.1) and (7.4))

$$v_1(d,1h) = (\frac{13}{15} + \frac{1}{30})\, u_1(d,1h)/(0.9) = u_1(d,1h),$$

$$v_1(d,1\ell) = \left(\frac{2}{15}\, u_1(d,1\ell) - \frac{1}{30}\, u_1(d,1h)\right)/(0.1) = \frac{4}{3}\, u_1(d,1\ell) - \frac{1}{3}\, u_1(d,1h),$$

$$v_2(d,t_1) = u_2(d,t_1), \qquad \forall t_1 \varepsilon T_1, \quad \forall d \varepsilon D.$$

That is, the virtual utility functions are as follows:

(v_1,v_2)	d_0:	d_1:	d_2:
t_1=1h:	(0,0)	(-10,90)	(90,-10)
t_1=1ℓ:	(0,0)	(-90,90)	(10,-10)

Recalling that d_1 has 1 paying \$100 for the road, and d_2 has 2 paying \$100 for the road, we see that these virtual utilities differ from the real utilities only in that 1's virtual valuation for the road is \$10 (instead of \$30) when his type is 1ℓ. This low virtual valuation for 1ℓ exaggerates type 1ℓ's difference from type 1h, which jeopardizes 1ℓ. When 1's type is 1ℓ, the players' total virtual valuation for the road just equals its cost (\$10+\$90 = \$100), so that players can randomize between building the road or not in an incentive-efficient mechanism. If the road is built when 1's type is 1ℓ, then virtual equity would require that 1 pay \$10 and 2 pay \$90, so that each pays his virtual valuation. When 1's type is 1h, building the road and having each player pay \$50 is efficient and equitable in both the real and virtual utility scales. Type 1h is just indifferent between getting the road at a (personal) cost of \$50, and getting the road at a cost of \$10 with probability 0.5 (his expected gain is \$40 in either case), so 1h can jeopardize 1ℓ. Thus μ_4 is the unique virtually-equitable incentive-efficient mechanism.

Formally, the warrant equations are

$$\left(\frac{13}{15} + \frac{1}{30}\right) \omega_1(1h) = \frac{1}{2}\big((0.9)(90 - 10)\big)$$

$$\frac{2}{15}\,\omega_1(1\ell) - \frac{1}{30}\,\omega_1(1h) = 0$$

$$\omega_2 = \frac{1}{2}\big((0.9)(90 - 10) + 0\big)$$

and so $\omega_1(1h) = 40 = U_1(\mu_4|1h)$, $\omega_1(1\ell) = 10 = U_1(\mu_4|1\ell)$, $\omega_2 = 36 = U_2(\mu_4)$. Then, by Theorem 6, μ_4 is a neutral bargaining solution.

11. Proofs

We begin with a key lemma.

Lemma 1. Suppose that $e_1 \varepsilon D$, $e_2 \varepsilon D$, and $u_2(e_1,t) = 0 = u_1(e_2,t)$, $\forall t \varepsilon T$. Let $\bar{\mu}$ be the mechanism defined by

$$\bar{\mu}(e_1|t) = \bar{\mu}(e_2|t) = .5 \ , \quad \forall t \varepsilon T.$$

Then $\bar{\mu}$ is incentive-efficient if and only if there exist vectors λ in Ω_{++} and α in $\underset{\sim}{A}$ such that

$$\Big(\lambda_i(t_1) + \sum_{s_i \varepsilon T_i} \alpha_i(s_i|t_i)\Big) U_i(\bar{\mu}|t_i) - \sum_{s_i \varepsilon T_i} \alpha_i(t_i|s_i) U_i(\bar{\mu}|s_i)$$

$$= \sum_{t_{-i} \varepsilon T_{-i}} \max_{d \varepsilon D} \sum_{j=1}^{2} V_j(d,t,\lambda,\alpha)/2, \qquad \forall i \varepsilon \{1,2\}, \ \forall t_i \varepsilon T_i.$$

Furthermore, if $\bar{\mu}$ is incentive-efficient then each e_i is a strongly optimal decision for player i.

Proof. If $\bar{\mu}$ satisfies the equations in the lemma for some λ and α, then summation implies

$$\sum_{i=1}^{2} \sum_{t_i \varepsilon T_i} \lambda_i(t_i) U_i(\bar{\mu}|t_i) = \sum_{t \varepsilon T} \max_{d \varepsilon D} \sum_{j=1}^{2} V_j(d,t,\lambda,\alpha) \ .$$

Thus, since $\bar{\mu}$ is feasible in the primal problem for λ, and α is feasible in the dual problem for λ, both $\bar{\mu}$ and α must be optimal solutions in their respective problems. Thus $\bar{\mu}$ must be incentive-efficient.

Conversely, suppose that $\bar{\mu}$ is incentive-efficient. Then there exists some λ in Ω_{++} such that $\bar{\mu}$ is an optimal solution of the primal for λ. Let α be any optimal solution of the dual for λ. By duality, $\bar{\mu}$ must put weight only on decisions that maximize the sum of the virtual evaluations. Thus, for every t in T,

$$V_1(e_1,t,\lambda,\alpha) = V_2(e_2,t,\lambda,\alpha) = \max_{d \in D} \sum_{j=1}^{2} V_j(d,t,\lambda,\alpha).$$

(We use here the fact that $V_2(e_1,t,\lambda,\alpha) = 0 = V_1(e_2,t,\lambda,\alpha) \quad \forall t.$) Thus, e_1 and e_2 are incentive-efficient as mechanisms (as they also solve the primal for λ), which implies the last sentence of the lemma. Furthermore, by definition of V_i and U_i ((7.4) and (1.3)),

$$\Big(\lambda_i(t_i) + \sum_{s_i} \alpha_i(s_i|t_i)\Big) U_i(\bar{\mu}|t_i) - \sum_{s_i} \alpha_i(t_i|s_i) U_i(\bar{\mu}|s_i)$$

$$= \sum_{t_{-i} \in T_{-i}} V_i(e_i,t,\lambda,\alpha)/2, \qquad \forall i,\ \forall t_i \in T_i,$$

which implies that $\bar{\mu}$ satisfies the equations in the lemma. Q.E.D.

We prove Theorem 4 first and then return to prove Theorems 2 and 3.

Proof of Theorem 4. (Characterization of neutral bargaining solutions.)

We show first that the set of incentive-efficient mechanisms for which (8.1)-(8.3) can be satisfied constitutes a solution concept that satisfies the three axioms. The definitions of V_j and U_i both depend only on the $p_i u_i$ product, so this solution concept satisfies the probability-invariance axiom. By the Lemma, this solution concept satisfies the random-dictatorship axiom. To check the extension axiom suppose that, for each k, there is some extension Γ^k in which a mechanism μ^k satisfies these conditions (8.1)-(8.3) in Γ^k and such that

$$U_i^k(\mu^k|t_i) \leq U_i(\mu|t_i) + 1/k, \quad \forall i,\ \forall t_i \in T_i.$$

Select λ^k in Ω_{++} and α^k in $\underset{\sim}{A}$ and $\hat{\omega}^k$ so that $\hat{\omega}_i^k(t_i) \leq U_i^k(\pi^k|t_i) + 1/k$ for all i and t_i, and so that the warrant equations (8.2) for λ^k and α^k are satisfied in Γ^k. (That is, use the maximum over d in the extended decision domain D^k, instead of D.) Now let ω^k satisfy these same warrant equations in Γ; that is, with the maximum only over d in D, instead of over d in D^k. It is easy to show (see Lemma 1 of Myerson [1981]) that the warrant equations (8.2) have a unique solution ω^k, and that it is increasing in the right-hand sides. Thus (since a maximum over D is less than or equal to a maximum over D^k) we must have $\omega_i^k(t_i) \leq \hat{\omega}_i^k(t_i)$ for every i and t_i. This sequence $\{\lambda^k,\alpha^k,\omega^k\}_k$ satisfies (8.1)-(8.3) for μ in Γ. So the set of incentive-efficient mechanisms for which (8.1)-(8.3) can be satisfied does obey the extension axiom, as a solution concept.

Because $\bar{S}$ is the smallest correspondence satisfying the three axioms, these conditions (8.1)-(8.3) can be satisfied for every neutral bargaining solution in $\bar{S}(\Gamma)$.

Conversely, suppose that conditions (8.1)-(8.3) are satisfied in Γ for the incentive-efficient mechanism μ, together with some sequences $\{\lambda^k,\alpha^k,\omega^k\}_{k=1}^{\infty}$. We define quantities $y_i^k(t)$ so that

(11.1)
$$\Big(\lambda_i^k(t_i) + \sum_{s_i\varepsilon T_i} \alpha_i^k(s_i|t_i)\Big)\, y_i^k(t) - \sum_{s_i\varepsilon T_i} \alpha_i^k(t_i|s_i)\, y_i^k(t_{-i}, s_i)$$

$$= \max_{d\varepsilon D} \sum_{j=1}^{2} V_j(d,t,\lambda^k,\alpha^k)\ , \qquad \forall k,\ \forall i\varepsilon\{1,2\},\ \forall t\varepsilon T.$$

As long as $\lambda^k \varepsilon \Omega_{++}$, these equations have a unique solution (again, see Lemma 1 of Myerson [1981]). Assume that $p_i(t_{-i}|t_i) > 0$ for every i and t.

Then let us define Γ^k as an extension of Γ with two additional decision-options e_1 and e_2 such that

$$p_1(t_2|t_1)\ u_1^k(e_1,t) = y_1^k(t), \qquad u_2^k(e_1,t) = 0,$$

$$p_2(t_1|t_2)\ u_2^k(e_2,t) = y_2^k(t), \qquad u_1^k(e_2,t) = 0.$$

Let μ^k be the mechanism that randomizes equally between these two decisions e_1 and e_2 in the extension Γ^k. Then summing (11.1) and using the fact that ω^k is the unique solution to (8.2) (given λ^k and α^k), we conclude that $U_i^k(\mu^k|t_i) = \omega_i^k(t_i)$ for all i and t_i. Thus

$$\limsup_{k\to\infty} U_i^k(\mu^k|t_i) \leq U_i(\mu|t_i), \quad \forall i\varepsilon\{1,2\}, \quad \forall t_i \varepsilon T_i.$$

But by the Lemma and the random-dictatorship axiom, each $\mu^k \varepsilon \bar{S}(\Gamma^k)$. Therefore, by the extension axiom, $\mu \varepsilon \bar{S}(\Gamma)$. That is, any incentive-efficient mechanism μ that satisfies (8.1)-(8.3) for some sequence is a neutral bargaining solution.

In the preceding paragraph, we needed to assume that all $p_i(t_{-i}|t_i) > 0$, to construct $u_i^k(e_i,t)$. However, for any problem Γ that violates this assumption, there are other problems $\hat{\Gamma}$ (as in (3.2)) that satisfy this assumption, and are equivalent in terms of the probability-invariance axiom, and have the same set of incentive-efficient mechanisms satisfying (8.1)-(8.3). So by probability-invariance, (8.1)-(8.3) imply $\mu \varepsilon \bar{S}(\Gamma)$, for any Bayesian bargaining problem Γ. (This is the only use made of the probability-invariance in these proofs.) Q.E.D.

Proof of Theorem 2. (Existence of neutral bargaining solutions.)

We begin with some definitions. For any k larger than $|T_1|+|T_2|$, let

$$\Lambda^k = \{\lambda \in \Omega_+ \mid \sum_{i=1}^{2} \sum_{t_i \in T_i} \lambda_i(t_i) = 1,\ \lambda_i(t_i) \geq 1/k,\ \forall i,\ \forall t_i\}.$$

We let Λ denote the unit simplex in Ω_+, containing these sets Λ^k, and let F denote the set of all incentive-compatible mechanisms for Γ.

There exists a compact convex set $\underset{\sim}{A}^*$ such that $\underset{\sim}{A}^* \subseteq \underset{\sim}{A}$ and, for every λ in Λ, there is some α in $\underset{\sim}{A}^*$ such that α is an optimal solution of the dual for λ. To prove this fact, observe first that the feasible set of the primal for λ is compact and independent of λ. So the unit simplex Λ can be covered by a finite collection of sets (each corresponding to the range of optimality of one basic feasible solution in the primal) such that, within each set, an optimal solution of the dual can be given as a linear function of λ. Each of these linear functions is bounded on the compact unit simplex Λ, so we can choose $\underset{\sim}{A}^*$ to contain the union of the ranges of these linear functions on Λ.

Let $B = \max_i \max_{t_i} \max_{\mu \in F} |U_i(\mu|t_i)|+1.$

Let $X = \{\omega \in \Omega \mid\ 0 \leq \omega_i(t_i) \leq B,\quad \forall i,\ \forall t_i \in T_i\}.$

For each k greater than $|T_1|+|T_2|$, we now define a correspondence

$$Z^k\colon\ F \times \underset{\sim}{A}^* \times X \times \Lambda^k \Rightarrow F \times \underset{\sim}{A}^* \times X \times \Lambda^k$$

so that $(\hat{\mu},\hat{\alpha},\hat{\omega},\hat{\lambda}) \in Z^k(\mu,\alpha,\omega,\lambda)$ iff the following conditions are satisfied

(11.2) $\hat{\mu}$ is an optimal solution of the primal for λ;

(11.3) $\hat{\alpha}$ is an optimal solution of the dual for λ;

(11.4) $\hat{\omega}_i(t_i) = \min\{\tilde{\omega}_i(t_i),\ B\}$,

where $\tilde{\omega}$ is the unique solution to the equations

(11.5) $$\Big(\lambda_i(t_i) + \sum_{s_i} \alpha_i(s_i|t_i)\Big)\, \tilde{\omega}_i(t_i) - \sum_{s_i} \alpha_i(t_i|s_i)\, \tilde{\omega}_i(s_i)$$

$$= \sum_{t_{-i}} \max_d \sum_{j=1}^{2} V_j(d,t,\lambda,\alpha)/2\ , \quad \forall i\varepsilon\{1,2\},\ \forall t_i \varepsilon T_i; \quad \text{and}$$

(11.6) $\hat{\lambda}_i(t_i) = 1/k$ for every t_i such that

$$\omega_i(t_i) - U_i(\mu|t_i) < \max_{j,s_j} \ (\omega_j(s_j) - U_j(\mu|s_j)).$$

Lemma 1 of Myerson [1981] guarantees that $\tilde{\omega}$ is uniquely determined by (11.5), and that each $\tilde{\omega}_i(t_i) \geqslant 0$ because each

$$\max_d \sum_{j=1}^{2} V_j(d,t,\lambda,\alpha) \geqslant \sum_{j=1}^{2} V_j(d^*,t,\lambda,\alpha) = 0.$$

Thus $\hat{\omega}$ is in X. Condition (11.6) asserts that $\hat{\lambda}$ should put as much weight as possible (within Λ^k) on types for which $\omega_i(t_i)$ most exceeds $U_i(\mu|t_i)$.

By the Kakutani Fixed Point Theorem, for each k there exists some $(\mu^k,\alpha^k,\omega^k,\lambda^k)$ such that

$$(\mu^k,\alpha^k,\omega^k,\lambda^k)\ \varepsilon\ Z^k(\mu^k,\alpha^k,\omega^k,\lambda^k).$$

There also exists a convergent subsequence of these fixed points, converging to some $(\bar{\mu},\bar{\alpha},\bar{\omega},\bar{\lambda})$ in the compact set $F \times \underset{\sim}{A}^* \times X \times \Lambda$. We now show that $\bar{\mu}$ is in $\bar{S}(\Gamma)$.

Let $\tilde{\omega}^k$ be the vector determined by (11.5) for λ^k and α^k. Summing (11.5) and using duality theory, we get

$$(11.7) \qquad \sum_{i=1}^{2} \sum_{t_i} \lambda_i^k(t_i)\, \tilde{\omega}_i^k(t_i) = \sum_t \max_d \sum_{j=1}^{2} V_j(d,t,\lambda^k,\alpha^k)$$

$$= \sum_{i=1}^{2} \sum_{t_i} \lambda_i^k(t_i)\, U_i(\mu^k|t_i).$$

For any t_i, if $U_i(\mu^k|t_i) > \tilde{\omega}_i^k(t_i)$ then $\tilde{\omega}_i^k(t_i) = \omega_i^k(t_i)$ and $\lambda_i^k(t_i) = 1/k$, because by (11.7) there must be some type s_j for which $\tilde{\omega}_j^k(s_j) \geqslant U_j(\mu^k|s_j)$ and so $\omega_j^k(s_j) \geqslant U_j(\mu^k|s_j)$. So for any type t_i in $T_1 \cup T_2$,

$$\limsup_{k \to \infty} \lambda_i^k(t_i)(U_i(\mu^k|t_i) - \tilde{\omega}_i^k(t_i)) \leqslant 0,$$

because the $\lambda_i^k(t_i)$ coefficient must go to zero whenever $U_i(\mu^k|t_i)-\tilde{\omega}_i^k(t_i)$ is positive, and this term is bounded above by B.

Now suppose that there were some i and t_i such that

$$\limsup_{k \to \infty} (\tilde{\omega}_i^k(t_i) - U_i(\mu^k|t_i)) > 0.$$

Any such t_i would also satisfy

$$\bar{\omega}_i(t_i) - U_i(\bar{\mu}|t_i) = \lim_{k\to\infty} (\omega_i^k(t_i) - U_i(\mu^k|t_i)) > 0$$

by (11.4). Then by (11.6), we could choose such a t_i that also satisfies

$$\bar{\lambda}_i(t_i) = \lim_{k\to\infty} \lambda_i^k(t_i) > 0.$$

But this would imply

$$0 < \limsup_{k \to \infty} \lambda_i^k(t_i)(\tilde{\omega}_i^k(t_i) - U_i(\mu^k|t_i))$$

$$= \limsup_{k \to \infty} \sum_{(j,s_j)\neq(i,t_i)} \lambda_j^k(s_j)(U_j(\mu^k|s_j) - \tilde{\omega}_j^k(s_j)) \leqslant 0,$$

which is impossible.

Thus, for every i and t_i,

$$\limsup_{k\to\infty} \tilde{\omega}_i^k(t_i) \leq U_i(\bar{\mu}|t_i) = \lim_{k\to\infty} U_i(\mu^k|t_i).$$

and so $\{\lambda^k,\alpha^k,\tilde{\omega}^k\}_k$ satisfy the conditions of Theorem 4 for $\bar{\mu}$. (Also, $\tilde{\omega}^k$ must equal ω^k for all sufficiently large k, since the B-bound in (10.4) is not binding.) The mechanism $\bar{\mu}$ is incentive compatible, but not necessarily incentive-efficient. However, there must exist an incentive-efficient mechanism μ such that $U_i(\mu|t_i) \geq U_i(\bar{\mu}|t_i)$ for all i and t_i. Then all the conditions in Theorem 4 are also satisfied for μ, so $\mu \varepsilon \bar{S}(\Gamma)$. Q.E.D.

<u>Proof of Theorem 3. (Extension of Nash solution.)</u>

If there is no mechanism giving both players strictly more than zero, in a bargaining problem with complete information, then there is only one efficient individually-rational utility allocation, which must correspond to the Nash bargaining solution and the neutral bargaining solution. So let us assume that Γ is a bargaining problem with complete information in which both players can get positive payoffs together.

Then μ (now just a randomization over D) is a Nash bargaining solution iff there exist nonnegative numbers λ_1 and λ_2, not both zero, such that μ maximizes $\lambda_1 U_1 + \lambda_2 U_2$, and $\lambda_1 U_1(\mu) = \lambda_2 U_2(\mu)$. Under our positivity assumption, these conditions can only be satisfied when both λ_1 and λ_2 are strictly positive.

Since each player has only one possible type, there are no incentive constraints, and the vector α must be just (0,0). (Recall the definition of $\underset{\sim}{A}$ in Section 7.) So the conditions (8.6)-(8.10) in Theorems 5 and 6 reduce to:

$(\lambda_1,\lambda_2) \neq (0,0)$, with $\lambda_1 \geq 0$, $\lambda_2 \geq 0$;

μ maximizes $\lambda_1 U_1 + \lambda_2 U_2$;

$$\lambda_i \omega_i = \frac{1}{2} \max_d \sum_{j=1}^{2} \lambda_j U_j(\mu), \quad \forall i;$$

$\omega_i \leq U_i(\mu)$, $\forall i$.

These conditions can only be satisfied when each $\lambda_i > 0$ and each $\omega_i = U_i(\mu)$, because of our positivity assumption on Γ. Thus the necessary and sufficient conditions in Theorems 5 and 6 both coincide with the conditions for a Nash bargaining solution. Q.E.D.

REFERENCES

Aumann, R. J. Maschler [1967], "Repeated Games with Incomplete Information: A Survey of Recent Results," in _Models of Gradual Reduction of Arms_ (Mathematica, Princeton, N. J.), 289-403.

Harsanyi, J. C. [1967-8], "Games with Incomplete Information Played by 'Bayesian' Players," _Management Science 14_, 159-189, 320-334, 486-502.

Harsanyi, J. C. and R. Selten [1972], "A Generalized Nash Bargaining Solution for Two-Person Bargaining Games with Incomplete Information," _Management Science 18_, P80-P106.

Holmström B., and R. B. Myerson [1981], "Efficient and Durable Decision Rules with Incomplete Information," Northwestern University discussion paper, to appear in _Econometrica._

Myerson, R. B. [1976], _A Theory of Cooperative Games_, Ph.D. dissertation, Harvard University.

Myerson, R. B. [1979a], "Incentive Compatibility and the Bargaining Problem," _Econometrica 47_, 61-73.

Myerson, R. B. [1979b], "An Axiomatic Derivation of Subjective Probability, Utility and Evaluation Functions," _Theory and Decision 11_, 339-352.

Myerson, R. B. [1981], "Mechanism Design by an Informed Principal," Northwestern University discussion paper, to appear in _Econometrica_.

Myerson, R. B. [1982], "Cooperative Games with Incomplete Information," Northwestern University discussion paper, to appear in _International Journal of Game Theory._

Nash, J. F. [1950], "The Bargaining Problem," _Econometrica 18_, 155-162.

Nash, J. F. [1953], "Two-Person Cooperative Games," _Econometrica 21_, 128-140.