Discussion Paper No. 432

SOLUTIONS FOR TWO-PERSON BARGAINING PROBLEMS WITH INCOMPLETE INFORMATION

by

Roger B. Myerson*

July, 1980

J.L. Kellogg Graduate School of Management
Northwestern University

Abstract. A new solution concept is derived axiomatically for two-person Bayesian bargaining problems with incomplete information. The solution factors through the feasibility graph, which is the graph of a correspondence specifying which utility allocations are feasible as a function of the probabilities of cooperation in each state of the players' information. In the case of complete information, the feasibility graph just specifies the feasible set of the bargaining problem, and our solution reduces to the Nash bargaining solution.

*Research for this paper was supported by the J.L. Kellogg Center for Advanced Study in Managerial Economics.

Solutions for Two-Person Bargaining Problems with Incomplete Information

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 an 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 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 to games with incomplete information.

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 [1967-8]. Formally, a two-person Bayesian bargaining problem is an object of the form

(1.1) $(C,c^*,T_1,T_2\ ,\ U_1,U_2,P_1,P_2)$

whose components are interpreted as follows. C is the set of collective choices or feasible outcomes available to the two players if they cooperate, and $c^* \varepsilon C$ is the conflict outcome which the players must get by default if they fail to cooperate. For each player i (i=1,2), T_i is the set of possible types for player i. That is, each $t_i \varepsilon T_i$ represents a complete description of player i's relevant characteristics: his preferences, beliefs, and endowments. T_1 and T_2 are disjoint sets. Each U_i is a function from $C \times T_1 \times T_2$ into the real numbers, such that $U_i(c,t_1,t_2)$ is the payoff which player i would get if c in C 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 $U_i(c^*,t_1,t_2) = 0$ for all i, t_1, t_2. P_1 and P_2 are the conditional probability distributions which each player assesses over the other players' type. That is, $P_1(t_1,t_2)$ 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. For mathematical simplicity, we shall assume that C, T_1, and T_2 are finite sets; throughout most of this paper.

The players in a bargaining problem do not have to agree on a specific outcome in C, instead they may agree on some choice mechanism, which is a contract specifying how the choice should depend on the players' types. Since we will allow randomized strategies, a choice mechanism is here defined to be any real-valued function π on the domain $C \times (T_1 x T_2)$ such that

$$(1.2) \qquad \sum_{c' \varepsilon C} \pi(c'|t_1,t_2) = 1 \text{ and } \pi(c|t_1,t_2) \geq 0,$$
$$\forall c \varepsilon C, \forall t_1 \varepsilon T_1, \forall t_2 \varepsilon T_2.$$

That is, $\pi(c|t_1,t_2)$ is the probability of choosing outcome c in the mechanism π , if t_1 and t_2 are the players' types.

Since the players can agree on a choice 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 choice c" if he is type t_1 and saying "I demand choice $\hat{c}$ " if he is type $\hat{t}_i$, he can say "I demand a mechanism with $\pi(c|t_1,t_2)=1$ and $\pi(\hat{c}|\hat{t}_1,t_2)=1$" 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 choice mechanism is agreed upon. One might call this the poker-face assumption.

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 poker-face requirement can create a new kind of dilemma for a player, because it can easily happen that player 1 knowing only his own type would be indifferent between two choice mechanisms π and $\hat{\pi}$ (for any type t_1), while player 2 might prefer π over $\hat{\pi}$ if t_2 is his type and $\hat{\pi}$ over π if $\hat{t}_2$ is his type. So player 2 would prefer to argue for π if t_2 is true and for $\hat{\pi}$ if $\hat{t}_2$ is true; but such a policy would certainly reveal information to player 1, which could destroy player 1's indifference between π and $\hat{\pi}$. For example, 1 might be indifferent between betting that 2 can or cannot speak French until 1 learns that 2 wants to bet that he can.

Thus, each player must be careful to use a bargaining strategy which maintains a balance between the conflicting goals which he would have if he were of different types, 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 Choice 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 C, 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. Nevertheless, we shall find it convenient to treat both of these two cases separately in this paper. Thus, to completely define a Bayesian bargaining problem, we must add to the structures in (1.1) a specification as to whether the players' types are verifiable or unverifiable.

To simplify our notation, we let T denote the set of all possible type-pairs $t=(t_1,t_2)$; that is:

$$T = T_1 \times T_2.$$

Given any choice mechanism π satisfying (1.2), we let $\bar{U}_i(\pi|t_i)$ denote conditionally expected utility for player i, given that he is of type t_i, if the mechanism π is implemented. That is, for any $i\varepsilon\{1,2\}$ and any $t_i \varepsilon\ T_i$,

$$(2.1) \qquad \bar{U}_i(\pi|t_i) = \sum_{t_{-i}\varepsilon T_{-i}} \sum_{c\varepsilon C} P_i(t)\ U_i(c,t)\ \pi(c|t)$$

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 can now formally define the set of feasible choice mechanisms under each of the two assumptions about verifiability.

For a Bayesian bargaining problem with unverifiable types, we say that a choice mechanism π: C x T → $\mathbb{R}$ is feasible if it satisfies the following conditions:

(2.2) $\sum_{c \varepsilon C} \pi(c|t) = 1, \forall t \varepsilon T,$

(2.3) $\pi(c|t) \geq 0, \forall t \varepsilon T, \forall c \varepsilon C,$

and

(2.4) $\bar{U}_i(\pi|t_i) \geq \sum_{t_{-i} \varepsilon T_{-i}} \sum_{c \varepsilon C} P_i(t)\, U_i(c,t)\, \pi(c|t_{-i},s_i),$

$\forall i \varepsilon \{1,2\}, \forall t_i \varepsilon T_i, \forall s_i \varepsilon T_i.$

Conditions (2.2) and (2.3) simply repeat (1.2), asserting that $\pi(.|t)$ must be a valid probability distribution over C, for any types pair t.

Condition (2.4) is an incentive-compatibility condition. It says that, if player i is of type t_i, then his expected utility $\bar{U}_i(\pi|t_i)$ from participating honestly in mechanism π cannot be less than his expected utility from pretending to be of any other type s_i. That is, (2.4) asserts that honest participation in the choice mechanism π is a Nash equilibrium for the two players. If (2.4) 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 π could not be implemented. (It can be shown that even dishonest equilibria or equilibrium behavior in more general mechanisms cannot achieve any expected utility allocations which are not also achieved by incentive-compatible mechanisms satisfying (2.4); see Myerson [1979a], for example. Thus there is no loss of generality in restricting our attention to such incentive-compatible direct revelation mechanisms.)

For a Bayesian bargaining problem with verifiable types, we say that a choice mechanism π is feasible iff it satisfies conditions (2.2) and (2.3). That is, only the probability conditions are required in the case of verifiable types. With verifiable types, one can compel players to reveal their types honestly, without incentive-compatibility, so (2.4) can be dropped.

Given any mechanism π, we let $\bar{U}(\pi)$ denote the vector of all $\bar{U}_i(\pi|t_i)$ conditionally expected utility levels for each player, given each of his possible types. That is,

$$\bar{U}(\pi) = ((\bar{U}_i(\pi|t_i))_{t_i \varepsilon T_i})_{i\varepsilon\{1,2\}},$$

so $\bar{U}(\pi)$ is a vector with $|T_1| + |T_2|$ components.

If player i is of type t_i, then he would prefer mechanism $\hat{\pi}$ over π if and only if $\bar{U}_i(\hat{\pi}|t_i) > \bar{U}_i(\pi|t_i)$. Thus, an arbitrator could be sure that both players prefer $\hat{\pi}$ over π only if

(2.5) $\quad \bar{U}_i(\hat{\pi}|t_i) > \bar{U}_i(\pi|t_i) \qquad \forall i\varepsilon \{1,2\}, \forall t_i \varepsilon T_i.$

We say that a feasible mechanism π is efficient iff there does not exist any feasible mechanism $\hat{\pi}$ such that (2.5) holds. (Whenever we refer to a mechanism as "feasible", it is understood to be in the sense appropriate to the problem, that is, satisfying (2.2)-(2.3) in the verifiable case, and satisfying (2.2)-(2.4) in the unverifiable case.)

Notice that our definition of efficiency implicitly makes $\bar{U}(\pi)$ 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 $U_i(\pi|t_i)$ numbers, for all possible t_i. Even if the players bargain without the help of an arbitrator, all of the components of $\bar{U}(\pi)$ may be significant in determining whether mechanism π is chosen (not just the components corresponding to the two true 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.

Notice that (2.1)-(2.4) are all linear in π, so the set of feasible mechanisms (in either the verifiable or unverifiable case) is compact and convex. Furthermore, by the Separating Hyperplane Theorem, a mechanism π is efficient iff there exists some vector $\lambda = (\lambda_{t_i})_{t_i \varepsilon T_1 \cup T_2}$ in $\mathbb{R}^{T_1 \cup T_2}$ such that

$$(2.6) \qquad \lambda_{t_i} \geq 0 \; \forall t_i \varepsilon T_1 \cup T_2 \, , \quad \sum_{i=1}^{2} \sum_{t_i \varepsilon T_i} \lambda_{t_i} = 1,$$

and such that π maximizes

$$(2.7) \qquad \sum_{i=1}^{2} \sum_{t_i \varepsilon T_i} \lambda_{t_i} U_i(\pi | t_i)$$

over all feasible mechanisms. Thus, the problem of finding all efficient mechanisms is a parametric linear programming problem: to maximize (2.7) subject to (2.2)-(2.3) (in the verifiable case) or (2.2)-(2.4) (in the unverifiable case), as the vector λ varies over the range defined by (2.6). We shall see more of these linear programs in Section 6.

3. The probability-invariance axiom

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

$$(3.1) \qquad [\prod_{t_1 \varepsilon T_1} \bar{U}_1(\pi | t_1)^{P(t_1)}] \, [\prod_{t_2 \varepsilon T_2} \bar{U}_2(\pi | t_2)^{P(t_2)}]$$

over the set of all feasible mechanisms (although they defined the set of feasible mechanisms somewhat differently from in this paper). In formula (3.1), $P(t_i)$ denotes the marginal probability (as would be assessed ex ante by an outside observer) that player i is type t_i. 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:

Probability-invariance: Consider any two Bayesian bargaining problems $(C,c^*,T_1,T_2,U_1,U_2,P_1,P_2)$ and $(C,c^*,T_1,T_2,\hat{U}_1,\hat{U}_2,\hat{P}_1,\hat{P}_2)$ having the same choice sets, type sets, and conflict outcome, and with the same assumptions about type-verifiability. Suppose that $P_i(t)\ U_i(c,t) = \hat{P}_i(t)\ \hat{U}_i(c,t)$ for every i, c in C and t in $T_1 x\ T_2$. Then these two bargaining problems must have the same solutions.

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 π generates the same vector $\bar{U}(\pi)$ of conditionally expected utilities in both problems.

In effect, the probability-invariance axiom states that probabilities cannot be meaningfully defined separately 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, for example, that there is no loss of generality in considering only problems in which the two players' types are stochastically independent provided that $U_i(c,t)$ is allowed to depend on both components of t in any arbitrary way. In Myerson [1976], this axiom was extended to n-person dynamic games, in which the probabilities of some players' types may depend on the choices of earlier players. The general probability-invariance axiom can be used to reduce any dynamic problem, in which $\hat{P}_i(t|c)$ depends on c, to an equivalent static problem in which all players' types are independent.

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. (In fact, our Bayesian bargaining problems do not even specify unconditional marginal probabilities for types, although we could have easily revised the definitions in Section 1 to include such a specification.) 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 axioms.

4. The feasibility graph.

Given a Bayesian bargaining problem, the set of all feasible conditionally-expected utility allocations is

$$(4.1) \qquad F(1) = \{\bar{U}(\pi) \mid \pi \text{ is a feasible mechanism}\}$$

(Recall that "feasible" may mean either satisfying (2.2)-(2.3) or satisfying (2.2)-(2.4), depending on the assumptions about type-verifiability. The significance of the "1" in F(1) will become evident shortly.) The set F(1) is a closed convex subset of $\mathbb{R}^{T_1 \cup T_2}$. The efficient mechanisms are the ones which give utility allocations on the upper boundary of F(1). Nash [1950] assumed that the bargaining solution should depend only on the set of feasible utility allocations and on the conflict playoffs (here normalized to zero). Extending this idea, Harsanyi and Selten [1972] assumed that their bargaining solution concept must be defined on the set of feasible utility allocations in $\mathbb{R}^{T_1 \cup T_2}$, together with a specification of all the types' marginal probabilities. That is, they assumed that the set F(1) carries all relevant information about the relative power of each type of each player, so that a fair allocation on the upper boundary of F(1) can be chosen only with reference to F(1) and the vector of margnal probabilities. Since the probability-invariance axiom disallows using the marginal probability vector, we must find some way to extract more information about the structure of the bargaining problem than is contained in F(1).

To develop our ideas, it will be useful to also consider bargaining problems with transferable utility and free disposal of utility, even though such problems cannot be formally modelled in the format of (1.1), without an infinite choice set C. (With bounds on the transfers and disposals, such problems could be modelled with finite C, however.)

Let us consider two examples. In Example 1, $T_1 = \{1a, 1b\}$ and $T_2 = \{2\}$. The two types of player 1 are equally likely, and they are verifiable. Both players measure utility in dollars, and the players can transfer money between themselves by sidepayments. If the two players cooperate then they can get \$100 together from an outside source, paid to player 2, who can transfer any part to player 1. Players 1's type has nothing to do with their ability to get money, but the players could still agree to transfer different amounts of money depending on 1's type, since it is verifiable. Thus, the set of feasible utility allocations (with free disposal) is:

$$F(1) = \{\omega \in \mathbb{R}^{T_1 \cup T_2} | \frac{1}{2}\omega_{1a} + \frac{1}{2}\omega_{1b} + \omega_2 \leq 100\}$$

In Example 2 everything is the same as in Example 1, except that now the players can get \$200 together if 1's type is 1a, but they cannot earn any money together if 1's type is 1b. So type 1b has nothing to contribute to player 2, but their ex ante expected income is still \$100. The feasible set F(1) for this example is the same as for Example 1. To see this, observe that, when they cooperate and divide the available income, giving ω_{1a} to 1 if t_1=1a and giving ω_{1b} to 1 if t_1=1b, then 2's expected payoff is $\omega_2 = (.5)(200 - \omega_{1a}) + (.5)(0 - \omega_{1b}) = 100 - (.5)(\omega_{1a} + \omega_{1b})$.

In both of these examples, the Harsanyi-Selten solution would be $(\omega_{1a}, \omega_{1b}, \omega_2) = (50,50,50)$. That is, player 1 must get \$50 whether he is type 1a or type 1b. This seems like a reasonable solution in Example 1, since both types of 1 contribute equally to 2's ability to get the outside money. But in Example 2, this solution is not so reasonable; instead we might expect player 1 to demand \$100 (half of \$200) if t_1=1a, and \$0 if t_1=1b, so that $(\omega_{1a}, \omega_{1b}, \omega_2) = (100, 0, 50)$ is the utility allocation. After all, player 1 can always prove his type, and type 1b has no power to contribute anything to player 2. In the solution theory to be developed in this paper, (50,50,50) will be the unique solution for Example 1, and (100,0,50) will be the unique solution for Example 2.

To distinguish between Examples 1 and 2, we need more information than is contained in F(1). We need to use the fact that, if player 1 refused to cooperate when he is of type 1b, then player 2 would lose some ability to earn money in Example 1, but he would lose nothing in Example 2.

Returning to the general Bayesian bargaining problem, we let a participation vector be any vector q in $[0,1]^T$. (Recall $T = T_1 \times T_2$.) That is, $q = (q_t)_{t\varepsilon T}$ is a participation vector if each q_t is a number between 0 and 1, to be interpreted as a probability that the two players will cooperate if $t = (t_1,t_2)$ is their vector of types. For any q in $[0,1]^T$, let

$$(4.2) \qquad F(q) = \left\{ \bar{U}(\pi) \;\middle|\; \begin{array}{l} \pi \text{ is a feasible mechanism,} \\ \text{and } \pi(c^*|t) \geq 1-q_t,\ \forall t\varepsilon T \end{array} \right\}$$

That is, F(q) is a subset of $\mathbb{R}^{T_1 \cup T_2}$, representing the set of all utility allocations which can be achieved using mechanisms which would still be feasible even if the players were only going to be available for cooperation with probability q_t when t is the vector of types. When $q = \underset{\sim}{1} = (1,\ldots,1)$ (4.2) reduces to (4.1). With no danger of confusion, we let F denote the graph of this feasibility correspondence; that is,

$$(4.3) \qquad F = \{(q,\omega) \mid q\varepsilon[0,1]^T \text{ and } \omega\varepsilon F(q)\}.$$

We shall refer to F as the feasibility graph of the Bayesian bargaining problem.

In our two examples, we get (for any $q\varepsilon(0,1]^T$)

$$F(q) = \{\omega \mid \tfrac{1}{2}\,\omega_{1a} + \tfrac{1}{2}\,\omega_{1b} + \omega_2 \leq \tfrac{1}{2}\,(100q_{(1a,2)} + 100q_{(1b,2)})\}$$

in Example 1, and we get

$$F(q) = \{\omega \mid \tfrac{1}{2}\,\omega_{1a} + \tfrac{1}{2}\,\omega_{1b} + \omega_2 \leq \tfrac{1}{2}\,(200q_{(1a,2)} + 0)\}$$

in Example 2. Thus, the feasibility graph F does distinguish these two examples, even though the simple feasible set $F(\underset{\sim}{1})$ does not.

When there is no uncertainty ($|T|=1$), the feasibility graph reduces to

$$F = \{(\alpha,\alpha\omega) \mid 0 \leq \alpha \leq 1 \text{ and } \omega\varepsilon F(\underset{\sim}{1})\},$$

so the feasibility graph F can be entirely derived from the feasible set

F(1). Thus, F contains no more information than F(1) when $|T|=1$. But when there is proper uncertainty, then F cannot be derived from F(1), as our examples have shown.

In this paper, we shall develop a bargaining solution concept which depends on the entire feasibility graph F, rather than just on F(1). In doing so, we do not mean to imply that we expect some participation vector other than q=1 might be imposed on the players or on their arbitrator. The role of F(q) for $q \neq 1$ in our theory will be analogous to the role of v(S) for $S \neq N$ in classical n-person game theory. In n-person game theory, one expects all the players to cooperate in the grand coalition N, but one also expects that the allocation chosen by the grand coalition will depend on what might have been achieved by small coalitions. In a Bayesian bargaining problem, we expect that the players will be fully available for cooperation, but we also expect that the allocation chosen may depend on what might have been achieved with only limited participation of the players.

We now list some of the basic properties which any feasibility graph must satisfy. First, any type which is sure to not cooperate must get zero utility. So we say that (q,ω) is <u>admissible</u> iff, for every player i and every type t_i in T_i, if $\omega_{t_i} \neq 0$ then there must be some t_{-i} in T_{-i} such that $q_t > 0$. We let A denote the set of admissible points:

$$(4.4) \qquad A = \{(q,\omega) \in [0,1]^T \times \mathbb{R}^{T_1 \cup T_2} \mid (q,\omega) \text{ is admissible}\}.$$

For example $(0,\omega) \in A$ only if $\omega = 0$.

Any feasibility graph F which is derived from a Bayesian bargaining problem (as in Sections 1 and 2) by (4.2) and (4.3) must satisfy the following properties:

(4.5) F is a nonempty convex subset of A;

(4.6) if $0 \leq \hat{q}_t \leq q_t \leq 1 \quad \forall t \varepsilon T$ then $F(\hat{q}) \subseteq F(q)$;

(4.7) if $0 \leq \gamma \leq 1$ then $F(\gamma q) = \{\gamma\omega \mid \omega \varepsilon F(q)\}$, $\forall q \varepsilon [0,1]^T$;

(4.8) $\{\omega \varepsilon F(1) \mid \omega_{t_i} \geq z \quad \forall t_i \varepsilon T_1 \cup T_2\}$ is a compact set, $\forall\, z \,\varepsilon\, \mathbb{R}$.

To verify these conditions, observe that (q,ω) is in F iff there exists some π satisfying (2.2), (2.3), (2.4) if types are unverifiable, and

(4.9) $$\sum_{c \neq c^*} \pi(c|t) \leq q_t, \ \forall t \varepsilon T,$$

such that $\omega = \bar{U}(\pi)$. Since $\bar{U}(\cdot)$ is linear, and all of these conditions on π are linear inequalities, F is actually a compact polyhedron. However, we only note compactness of the intersection of $F(1)$ with an orthant above z, because we will also want to consider bargaining problems with transferable utility and free disposal of utility (conditions which were not allowed in the framework of Sections 1 and 2) which do not give compact feasibility graphs. Conditions (4.6) and (4.8) use our assumption that $U_i(c^*,t) = 0$ for all i and t.

Henceforth, we may use the term feasibility graph to refer to any set F satisfying (4.5)-(4.8), with $F(q) = \{\omega | (q,\omega) \ \varepsilon \ F\}$.

5. Axioms for the bargaining solutions

Assuming that the feasibility graph carries all of the relevant information about the relative power of each type of each player, we can now develop a solution theory directly on the set of feasibility graphs. For any F which satisfies (4.5)-(4.8), we shall let $S(F)$ denote the solution set for

F. That is, S(F) will be some subset of F(1), denoting the set of utility allocations which should be considered fair outcomes for the bargaining problem. Once S(F) is defined for every F satisfying (4.5)-(4.8), one can return to the strategic structure of the Bayesian bargaining problem (1.1), to study feasible mechanisms which implement the solution set, in the sense that $\bar{U}(\pi) \varepsilon S(F)$.

The simplest feasibility graphs to study are those defined by a single linear constraint. So, suppose that $p=(p_t)_{t\varepsilon T}$ is any probability distribution over $T = T_1 \times T_2$. Then we define F^p to be the feasibility graph

$$(5.1) \qquad F^p = \{(q,\omega) \varepsilon A \mid \sum_{i=1}^{2} \sum_{t\varepsilon T} p_t \omega_{t_i} \le \sum_{t\varepsilon T} p_t q_t\}.$$

(Recall (4.4).)

To interpret F^p, consider the following story. The two players' types are first determined by a chance move, according to the probability distribution p. The two players can then earn one dollar together if they cooperate, regardless of their types. In addition, the players can make arbitrary contracts (or "bets") to transfer money as a function of their types, which are verifiable. Both players have utility which is linear in money. If q is the participation vector, so that q_t is the probability of cooperation for types t, then $\sum_{t\varepsilon T} p_t q_t$ is the expected income which the players can get. For the utility allocation ω, $\sum_{t\varepsilon T} p_t \omega_{t_i}$ is the ex ante expected payoff to player i, before his type is known. So the constraint in (5.1) says that the ex ante expected payoff to the two players must be less than or equal to the players' expected income. Suppose that the players are working with an arbitrator who is willing to involve himself in any system of sidebets with the players when they cooperate, provided that the arbitrator's expected payoff must be nonnegative. Then this arbitrator can implement a utility allocation ω consistently with a participation vector q iff $(q,\omega) \in F^p$.

Thus, F^p is the feasibility graph for a situation in which the two players earn a dollar together, and can also make bets about some otherwise inconsequential types which are generated according to the consistent prior probability distribution p. For such a situation, there is an obvious bargaining solution: divide the dollar equally, and ignore the inconsequential types. That is, since both players contribute equally to their earning power, regardless of type, the allocation $\bar{\omega}$ such that $\bar{\omega}_{t_i} = .5$ for all t_i in $T_1 \cup T_2$ is the obvious fair allocation. If player 1 were to advocate any other allocation (e.g.: "you take all our income if I cannot speak French, but I take it all if I can," when the cooperative task has nothing to do with speaking French), then player 2 would probably interpret this offer as an indication that 1 was of a type which would gain more than half under the plan, and so player 2 would prefer $\bar{\omega}$. As Milgrom and Stokey [1980] have shown, it can never be common knowledge that both players expect to gain from a system of bets when their types come from a commonly accepted prior. Thus, we are led to the followed axiom.

<u>Axiom 1</u> (<u>Equal division</u>). Let $\bar{\omega}$ in $\mathbb{R}^{T_1 \cup T_2}$ satisfy $\bar{\omega}_{t_i} = .5$ for all t_i in $T_1 \cup T_2$. Then $\bar{\omega} \in S(F^p)$, for any probability distribution p over T satisfying $p_t > 0$ for all t.

Our second axiom is an extension of Nash's axiom of <u>independence of irrelevant alternatives</u> (<u>IIA</u>).

Axiom 2 (IIA). If $\hat{F} \subseteq F$ and $\omega \varepsilon S(F)$ and $\omega \varepsilon \hat{F}(1)$, then $\omega \varepsilon S(\hat{F})$.

This axiom asserts that, if we reduce the range of feasible alternatives available to the players under all participation vectors, then a former bargaining solution which is still feasible should still be a bargaining solution (since there are now fewer feasible alternatives to be proposed against it).

Our third axiom is an extension of Nash's axiom of scale invariance. Since we are assuming that a player already knows his type at the time of bargaining, there is then no way to test assertions comparing a utility value for i conditional on type $\hat{t}_i$ with a utility value for i conditional on type $t_i \neq \hat{t}_i$. That is, if we doubled all $U_1(c,\hat{t}_1, t_2)$ numbers but left all $U_1(c, t_1,t_2)$ numbers unchanged (for all c and t_2), then we could not distinguish the new utility function from the old one in any decision problem which player 1 could face at the time of bargaining. If he knows that $\hat{t}_1$ is true then the $U_1(c, t_1, t_2)$ numbers are irrelevant to his decision-making behavior (and doubling a utility scale cannot affect behavior); and if he knows that $\hat{t}_1$ is false then the $U_1(c, \hat{t}_1, t_2)$ numbers are irrelevant to his decision-making. Before he learned his type (if there ever was such a time), such a utility transformation would have had decision-theoretically testable implications, but that is all past history at the time of bargaining. If our solution concept is to depend only on the properties of the utility functions which are decision-theoretically observable at the time of bargaining, then we must avoid intertype comparisons of utility, as well as interpersonal comparisons. This idea is formalized as follows.

For any vectors μ and ω in $\mathbb{R}^{T_1 \cup T_2}$, we define $\mu * \omega$ to be the vector in

$\mathbb{R}^{T_1 \cup T_2}$ such that

$$(\mu * \omega)_{t_i} = \mu_{t_i} \omega_{t_i}, \ \forall t_i \varepsilon T_1 \cup T_2$$

If we think of μ as a vector of factors for transforming the utility scales of each type of each player, then μ transforms the feasibility graph F into

$$\mu * F = \{(q, \mu * \omega) | (q,\omega) \varepsilon F\}$$

and transforms the solution set S(F) into

$$\mu * S(F) = \{\mu * \omega | \omega \varepsilon S(F)\}$$

in the new utility scales.

Axiom 3 (Scale invariance). For any vector μ in $\mathbb{R}^{T_1 \cup T_2}$ such that $\mu_{t_i} > 0$ for every t_i in $T_1 \cup T_2$, and for any feasibility graph F, $S(\mu * F) = \mu * S(F)$.

We let $\text{conv}(F \cup \{(\underset{\sim}{1}, \omega)\})$ denote the smallest convex set containing $F \cup \{(\underset{\sim}{1}, \omega)\}$. It is straightforward to check that $\text{conv}(F \cup \{(\underset{\sim}{1}, \omega)\})$ satisfies (4.5)-(4.8) if F does. With this notation, we can state the following weak continuity assumption.

Axiom 4 (Continuity). Suppose that $\omega \varepsilon F(\underset{\sim}{1})$ and there exists some sequence $\{\omega^k\}_{k=1}^{\infty}$ converging to ω, such that $\omega^k \varepsilon S(\text{conv}(F \cup \{(\underset{\sim}{1}, \omega^k)\}))$ for all k. Then $\omega \ \varepsilon \ S(F)$.

There certainly do exist solution correspondences which satisfy these axioms, since letting S(F) equal the entire efficient frontier of $F(\underset{\sim}{1})$ would satisfy all four axioms. Our goal is to find the strongest (smallest) solution concept consistent with these axioms.

Axiom 5 (Minimality). If S'(.) is any other solution correspondence which satisfies Axioms 1 throught 4, then $S'(F) \supseteq S(F)$ for every feasibility graph F.

It is easy to verify that there exists a unique solution correspondence S(.) satisfying Axioms 1-5; Let $\underset{\sim}{H}$ be the collection of all solution correspondences S'(.) which satisfy Axioms 1-4, and then let

$$(5.2) \qquad S(F) = \bigcap_{S' \varepsilon \underset{\sim}{H}} S'(F)$$

It is straightforward to check that S(.) must also satisfy Axioms 1-4, and it satisfies minimality as well.

Henceforth, we let S(.) denote the solution correspondence satisfying Axioms 1-5, and we refer to S(F) as the set of bargaining solutions for the feasibility graph F.

Formula (5.2) is a rather abstract way to define the bargaining solutions. Our main results are to give a more practical characterization of the bargaining solutions, and to show nonemptiness.

We let Λ denote the unit simplex in $\mathbb{R}^{T_1 \cup T_2}$, and we let Λ^0 denote the relative interior of Λ. That is,

$$(5.3) \qquad \Lambda = \{\lambda \varepsilon \ \mathbb{R}^{T_1 \cup T_2} | \sum_{t_i \varepsilon T_1 \cup T_2} \lambda_{t_i} = 1, \lambda_{t_i} \geq 0 \ \ \forall t_i\},$$

$$(5.4) \qquad \Lambda^0 = \{\lambda \varepsilon \Lambda | \lambda_{t_i} > 0 \ \forall t_i\}.$$

$\mathbb{R}^T_+$ denotes the nonnegative orthant of $\mathbb{R}^T$.

Theorem 1. Let F be any feasibility graph satisfying (4.5)-(4.8). Then $\omega \varepsilon S(F)$ iff $\omega \varepsilon F(1)$ and there exist sequences $\{\lambda^k\}_{k=1}^{\infty}$ in Λ^0 and $\{\alpha^k\}_{k=1}^{\infty}$ in $\mathbb{R}_+^T$ such that

$$(5.5) \qquad \sum_{t_i \varepsilon T_1 \cup T_2} \lambda_{t_i}^k \upsilon_{t_i} \leq \sum_{t \varepsilon T} \alpha_t^k q_t, \quad \forall (q,\upsilon)\varepsilon F;$$

and

$$(5.6) \qquad \omega_{t_i} = \lim_{k\to\infty} \left[\sum_{t_{-i}\varepsilon T_{-i}} (\alpha_t^k / (2\lambda_{t_i}^k))\right], \quad \forall i\varepsilon\{1,2\}, \quad \forall t_i \varepsilon T_i.$$

Notice that (5.6) implies that $\omega_{t_i} \geq 0$, so our solutions are individually rational. It is also straightforward to check that our solutions satisfy the probability-invariance axiom of Section 3,since they are defined on the feasibility graph, which is itself probability-invariant.

Theorem 2 For any feasibility graph F satisfying (4.5)-(4.8), $S(F) \neq \emptyset$.

The proofs are deferred until Section 7.

6. Analysis of the solutions

To get a clearer understanding of our bargaining solutions, consider first the following corollary of Theorem 1, proven in Section 7.

Corollary 1. Suppose $\omega \varepsilon F(1)$, where F is a feasibility graph. If $\omega \varepsilon S(F)$ then there exist vectors λ in Λ and α in $\mathbb{R}_+^T$ such that

$$(6.1) \qquad \underset{\upsilon \varepsilon F(q)}{\text{maximum}} \Big(\sum_{t_i \varepsilon T_1 \cup T_2} \lambda_{t_i} \upsilon_{t_i} \Big) \leq \sum_{t \varepsilon T} \alpha_t q_t, \quad \forall q \varepsilon [0,1]^T,$$

and

$$(6.2) \qquad \lambda_{t_i} \omega_{t_i} = \sum_{t_{-i} \varepsilon T_{-i}} \alpha_t / 2, \quad \forall i \varepsilon \{1,2\}, \forall t_i \varepsilon T_i.$$

Conversely, if there exist vectors λ in Λ^0 and α in $\mathbb{R}_+^T$ satisfying (6.1) and (6.2) then $\omega \varepsilon S(F)$.

Thus, (6.1) and (6.2) are necessary conditions for ω to be a bargaining solution, and they are sufficient conditions if λ is in the interior of the unit simplex. These are the conditions which we must try to interpret.

Since $\omega \varepsilon F(1)$, (6.1) and (6.2) imply

$$(6.3) \qquad \underset{\upsilon \varepsilon F(1)}{\text{maximum}} \Big(\sum_{t_i \varepsilon T_1 \cup T_2} \lambda_{t_i} \upsilon_{t_i} \Big) = \sum_{t \varepsilon T} \alpha_t = \sum_{t_i \varepsilon T_1 \cup T_2} \lambda_{t_i} \omega_{t_i} .$$

Then (6.1) and (6.3) imply that α_t can be interpreted as the <u>shadow cost</u> (at $q=1$) of decreasing the probability of cooperation in state t, when the objective is to maximize the weighted sum of expected utilities $\sum_{t_i} \lambda_{t_i} \upsilon_{t_i}$.

If α_t is the shadow cost of disagreement when t_1 and t_2 are the players' types, then a fair arbitrator might credit types t_1 and t_2 each with half of this "shadow value" of their agreement. After all, type t_1 of player 1 and type t_2 of player 2 both have equal power to force disagreement in state t. Then the right side of (6.2) is the total credit owed to type t_i for not forcing disagreement in any state $t=(t_{-i},t_i)$, for any t_{-i}. Condition (6.2) asserts that the weighted utility payoff expected by player i if t_i is his type must equal the total credit owed to type t_i by this fair arbitrator. So we may interpret (6.2) as a fairness condition, where fairness means that each type of each player should gain as much from cooperation as he contributes to it, as measured in the λ -weighted utility scales.

Given any λ, we can now show how to compute α so that (6.1) and (6.3) are satisfied.

Suppose first that F is the feasibility graph for a Bayesian bargaining problem as in (1.1) with verifiable types. Then

$$\underset{\upsilon\varepsilon F(q)}{\text{maximum}} \sum_{t_i \varepsilon T_1 \cup T_2} \lambda_{t_i} \upsilon_{t_i}$$

$$= \underset{\pi}{\text{maximum}} \left(\sum_i \sum_t \sum_c \lambda_{t_i} P_i(t)\, U_i(c,t)\, \pi(c|t)\right)$$
$$\text{subject to (2.3) and (4.9)}$$

$$= \left[\sum_{t\varepsilon T} \left(\max_{c\varepsilon C} \left(\sum_{i=1}^{2} \lambda_{t_i} P_i(t)\, U_i(c,t)\right)\right) q_t\right].$$

Thus, for a Bayesian bargining problem with verifiable types, (6.1) and (6.3) imply that

$$(6.4) \qquad \alpha_t = \max_{c\varepsilon C} \left(\sum_{i=1}^{2} \lambda_{t_i} P_i(t)\, U_i(c,t)\right).$$

For a Bayesian bargaining problem with unverifiable types, things are more complicated. The left side of (6.1) is now the optimum value of the linear programming problem to maximize (2.7) over π, subject to (4.9),(2.3), and (2.4). When we set q=1, the dual of this linear program is to choose

nonnegative vectors α in $\mathbb{R}_+^T$, β^1 in $\mathbb{R}_+^{T_1 \times T_1}$, and β^2 in $\mathbb{R}_+^{T_2 \times T_2}$ such that, for all t and c,

$$(6.5) \qquad \alpha_t \geq \sum_{i=1}^{2} [(\lambda_{t_i} + \sum_{s_i \varepsilon T_i} \beta^i_{t_i,s_i}) P_i(t)\, U_i(c,t)$$

$$- \sum_{s_i \varepsilon T_i} \beta^i_{s_i,t_i} P_i(t_{-i},s_i)\, U_i(c,(t_{-i}, s_i))],$$

so as to

$$(6.6) \qquad \text{minimize} \sum_{t \varepsilon T} \alpha_t .$$

In the dual problem, $\beta^i_{t_i,s_i}$ is the shadow price of the primal constraint (2.4), which says that player i should not be tempted to claim that s_i is his type when his type is really t_i. The dual variable α_t is the shadow price of the primal constraint $\sum_{c \neq c^*} \pi(c|t) \leq 1$. From the theory of duality in linear programming, it follows that α satisfies (6.1) and (6.3) for λ iff α is an optimal solution to the dual problem (6.5)-(6.6) together with some β^1 and β^2.

Thus, given any vector of utility-weights λ, it is a straightforward linear programming problem to compute the shadow costs α satisfying (6.1) and (6.3). The hard part of computing solutions is to find some λ such that the corresonding α will also satisfy the fairness condition (6.2). A fixed - point argument will be required to show that such a λ can be found.

7. Proofs.

Lemma 1 Suppose that

$$F = \{(q,\omega)\varepsilon A \mid \lambda \cdot \omega \leq \alpha \cdot q\}$$

for some λ in $\mathbb{R}_+^{T_1 \cup T_2}$ and α in $\mathbb{R}_+^T$, where all components of α and λ are strictly positive. (We use here the usual dot product in $\mathbb{R}^{T_1 \cup T_2}$ and $\mathbb{R}^T$.)

Let $\omega_{t_i} = \sum_{t_{-i}\varepsilon T_{-i}} \alpha_t/(2\lambda_{t_i}),\ \forall i,\ \forall t_i \varepsilon T_i.$

Then $\omega \varepsilon S(F)$.

Proof Let $p_t = \alpha_t/(\sum_{s\varepsilon T} \alpha_s)$, and let $\mu_{t_i} = \sum_{t_{-i}\varepsilon T_{-i}} \alpha_t/\lambda_{t_i}$.

Then it is straightforward to check that $F = \mu * F^p$ and $\omega = \mu * \bar{\omega}$, where F^p is as in (5.1) and $\bar{\omega}$ is as in Axiom 1. So $\omega \varepsilon S(F)$, by Equal Division and Scale Invariance. Q.E.D.

Lemma 2 If ω satisfies the conditions of Theorem 1 for F, then ω must be a solution for F, for any solution correspondence satisfying Axioms 1-4.

Proof Given the sequences $\{\lambda^k\}$ and $\{\alpha^k\}$ as in Theorem 1, let

$$\hat{\alpha}_t^k = \alpha_t^k + \frac{1}{k}(\lambda_{t_1}^k \lambda_{t_2}^k) > 0,$$

and

$$\omega_{t_i}^k = \sum_{t_{-i}\varepsilon T_{-i}} \hat{\alpha}_t^k/(2\lambda_{t_i}^k)$$

Then (5.5) implies that $F \subseteq \{(q,\upsilon)\varepsilon A| \lambda^k \cdot \upsilon \leq \hat{\alpha}^k \cdot q\}$, and (5.6) implies that the ω^k converge to ω. By Lemma 1, $\omega^k \varepsilon S(\{(q,\upsilon)\varepsilon A| \lambda^k \cdot \upsilon \leq \hat{\alpha}^k \cdot q\})$.

By IIA, $\omega^k \varepsilon S(\text{conv}(F \cup \{(1,\omega^k)\}))$. Thus, we have constructed the sequence required by the Continuity axiom, so ω must be a solution. Q.E.D.

Lemma 3. The solution set defined by the conditions in Theorem 1 satisfies Axioms 1-4.

Proof. To check Equal Division, simply let $\lambda^k_{t_i} = \sum_{t_{-i}\varepsilon T_{-i}} p_t/2$ and $\alpha^k_t = p_t/2$ for all k. Then $\lambda^k \varepsilon \Lambda^0$, (5.5) is satisfied for F^p, and(5.6) is satisfied for $\bar{\omega}$.

IIA is satisfied, because making F smaller only makes it easier to satisfy (5.5).

To check Scale Invariance, observe that, if $\{\lambda^k\}$ and $\{\alpha^k\}$ satisfy (5.5)-(5.6) for ω and F ,then $\{\hat{\lambda}^k\}$ and $\{\hat{\alpha}^k\}$ satisfy (5.5)-(5.6) for $\mu * \omega$ and $\mu * F$, where

$\hat{\lambda}^k_{t_i} = \lambda^k_{t_i}/(M^k \mu_{t_i})$, $\hat{\alpha}^k_t = \alpha^k_t/M^k$, and $M^k = \sum_{t_i \varepsilon T_1 \cup T_2} (\lambda^k_{t_i}/\mu_{t_i})$.

To check Continuity, let $\{\omega^k\}$ be a sequence converging to ω. If ω^k satisfies the conditions of Theorem 1 for $\text{conv}(F \cup \{(1,\omega^k)\})$ then we can find λ^k and α^k such that

$$F \subseteq \text{conv}(F \cup \{(1,\omega^k)\}) \subseteq \{(q,\upsilon)\varepsilon A \mid \lambda^k \cdot \upsilon \leq \alpha^k \cdot q\}$$

and $$|\omega^k_{t_i} - \sum_{t_{-i}\varepsilon T_{-i}} \alpha^k_t/(2\lambda^k_{t_i})| \leq |\omega^k_{t_i} - \omega_{t_i}| \qquad \text{for all } t_i.$$

Then $\{\lambda^k\}$ and $\{\alpha^k\}$ verify (5.5) and (5.6) for ω as a solution for F. Q.E.D.

Theorem 1 immediately follows from Lemmas 2 and 3.

Proof of Theorem 2. We begin with some definitions.

Given the feasibility graph F, let

$$E = \{(q,\omega)\varepsilon F \mid \omega_{t_i} \geq -(\max_{s\varepsilon T} q_s),\ \forall t_i \varepsilon T_1 \cup T_2\}.$$

Let $M = \underset{\omega\varepsilon E(1)}{\text{maximum}} \ (\sum_{t_i \varepsilon T_1 \cup T_2} |\omega_{t_i}|)$.

Let $B = \{\alpha\varepsilon\ \mathbb{R}^T \mid 0 \leq \alpha_t \leq M,\ \forall t\varepsilon T\}$.

For any λ in Λ, let $L(\lambda) = \underset{\omega\varepsilon E(1)}{\text{maximum}} \ \lambda \cdot \omega$.

For any $k > |T_1 \cup T_2|$, let

$$\Lambda^k = \{\lambda\varepsilon\Lambda \mid \lambda_{t_i} \geq \frac{1}{k} \quad \forall t_i \varepsilon T_1 \cup T_2\}.$$

We can now begin to construct a Kakutani correspondence. For any λ in Λ, let

$$Z_1(\lambda) = \{\omega\varepsilon E(1) \mid \lambda \cdot \omega = L(\lambda)\},$$

$$Z_2(\lambda) = \{\alpha\varepsilon B \mid \alpha \cdot 1 = L(\lambda) \text{ and } \alpha \cdot q \geq \lambda \cdot \omega, \quad \forall (q,\omega)\varepsilon E\}.$$

Let $W_{t_i}(\alpha,\lambda) = \sum_{t_{-i}\varepsilon T_{-i}} (\alpha_t/2\lambda_{t_i})),$ if $\lambda_{t_i} > 0.$

For any $k > |T_1 \cup T_2|$, if $\lambda\varepsilon\Lambda^k$, let $Z_3^k(\omega,\alpha,\lambda)$ be the set of all $\hat{\lambda}$ in Λ^k such that, for any t_i in $T_1 \cup T_2$,

$$\text{if} \quad W_{t_i}(\alpha,\lambda) - \omega_{t_i} < \underset{s_j\varepsilon T_1 \cup T_2}{\text{maximum}} \ (W_{s_j}(\alpha,\lambda) - \omega_{s_j}) \quad \text{then} \quad \lambda_{t_i} = \frac{1}{k}.$$

It is straightforward to verify that Z_1, Z_2, and Z_3^k are convex-valued upper-semicontinuous correspondences, and that Z_1 and Z_3^k are nonempty-valued. The only fine point is to verify that $Z_2(\lambda)$ is a nonempty set, for any λ in Λ.

Let $\quad H(\lambda) = \{(q,h)\varepsilon\ \mathbb{R}^T \times \mathbb{R} \mid h > L(\lambda) \max_{t\varepsilon T} q_t\}.$

Using (4.6) and (4.7) for F, if

$(q,\omega)\varepsilon E$ then $(\gamma 1,\omega)\varepsilon E$ and $(1,\frac{1}{\gamma}\,\omega)\varepsilon E$, where $\gamma = \max_{t\varepsilon T} q_t$,

and so $\lambda \cdot \omega \leq L(\lambda)\gamma$. So $H(\lambda)$ and

$G(\lambda) = \{(q,\ \lambda \cdot \omega) \mid (q,\omega)\varepsilon E\}$

are disjoint convex sets. By the Separating Hyperplane Theorem, there exists some α in $\mathbb{R}^T$ such that $(-\alpha,1)$ separates $H(\lambda)$ from $G(\lambda)$. (The last component of the separating vector cannot be zero, since the projection of $H(\lambda)$ onto $\mathbb{R}^T$ covers $\mathbb{R}^T$.) Since $(0,0)\varepsilon G(\lambda)$ and $H(\lambda)$ is an open cone, $-\alpha \cdot q + \lambda \cdot \omega \leq 0$ for all (q,ω) in E and $-\alpha \cdot 1 + L(\lambda) = 0$. It is straightforward to check that $0 \leq \alpha_t \leq L(\lambda) \leq M$ for all t in T. So $\alpha\varepsilon Z_2(\lambda) \neq \emptyset$.

By the Kakutani Fixed Point Theorem, for any $k > |T_1 \cup T_z|$, there exist some ω^k in $E(1)$, α^k in B, and λ^k in Λ^k such that $(\omega^k,\alpha^k,\lambda^k)\varepsilon\ Z_1(\lambda^k) \times Z_2(\lambda^k) \times Z_3^k\ (\omega^k,\alpha^k,\lambda^k)$. Since $E(1)$ is compact, we can choose a convergent subsequence of the $\{\omega^k\}$, converging to some ω^* in $E(1) \subset F(1)$. We will show that $\omega^*\varepsilon S(F)$.

Using $\omega^k\varepsilon\ Z_1(\lambda^k)$ and $\alpha^k\varepsilon\ Z_2(\lambda^k)$, we get

$$\sum_{t_i} \lambda^k_{t_i} \omega^k_{t_i} = \sum_t \alpha^k_t = \sum_{t_i} \lambda^k_{t_i}\ W_{t_i}(\alpha^k,\ \lambda^k).$$

Because $\lambda^k\varepsilon Z_3^k(\omega^k,\alpha^k,\lambda^k)$, if $\omega^k_{t_i} < W_{t_i}(\alpha^k,\lambda^k)$ then $\lambda^k_{t_i} = \frac{1}{k}$.

So $$\sum_{t_i} \lambda^k_{t_i} \max\{0, W_{t_i}(\alpha^k,\lambda^k) - \omega^k_{t_i}\} = \sum_{t_i} \lambda^k_{t_i} \max\{0, \omega^k_t - W_{t_i}(\alpha^k,\lambda^k)\}$$

$$\leq \frac{1}{k}\left(\sum_{t_i} \omega^k_{t_i}\right) \leq \frac{M}{k}.$$

So, for all k and t_i

(7.1) $$W_{t_i}(\alpha^k,\lambda^k) \leq \omega^k_{t_i} + \frac{M}{k}.$$

Suppose $k > M$. Then (7.1) implies that $\omega^k_{t_i} > -1$, $\forall t_i$. If $(q,\upsilon)\varepsilon F$ and $\lambda^k \cdot \upsilon > \alpha^k \cdot q$, then some $\gamma > 0$ sufficiently small, $(\hat{q},\hat{\upsilon}) = (\gamma\, q + (1-\gamma)\, \underset{\sim}{1},\ \gamma\upsilon + (1-\gamma)\omega^k)\varepsilon\ E$ and $\lambda^k \cdot \hat{\upsilon} > \hat{\alpha}^k \cdot \hat{q}$, which contradicts the fact that $\alpha^k \varepsilon Z_2(\lambda^k)$. So $\lambda^k \cdot \upsilon \leq \alpha^k \cdot q$, for every (q,υ) in F.

Now let $$\hat{\alpha}^k_t = \alpha^k_t + \left(\frac{\lambda_{t_1}\lambda_{t_2}}{k}\right) > 0,\ \forall t\varepsilon T.$$

Then $$0 < W_{t_i}(\hat{\alpha}^k, \lambda^k) \leq W_{t_i}(\alpha^k,\lambda^k) + \frac{1}{k} \leq \omega^k_{t_i} + \frac{M+1}{k}.$$

Let $$\hat{\lambda}^k_{t_i} = \left(\frac{\lambda^k_{t_i} W_{t_i}(\hat{\alpha}^k,\lambda^k)}{\omega^k_{t_i} + \frac{M+1}{k}}\right).$$

Then $0 < \hat{\lambda}^k_{t_i}$ and $W_{t_i}(\hat{\alpha}^k, \hat{\lambda}^k) = \omega^k_{t_i} + \frac{M+1}{k}$ for all t_i. Furthermore, since

$\hat{\lambda}^k_{t_i} \leq \lambda^k_{t_i}$ $\forall t_i$ and $\hat{\alpha}^k_t \geq \alpha^k_t$ $\forall t$, $\hat{\lambda}^k \cdot \upsilon \leq \hat{\alpha}^k \cdot q$ $\forall (q,\upsilon)\varepsilon F$.

Unfortunately, $\hat{\lambda}^k$ is not Λ^0 because $\sum_{t_i} \hat{\lambda}^k_{t_i} < 1$, but this is easily remedied.

Let $$\delta_k = \sum_{t_i \varepsilon T_1 \cup T_2} \hat{\lambda}^k_{t_i},\quad \bar{\lambda}^k = \frac{1}{\delta_k}\hat{\lambda}^k,\quad \bar{\alpha}^k = \frac{1}{\delta_k}\hat{\alpha}^k.$$

Then $\bar{\lambda}^k \varepsilon \Lambda^0$, $\bar{\lambda}^k \cdot \upsilon \leq \bar{\alpha}^k \cdot q \quad \forall (q,\upsilon)\varepsilon F$,

and
$$\lim_{k\to\infty} W_{t_i}(\bar{\alpha}^k,\bar{\lambda}^k) = \lim_{k\to\infty} W_{t_i}(\hat{\alpha}^k,\hat{\lambda}^k)$$

$$=\lim_{k\to\infty}(\omega^k_{t_i} + \frac{M+1}{k}) = \lim_{k\to\infty}\omega^k_{t_i} = \omega^*_{t_i}$$

for all t_i. Thus, recalling the definition of $W_{t_i}(.)$, we see that the $\bar{\alpha}^k$ and $\bar{\lambda}^k$ sequences verify the conditions of Theorem 1 for ω^*. So $\omega^* \varepsilon S(F) \neq \emptyset$. Q.E.D.

<u>Proof of Corollary 1.</u> In Theorem 1, each $\lambda^k \varepsilon \Lambda$, and (5.6) implies $0 \leq \alpha^k_t \leq 2\omega_{t_i}+1$ for all k sufficiently large. Thus there must exist some cluster points λ in Λ and α in $\mathbb{R}^T_+$ for the $\{\lambda^k\}$ and $\{\alpha^k\}$ sequences. Then (5.5) and (5.6) imply (6.1) and (6.2) in the limit for λ and α.

Conversely, if (6.1) and (6.2) are satisfied with λ in Λ^0, then letting $\lambda^k=\lambda$ and $\alpha^k=\alpha$ for all k satisfies (5.5) and (5.6). Q.E.D.

8. Example

In Section 4, we discussed Examples 1 and 2, two simple examples with verifiable types. It follows easily from Lemma 1 that $\omega = (50,50,50)$ is a solution for Example 1, and that $\omega = (100,0,50)$ a solution for Example 2, as asserted in Section 4.

For a third example, with unverifiable types, let us consider the example in Myerson [1979a]. In this example, $T_1 = \{1a,1b\}$, $T_2 = \{2\}$, and player 2 assigns a probability of .9 to t_1=1a and a probability of .1 to t_1=1b. The two players can jointly undertake a project (say, a new road which both players would use) which costs $100. The road is worth $90 to player 2, and it is worth $90 to player 1 if he is type 1a, but it is only worth $30 to player 1 if he is type 1b. The problem is to decide if the project should be undertaken, and if so, how much should each player pay.

If the types were verifiable, then an obvious plan would be to always undertake the project, with each paying $50 if t_1 = 1a, and with 1 paying $20 and 2 paying $80 if t_1 = 1b. In this way, both players gain equally in each state ($40 if 1a, $10 if 1b) and the expected utility allocation is

(8.1) $\quad (\omega_{1a},\omega_{1b},\omega_2) = (40,10,36).$

In fact, it is easy to verify that (8.1) is our solution for this problem, with verifiable types. (Use $\lambda_{1a} = .9$, $\lambda_{1b} = .1$, $\lambda_2 = 1$, $\alpha_{(1a,2)} = 80$, $\alpha_{(1b,2)} = 20$, and apply Lemma 1.)

However, with unverifiable types, the above solution is infeasible, since it would induce 1a to pretend he was 1b. Our solution for this problem with unverifiable types is

(8.2) $\quad (\omega_{1a},\omega_{1b},\omega_2) = (41\frac{7}{13}, 0, 36).$

This allocation is implemented by the following choice mechanism: if t_1 = 1a then the project is undertaken and 1 pays \$48.46 and 2 pays \$51.54; if t_1 = 1b then the project is undertaken with probability $\frac{9}{13}$, in which case 1 pays \$30 (his full value for the project) 2 pays \$70. If the project is not undertaken, then neither player pays anything.

This mechanism is incentive-compatible. Player 1 in type 1a would prefer to pay only \$30, but the $\frac{4}{13}$ chance of losing the project prevents him from claiming to be 1b. This mechanism is efficient, in spite of the fact that there is a positive probability of not undertaking a project which is certainly worth more to the players than it costs. Without the positive risk of losing the project, type 1a could not be induced to bear his fair share of the cost, and this would hurt player 2.

To check that (8.2) is indeed a solution, one must examine geometry of the problem in greater detail. The feasible set $F(\underset{\sim}{1})$ is described in Myerson [1979a] as the convex hull of five points in $\mathbb{R}^3$, and the efficient frontier is a triangle perpendicular to the vector

(8.3) $$(\lambda_{1a},\lambda_{1b},\lambda_2) = (\frac{13}{30}, \frac{2}{30}, \frac{15}{30}).$$

Thus, choosing λ as in (8.3) would allow (6.3) to be satisfied for any efficient ω. Also, for any efficient mechanism, either the costs are divided independently of t_1, or else there must be a positive probability not undertaking the project when t_1 = 1b. But if there is a positive probability of no project when t_1 = 1b, then we must have zero shadow cost of conflict when t_1 = 1b; that is,

(8.4) $$\alpha_{(1b,2)} = 0.$$

Then (6.2) implies $\lambda_{1a}\omega_{1a} = \frac{1}{2}\alpha_{(1a,2)} = \lambda_2\omega_2$ and $\lambda_{1b}\omega_{1b} = 0$. (8.2) is the only efficient allocation satisfying these equations, for λ as in (8.3).

In fact, (8.2) is the unique solution for this problem with unverifiable types. However, if we simply charged both players \$50 for the project, independently of t_1, then we get

(8.5) $\quad (\omega_{1a},\omega_{1b},\omega_2) = (40,-20,40)$

With $\lambda_{1a} = \frac{1}{2}$, $\lambda_{1b} = 0$, $\lambda_2 = \frac{1}{2}$, $\alpha_{(1a,2)} = 40$, $\alpha_{(1b,2)} = 0$, (8.5) can be shown to satisfy (6.1) and (6.2); however this λ is not in Λ^0. In fact, (8.5) is not even individually rational for 1b, so we know that (8.5) cannot satisfy the conditions for a solution, given in Theorem 1. Thus, (8.5) shows why (6.1) and (6.2) are only necessary conditions and not sufficient for a solution, when some $\lambda_{t_i} = 0$.

As stated in Myerson [1979a], the feasible allocation (with unverifiable types) which maximizes the Harsanyi-Selten criterion (3.1) is

(8.6) $\quad \omega = (39.5,\ 13.2,\ 36)$.

This allocation is implemented by the following incentive-compatible mechanism: if $t_1 = 1a$ then the project is undertaken and 1 pays \$50.50 and 2 pays \$49.50; if $t_1 = 1b$ then the project is undertaken with probability .439, in which case 2 pays the entire cost of \$100.

Notice that player 2 gets the same expected utility in both (8.6) and (8.2). The only difference is how well the two types of player 2 do. From the point of view of our new solution concept, player 1 in type 1b has no bargaining power, since his only threat is to force the project to be abandoned, and that is already going to happen with positive probability (both in the mechanism which implements (8.2) and in the mechanism which implements (8.6)). Thus, if player 1 were to argue for (8.6) rather than our solution (8.2) then player 2 might interpret this as evidence that $t_1 = 1b$. But if 2 believes that $t_1 = 1b$, then he would no longer be indifferent between the mechanisms which implement (8.2) and (8.6). In fact, player 2's utility from the mechanism which implements (8.6) would be -4.39 if $t_1 = 1b$, and so he would be inclined to reject this mechanism.

It is significant that our solution requires a positive probability of the conflict outcome, which is ex post Pareto-inefficient. When types are unverifiable, ex ante efficiency does not imply ex post efficiency, because the incentive-compatibility constraints must be satisfied before the players can be made to reveal their information for collective use. We may often find a positive probability of conflict in the solutions for bargaining problems with unverifiable types. Thus, the theory of bargaining with incomplete information can offer us basic insights as to why cooperation must sometimes break down into a conflict in which both players lose.

REFERENCES

Aumann, R. J. and M. 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.

Milgrom, P., and N. Stokey [1979], "Information, Trade, and Common Knowledge"," Center for Math Studies Discussion Paper No. 377R, Northwestern University.

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.

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

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