Discussion Paper No. 433

A GENERAL THEORY OF COOPERATIVE SOLUTIONS FOR GAMES WITH INCOMPLETE INFORMATION

by

Roger B. Myerson*

July, 1980

J.L. Kellogg Graduate School of Management
Northwestern University

Abstract. This paper develops a generalization of the λ-transfer value for cooperative strategic-form games with incomplete information. A linear programming approach is used to assess shadow prices for the incentive-compatibility constraints. Every coalition must take these incentive-compatibility costs into account in computing its optimal threat. Cooperative solutions specify a threat equilibrium and a cooperative agreement which is efficient and individually rational, and which satisfies a fairness condition for every possible type of every player.

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

A GENERAL THEORY OF COOPERATIVE SOLUTIONS FOR GAMES WITH INCOMPLETE INFORMATION

1. Introduction

The basic goal of a cooperative solution concept must be to systematically determine Pareto-efficient outcomes for games in such a way that, the more power a player has, the more his payoff should be. Here power is taken to mean the ability to alternatively help or hurt other players at will, and to defend oneself against the attacks of others. The λ-transfer value was proposed by Harsanyi [1963] and Shapley [1969] as a cooperative solution concept, to extend the Shapley [1953] value to n-person games without sidepayments. The λ-transfer concept has not been formally derived from any basic axioms or assumptions; its justification has come from the fact that it generalizes the Shapley value (for n-person games with sidepayments) and the Nash [1950,1953] bargaining solution (for two-person games without sidepayments), both of which have been axiomatically derived. Although recent examples (see Roth [1980] and Shafer [1980]) have suggested difficulties with the λ-transfer value, it remains an important solution concept, if only because of the lack of alternative solution concepts with comparable generality and precision.

A game with incomplete information is a game in which, at the time of play, each player may have information which the other players do not know. Harsanyi [1967-8] has developed the basic structures for modelling such games, which we shall use in this paper. In Myerson [1980], a new generalization of the Nash bargaining solution was axiomatically derived for two-person games with incomplete information. In this paper, we will define a solution concept for general n-person cooperative games with incomplete information. Our solution concept will generalize both the

λ-transfer value and the bargaining solution of Myerson [1980].

A cooperative solution in our terms will specify a system of threats, one threat for every possible coalition, to represent to represent what it would do if it were to form. We anticipate that only the "threat" of the grand coalition will actually be carried out (it is in fact the cooperative agreement), but the other threats are important as an expression of the power as an expression of the power structure of the game. This idea of measuring power in strategic form games through a collection of endogenous optimal threats for all coalitions, was suggested by Harsanyi [1963]; for more on this idea, see Myerson [1978]. By having every set of players generate a threat, our model can be sensitive to the abilities of players to threaten or assist each other in all possible combinations.

However, before we can compute the optimal threats for the various coalitions, we need to establish the criteria which each coalition should use to assess trade-offs between the payoffs for the various players in the coalition, in all of their possible types. To do this we follow the λ-transfer idea of Shapley [1969], and we admit into out model a system of endogenous weighting factors for the players' utility scales. These weighting factors can be thought of as determining a new set of weighted-utility scales in which interpersonal comparisons may be appropriate. Given these weights, the coalitions can determine their optimal threats, and then the Shapley value can be used to derive each player's justifiable demands, in the context of the power structure expressed by these threats. In equilibrium, the weighting factors are determined by the condition that the players' demands must coincide with what the grand coalition can give them.

With incomplete information, players who cooperate must be concerned with the issue of incentive-compatibility; that is, the cooperative agreement must not create incentives for any player to lie about his information. Thus, coalitions must take into account the impact which their threats might have on the incentive-compatibiltiy of the cooperative agreement. To account for these incentive-compatibility constraints in each coalition's decision problem, we shall use the shadow prices for the incentive-compatibility constraints. These shadow prices will be derived from a linear programming approach in Section 3, after we formulate the basic model in Section 2.

Our solution concept will be developed first for a simpler special case in Section 4. Then, after reviewing some basic facts about partition function games in Section 5, we shall define solutions for the general case in Section 6 and 7. The proof of our basic existence theorem is in Section 8.

We do not derive our solution concept from any formal axioms or assumptions about what a cooperative solution concept should be like. Instead, each definition in the model is only motivated by a discussion of why it is reasonable to measure power in this way. However, our solution concept can also be justified by the fact that, as a generalized λ-transfer theory, it does include many axiomatically derived solution concepts as special cases, such as the solution concepts of Nash [1950], Shapley [1953], Myerson [1978] and Myerson [1980]. However, the ultimate test of our solution theory must be whether it succeeds in generating useful insights into the power structure and the fair demands of players in a wide variety of examples with applied interest.

2. Basic definitions

We let $N = \{1,2,\ldots,n\}$ denote the set of players, and we let

$$CL = \{S | S \subseteq N \text{ and } S \neq \emptyset\}$$

denote the set of coalitions or nonempty subsets of N. For any coalition S, we let D_S denote the set of collective actions or decisions feasible for the members of S if they work together. (E.g., in a market game, D_S might be the set of trades among the members of S.) For any two disjoint coalitions S and R we assume that

$$D_S \times D_R \subseteq D_{S \cup R}.$$

That is, $S \cup R$ can implement any decisions feasible for S and R separately, if $S \cap R = \emptyset$.

For any player i in N, we let T_i denote the set of possible types for player i, where each type t_i in T_i is a complete description of i's beliefs, preferences, endowments, and other private information. For any coalition S, we let

$$T_S = \times_{i \in S} T_i,$$

so any t_S in T_S denotes a possible combination of types for the members of S. We let

$$T_* = \bigcup_{i \in N} T_i.$$

We assume that the players' types-sets are disjoint, so $|T_*| = \sum_{i \in N} |T_i|$. Throughout this paper, we assume that all T_i and D_S are finite sets.

For any d_N in D_N and t_N in T_N, we let $U_i(d_N,t_N)$ denote the payoff to player i, measured in some vonNeumann-Morgenstern utility scale, if d_N represents the decisions made and t_N represents the types of all players.

Throughout this paper, whenever a type-vector t_N in T_N has been specified, then t_i (for any player i) must denote the i^{th} component of t_N, and (for any $S \subset N$) $t_S = (t_j)_{j \in S}$. Conversely, whenever some type-vectors t_S in T_S and $t_{N \setminus S}$ in $T_{N \setminus S}$ (or some t_i and t_{N-i}) have been specified, then t_N must denote the type-vector which combines these two lists of the players' types.

For any t_N in T_N, we let $P_i(t_{N-i}|t_i)$ the conditional probability that t_{N-i} is the combination of types for all players other than i, as would be assessed by player i if t_i were his type. (We use the notation $N-i = N \setminus \{i\}$.)

Thus, a cooperative game with incomplete information is defined by these structures:

$$((D_S)_{S \in CL}, (T_i)_{i \in N}, (U_i)_{i \in N}, (P_i)_{i \in N}). \tag{2.1}$$

We let $\bar{D}_S$ denote the set of all probability distributions over D_S; that is,

$$\bar{D}_S = \{\gamma : D_S \to \mathbb{R}_+ | \sum_{d_S \in D_S} \gamma(d_S) = 1\},$$

where $\mathbb{R}_+$ denotes the nonnegative real numbers. A <u>randomized strategy</u> for coalition S is any mapping from T_S to $\bar{D}_S$. We let Δ_S denote the set of mixed strategies for S. Thus, for any δ_S in Δ_S, we have

$$\delta_S(t_S) \in \bar{D}_S, \ \forall t_S \in T_S,$$

or, equivalently,

$$\sum_{d_S \in D_S} \delta_S(d_S|t_S) = 1 \text{ and } \delta_S(d_S'|t_S) \geq 0, \quad \forall t_S \in T_S, \quad \forall d_S' \in D_S. \tag{2.2}$$

In the strategy δ_S, $\delta_S(d_S|t_S)$ represents the probability of coalition S choosing the action d_S when its members' types are t_S.

In our notation, $\delta_S(t_S)$ is equivalent to $\delta_S(\cdot|t_S)$, for any δ_S in Δ_S.

We let $\Delta = \underset{S \in CL}{X} \Delta_S$. A typical point $\delta = (\delta_S)_{S \in CL}$ in Δ is to be interpeted as a <u>vector of threats</u>, describing what each coalition would do if it were to form. In this context, we do not need to assume any special relationship between the threats of two nondisjoint coalitions S and R, since

δ_S describes what the members of S would do if coalition S were to form, and δ_R describes what the members of R would do if R were to form. If $S \cap R \neq \emptyset$, then S and R could not both "form" as effective coalitions. In fact, we anticipate that only the grand coalition N will actually form, so that only δ_N in the vector of threats δ will be actually carried out. The other components of δ are relevant, however, because these threats determine the justifiable demands of each player in a fair bargaining process.

For any player i, we define the <u>evaluation function</u> $Y_i: D_N \times T_N \to \mathbb{R}$ by

(2.3) $Y_i(d_N,t_N) = P_i(t_{N-i}|t_i)\ U_i\ (d_N,t_N)$.

That is, $Y_i(d_N,t_N)$ measures how much the prospect of decision d_N being made in the case of T_N being the types would contribute to player i's expected utility, when he knows that t_i is his type. It has been argued elsewhere (see Myerson [1980]) that probabilities and utilities should be significant only through the evaluation function. To shorten and simplify our formulas, we will henceforth work mainly with the Y_i functions, instead of P_i and U_i.

We let $\bar{Y}_i: \bar{D}_N \times T_N \to \mathbb{R}$ be the extension of Y_i to randomized decisions. Thus, for any δ_N in Δ_N, and for any type-vectors t_N and $\hat{t}_N$ in T_N

(2.4) $\bar{Y}_i(\delta_N(\hat{t}_N),\ t_N) = \sum_{d_N \in D_N} \delta_N(d_N|\hat{t}_N)\ Y_i(d_N,t_N)$.

We let $\bar{U}_{t_i}(\delta_N)$ denote the conditionally expected utility for player i from the collective strategy δ_N, given that i's type is t_i. Thus,

(2.5) $\bar{U}_{t_i}(\delta_N) = \sum_{t_{N-i} \in T_{N-i}} \bar{Y}_i(\delta_N(t_N),t_N)$.

(Recall $t_N = (t_{N-i},t_i)$.)

3. The Primal and Dual Problems

In this paper, we assume that players' types are not observable to other players, and hence are unverifiable. Thus, if a player had some incentive to lie about his type when the grand coalition N implements its strategy δ_N, then he would do so. To be feasible, therefore, a strategy δ_N in δ_N must also satisfy the following incentive-compatibility constraint:

$$(3.1) \quad \bar{U}_{t_i}(\delta_N) \geq \sum_{t_{N-i} \varepsilon T_{N-i}} \bar{Y}_i(\delta_N(t_{N-i}, \hat{t}_i), t_N), \quad \forall i \varepsilon N,\ \forall t_i \varepsilon T_i,\ \forall \hat{t}_i \varepsilon T_i$$

The right-hand side of (3.1) is player i's conditionally expected utility from the collective strategy δ_N, if i's true type is t_i but he is planning to report $\hat{t}_i$ to the coalition, and all other player are expected to report their types honestly to the coalition in implementing δ_N. Thus, (3.1) asserts that it must be an equilibrium for all players to report their types honestly to the coalition N.

An incentive-compatible strategy δ_N in Δ_N is efficient iff there does not exist any other incentive-compatible strategy $\hat{\delta}_N$ in Δ_N such that

$$\bar{U}_{t_i}(\hat{\delta}_N) > \bar{U}_{t_i}(\delta_N), \ \forall i \varepsilon N, \ \forall t_i \varepsilon T_i.$$

That is, δ_N is efficient iff it is not possible to find any other incentive-compatible strategy which is sure to increase the expected utility of every player, no matter what his type is. It is a minimal requirement that our solution theory must select efficient an strategy for the grand coalition N.

We let Λ denote the unit simplex in $\mathbb{R}^{T*}$

$$\Lambda = \{\lambda \ \varepsilon \ \mathbb{R}^T \mid \sum_{t_i \varepsilon T_*} \lambda_{t_i} = 1, \ \lambda_{t_i} \geq 0 \ \ \forall t_i \varepsilon T_*\}.$$

(Recall $T_* = \bigcup_{i \varepsilon N} T_i$). By the Separating Hyperplane Theorem, an incentive-compatible strategy δ_N in Δ_N is efficient iff there exists some nonnegative vector $\lambda = (\lambda_{t_i})_{t_i \in T_*}$ in Λ such that δ_N is an optimal solution to the problem

$$(3.2)\qquad \text{maximize} \sum_{i\varepsilon N} \sum_{t_i \varepsilon T_i} \lambda_{t_i} U_{t_i}(\delta_N)$$

subject to (2.2) and (3.1).

Since $\bar{U}_{t_i}(\delta_N)$ is linear in δ_N, (3.2) is a linear programming problem. Thus, our search for a fair cooperative solution among all efficient strategies may be rephrased as a search among all solutions to (3.2) as λ varies over Λ.

The dual to (3.2) will be very important in our analysis. We let $\alpha(t_N)$ denote the shadow price of the constraint $\sum_{d_N} \delta_N(d_N|t_N) = 1$. For any player i, and for any two types t_i and $\hat{t}_i$ in T_i, we let $\beta(t_i,\hat{t}_i)$ denote the shadow price of the incentive-compatibility constraint (3.1), asserting that i should not prefer to report $\hat{t}_i$ when t_i is the true type. Then it is straightforward to show that the dual may be written as follows:

$$(3.3)\qquad \text{minimize} \sum_{t_N \varepsilon T_N} \alpha(t_N) \qquad \text{subject to:}$$

$$(3.4)\qquad \alpha(t_N) = \underset{d_N \varepsilon D_N}{\text{maximum}} \Big[\sum_{i \varepsilon N} \lambda_{t_i} Y_i(d_N,t_N)$$

$$+ \sum_{i\varepsilon N} \sum_{\hat{t}_i \varepsilon T_i} \beta(t_i,\hat{t}_i)\, Y_i(d_N,t_N)$$

$$- \sum_{i\varepsilon N} \sum_{\hat{t}_i \varepsilon T_i} \beta(\hat{t}_i,t_i)\, Y_i(d_N,(t_{N-i},\hat{t}_i))\Big], \qquad \forall t_N \varepsilon T_N ,$$

and

$$(3.5)\qquad \beta(t_i,\hat{t}_i) \geq 0,\ \forall i\varepsilon N,\ \forall t_i \varepsilon T_i,\ \forall \hat{t}_i \varepsilon T_i.$$

To interpret (3.4), notice that the first term in the maximand measures how much the prospect of using d_N in state t_n would contribute directly to the primal objective function (3.2). The second term measures the value of d_N in state t_N for increasing incentive-compatibility, by rewarding each player i with $Y_i(d_N,t_N)$ if his type is t_i and he is honest. The third term measures the cost of d_N in state t_N for reducing incentive-compatibility, by rewarding each player i with $Y_i(d_N,(t_{N-i},\hat{t}_i))$ if he reports t_i when $\hat{t}_i$ is really his type (and all others report t_j honestly). Then $\alpha(t_N)$ equals the total contribution of the best decision in state t_N.

In some examples, it may be possible for some types to costlessly prove that other types are false. For example, if a person can play the piano, then he can prove that he is not a non-pianist simply by playing a few bars. On the other hand, the non-pianist cannot prove that he is not really a pianist, unless he is given the proper incentives.[1] In general, if player i could costlessly verify that t_i was false if $\hat{t}_i$ were true, then we must drop from the primal problem the constraint which says that i should not be tempted to report $\hat{t}_i$ if t_i were true, and we must set the dual variable $\beta(t_i,\hat{t}_i)$ equal to zero. With these revisions, our model can easily accomodate the case of verifiable types. Nevertheless, throughout the rest of this paper, we shall consider only the case of unverifiable types.

If there were no incentive-compatibility constraints, then the solution to the primal problem would be simple: for every t_N, $\delta_N(\cdot|t_N)$ would put all probability weight on the decisions which maximize $\sum_i \lambda_{t_i} Y_i(d_N,t_N)$. The incentive-compatibility constraints complicate matters by interconnecting the decisions in different states t_N. However, we can decompose the primal problem by using the dual variables to remove the incentive-compatibility constraints. That is, by the duality theorem of linear programming, if δ_N is optimal for the primal then, for every t_N in T_N, $\delta_N(\cdot|t_N)$ must put all probability weight on the decisions which achieve the maximum in (3.4).

1. I am indebted to Paul Milgrom for pointing out this issue.

4. Solutions in the special case.

In this section we restrict our attention to games satisfying two special assumptions.

Assumption A For any coalition S and any player i in S,

$$U_i((d_S, d_{N\setminus S}), t_N) = U_i((d_S, \hat{d}_{N\setminus S}), t_N),$$

$$\forall d_S \in D_S, \forall d_{N\setminus S} \in D_{N\setminus S}, \forall \hat{d}_{N\setminus S} \in D_{N\setminus S}, \forall t_N \in T_N.$$

Assumption B. For any play i,

$$U_i((d_i, d_{N-i}), t_N) = 0$$

$$\forall d_i \in D_{\{i\}}, \forall d_{N-i} \in D_{N-i}, \forall t_N \in T_N.$$

Assumption A asserts that, when coalition S chooses an action which is feasible for it, then the payoff to the members of S does not depend on the actions of the complementary coalition $N\setminus S$. Shapley and Shubik [1973] have called this the assumption of orthogonal coalitions. Assumption B asserts that, without any other player's help, each player i can only get zero utility by choosing an action in $D_{\{i\}}$.

With Assumption A, we can define functions $Y_i^S: D_S \times T_N \to \mathbb{R}$, for any coalition S and any player i in S, so that

$$Y_i^S(d_S, t_N) = Y_i((d_S, d_{N\setminus S}), t_N), \forall d_{N\setminus S} \in D_{N\setminus S}.$$

A general cooperative solution concept must take into account the strategic possibilities of each coalition, so we need to determine an optimal bargaining threat for each coalition. That is, for each coalition S, we must find some strategy δ_S in Δ_S, which represents what the members of S threaten to do if the members of $N\setminus S$ were to refuse to cooperate with them.

We anticipate that the grand coalition N will actually form, so that there is only an infinitesimal probability that any smaller coalition S will have to carry out its threat. We want to know, how much would the coalition S have to lose if there were some small positive probability that S might have to carry out its threat. The more that S would lose, the less power S has to demand high payoffs for its members.

Suppose that the grand coalition N were acting so as to solve the primal problem (3.2), for some λ. In this context, suppose that we imposed some small positive probability that, in state t_N, all players j not in S might act alone, getting zero utility (by Assumption B), and leaving coalition S to carry out its threat in D_S. We may ask, how much would be lost, in terms of the objective function in (3.2), if such an event were given postive probability? This question may be formalized as follows. Suppose that S threatened to use action d_S when t_S are its members' types. Choose $d^0_{N\setminus S}$ in $\times_{j\in N\setminus S} D_{\{j\}}$. (All such $d^0_{N\setminus S}$ are equivalent, by Assumptions A and B.) Consider the constraint in the primal problem

(4.1) $\quad \delta_N((d_S,d^0_{N\setminus S}) \mid t_N) \geq 0.$

Compelling the members of S to carry out the threat d_S with positive probability is equivalent to increasing the right-hand side of (4.1) above zero. So we want to know, what is the shadow cost for the right-hand side of constraint (4.1) in the linear program (3.2)? The higher this shadow cost is, the less power we should attribute to S.

To compute shadow prices, we must look at the dual. Given β which solves the dual (3.3) - (3.5), let

(4.2)
$$v_S(d_S,t_N,\lambda,\beta) = [\sum_{i\in S} \lambda_{t_i} Y^S_i(d_S,t_N)$$

$$+ \sum_{i\in S} \sum_{\hat{t}_i\in T_i} \beta(t_i,\hat{t}_i)\, Y^S_i(d_S,\, t_N)$$

$$- \sum_{i\in S} \sum_{\hat{t}_i\in T_i} \beta(\hat{t}_i,t_i)\, Y^S_i(d_S,(t_{N-i},\hat{t}_i))].$$

The shadow cost of the nonnegativity constraint for $\delta_N((d_S,d^0_{N\setminus S}),t_N)$ in the primal is just the slack in the dual constraint

$$\alpha(t_N) \geq v_N((d_S,d^0_{N\setminus S}),t_N,\lambda,\beta) = v_S(d_S,t_N,\lambda,\beta).$$

So the shadow cost of carrying out the S-threat d_S in state t_N is $\alpha(t_N) - v_S(d_S,t_N,\lambda,\beta)$. Coalition S should choose its threat so as to minimize this shadow cost, or to maximize $v_S(d_S,t_N,\lambda,\beta)$.

The threat for coalition S can only depend on the information (t_S) available to the members of S. Thus, the optimal threat δ_S for coalition S should be chosed so as to maximize

$$\sum_{t_N \varepsilon T_N} \sum_{d_S \varepsilon D_S} \delta_S(d_S|t_S)\ v_S(d_S,t_N,\lambda,\beta)$$

Equivalently, we say that δ_S is an <u>optimal threat</u> for S iff, for every t_S in T_S,

$$\text{(4.3)} \qquad \sum_{d_S \varepsilon D_S} \delta_S(d_S|t_S)\ \Big(\sum_{t_{N\setminus S} \varepsilon T_{N\setminus S}} v_S(d_S,t_N,\lambda,\beta)\Big)$$

$$= \underset{d_S \varepsilon D_S}{\text{maximum}} \Big[\sum_{t_{N\setminus S} \varepsilon T_{N\setminus S}} v_S(d_S,t_N,\lambda,\beta)\Big]$$

For the case of S = N, when δ_N solves the primal and β solves the dual, then (4.3) must hold, as we remarked at the end of Section 3.

Given any vector of coalitions' threats $\delta = (\delta_S)_{S\varepsilon CL}$ in Δ, we let

$$\text{(4.4)} \qquad V_S(\delta,t_N,\lambda,\beta) = \sum_{d_S \varepsilon D_S} \delta_S(d_S|t_S)\ v_S(d_S,t_N,\lambda,\beta).$$

Then $V(\delta,t_N,\lambda,\beta) = (V_S(\delta,t_N,\lambda,\beta))_{S\varepsilon CL}$ is a characteristic function game, attributing to each coalition S the worth of its threat in state t_N, as measured by the utility weights λ and the dual variables β. We let $\phi(.)$

denote the Shapley [1953] value. The Shapley value $\phi_i(V(\delta,t_N,\lambda,\beta))$ provides a natural measure of player i's cooperative contribution in state t_N. With this measure, the total expected contribution of player i if he is of type t_i is $\sum_{t_{N-i}\in T_{N-i}} \phi_i(V(\delta,t_N,\lambda,\beta))$. Thus, a natural fairness condition is that each player's expected utility for each of his types, when weighted by the appropriate λ_{t_i} factor, should equal his expected cooperative contribution in this type, measured with respect to λ and β. That is, we say that δ_N is <u>fair</u> for type t_i of player i (with respect to λ and β and the other threats in δ) iff

$$(4.5) \qquad \lambda_{t_i} \bar{U}_{t_i}(\delta_N) = \sum_{t_{N-i}\in T_{N-i}} \phi_i(V(\delta,t_N,\lambda,\beta)).$$

We say that (δ,λ,β) is a <u>cooperative solution</u> for the game (2.1) satisfying Assumption A and B iff $\lambda\epsilon\Lambda$, δ_N solves the primal problem (3.2) for λ, β solves the dual problem (3.3.)-(3.5) for λ, each δ_S for $S \subset N$ satisfies the optimal-threat condition (4.3), and the fairness condition (4.5) is satisfied for every type t_i of every player i.

<u>Proposition.</u> Solutions as defined above exist, for any game (2.1) satisfying Assumptions A and B.

We omit the proof of the proposition, since the theorem in Section 7 will be strictly more general. However, a number of other facts about the solutions do merit discussion now, before we go to the general case.

In the case of n = 2, the fairness condition (4.5) for our solutions reduces to

$$\lambda_{t_i} \bar{U}_{t_i}(\delta_N) = \sum_{t_{N-i}} V_N(\delta,t_N,\lambda,\beta)/2 = \sum_{t_{N-i}} \alpha(t_N)/2,$$

which is the same as the fairness condition used in the two-person bargaining solution derived axiomatically in Myerson [1980].

Given any threats-vector δ in Δ, we let (δ_{-t_S}, d_S) denote the threats-vector in which differs from δ only in that coalition S changes its threat to d_S when its members' types are t_S. That is, $\hat{\delta} = (\delta_{-t_S}, d_S)$ iff

(4.6) $$\hat{\delta}_R(\hat{d}_R|\hat{t}_R) = \delta_R(\hat{d}_R|\hat{t}_R), \text{ if } R \neq S \text{ or if } \hat{t}_R \neq t_S,$$

$$= 1, \text{ if } R = S \text{ and } \hat{t}_R = t_S \text{ and } \hat{d}_R = d_S,$$

$$= 0, \text{ if } R = S \text{ and } \hat{t}_R = t_S \text{ and } \hat{d}_R \neq d_S.$$

Notice that the criterion for optimal threats (4.3) implies that, for any player i in S, t_S in T_S, and d_S in D_S,

(4.7) $$\sum_{t_{N\setminus S}\in T_{N\setminus S}} \phi_i(V(\delta, t_N, \lambda, \beta))$$

$$\geq \sum_{t_{N\setminus S}\in T_{N\setminus S}} \phi_i(V((\delta_{-t_S}, d_S), t_N, \lambda, \beta))$$

This is true because changing $\delta_S(t_S)$ from the optimal threat can only reduce $\sum_{t_{N\setminus S}} V_S(\delta, t_N, \lambda, \beta)$, and the Shapley value for player i is an increasing linear function of the worth of coalition S. Thus, the fairness condition justifies our definition of optimal threats because, for any player i in coalition S, the optimal threat for S maximizes player i's fair payoff demand, as measured by the right-hand side of (4.5).

5. Partition function games.

Before considering the general case, without Assumptions A and B, we must review some basic ideas about partition function games.

An embedded coalition is any pair (S,Q) such that Q is a partition of N and S is one of the coalitions in Q. We let ECL denote the set of embedded coalitions.

A partition function game is any vector w in $\mathbb{R}^{ECL}$, where the (S,Q)-component $w_{S,Q}$ is interpreted as the worth of coalition S if it forms and the other players align themselves into coalition as described by the partition Q. We shall need to consider partition function games (rather than characteristic function games) because, when we drop Assumption A, the payoffs to members of coalition S may depend on which coalitions among the other players are carrying out their threats.

We let $\Phi: \mathbb{R}^{ECL} \to \mathbb{R}^N$ denote the value for partition function, derived in Myerson [1977] as the natural extension of the Shapley value to $\mathbb{R}^{ECL}$. In Myerson [1978] this partition function value was also derived as a fair settlement function for cooperative strategic-form games with endogenous threats.

For any two partitions Q and $\hat{Q}$, let $Q \wedge \hat{Q} = \{S \cap \hat{S} | S \in Q, \hat{S} \in \hat{Q}, \text{ and } S \cap \hat{S} \neq \emptyset\}$. For any (S,Q) in ECL, let $w^{S,Q}$ be the partition function game such that

$$(w^{S,Q})_{\hat{S},\hat{Q}} = 1, \text{ if } \hat{S} \supseteq S \text{ and } \hat{Q} \wedge Q = Q,$$
$$= 0, \text{ otherwise.}$$

$\Phi(.)$is the linear mapping from $\mathbb{R}^{ECL}$ to $\mathbb{R}^N$ such that, for any $w^{S,Q}$ as above,

$$\Phi_i(w^{S,Q}) = 1/|S|, \text{ if } i \in S,$$
$$= 0, \text{ if } i \notin S.$$

Since these $w^{S,Q}$ games form a basis for $\mathbb{R}^{ECL}$, this result completely characterizes Φ.

For any partition Q, we say that a partition function game w is <u>Q-decomposable</u> iff, for any $(\hat{S},\hat{Q})$ in ECL,

$$w_{\hat{S},\hat{Q}} = \sum_{S\varepsilon Q} w_{\hat{S}\cap S,\hat{Q}\wedge Q}.$$

(We use the convention $w_{\emptyset,Q} = 0$ here.) For any player i, we say that w is <u>i-decomposable</u> iff w is $\{N-i,\{i\}\}$-decomposable. If w is an i-decomposable partition function game, then the value Φ gives player i what he can get by himself against the coalition $N-i$, that is,

(5.1) $$\Phi_i(w) = w_{\{i\},\{N-i,\{i\}\}}.$$

For any partition function game w, Φ divides all the worth of N among the players; that is,

(5.2) $$\sum_{i\varepsilon N} \Phi_i(w) = w_{N,\{N\}}.$$

The partition function value can be rewritten in the form

(5.3) $$\Phi_i(w) = \sum_{S\supseteq\{i\}} f_S(w), \quad \forall i\varepsilon N, \ \forall w\varepsilon \mathbb{R}^{ECL},$$

where each $f_S\colon \mathbb{R}^{ECL} \to \mathbb{R}$ is a linear function which depends only on the components $w_{R,Q}$ such that $S \varepsilon Q$.

These are the basic properties of Φ which we shall need in this paper. For more detailed discussion and proofs, see Myerson [1977] and Myerson [1978].

6. The general case.

We now return to the general cooperative game with incomplete information (2.1), dropping the special Assumptions A and B used in Section 4.

For any partition Q of the set of players, we let

$$D_Q = \underset{S\varepsilon Q}{X} D_S \subseteq D_N.$$

That is D_Q is the set of all actions which would be feasible if the coalitions in Q acted separately. We may use $d_Q = (d_S)_{S\varepsilon Q}$ to denote a typical action in D_Q.

Given any vector of threats $\delta = (\delta_S)_{S\varepsilon CL}$ in Δ, and given any partition Q, we let δ_Q be the randomized strategy in Δ_N such that

$$(6.1) \qquad \delta_Q(d_Q|t_N) = \prod_{S\varepsilon Q} \delta_S(d_S|t_S) \quad \text{if } d_Q = (d_S)_{S\varepsilon Q} \varepsilon\, D_Q,$$

$$\text{and } \delta_Q(d_N|t_N) = 0 \quad \text{if } d_N \not\varepsilon\, D_Q.$$

That is, δ_Q is the strategy (for N) in which the players break up into the coalitions of Q and carry out their δ_S threats separately.

We let $X_i(\delta,\hat{t}_i,t_N)$ denote player i's evaluation of the prospect of using strategy $\delta_{\{i\}}(\cdot|\hat{t}_i)$ against the coalition $N-i$ when t_N are the players' types. That is, for any i in N, δ in Δ, t_N in T_N, and $\hat{t}_i$ in T_i,

$$(6.2) \qquad X_i(\delta,\hat{t}_i,t_N) = \bar{Y}_i(\delta_{\{N-i,\{i\}\}}(t_{N-i},t_N)$$
$$= \sum_{d_{N-i}} \sum_{d_i} \delta_{N-i}(d_{N-i}|t_{N-i})\, \delta_{\{i\}}(d_i|\hat{t}_i)\, Y_i((d_{N-i},d_i),t_N)$$

where d_{N-i} ranges over D_{N-i} and d_i ranges over $D_{\{i\}}$.

The characteristic function game $V(\delta,t_N,\lambda,\beta)$ was the central feature of the model in Section 4, and finding the correct way to generalize this structure is a subtle problem. Without Assumption A, the payoffs to members of a coalition may depend on which other coalitions form, so it is natural to look for a partition function structure to generalize $V(\delta,t_N,\lambda,\beta)$. In order to prove existence and individual rationality of our cooperative solutions we shall need to use the following definition:

(6.3) $W_{S,Q}(\delta,t_N,\lambda,\beta) =$

$$= \sum_{i\varepsilon S} \sum_{d_Q \varepsilon D_Q} \delta_Q(d_Q|t_N) \, [\lambda_{t_i} \, Y_i(d_Q,t_N)$$

$$+ \sum_{\hat{t}_i \varepsilon T_i} \beta(t_i,\hat{t}_i) \, (Y_i(d_Q,t_N) - X_i(\delta,\hat{t}_i,t_N))$$

$$- \sum_{\hat{t}_i \varepsilon T_i} \beta(\hat{t}_i,t_i) \, (Y_i(d_Q,(t_{N-i},\hat{t}_i)) - X_i(\delta,t_i,(t_{N-i},\hat{t}_i))]$$

for any (S,Q) in ECL, any δ in Δ, any any t_N in T_N. It is straightforward to check that this definition generalizes the construction of $V_S(\delta,t_N,\lambda,\beta)$ in (4.2) and (4.4). (Notice that the X_i terms are zero when Assumption B holds.) The term $Y_i(d_Q,t_N) - X_i(\delta,\hat{t}_i,t_N)$ measures how much i would gain from d_Q over the alternative of claiming to be type $\hat{t}_i$ and acting alone against N-i when t_N is the true vector of types; and this gain is weighted by $\beta(t_i,\hat{t}_i)$, the shadow price for keeping t_i from claiming to be $\hat{t}_i$. The term $Y_i(d_Q,(t_{N-i},\hat{t}_i)) - X_i(\delta,t_i,(t_{N-i},\hat{t}_i))$ measures how much i would gain from d_Q over the alternative of claiming to be type t_i and acting alone against N-i when i is really type $\hat{t}_i$ and the others' types are t_{N-i}; and this term is weighted by $\beta(\hat{t}_i,t_i)$, the shadow price of keeping $\hat{t}_i$ from claiming to be t_i. Thus, the bracketed expression in (6.3) may be interpreted as the net worth to player i of d_Q in state t_N, when we take into account both the direct payoff contribution to i $(\lambda_{t_i} Y_i(d_Q,t_N))$ and the indirect contribution to helping player i to satisfy his incentive-compatibility constraints.

We can now generalize our fairness condition (4.5). We say that δ_N is <u>fair</u> for type t_i of player i (with respect to λ and β and the other threats in δ) iff

$$(6.4) \qquad \lambda_{t_i} \bar{U}_{t_i}(\delta_N) = \sum_{t_{N-i} \varepsilon T_{N-i}} \Phi_i(W(\delta, t_N, \lambda, \beta))$$

The left-hand side of (6.4) is the weighted utility payoff to player i if t_i is his type; and the right-hand side (6.4) is a measure of the expected cooperative contribution of player i if t_i is his type, using the value of the partition function games $W(\delta, t_N, \lambda, \beta)$. We may refer to the right-hand side of (6.4) as the <u>fair demand</u> for type t_i.

We say that a threats-vector δ is <u>i-decomposable</u> in state t_N iff, for any coalition S such that $i \varepsilon S$:

$$\delta_S(d_S | t_S) = 0 \quad \text{if } d_S \notin D_{S-i} \times D_{\{i\}}, \qquad \text{and}$$

$$\delta_S((d_{S-i}, d_i) | t_S) = \delta_{S-i}(d_{S-i} | t_{S-i})\ \delta_{\{i\}}(d_i | t_i), \quad \forall d_{S-i} \varepsilon D_{S-i}, \quad \forall d_i \varepsilon D_{\{i\}}.$$

That is, if δ is i-decomposable then there is no coalition which would change its threat if i were to leave the coalition, so i is not effectively cooperating with any other players.

The following lemma summarizes three desirable properties of $\Phi(W(\delta, t_N, \lambda, \beta))$, which make it useful to us as measure of players' contributions.

Lemma 1. For any two players i and j who are both in coalition S, the difference $\Phi_i(W(\delta,t_N,\lambda,\beta)) - \Phi_j(W(\delta,t_N,\lambda,\beta))$ does not depend on the threat δ_S of coalition S. If δ_N solves the primal and β solves the dual for λ, then

(6.5) $$\sum_{i\in N} \sum_{t_N \in T_N} \Phi_i(W(\delta,t_N,\lambda,\beta)) = \sum_{t_i \in T_*} \lambda_{t_i} \bar{U}_{t_i}(\delta_N).$$

If δ is i-decomposable in state t_N then

(6.6) $$\Phi_i(W(\delta,t_N,\lambda,\beta)) =$$
$$= [\lambda_{t_i} X_i(\delta,t_i,t_N) + \sum_{\hat{t}_i \in T_i} \beta(t_i,\hat{t}_i)\,(X_i(\delta,t_i,t_N) - X_i(\delta,\hat{t}_i,t_N))]$$

(All proofs are in Section 8.)

Equation (6.5) guarantees that the sum of the fair demands of all types of all players is always equal to the total weighted-utility actually expected by all types of all players. (So (6.5) is a sort of "Walras' Law" for our bargaining model.) Equation (6.6) asserts that, if a player is not effectively cooperating with any other players in state t_N, then his cooperative contribution in that state is just the weighted-utility payoff he gets alone plus the value (for incentive compatibility) of his not being tempted to pretend that some other type $\hat{t}_i$ is true when he acts alone. ($[X_i(\delta,t_i,t_N) - X_i(\delta,\hat{t}_i,t_N)]$ is player i's gain in state t_N from honestly implementing his threat $\delta_{\{i\}}(\cdot|t_i)$ alone against N-i, rather than dishonestly using his threat $\delta_{\{i\}}(\cdot|\hat{t}_i)$, intended for a different state.)

The first sentence of Lemma 1 assures us that all members of coalition S can agree about which threat δ_S is best, because maximizing $\Phi_i(W(\delta,t_N,\lambda,\beta))$ with respect to δ_S (keeping all other δ_R fixed, for $R \neq S$) is equivalent to maximizing $\Phi_i(W(\delta,t_N,\lambda,\beta))$ with respect to δ_S. Equivalently, all players in S would agree that $\delta_S(\cdot|t_S)$ should be chosen so as to maximize

$$\sum_{t_{N\setminus S}\varepsilon T_{N\setminus S}} \sum_{i\varepsilon S} \Phi_i(W(\delta,t_N,\lambda,\beta)),$$

if each player wants to maximize his fair demand.

We can now generalize the concept of <u>optimal threats</u> from Section 4, except that now it is better to speak of a threat-equilibrium, since the optimality criterion of one coalition may depend on what the others plan to do. Given λ and β, we say that δ in Δ is a <u>threat-equilibrium</u> iff, for any coalition S and any t_S in T_S,

$$(6.7) \qquad \sum_{t_{N\setminus S}} \sum_{i\varepsilon S} \Phi_i(W(\delta,t_N,\lambda,\beta))$$

$$= \max_{d_S\varepsilon D_S} \sum_{t_{N\setminus S}} \sum_{i\varepsilon S} \Phi_i(W((\delta_{-t_S},d_S),\ t_N,\lambda,\beta)).$$

where $t_{N\setminus S}$ ranges over $T_{N\setminus S}$. (Recall (4.6).) This optimality criterion (6.7) is equivalent to (4.7) in the special case of Section 4. It can be shown that, if δ_N solves the primal and β solves the dual for λ, then (6.7) must hold when S=N, for any t_N. (We omit the summation over $t_{N\setminus S}$ when S=N).

We say that (δ,λ,β) is a <u>cooperative solution</u> for the game (2.1) iff $\lambda\varepsilon\Lambda$, β solves the dual problem (3.3)-(3.5) for λ, δ_N solves the primal problem (3.2) for λ, each δ_S for $S \subset N$ satisfies the threat-equilibrium condition (6.7) for all t_S in T_S, and the fairness condition (6.4) is satisfied for every type t_i of every player i. Notice that there are as many fairness equations in (6.4) as there are components of the λ vector. Of course (6.5) guarantees that one fairness equation is redundant, but we have also removed one degree of freedom by assuming that λ is in the simplex Λ. One may think of the fairness equations as determining the λ weights.

Thus, our cooperative solution concept is well-determined, in the sense of having as many independent equations as variables.

When Assumptions A and B hold, the definitions of cooperative solutions given in this section and in Section 4 are equivalent. However, before we state and prove the existence theorem, we must define a further refinement of this solution concept in Section 7.

7. Proper solutions and individual rationality.

In order to guarantee existence of cooperative solutions we must allow for the possibility that some of the λ_{t_i} factors may equal zero. (Recall that Λ is the entire closed unit simplex in $\mathbb{R}^{T_*}$.) However, if $\lambda_{t_i} = 0$ then the fairness equation (6.4) imposes no restriction on $\bar{U}_{t_i}(\delta_N)$. In fact, if we added a dummy player to any game, and gave λ-weights equal to zero to every player except the dummy, then any feasible outcome could be a solution. This problem is not special to the incomplete information case; it arises in the λ-transfer values of games with complete information just as easily.

We say that (δ,λ,β) is a semi-solution for the game (2.1) iff $\lambda \varepsilon \Lambda, \beta$ solves the dual problem, δ_N solves the primal problem for λ, and δ is a threat-equilibrium for λ and β. (That is, a semi-solution lacks only the fairness condition to be a solution.) We say that a cooperative solution (δ,λ,β) is proper iff there exists some sequence of semi-solutions $\{(\delta^k,\lambda^k,\beta^k)\}_{k=0}^{\infty}$ such that

(7.1) $\quad \lambda^k_{t_i} > 0 \quad \forall t_i \;\varepsilon\; T_*, \; \forall k,$

(7.2) $\quad \lim\limits_{k\to\infty} (\delta^k,\lambda^k,\beta^k) = (\delta,\lambda,\beta),$

and

(7.3) $$\bar{U}_{t_i}(\delta_N) = \lim_{k\to\infty} \bar{U}_{t_i}(\delta^k_N) \geq \lim_{k\to\infty} \Big(\sum_{t_{N-i}\varepsilon T_{N-i}} \Phi_i(W(\delta^k,t_N,\lambda^k,\beta^k))/\lambda^k_{t_i} \Big),$$

$$\forall i \varepsilon N, \quad \forall t_i \varepsilon T_i.$$

That is, each types's actual expected utility must be at least the limit of its fair demands, when measured in unweighted utility units.

In the special case of two-person bargaining problems, the bargaining solutions derived in Myerson [1980] are precisely these proper cooperative solutions.

We say that a cooperative solution (δ,λ,β) is individually rational iff, for every i in N, every t_i in T_i, and every d_i in $D_{\{i\}}$,

$$(7.4) \qquad \bar{U}_{t_i}(\delta_N) \geq \sum_{t_{N-i}} \sum_{d_{N-i}} \delta_{N-i}(d_{N-i}|t_{N-i})\, Y_i((d_{N-i},d_i),t_N).$$

The right-hand side of (7.4) is the expected utility for player i if his type is t_i and he uses the action d_i against the threat δ_{N-i} of coalition N-i. So the solution is individually rational iff no player, in any type, could expect to do better alone than he can do by participating in the grand coalition N.

We can now state the main result of this paper.

Theorem There exists a proper cooperative solution for any game of the form (2.1). Furthermore, any proper cooperative solution is individually rational.

8. Proofs

Proof of Lemma 1. To prove (6.5), we first use (5.2) to get

$$\sum_{t_N} \sum_{i\varepsilon N} \Phi_i(W(\delta,t_N,\lambda,\beta)) = \sum_{t_N} W_{N,\{N\}}(\delta,t_N,\lambda,\beta).$$

We can expand $W_{N,\{N\}}$ using the definition (6.3) and rearrange the sums to get

$$\sum_{t_N} W_{N,\{N\}}(\delta,t_N,\lambda,\beta) = \sum_{i\varepsilon N} \sum_{t_N} [\lambda_{t_i} \bar{Y}_i(\delta_N(t_N),t_N)$$
$$+ \sum_{\hat{t}_i\varepsilon T_i} \beta(t_i,\hat{t}_i)\ (\bar{Y}_i(\delta_N(t_N),t_N) - X_i(\delta,\hat{t}_i,t_N))$$
$$- \sum_{\hat{t}_i\varepsilon T_i} \beta(t_i,\hat{t}_i)\ (\bar{Y}_i(\delta_N(t_{N-i},\hat{t}_i),t_N) - X_i(\delta,\hat{t}_i,t_N))].$$

(The third term in this sum follows from (6.3) by reversing the roles of t_i and $\hat{t}_i$.) If δ_N solves the primal and β solves the dual, then

$$\beta(t_i,\hat{t}_i) \sum_{t_{N-i}} (Y_i(\delta_N(t_N),t_N) - Y_i(\delta_N(t_{N-i},\hat{t}_i),t_N)) = 0$$

by complementary slackness. Thus, since the X_i terms cancel out, (6.5) follows.

If δ is i-decomposable, then $W(\delta,t_N,\lambda,\beta)$ must be an i-decomposable partition function game, and so (6.6) follows from (5.1), (6.3), and (6.2).

Let $\hat{W}(\delta,t_N,\lambda,\beta)$ be the partition function game such that, for all (S,Q) in ECL

$$\hat{W}_{S,Q}(\delta,t_N,\lambda,\beta) = \sum_{i\varepsilon S} [\lambda_i\ \bar{Y}_i(\delta_Q(t_N),t_N)$$

$$+ \sum_{\hat{t}_i\varepsilon T_i} (\beta(t_i,\hat{t}_i)\bar{Y}_i(\delta_Q(t_N),t_N) - \beta(\hat{t}_i,t_i)\bar{Y}_i(\delta_Q(t_N),(t_{N-i},\hat{t}_i)))]$$

That is $\hat{W}$ is the same as W, except that the X_i terms are omitted. Using the linearity and decomposibility properties of Φ, we get

$$(8.1) \qquad \Phi_i(W(\delta,t_N,\lambda,\beta)) = [\Phi_i(\hat{W}(\delta,t_N,\lambda,\beta)) + \sum_{\hat{t}_i \in T_i} (\beta(t_i,\hat{t}_i)X_i(\delta,\hat{t}_i,t_N) - \beta(\hat{t}_i,t_i)X_i(\delta,t_i,(t_{N-i},\hat{t}_i)))],$$

because $W(\delta,t_N,\lambda,\beta)-\hat{W}(\delta,t_N,\lambda,\beta)$ is a completely decomposable partition function game.

To prove the first sentence of Lemma 1, use (5.3) to get

$$\Phi_i(\hat{W}(.))-\Phi_j(\hat{W}(.)) = \sum_{R \supseteq \{i\}} f_R(\hat{W}(.)) - \sum_{R \supseteq \{j\}} f_R(\hat{W}(.)).$$

If $\{i,j\} \subseteq S$, then δ_S only influences $\Phi_i(\hat{W}(.))$ and $\Phi_j(\hat{W}(.))$ through the $f_S(\hat{W}(.))$ terms, which cancel out in the difference. The X_i terms in (8.1) depend only on δ_{N-i} and $\delta_{\{i\}}$, and thus can not depend on δ_S if $\{i,j\} \subseteq S$. This completes the proof of the lemma.

<u>Proof of the Theorem.</u> We begin with some definitions. Let $T_*^2 = \bigcup_{i \in N} (T_i \times T_i)$. There exists some number $M > 0$ such that, for every λ in Λ, there exists some β in $\mathbb{R}_+^{T_*^2}$ such that β is an optimal solution to the dual for λ and $\|\beta\| \leq M$. To prove this fact, observe first that the feasible set for the primal is compact and is independent of λ. So the simplex Λ can be covered by a finite collection of closed convex regions (each region corresponding to an optimal basis for the primal) such that, within each region, an optimal solution to the dual can be given as a linear function of λ. Each of these linear mappings is bounded on its compact domain; let M be

the smallest bound which works for all of the mappings. We let

$$B = \{\beta \in \mathbb{R}_+^{T_*^2} \mid \|\beta\| \le M\}.$$

Then B is a compact domain within which we can always find optimal solutions for the dual.

For any player i and any t_i in T_i, let

$$\bar{\Phi}_{t_i}(\delta,\lambda,\beta) = \sum_{t_{N-i} \varepsilon T_{N-i}} \Phi_i(W(\delta,t_N,\lambda,\beta)).$$

So $\bar{\Phi}_{t_i}(\delta,\lambda,\beta)$ is the fair demand of type t_i.

Let $G_{t_i}(\delta_{N-i})$ be the highest expected utility which type t_i of player i can get against strategy δ_{N-i}, that is

$$G_{t_i}(\delta_{N-i}) = \underset{d_i \varepsilon D_{\{i\}}}{\text{maximum}} \sum_{t_{N-i}} \sum_{d_{N-i}} \delta_{N-i}(d_{N-i}|t_{N-i})\, Y_i((d_{N-i},d_i),t_N).$$

For any semi-solution (δ,λ,β) and any t_i in T_*,

$$(8.2) \qquad \bar{\Phi}_{t_i}(\delta,\lambda,\beta) \ge \lambda_{t_i}\, G_{t_i}(\delta_{N-i}).$$

That is, the fair demand of any type is not less than the weighted-utility which the type can get alone against the threat of the complementary coalition. To prove claim (8.2), suppose that it were not true for type t_i of player i. Let $\hat{\delta}$ be such that $\hat{\delta}_{\tilde{S}} = \delta_{\tilde{S}}$ if $i \notin \tilde{S}$, $\hat{\delta}_S(\hat{t}_S) = \delta_S(\hat{t}_S)$ if $i \in S$ but $\hat{t}_i \ne t_i$, $\hat{\delta}_{\{i\}}(\bar{d}_i|t_i) = 1$ for some $\bar{d}_i$ which achieves the maximum in the definition of $G_{t_i}(\delta_{N-i})$, and $\hat{\delta}$ is i-decomposible whenever i is type t_i.

Using (6.6),

$$\bar{\Phi}_{t_i}(\hat{\delta},\lambda,\beta) = \sum_{t_{N-i}} \Phi_i(W(\hat{\delta},t_N,\lambda,\beta))$$

$$= \sum_{t_{N-i}} [\lambda_{t_i} X_i(\hat{\delta},t_i,t_N) + \sum_{\hat{t}_i} \beta(t_i,\hat{t}_i)\ (X_i(\hat{\delta},t_i,t_N) - X_i(\hat{\delta},\hat{t}_i,t_N))]$$

$$= \lambda_{t_i} G_{t_i}(\delta_{N-i}) + \sum_{\hat{t}_i} \beta(t_i,\hat{t}_i)\ (G_{t_i}(\delta_{N-i}) - \sum_{t_{N-i}} X_i(\hat{\delta},\hat{t}_i,t_N))$$

$$\geq \lambda_{t_i} G_{t_i}(\delta_{N-i}).$$

So if (8.2) is violated, then $\bar{\Phi}_{t_i}(\hat{\delta},\lambda,\beta) > \bar{\Phi}_{t_i}(\delta,\lambda,\beta)$. But $\hat{\delta}$ differs from δ only in that coalitions including player i in type t_i have changed their threats. Thus, there must exist at least one coalition S such that $i\varepsilon S$ and the change from δ_S to $\hat{\delta}_S$ is strictly prefered by i, even if all other coalitions' threats remained fixed. (We use here the fact that $\bar{\Phi}_{t_i}$ is additively separable with respect to the threats of coalitions containing i, by (5.3).) But by Lemma 1, all members of S must also agree with i that this change would improve their fair demand, which contradicts the fact that δ is a threat equilibrium. Thus (8.2) must hold.

Notice that (7.3) and (8.2) immediately imply the individual-rationality condition (7.4), so all proper solutions are individually rational.

For any $k > |T_*|$, we let

$$\Lambda^k = \{\lambda\varepsilon\Lambda |\ \lambda_{t_i} \geq \frac{1}{k},\ \forall t_i\ \varepsilon\ T_*\}.$$

We now define a series of Kakutani mappings on $\Delta \times \Lambda^k \times B$.

We define $Z_1\colon \Delta \times \Lambda \times B \Rightarrow \Delta$ so that $\hat{\delta}\ \varepsilon\ Z_1(\delta,\lambda,\beta)$ iff $\hat{\delta}_N$ solves the primal for λ, and for every $S \subset N$ and every t_S in T_S,

$$\sum_{d_S\varepsilon D_S} \hat{\delta}_S(d_S|t_S) \sum_{t_{N\setminus S}} \sum_{i\varepsilon S} \Phi_i(W((\delta_{-t_S},d_S),t_N,\lambda,\beta))$$

$$= \max_{d_S\varepsilon D_S} \sum_{t_{N\setminus S}} \sum_{i\varepsilon S} \Phi_i(W((\delta_{-t_S},d_S),t_N,\lambda,\beta))$$

We define $Z_2^k: \Delta \times \Lambda^k \times B \Rightarrow \Lambda^k$ so that $\hat{\lambda} \in Z_2^k(\delta,\lambda,\beta)$ iff, for every t_i in T_*,

$$\text{if } t_i \notin \underset{\hat{t}_j \in T_*}{\operatorname{argmax}} \left[(\bar{\Phi}_{\hat{t}_j}(\delta,\lambda,\beta)/\lambda_{\hat{t}_j}) - \bar{U}_{\hat{t}_j}(\delta_N)\right] \quad \text{then } \hat{\lambda}_{t_i} = \tfrac{1}{k}.$$

We define $Z_3: \Lambda \Rightarrow B$ so that $\hat{\beta} \in Z_3(\lambda)$ iff $\hat{\beta} \in B$ and β solves the dual for λ.

We define $Z^k: \Delta \times \Lambda^k \times B \Rightarrow \Delta \times \Lambda^k \times B$ by

$$Z^k(\delta,\lambda,\beta) = Z_1(\delta,\lambda,\beta) \times Z_2^k(\delta,\lambda,\beta) \times Z_3(\lambda).$$

It is straightforward to check that Z^k is nonempty-valued, convex-valued, and upper-semicontinuous. (We use the construction of M and B to guarantee $Z_3(\lambda) \neq \emptyset$.) $\Delta \times \Lambda \times B$ is compact and convex, so by the Kakutani Fixed-Point Theorem we can find some $(\delta^k,\lambda^k,\beta^k)$ such that

$$(\delta^k,\lambda^k,\beta^k) \in Z^k(\delta^k,\lambda^k,\beta^k).$$

For each k, $(\delta^k,\lambda^k,\beta^k)$ is a semisolution, by the definitions of the Z_1 and Z_3 correspondences. Furthermore, using the definition of Z_2^k, we get

(8.3)
$$\begin{aligned}
&\max_{t_i \in T_*} ((\bar{\Phi}_{t_i}(\delta^k,\lambda,\beta^k)/\lambda^k_{t_i}) - \bar{U}_{t_i}(\delta^k_N)) \\
&\leq |T_*| \max_{t_i \in T_*} (\bar{\Phi}_{t_i}(\delta^k,\lambda^k,\beta^k) - \lambda^k_{t_i}\bar{U}_{t_i}(\delta^k_N)) \\
&\leq |T_*|^2 \max_{t_i \in T_*} (\lambda^k_{t_i}\bar{U}_{t_i}(\delta^k_N) - \bar{\Phi}_{t_i}(\delta^k,\lambda^k,\beta^k)) \\
&\leq \frac{|T_*|^2}{k} \max_{t_i \in T_*} (\bar{U}_{t_i}(\delta^k_N) - (\bar{\Phi}_{t_i}(\delta^k,\lambda^k,\beta^k)/\lambda^k_{t_i})) \\
&\leq \frac{|T_*|^2}{k} \max_{t_i \in T_*} (\bar{U}_{t_i}(\delta^k_N) - G_{t_i}(\delta^k_{N-i})) \\
&\leq \frac{M_* \, |T_*|^2}{k}
\end{aligned}$$

where $M_* = \max_{\delta \in \Delta} \max_{t_i \in T_*} (\bar{U}_{t_i}(\delta_N) - G_{t_i}(\delta_{N-i}))$. In this chain of inequalities, the second line uses the fact that we must have $\lambda_{t_i} \geq \frac{1}{|T_*|}$ for some t_i achieving the maximum in the first line. The third line follows from equation (6.5) ("Walras' Law"). The fourth line follows from the fact that $\lambda^k_{t_i}$ must equal $1/k$ for t_i achieving the maximum on the third line. The fifth line uses (8.2). Continuity of $\bar{U}_{t_i}$ and G_{t_i} and compactness of Δ guarantee that M_* is finite.

Since $\Delta \times \Lambda \times B$ is compact, we can find some subsequence of the $(\delta^k,\lambda^k,\beta^k)$ converging to some point (δ,λ,β). We now show that this (δ,λ,β) is a proper solution. δ is a threat equilibrium and β solves the dual for λ, because each δ^k and β^k have these properties for λ^k (and Z_1 and Z_3 are upper-semicontinuous). Condition (7.3) follows from (8.3) as $k \to \infty$. (7.3) also implies

$$\lambda_{t_i}\bar{U}_{t_i}(\delta_N) \geq \bar{\Phi}_{t_i}(\delta,\lambda,\beta),$$

and (6.5) guarantees that none of these inequalities can be strict. So (δ,λ,β) is a proper solution.

Q.E.D.

References

Harsnyi, J. C. [1963], "A Simplified Bargaining Model for the n-Person Cooperative Game," International Economic Review 4, 194-220.

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

Myerson, R. B. [1977], "Values of Games in Partition Function Form," International Journal of Game Theory 6, 23-31.

Myerson, R. B. [1978], "Threat Equilibria and Fair Settlements in Cooperative Games," Mathematics of Operations Research 3, 265-274.

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

Myerson, R. B. [1980], "Solutions for Two-Person Bargaining Problems with Incomplete Information," Center for Math. Studies, DP No. 432, Northwestern University.

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

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

Roth, A. E. [1980], "Values for Games Without Sidepayments: Some Difficulties with Current Concepts,"Econometrica 48, 457-465.

Shafer, W. J. [1980], "On the Existence and Interpretation of Value Allocation," Econometrica 48, 467-476.

Shapley, L. S. [1953], "A Value for n-Person Games," in Contributions to the Theory of Games II, H. W. Kuhn and A, W, Tucker eds., Princeton: Princeton University Press, 307-317.

Shapley, L. S. [1969], "Utility Comparison and the Theory of Games," in La Décision, Paris: Edition du Centre National de la Recherche Scientifique, France, 251-263.

Shapley, L. S. and M. Shubik [1973], "Game Theory in Economics - Chapter 6: Characteristic Function, Core, and Stable Set," Santa Monica: Rand Corporation Report R-904-NSF/6.

specified, then t_i (for any player i) must denote the i^{th} component of t_N, and (for any $S \subset N$) $t_S = (t_j)_{j \varepsilon S}$. Conversely, whenever some type-vectors t_S in T_S and $t_{N\setminus S}$ in $T_{N\setminus S}$ (or some t_i and t_{N-i}) have been specified, then t_N must denote the type-vector which combines these two lists of the players' types.

For any t_N in T_N, we let $P_i(t_{N-i}|t_i)$ the conditional probability that t_{N-i} is the combination of types for all players other than i, as would be assessed by player i if t_i were his type. (We use the notation $N-i = N\setminus\{i\}$.)

Thus, a cooperative game with incomplete information is defined by these structures:

(2.1) $$((D_S)_{S\varepsilon CL},(T_i)_{i\varepsilon N},(U_i)_{i\varepsilon N},(P_i)_{i\varepsilon N}).$$

We let $\bar{D}_S$ denote the set of all probability distributions over D_S; that is,

$$\bar{D}_S = \{\gamma:D_S \to \mathbb{R}_+ | \sum_{d_S \varepsilon D_S} \gamma(d_S) = 1\},$$

where $\mathbb{R}_+$ denotes the nonnegative real numbers. A randomized strategy for coalition S is any mapping from T_S to $\bar{D}_S$. We let Δ_S denote the set of mixed strategies for S. Thus, for any δ_S in Δ_S, we have

$$\delta_S(t_S) \varepsilon \bar{D}_S, \forall t_S \varepsilon T_S,$$

or, equivalently,

(2.2) $$\sum_{d_S \varepsilon D_S} \delta_S(d_S|t_S) = 1 \text{ and } \delta_S(d_S'|t_S) \geq 0, \quad \forall t_S \varepsilon T_S, \quad \forall d_S' \varepsilon D_S.$$

In the strategy δ_S, $\delta_S(d_S|t_S)$ represents the probability of coalition S choosing the action d_S when its members' types are t_S.
In our notation, $\delta_S(t_S)$ is equivalent to $\delta_S(\cdot|t_S)$, for any δ_S in Δ_S.

We let $\Delta = \underset{S\varepsilon CL}{X} \Delta_S$. A typical point $\delta = (\delta_S)_{S\varepsilon CL}$ in Δ is to be interpeted as a vector of threats, describing what each coalition would do if it were to form. In this context, we do not need to assume any special relationship between the threats of two nondisjoint coalitions S and R, since primal objective function (3.2). The second term measures the value of d_N in state t_N for increasing incentive-compatibility, by rewarding each player i with $Y_i(d_N,t_N)$ if his type is t_i and he is honest. The third term measures the cost of d_N in state t_N for reducing incentive-compatibility, by rewarding each player i with $Y_i(d_N,(t_{N-i},\hat{t}_i))$ if he reports t_i when $\hat{t}_i$ is really his type (and all others report t_j honestly). Then $\alpha(t_N)$ equals the total contribution of the best decision in state t_N.

In some examples, it may be possible for some types to costlessly prove that other types are false. For example, if a person can play the piano, then he can prove that he is not a non-pianist simply by playing a few bars. On the other hand, the non-pianist cannot prove that he is not really a pianist, unless he is given the proper incentives.[1] In general, if player i could costlessly verify that t_i was false if $\hat{t}_i$ were true, then we must drop from the primal problem the constraint which says that i should not be tempted to report $\hat{t}_i$ if t_i were true, and we must set the dual variable $\beta(t_i,\hat{t}_i)$ equal to zero. With these revisions, our model can easily accomodate the case of verifiable types. Nevertheless, throughout the rest of this paper, we shall consider only the case of unverifiable types.

If there were no incentive-compatibility constraints, then the solution to the primal problem would be simple: for every t_N, $\delta_N(\cdot|t_N)$ would put all probability weight on the decisions which maximize $\sum_i \lambda_{t_i} Y_i(d_N,t_N)$. The incentive-compatibility constraints complicate matters by interconnecting the decisions in different states t_N. However, we can decompose the primal problem by using the dual variables to remove the incentive-compatibility constraints. That is, by the duality theorem of linear programming, if δ_N is optimal for the primal then, for every t_N in T_N, $\delta_N(\cdot|t_N)$ must put all probability weight on the decisions which achieve the maximum in (3.4).

1. I am indebted to Paul Milgrom for pointing out this issue.

We define $Z_2^k: \Delta \times \Lambda^k \times B \Rightarrow \Lambda^k$ so that $\hat{\lambda} \in Z_2^k(\delta,\lambda,\beta)$ iff,

for every t_i in T_*,

$$\text{if } t_i \notin \underset{\hat{t}_j \in T_*}{\text{argmax}} \left[(\bar{\Phi}_{\hat{t}_j}(\delta,\lambda,\beta)/\lambda_{\hat{t}_j}) - \bar{U}_{\hat{t}_j}(\delta_N)\right] \quad \text{then } \hat{\lambda}_{t_i} = \tfrac{1}{k}.$$

We define $Z_3: \Lambda \Rightarrow B$ so that $\hat{\beta} \in Z_3(\lambda)$ iff $\hat{\beta} \in B$ and β solves the dual for λ.

We define $Z^k: \Delta \times \Lambda^k \times B \Rightarrow \Delta \times \Lambda^k \times B$ by

$$Z^k(\delta,\lambda,\beta) = Z_1(\delta,\lambda,\beta) \times Z_2^k(\delta,\lambda,\beta) \times Z_3(\lambda).$$

It is straightforward to check that Z^k is nonempty-valued, convex-valued, and upper-semicontinuous. (We use the construction of M and B to guarantee $Z_3(\lambda) \neq \emptyset$.) $\Delta \times \Lambda \times B$ is compact and convex, so by the Kakutani Fixed-Point Theorem we can find some $(\delta^k,\lambda^k,\beta^k)$ such that

$$(\delta^k,\lambda^k,\beta^k) \in Z^k(\delta^k,\lambda^k,\beta^k).$$

For each k, $(\delta^k,\lambda^k,\beta^k)$ is a semisolution, by the definitions of the Z_1 and Z_3 correspondences. Furthermore, using the definition of Z_2^k, we get

$$
\begin{aligned}
(8.3)\qquad & \max_{t_i \in T_*} \left((\bar{\Phi}_{t_i}(\delta^k,\lambda,\beta^k)/\lambda^k_{t_i}) - \bar{U}_{t_i}(\delta^k_N)\right) \\
& \leq |T_*| \max_{t_i \in T_*} \left(\bar{\Phi}_{t_i}(\delta^k,\lambda^k,\beta^k) - \lambda^k_{t_i}\bar{U}_{t_i}(\delta^k_N)\right) \\
& \leq |T_*|^2 \max_{t_i \in T_*} \left(\lambda^k_{t_i}\bar{U}_{t_i}(\delta^k_N) - \bar{\Phi}_{t_i}(\delta^k,\lambda^k,\beta^k)\right) \\
& \leq \frac{|T_*|^2}{k} \max_{t_i \in T_*} \left(\bar{U}_{t_i}(\delta^k_N) - (\bar{\Phi}_{t_i}(\delta^k,\lambda^k,\beta^k)/\lambda^k_{t_i})\right) \\
& \leq \frac{|T_*|^2}{k} \max_{t_i \in T_*} \left(\bar{U}_{t_i}(\delta^k_N) - G_{t_i}(\delta^k_{N-i})\right) \\
& \leq \frac{M_* \, |T_*|^2}{k}
\end{aligned}
$$

where $M_* = \max_{\delta \in \Delta} \max_{t_i \in T_*} (\bar{U}_{t_i}(\delta_N) - G_{t_i}(\delta_{N-i}))$. In this chain of inequalities, the second line uses the fact that we must have $\lambda_{t_i} \geq \frac{1}{|T_*|}$ for some t_i achieving the maximum in the first line. The third line follows from equation (6.5) ("Walras' Law"). The fourth line follows from the fact that $\lambda^k_{t_i}$ must equal $1/k$ for t_i achieving the maximum on the third line. The fifth line uses (8.2). Continuity of $\bar{U}_{t_i}$ and G_{t_i} and compactness of Δ guarantee that M_* is finite.

Since $\Delta \times \Lambda \times B$ is compact, we can find some subsequence of the $(\delta^k,\lambda^k,\beta^k)$ converging to some point (δ,λ,β). We now show that this (δ,λ,β) is a proper solution. δ is a threat equilibrium and β solves the dual for λ, because each δ^k and β^k have these properties for λ^k (and Z_1 and Z_3 are upper-semicontinuous). Condition (7.3) follows from (8.3) as $k \to \infty$. (7.3) also implies

$$\lambda_{t_i}\bar{U}_{t_i}(\delta_N) \geq \bar{\Phi}_{t_i}(\delta,\lambda,\beta),$$

and (6.5) guarantees that none of these inequalities can be strict. So (δ,λ,β) is a proper solution.

Q.E.D.