Myerson, Roger B.

Working Paper

Cooperative Games with Incomplete Information

Discussion Paper, No. 528

Provided in Cooperation with:
Kellogg School of Management - Center for Mathematical Studies in Economics and Management Science, Northwestern University

Suggested Citation: Myerson, Roger B. (1982) : Cooperative Games with Incomplete Information, Discussion Paper, No. 528, Northwestern University, Kellogg School of Management, Center for Mathematical Studies in Economics and Management Science, Evanston, IL

This Version is available at:
https://hdl.handle.net/10419/220888

RR

Discussion Paper No. 528

COOPERATIVE GAMES WITH INCOMPLETE INFORMATION

by

Roger B. Myerson

June 1982

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

Abstract. A bargaining solution concept which generalizes the Nash bargaining solution and the Shapley NTU value is defined for cooperative games with incomplete information. These bargaining solutions are efficient and equitable when interpersonal comparisons are made in terms of certain virtual utility scales. A player's virtual utility differs from his real utility by exaggerating the difference from the preferences of false types that jeopardize his true type. In any incentive-efficient mechanism, the players always maximize their total virtual utility ex post. Conditionally-transferable virtual utility is the strongest possible transferability assumption for games with incomplete information.

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

COOPERATIVE GAMES WITH INCOMPLETE INFORMATION

by Roger B. Myerson

1. Introduction

In a cooperative game, the players must bargain to select an outcome which is efficient for them. Each player wants to demand the outcome that is best for himself, so the players must moderate their demands to reach a feasible argreement. In general, the amount that a player can realistically demand in such bargaining will depend on his power in the game situation. Here, power means the ability to alternatively help or hurt other players at will, and to defend oneself against the threats of others. A solution concept in cooperative game theory is an attempt to systematically predict which outcomes on the Pareto frontier would be selected by the players, in any cooperative game, in such a way that each player's payoff is commensurate with his power. This paper will develop a general solution concept for games with incomplete information.

The Nash [1950,1953] bargaining solution, defined for two-person bargaining problems, and the Shapley [1953] value, defined for n-person games with transferable utility, are the most conceptually elegant and appealing solution theories in cooperative game theory. Each is derived as the unique fair allocation rule satisfying a set of compelling and (seemingly) weak axioms. Harsanyi [1963] showed that these two important solution concepts are special cases of a more general solution concept, called a nontransferable-utility value or NTU value, that is defined for all complete-information cooperative games, with any number of players, with or without transferable utility. Shapley [1969] developed a simplified version of the NTU value.

These solution concepts may be viewed as generalizations and extensions of the equal-gains principle (that any two players should gain equally from cooperating with each other) to games with more than two players and without obvious symmetries or interpersonally-comparable utility scales.

Harsanyi and Selten [1972] developed a generalized Nash solution for games with incomplete information, a modified version of which was presented by Myerson [1979]. However, this solution concept had serious theoretical drawbacks, and no n-person generalization value could be found. Myerson [1981] analyzed the problem of inscrutable selection of a mechanism by a player who has all of the bargaining ability but also needs to conceal his preferences and private information. This work led to a new generalization of the Nash bargaining solution for two-player games with incomplete information where both players have equal bargaining ability. This new generalized Nash solution was derived from axioms in Myerson [1982].

In this paper, we will construct a bargaining solution concept that will extend the solution concept of Myerson [1982] and the NTU value of Shapley [1969] to general cooperative games with incomplete information, using the Bayesian formulation of Harsanyi [1967-8]. Our bargaining solution will not be derived from axioms here. Its justification will be that it generalizes and unifies the three basic axiomatically-derived theories of Nash [1950], Shapley [1953], and Myerson [1982].

It is reasonable to ask why we should be interested in finding unified cooperative solution concepts of such great generality. One goal is to have a common framework within which to analyze and compare a wide variety of games. Another goal is to use generalizability as a test of solution concepts themselves. That is, if there are two solution concepts which appear equally plausible for a limited class of games, but only one is naturally

generalizable to a much broader class, then that is evidence in favor of the conceptual significance of the generalizable concept. In this sense, perhaps the bargaining solution concept in this paper should be viewed as a further justification of the Shapley value and the Nash bargaining solution.

But the most important gain from developing a unified solution concept for general cooperative games with incomplete information maybe that it forces us to systematically survey the basic logical issues involved in cooperation under uncertainty. In this paper, we will be developing conceptual structures and perspectives which may prove to have significance behond the specific solution concept to which they are applied in this paper. In particular, the ideas of virtual utility and maximal linear extensions, developed in Sections 3 and 4 respectively, might also be applied to develop alternative solution concepts for cooperative games with incomplete information. Also, the interpretation of the rational-threats criterion developed in Section 6 may also help justify the Shapley NTU value against the recent criticism of Roth [1980] and Shafer [1980].

In Section 2, the general structure of cooperative games with incomplete information is formalized. Incentive-efficient mechanisms for such games satisfy a parametric linear programming problem, which is characterized in Section 3. Virtual utility is defined so that the Lagrangian function for this parametric optimization problem can be expressed as the expected sum of the players' virtual utility. Thus, in an efficient agreement subject to incentive constraints, it may appear ex post that the players have maximized their virtual utilities, rather than their real utilities. This suggests the following virtual utility hypothesis: that when incentive constraints (necessary for players to trust each other) are binding in a bargaining situation, players may act as if they want to maximize their virtual

utilities, rather than their real utilities.

The concept of transferable utility was extremely important in the first development of cooperative game theory. However, for games with incomplete information, linear activities like side payments can serve a signalling purpose as well as a transfer purpose, which makes matters more complicated. In Section 4, it is shown that, for a game with incomplete information, the most transferability that can be allowed, without totally replacing the efficient frontier, is transferability of virtual utility conditionally on the state of information in the game.

In Section 5, the ideas of Sections 3 and 4 are applied to construct the general solution concept. With complete information, a Shapley NTU value is an allocation for which there exist nonnegative weighting factors for all players' utility scales such that the allocation would be both equitable (as evaluated by the Shapley value) and efficient if interpersonal comparisons and transfers could be made in terms of these weighted utility scales. For games with incomplete information, a bargaining solution is an incentive-compatible mechanism for which there exist virtual utility scales such that the mechanism would be both equitable and efficient if interpersonal comparisons and transfers could be made in terms of these virtual utility scales. The main results of this paper are the existence and individual rationality of these general bargaining solutions.

The rational-threat criterion used in our solution concept is reconsidered in Section 6. We show that the rational-threat criterion may be most appropriate in games where the coalitions can commit themselves to threats in advance, when they anticipate only a small probability of actually carrying out the threats. In such a situation, a single coalition's threat against its complement does not need to be either equitable or incentive

compatible. Instead, it should be evaluated as part of a plan of threat and agreement that must be equitable and incentive-compatible overall.

Section 7 contains the longer proofs.

2. Basic Definitions

Let $N = \{1,2,\ldots,n\}$ denote the set of players, and CL denote the set of possible conditions or nonempty subsets of N, so that

$$CL = \{S \mid S \subseteq N,\ S \neq \emptyset\}.$$

For any coalition S, we let D_S denote the set of collective actions or decisions feasible for the members of S if they cooperate with each other. For example, in a market game, D_S might be the set of possible trades among the members of S. For any two disjoint coalitions R and S, we assume that

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

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

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 private information about preferences, endowments, and any other factors relevant to the players. For any coalition S, we let

$$T_S = \underset{i \in S}{\times} T_i,$$

so any t_S in T_S denotes a possible combination of types for the members of S. For mathematical simplicity, we will assume that all T_i and D_S are nonempty finite sets. The decision spaces and type spaces for the grand

coalition N will play a major role here, so we may drop the subscript N for these sets; that is,

$$D = D_N, \quad T = T_N.$$

For any d in D and t in T, we let $u_i(d,t)$ denote the payoff to player i, measured in some vonNeumann-Morgenstern utility scale, if t is the combination of types for then players and d represents the decisions made by the players.

Throughout this paper, whenever t, t_S, and t_i appear in the same formula, then t_i denotes the i^{th} component of the vector t in T, and $t_S = (t_j)_{j\in S}$. We also use the notation $N-i = N\setminus\{i\}$, and we may write $t = (t_{N-i}, t_i)$. Similarly, (t_{N-i}, s_i) is the vector of types differing from t in that the i^{th} component is changed to s_i.

For any t in T, we let $p_i(t_{N-i}|t_i)$ denote the conditional probability that t_{N-i} is the combination of types for players other than i, as would be assessed by player i if t_i were his type. We will assume that these probabilities are <u>consistent</u> in the sense of Harsanyi [1967-8]. That is, there exists some probability distribution p over T such that

$$p_i(t_{N-i}|t_i) = p(t)/p^i(t_i) \qquad \forall i\in N, \ \forall t\in T, \tag{2.1}$$

where

$$p^i(s_i) = \sum_{s_{N-i}\in T_{N-i}} p(s) \qquad \forall i\in N, \ \forall s_i\in T_i. \tag{2.2}$$

we will also assume that no types-vector has zero probability, so

$$p(t) > 0, \qquad \forall t\in T. \tag{2.3}$$

(These consistency and positivity assumptions (2.1)-(2.3) will be needed only to simplify the interpretation of our results. The solution concept developed in this paper will satisfy the <u>probability-invariance axiom</u> described in Myerson [1982], and so it can be extended using this axiom to

games without consistent positive probability distributions.)

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

$$\Gamma = ((D_S)_{S\in CL},\ (T_i,u_i)_{i\in N},\ p).$$

We assume that this structure Γ is common knowledge among the players when they play the game, plus each player also knows his own true type. We may refer to a vector of the players' types as a state of the game.

Any coalition S, if it were to form, could plan to determine its collective decision randomly as a function of its members' information. We let M_S denote the set of all functions from T_S into the set of probability distributions over D_S. That is, $\mu_S \in M_S$ iff

$$(2.5)\qquad \mu_S(d_S|t_S) \geq 0 \text{ and } \sum_{c_S\in D_S} \mu_S(c_S|t_S) = 1 \qquad \forall d_S\in D_S,\ \forall t_S\in T_S.$$

Any such μ_S in M_S may be referred to as a mechanism for coalition S.

If $R\cap S = \emptyset$, then we can embed $M_R \times M_S$ in $M_{R\cup S}$ in the obvious way. That is, if $\mu_R\in M_R$ and $\mu_S\in M_S$, then (μ_R,μ_S) in $M_{R\cup S}$ is defined by

$$(\mu_R,\mu_S)(d_R,d_S|t_R,t_S) = \mu_R(d_R|t_R)\cdot\mu_S(d_S|t_S) \quad \text{if } (d_R,d_S)\in D_R\times D_S\subseteq D_{R\cup S},$$

and

$$(\mu_R,\mu_S)(d_{R\cup S}|t_R,t_S) = 0 \quad \text{if } d_{R\cup S}\notin D_R\times D_S.$$

We shall assume that, in the cooperative game, only the mechanism chosen by the grand coalition N will actually be implemented. As a threat during bargaining, each coalition S may commit itself to some mechanism μ_S in M_S, to be carried out if the other players refuse to cooperate with the members of S. Such threats will be significant only to the extent that they may

influence the mechanism μ_N chosen by the grand coalition. In the rest of this section and in Sections 3 and 4, we will only consider mechanisms in M_N, to develop the theory of efficient mechanisms for the grand coalition. In Section 5 we will reconsider the threats of all coalitions and construct our bargaining solution.

We let $U_i^*(\mu_N, s_i | t_i)$ denote the expected utility for player i from the mechanism μ_N in M_N, if i's true type is t_i but he reports type s_i, while all other players are expected to report their types truthfully. That is

$$U_i^*(\mu_N, s_i | t_i) = \tag{2.6}$$

$$= \sum_{t_{N-i} \in T_{N-i}} p_i(t_{N-i} | t_i) \sum_{d \in D} \mu_N(d | t_{N-i}, s_i)\, u_i(d,t).$$

We let

$$U_i(\mu_N | t_i) = U_i^*(\mu_N, t_i | t_i) \tag{2.7}$$

$$= \sum_{t_{N-i} \in T_{N-i}} p_i(t_{N-i} | t_i) \sum_{d \in D} \mu_N(d | t)\, u_i(d,t).$$

That is, $U_i(\mu_N | t_i)$ is the expected utility for player i from the mechanism μ_N, if i's true type is t_i and all players are expected to report their types truthfully in implementing μ_N.

We shall assume that each player's type is not observable by other players, so that the types are unverifiable. Thus, if a player had some incentive to lie about his type when the grand coalition N implements its mechanism μ_N, then he would do so. A mechanism is <u>incentive compatible</u> (or, more correctly, <u>Bayesian incentive compatible</u> in the sense of d'Aspremont and

Gerard-Varet [1979]) iff

$$(2.8) \qquad U_i(\mu_N|t_i) \geq U_i^*(\mu_N, s_i|t_i) \qquad \forall i \in N,\ \forall t_i \in T_i,\ \forall s_i \in T_i.$$

That is, μ_N is incentive compatible iff it would be a Bayesian Nash equilibrium for all players to plan to report their types honestly in the mechanism μ_N, assuming that they are asked to report their types simultaneously and confidentially. Thus, with unverifiable types, the players must choose an incentive-compatible mechanism if honest reporting of types is to be induced. It has been argued elsewhere (see Myerson [1979], for example) that any Bayesian equilibrium of possibly dishonest reporting strategies in any mechanism can be simulated by an equivalent incentive-compatible mechanism with honest reporting. So without loss of generality, we may assume that the mechanism selected by the grand coalition N must be incentive compatible.

In some games, it may be possible for some types to costlessly prove that other types are false.[1/] 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. If player i, when s_i is his true type, could costlessly prove that he is not type t_i, then we should drop the corresponding constraint (saying that t_i must not be tempted to report s_i) in (2.8). With this modification, our analysis in this paper can be extended to cover the case of verifiable or semi-verifiable types. Henceforth in this paper we will consider only the case of unverifiable types.

---

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

3. The Primal and Dual Problems and Virtual Utility

A mechanism μ_N in M_N is <u>incentive-efficient</u> iff it is incentive compatible and there does not exist any other incentive-compatible mechanism $\hat{\mu}_N$ such that

$$U_i(\hat{\mu}_N|t_i) \geq U_i(\mu_N|t_i), \qquad \forall i\in N, \ \forall t_i\in T_i, \tag{3.1}$$

with $U_j(\hat{\mu}_N|t_j) > U_j(\mu_N|t_j)$ for at least one type t_j of some player j. If the players can bargain effectively, then they should be able to ultimately agree on some incentive-efficient mechanism. Otherwise, it would be common knowledge that all players could agree to a change to some other mechanism $\hat{\mu}_N$ satisfying (3.1). See Holmström and Myerson [1981] for an analysis of this and other concepts of efficiency for games with incomplete information.

Let Λ be the following simplex in $\underset{i\in N}{\times} \mathbb{R}^{T_i}$,

$$\Lambda = \left\{\lambda\in \underset{i\in N}{\times} \mathbb{R}^{T_i} \,\middle|\, \lambda_i(t_i) \geq 0, \ \forall i\in N, \ \forall t_i\in T_i, \ \sum_{j\in N} \sum_{s_j\in T_j} \lambda_j(s_j) = n\right\} \tag{3.2}$$

Let Λ^0 denote the relative interior of Λ,

$$\Lambda^0 = \{\lambda\in\Lambda \,|\, \lambda_i(t_i) > 0, \ \forall i\in N, \ \forall t_i\in T_i\}. \tag{3.3}$$

Since we are assuming that D and T are finite sets, the set of all incentive-compatible mechanisms is a closed convex polyhedron in M_N, defined by the linear inequalities (2.5) and (2.8). (Notice that $U_i(\mu_N|t_i)$ and $U_i^*(\mu_N,s_i|t_i)$ are both linear functions of μ_N.) Thus, by the supporting hyperplane theorem, μ_N is incentive-efficient iff there exists some λ in Λ^0 such that μ_N is an optimal solution to the problem

$$(3.4) \qquad \underset{\mu_N \in M_N}{\text{maximize}} \sum_{i \in N} \sum_{t_i \in T_i} \lambda_i(t_i)\, U_i(\mu_N | t_i)$$

subject to the incentive constraints (2.8).

We shall refer to this optimization problem (3.4) as the primal problem for λ.

Given λ, the primal (3.4) is a linear programming problem. Let us now formulate its dual. We shall generally let $\alpha_i(s_i|t_i)$ denote the dual variable (or shadow price) corresponding to the incentive constraint (2.8) that asserts that player i should not be tempted to claim to be type s_i if his true type is t_i. We let

$$(3.5) \qquad \underset{\sim}{A} = \left\{ \alpha \in \underset{i \in N}{\times} \mathbb{R}^{T_i \times T_i} \,\middle|\, \alpha_i(s_i|t_i) \geq 0,\ \alpha_i(t_i|t_i) = 0,\ \forall i \in N,\ \forall t_i \in T_i,\ \forall s_i \in T_i \right\}$$

That is, $\underset{\sim}{A}$ is the set of all possible vectors of dual variables for the incentive constraints. ((2.8) holds trivially when $s_i = t_i$, so the shadow price $\alpha_i(t_i|t_i)$ will be zero.)

We now come to an important definition. Given any λ in Λ and α in $\underset{\sim}{A}$, let

$$(3.6) \qquad v_i(d,t,\lambda,\alpha) = \Big(\big(\lambda_i(t_i) + \sum_{s_i \in T_i} \alpha_i(s_i|t_i) \big)\, p_i(t_{N-i}|t_i)\, u_i(d,t) - \sum_{s_i \in T_i} \alpha_i(t_i|s_i)\, p_i(t_{N-i}|s_i)\, u_i(d,(t_{N-i},s_i)) \Big) / p(t)$$

for any i in N, d in D, and t in T. We shall refer to $v_i(d,t,\lambda,\alpha)$ as player i's virtual utility for decision d in state t, with respect to λ and α.

If we multiply the incentive constraints by their dual variables and add them into the primal objective function, then we get the following Lagrangian function:

$$(3.7)\qquad \sum_{i\in N}\sum_{t_i\in T_i} \lambda_i(t_i)\, U_i(\mu_N|t_i)$$

$$+ \sum_{i\in N}\sum_{t_i\in T_i}\sum_{s_i\in T_i} \alpha_i(t_i|s_i)\,(U_i(\mu_N|t_i)-U_i^*(\mu_N,s_i|t_i))$$

$$= \sum_{t\in T} p(t) \sum_{d\in D} \mu_N(d|t) \sum_{i\in N} v_i(d,t,\lambda,\alpha).$$

The equality in (3.7) follows by straightforward manipulation from the definitions (2.6), (2.7) and (3.6). So the Lagrangian function for the primal problem is just the expected sum of the players' virtual utilities.

By standard Lagrangian analysis, an incentive-compatible mechanism μ_N will be an optimal solution of the primal problem for λ if and only if there is some α in $\underset{\sim}{A}$ such that

$$\alpha_i(s_i|t_i)\,(U_i(\mu_N|t_i) - U_i^*(\mu_N,s_i|t_i)) = 0, \quad \forall i\in N,\ \forall t_i\in T_i,\ \forall s_i\in T_i,$$

and μ_N maximizes the Lagrangian function subject only to the probability constraints (2.3). Obviously, this Lagrangian function is maximized by putting all probability weight, in each $\mu_N(\cdot|t)$ distribution, on the decisions that maximize the sum of the players' virtual utilities. That is, μ_N maximizes the Lagrangian function over all mechanisms in M_N if and only if

$$(3.8)\qquad \sum_{d\in D} \mu_N(d|t) \sum_{i\in N} v_i(d,t,\lambda,\alpha)$$

$$= \text{maximum}_{d\in D} \sum_{i\in N} v_i(d,t,\lambda,\alpha), \quad \forall t\in T\ .$$

The appropriate vector α for use in this Lagrangian analysis is the vector that solves the dual of (3.4). This dual problem for λ can be written

$$(3.9) \qquad \underset{\alpha \in \underset{\sim}{A}}{\text{minimize}} \sum_{t \in T} p(t) \cdot \underset{d \in D}{\text{maximum}} \sum_{i \in N} v_i(d,t,\lambda,\alpha).$$

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

The virtual utility functions will play an important role in our theory of bargaining, so it is worthwhile to try to develop some intuitive understanding of them. So let us assume that μ_N is an incentive-efficient mechanism. Let λ in Λ^0 and α in $\underset{\sim}{A}$ be such that μ_N solves the primal for λ and α solves the dual for λ. We say that one type s_i <u>jeopardizes</u> another type t_i of player i, in the incentive-efficient mechanism μ_N, iff the constraint that says s_i should not gain by claiming to be t_i is binding (that is, $U_i(\mu_N|s_i) = U_i^*(\mu_N,t_i|s_i))$ and its shadow price $\alpha_i(t_i|s_i)$ is positive. Then player i's virtual utility when he is of type t_i differs from his real utility in a way that exaggerates the difference from the types that jeopardize t_i. That is, equation (3.6) defines i's virtual utility for d at t as a positive multiple of his real utility for d at t, minus a multiple of what his utility for d would be if his type were changed to a type that jeopardizes t_i. To see this more clearly, notice that (3.6) may be rewritten as

$$p^i(t_i)\, v_i(d,t,\lambda,\alpha) = \Big(\lambda_i(t_i) + \sum_{s_i} \alpha_i(s_i|t_i)\Big)\, u_i(d,t)$$

$$- \sum_{s_i} \alpha_i(t_i|s_i)\, u_i(d,(t_{N-i},s_i))\big(p_i(t_{N-i}|s_i)/p_i(t_{N-i}|t_i)\big),$$

where the probability-correction ratio in the last term vanishes to one if the players' types are stochastically independent.

For type t_i of player i, the expected virtual utility from the mechanism μ_N (honestly implemented) is

$$(3.10)\qquad \sum_{t_{N-i}\in T_{N-i}} \sum_{d\in D} p_i(t_{N-i}|t_i)\,\mu_N(d|t)\,v_i(d,t,\lambda,\alpha)$$

$$= \Big[\Big(\lambda_i(t_i) + \sum_{s_i} \alpha_i(s_i|t_i)\Big)\,U_i(\mu_N|t_i) - \sum_{s_i} \alpha_i(t_i|s_i)\,U_i^*(\mu_N,t_i|s_i)\Big]/p^i(t_i).$$

If μ_N solves the primal for λ and α solves the dual for λ, then by complementary slackness this formula can be further simplified to

$$(3.11)\qquad \Big[\Big(\lambda_i(t_i) + \sum_{s_i} \alpha_i(s_i|t_i)\Big)\,U_i(\mu_N|t_i) - \sum_{s_i} \alpha_i(t_i|s_i)\,U(\mu_N|s_i)\Big]/p^i(t_i).$$

Let us consider now an application of the virtual utility concept. An incentive-efficient mechanism need not be efficient ex post, after the players learn each other's type. That is because, in order to satisfy incentive constraints, it may be necessary to accept a positive probability of an outcome that is bad for both players. For example, in union-management negotiations, if the management of the "type" that can only afford to pay lower wages, then it might have to accept a positive probability of a strike before it can get a reduction in the wage rate. The strike is needed to prove to the workers that management is not of the type with high ability to pay. But it may be difficult to understand how the players can commit themselves to implement a strike of any duration, since management's low type is revealed as soon as the strike begins, and then both sides would prefer to settle at a low wage.

By (3.8), an incentive-efficient mechanism always maximizes the sum of the players' virtual utilities (with respect to the appropriate λ and α) in

every state t. Thus an incentive-efficient mechanism would appear efficient ex post if the players' payoffs were measured in virtual utility, instead of real utility. Instead of saying that the incentive constraints (2.8) force the players to accept ex post inefficiency, we may say that the incentive constraints force each player to transform his effective preferences from his real to his virtual utility function, to exaggerate the difference between his true type and the false types that jeopardize it. This idea, that players in bargaining may act as if they want to maximize their virtual utilities instead of their actual utilities, may be referred to as the virtual-utility hypothesis.

In Section 5, we will extend this virtual-utility hypothesis by assuming that the players also make interpersonal equity comparisons in terms of virtual utility, to compute their fair payoffs or warranted claims. But first, we consider generalizations of the classical transferable-utility assumption for games with incomplete information.

4. Transferable Utility and Linear Activities

The assumption of transferable utility has played an important role in the development of cooperative game theory. Of course, bounded utility transfers can be accommodated within the model described in Section 2 (by interpreting the decisions in each D_S as including specifications of how much utility should be tranferred between each pair of players in S), so there is very little loss of generality in restricting ourselves to this model. Nevertheless, to understand cooperation under uncertainty, it is useful to see how the assumption of transferable utility extends to games with incomplete information.

Transfer of utility between players is just a special kind of linear activity which could be permitted in a game. In general, a linear activity can be represented by a function $f:T \to \mathbb{R}^n$, such that $f_i(t)$ is the utility gained by player i in state t if the players do one unit of activity f. Let F be any finite set of such activities, so that F is a subset of $\mathbb{R}^{N\times T}$. Given Γ as in (2.1), the game Γ extended by F refers to the game in which the grand coalition can also use any linear combination of activities in F as a function of the players type-reports. (More generally, we could introduce a set of feasible linear acitivies F_S for each coalition S, with $F_S \subseteq F_R$ if $S \subseteq R$, but we will only be concerned with the grand coalition N in this section.)

In the extended game with linear activities, the set of mechanisms for N becomes $M_N \times \mathbb{R}^{F\times T}$. That is, a mechanism is a pair (μ_N, e) in $M_N \times \mathbb{R}^{F\times T}$ where $e(f|t)$ is interpreted as the level of activity f to be performed if the players report their vector of types as t. Notice that we allow that $e(f|t)$ may be positive or negative.

The expected utility gained by player i from linear activities in the mechanism (μ_N, e), given that i's type is t_i, is

$$(4.1) \qquad G_i(e|t) = \sum_{t_{N-i} \in T_{N-i}} p_i(t_{N-i}|t_i) \sum_{f\in F} f_i(t)\, e(f|t)$$

if all players are honest, and is

$$(4.2) \qquad G_i^*(e,s_i|t_i) = \sum_{t_{N-i}} p_i(t_{N-i}|t_i) \sum_{f} f_i(t)\, e(f|t_{N-i},s_i)$$

if all players are honest except for i who reports s_i. The mechanism (μ_N, e) is incentive compatible iff

$$(4.3) \qquad U_i(\mu_N|t_i) + G_i(e|t_i) \geq U_i^*(\mu_N,s_i|t_i) + G_i^*(e,s_i|t_i)$$
$$\forall i\in N,\ \forall t_i\in T_i,\ \forall s_i\in T.$$

With linear activities, the extended primal problem for λ is defined to be

$$(4.4) \qquad \underset{\mu_N,e}{\text{maximize}} \sum_{i\in N} \sum_{t_i\in T_i} \lambda_i(t_i)\big(U_i(\mu_N|t_i) + G_i(e|t_i)\big)$$
$$\text{subject to } \mu_N\in M_N \text{ and } (4.3).$$

This extended primal problem differs from the original primal problem (3.4) only in that it has more variables, in the vector e. Thus, the extended dual problem for λ differs from the original dual problem (3.9) in that it has more constraints, one new constraint for each variable $e(f|t)$, as follows:

$$(4.5) \qquad \sum_{i\in N} \Big(\big(\lambda_i(t_i) + \sum_{s_i\in T_i} \alpha_i(s_i|t_i)\big)\, p_i(t_{N-i}|t_i)\, f_i(t)$$
$$- \sum_{s_i\in T_i} \alpha_i(t_i|s_i)\, p_i(t_{N-i}|s_i)\, f_i(t_{N-i},s_i)\Big) = 0 \qquad \forall f\in F,\ \forall t\in T.$$

(Notice that (4.5) is a linear constraint on α.)

The assumption of transferable utility means that, for any two players j and k, F includes an activity of transfering one unit of utility from j to k. We may denote this activity by f^{jk}, where

$$f_j^{jk}(t) = -1,\ f_k^{jk}(t) = +1,\ f_i^{jk}(t) = 0 \text{ if } i\notin\{j,k\},\quad \forall t\in T.$$

Let us suppose that the players' types are stochastically independent random variables, so that

$$p(t) = \prod_{i\in N} p^i(t_i), \quad \forall t\in T.$$

Then (4.5) for $f=f^{jk}$ becomes

$$\Big(\lambda_j(t_j) + \sum_{s_j\in T_j} \alpha_j(s_j|t_j) - \sum_{s_j\in T_j} \alpha_j(t_j|s_j)\Big) \prod_{i\in N-j} p^i(t_i)$$

$$= \Big(\lambda_k(t_k) + \sum_{s_k\in T_k} \alpha_k(s_k|t_k) - \sum_{s_k\in T_k} \alpha_k(t_k|s_k)\Big) \prod_{i\in N-k} p^i(t_i), \quad \forall t\in T.$$

Dividing both sides of this equation by p(t) gives us

$$\Big(\lambda_j(t_j) + \sum_{s_j} \alpha_j(s_j|t_j) - \sum_{s_j} \alpha_j(t_j|s_j)\Big)/p^j(t_j) \tag{4.6}$$

$$= \Big(\lambda_k(t_k) + \sum_{s_k} \alpha_k(s_k|t_k) - \sum_{s_k} \alpha_k(t_k|s_k)\Big)/p^k(t_k), \quad \forall t_j\in T_j, \ \forall t_k\in T_k.$$

Thus, if the players' types are independent and if utility is transferable between all players then, for any λ in Λ, α satisfies the dual constraint (4.5) iff

$$\lambda_i(t_i) + \sum_{s_i\in T_i} \alpha_i(s_i|t_i) - \sum_{s_i\in T_i} \alpha_i(t_i|s_i) = p^i(t_i), \quad \forall i\in N, \ \forall t_i\in T_i. \tag{4.7}$$

(The constant ratio in (4.6) must be 1, because the $\lambda_i(t_i)$ sum to n, for λ in Λ. Recall (3.2).)

Equation (4.7) can be very helpful for solving applied problems. However, it also illustrates why transferable utility is less useful as an assumption for games with incomplete information than it was for games with complete information. With complete information each player has only one

type, so that (4.7) becomes simply $\lambda_i(t_i) = 1$; that is, all players must be given equal weight in the primal problem or else the dual is infeasible and the primal has no finite optimum. Thus, transferable utility with complete information implies that all efficient mechanisms solve the same primal problem, and the Pareto-efficient frontier is a hyperplane. With incomplete information, (4.7) implies that

$$\sum_{t_i \in T_i} \lambda_i(t_i) = 1, \qquad \forall i \in N,$$

but this still leaves $\sum_{i \in N} (|T_i| - 1)$ degrees of freedom in choosing λ, and the dual problems are generally nontrivial to solve. Under incomplete information, the incentive-efficient frontier in $\times_{i \in N} \mathbb{R}^{T_i}$ is generally <u>not</u> a hyperplane for games with transferable utility.

To get a conceptual simplification comparable to that offered by transferable utility under complete information, we must introduce a larger class of linear activities. So let us re-examine (4.5), but think of it now as a constraint on f for some given λ and α. The expression in brackets in (4.5) is just p(t) times i's virtual utility for one unit of activity f in state t. Thus, (4.5) asserts that f must transfer virtual utility between the players in each state.

Thus, instead of transferable utility, let us consider the assumption of <u>conditionally transferable virtual utility</u>. Given any two players j and k and given any type-vector s in T, let g^{jks} denote the activity that transfers one unit of virtual utility (with respect to λ and α) from player j to player k conditionally on s being the true vector of types, and that transfers zero units of virtual utility otherwise. That is, g^{jks} satisfies the following equations, for every i in N and t in T:

$$(4.8)\qquad \Big(\big(\lambda_i(t_i) + \sum_{r_i \in T_i} \alpha_i(r_i|t_i)\big)\, p_i(t_{N-i}|t_i)\, g_i^{jks}(t) - \sum_{r_i \in T_i} \alpha_i(t_i|r_i)\, p_i(t_{N-i}|r_i)\, g_i^{jks}(t_{N-i}, r_i)\Big)/p(t)$$

$$= \begin{cases} -1 & \text{if } i = j \text{ and } t = s \\ +1 & \text{if } i = k \text{ and } t = s \\ 0 & \text{if } i \notin \{j,k\} \text{ or } t \neq s \end{cases}$$

If $\lambda \in \Lambda^0$, $\alpha \in \underset{\sim}{A}$, and all $p_i(t_{N-i}|t_i) > 0$, then (4.8) has a unique solution g^{jks}, and this vector satisfies:

$$g_i^{jks}(t) = 0 \quad \text{if} \quad i \notin \{j,k\} \text{ or } t_{N-i} \neq s_{N-i}$$

$$g_j^{jks}(s_{N-j}, t_j) \leq 0 \qquad \forall t_j \in T_j$$

$$g_k^{jks}(s_{N-k}, t_k) \geq 0 \qquad \forall t_k \in T_k$$

(These properties follow from (4.8) using Lemma 1 in Section 7.) Notice that, although g^{jks} gives no <u>virtual utility</u> to player k except in state s, g^{jks} may in fact give him positive amounts of <u>real utility</u> in states where his own type differs from s_k.

Given any λ in Λ^0 and α in $\underset{\sim}{A}$, we let $\overline{F}^{\lambda\alpha}$ denote the set of all such g^{jks} generated by λ and α. That is:

$$\overline{F}^{\lambda\alpha} = \{ g^{jks} \mid j \in N,\ k \in N,\ s \in T,\ \text{and (4.8) is satisfied with } \lambda \text{ and } \alpha \}.$$

If Γ is extended by $\overline{F}^{\lambda\alpha}$ then we may say that virtual utility with respect to λ and α is conditionally transferable. (Here "conditionally transferable"

refers to the fact that the transfers can be conditioned on the players' true types, rather than just their reported types. So conditional transferabiltiy is a stronger property than simple transferability.)

The following theorem states that, if we try to extend a game in such a way as to preserve at least one of its incentive-efficient mechanisms, then the maximal extension is to allow conditionally-transferable virtual utility.

Theorem 1: Let Γ be as in (2.1) and let μ_N be an incentive-efficient mechanism for the grand coalition in Γ. Let F be any set of linear activities. Then $(\mu_N, \underset{\sim}{0})$ is incentive-efficient in Γ extended by F if and only if there exists some λ in Λ^0 and α in $\underset{\sim}{A}$ such that μ_N is an optimal solution of the primal problem for λ, α is an optimal solution of the dual problem for λ, and F is contained in the linear span of $\overline{F}^{\lambda\alpha}$.
(Here $(\mu_N, \underset{\sim}{0})$ is just μ_N without using any activities in F.)

Proof: $(\mu_N, \underset{\sim}{0})$ is incentive-efficient in Γ extended by F if and only if there is some λ in Λ^0 such that $(\mu_N, \underset{\sim}{0})$ is optimal in the extended primal for λ. But this holds if and only if there is some α in $\underset{\sim}{A}$ such that α is feasible in the extended dual for λ and the value of the primal objective function at μ_N equals the value of the dual objective function at α. This in turn holds if and only if μ_N is optimal in the (unextended) primal for λ, α is optimal in the (unextended) dual for λ, and α is feasible in the extended dual. But the linear span of $\overline{F}^{\lambda\alpha}$ is just the set of all activities f that satisfy (4.5) for λ and α. (To check this, observe that all activities in $\overline{F}^{\lambda\alpha}$ satisfy (4.5); there are $(n-1)|T|$ linearly independent vectors g^{jks} in $\overline{F}^{\lambda\alpha}$; and the set of vectors satisfying (4.5) has $(n-1)|T|$ dimensions.) So

α is feasible in the extended dual for λ if and only if F is contained in the linear span of $\overline{F}^{\lambda\alpha}$. Q.E.D.

To understand Theorem 1, it is helpful to recognize that linear activities in games with incomplete information can be used for signalling, that is, for helping to satisfy incentive compatibility, as well as for transferring utility. For example, if real utility (instead of virtual utility) were conditionally transferable, then a player could perfectly signal his type by agreeing to transfer large amounts of utility to other players conditionally on his type being anything other than what he reports. In general, any linear activity that affects different types of a player differently may be used for signalling, to help prove that the player is not of the type that loses more from the activity. The activity g^{jks}, which transfers virtual utility from j to k conditionally on state s, can affect the real utility payoffs of j or k in states other than s; so its potential for signalling purposes is less than that of an activity that transfers real utility from j to k conditionally on state s.

5. The General Bargaining Solution

The construction of Shapley's NTU value for games with complete information may be sketched as follows. First, select any outcome on the Pareto-efficient frontier for the grand coalition. Now extend the game by a maximal collection of linear activities such that the selected outcome is still on the efficient frontier of the extended game. These linear activities can be characterized as transfers of weighted utility between players, where each player's "weighted utility" payoff is some contstant λ_i times his

original utility. In the extended game, let each coalition choose a threat; let the worth of each coalition be the total weighted utility that would be earned by its members if it and its complement both carried out their threats; and let the grand coalition N act so as to give each player weighted utility equal to his Shapley-value allocation computed from these coalitional worths. If this hypothetical behavior in the extended game, when each coalition chooses its threat optimally for its members, turns out to give the players payoffs equal to what they were getting in the originally-selected outcome (which was feasible in the original game), then we say that that outcome is a Shapley NTU value for the original game. That is, the Shapley NTU value is defined as a Shapley value, for an extended game with transfers, that is also feasible in the original game without transfers.

In Section 3 we saw that, when players face binding incentive constraints, they may appear to act according to the preferences of their virtual utility functions. In Section 4 we saw that, with incomplete information, the maximal linear extension (without completely replacing the efficient frontier) is to let virtual utility (w.r.t. some λ and α) be conditionally transferable in every state. Thus, to follow the logic of the Shapley NTU value, we should let coalitional worths and Shapley values be computed in terms of virtual utilities. This key insight, to look at the game with transferable virtual utility rather than weighted utility, was not evident to this author until after eight years of search; but with it we can readily construct a bargaining solution which generalizes the Shapley-NTU value and has satisfactory mathematical properties, including individual rationality and existence.

In our model of bargaining, every coalition makes a threat against the complementary coalition, and then these threats form the basis for computing

the warranted claims of each player. We let

$$M = \underset{S \in CL}{\times} M_S$$

denote the set of possible combinations of mechanisms that the conditions might select as threats. That is, any vector $\mu = (\mu_S)_{S \in CL}$ in M includes a specification of the mechanism μ_S that each coalition $S \subset N$ threatens to use in the case that its complement $N \setminus S$ refuses to cooperate with it.

For any coalition S, we let $W_S(\mu,t,\lambda,\alpha)$ denote the sum of the virtual utilities (with respect to λ and α) that the members of S would expect in state t, if S and $N \setminus S$ carried out their threats. That is, if $S \neq N$,

$$(5.1) \qquad W_S(\mu,t,\lambda,\alpha) =$$

$$= \sum_{d_S \in D_S} \sum_{d_{N \setminus S} \in D_{N \setminus S}} \mu_S(d_S | t_S)\, \mu_{N \setminus S}(d_{N \setminus S} | t_{N \setminus S}) \sum_{i \in S} v_i((d_S, d_{N \setminus S}), t, \lambda, \alpha).$$

In the case of S = N, there is no complementary coalition to threaten, so (5.1) simply reduces to:

$$W_N(\mu,t,\lambda,\alpha) = \sum_{d \in D} \mu_N(d|t) \sum_{i \in N} v_i(d,t,\lambda,\alpha).$$

We let $W(\mu,t,\lambda,\alpha) = \left(W_S(\mu,t,\lambda,\alpha)\right)_{S \in CL}$ denote the characteristic function game with these conditional worths. Its Shapley value for player i is:

$$(5.2) \qquad \phi_i\left(W(\mu,t,\lambda,\alpha)\right) = \sum_{\substack{S \in CL \\ S \supseteq \{i\}}} \frac{(|S|-1)!(n-|S|)!}{n!} \left(W_S(\mu,t,\lambda,\alpha) - W_{N \setminus S}(\mu,t,\lambda,\alpha)\right)$$

(This formula is equivalent to the more familiar formula with S-i replacing $N \setminus S$. We let $W_\emptyset = 0$.)

Thus, if the coalitions make threats μ in the game with conditionally

transferable virtual utility, then the Shapley value gives type t_i of player i an expected virtual-utility payoff equal to:

$$\sum_{t_{N-i} \in T_{N-i}} p_i(t_{N-i} | t_i)\, \phi_i(W(\mu,t,\lambda,\alpha)).$$

We want to know what allocation of real utility corresponds to this allocation of virtual utility. By (3.11) we know that, if each type s_i of player i gets expected (real) utility $\omega_i(s_i)$ from an incentive-compatible mechanism which maximizes the sum of the virtual utilities, then the corresponding virtual utility expected by type t_i is:

$$\Big(\big(\lambda_i(t_i) + \sum_{s_i} \alpha_i(s_i | t_i)\big)\omega_i(t_i) - \sum_{s_i} \alpha_i(t_i | s_i)\omega_i(s_i)\Big)/p^i(t_i).$$

Equating these two formulas (and multiplying through by $p^i(t_i)$) we see that the allocation of real expected utilities corresponding to the Shapley value allocation of virtual utilities should satisfy:

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

$$= \sum_{t_{N-i} \in T_{N-i}} p(t)\, \phi_i(W(\mu,t,\lambda,\alpha)), \quad \forall i \in N,\ \forall t_i \in T_i. \tag{5.3}$$

A vector ω in $\underset{i \in N}{\times} \mathbb{R}^{T_i}$ which satisfies (5.3) is said to be <u>warranted</u> by λ,α, and μ; and $\omega_i(s_i)$ is then the <u>warranted claim</u> of type s_i. Thus, the warranted claims are real utility payoffs corresponding to an allocation which would give each type of each player his expected Shapley value, if the players made interpersonal equity comparisons in terms of their virtual utility scales.

For any λ in Λ^0 and α in $\underset{\sim}{A}$, equations (5.3) have a unique solution in ω, by Lemma 1 in Section 7. Furthermore, these solutions are monotone increasing (weakly) in the right-hand sides. That is, increasing the right-hand side of (5.3) for any type of player i weakly increases the warranted claims of all types of player i. Thus, to maximize i's warranted claim in any type, player i wants to maximize his expected virtual allocation from the Shapley value in all his types.

The threat μ_S affects the Shapley value allocation only through the difference $W_S - W_{N\setminus S}$, which all members of S want to maximize. Thus we say that μ in M is a vector of rational threats with respect to λ and α if

$$(5.4)\qquad \sum_{t\in T} p(t)\big(W_S(\mu,t,\lambda,\alpha) - W_{N\setminus S}(\mu,t,\lambda,\alpha)\big) =$$

$$= \max_{\nu_S\in M_S} \sum_{t\in T} p(t)\big(W_S(\mu_{-S},\nu_S),t,\lambda,\alpha) - W_{N\setminus S}((\mu_{-S},\nu_S),t,\lambda,\alpha)\big), \quad \forall S\in CL$$

(Here (μ_{-S},ν_S) is the vector where ν_S replaces μ_S in μ.) Notice that (5.4) really depends only on μ_S and $\mu_{N\setminus S}$, so the two complementary coalitions are involved in a two-person zero-sum game when they choosing their rational threats. We do not require that rational threats to be incentive compatible; we only require that μ_S must be in M_S, satisfying the probability constraints (2.5). (The set of incentive-compatible mechanisms for coalition S could depend discontinuously on the mechanism chosen by N\S. So the threat-selection game between S and N\S would be a pseudogame and would not necessarily have any equilibrium, if we required that each threat be incentive-compatible given the other.)

Condition (5.4) includes the case of S = N, using $W_\emptyset = 0$. Thus, if μ is a vector of rational threats with respect to λ and α then

$$W_N(\mu,t,\lambda,\alpha) = \max_{d\in D} \sum_{i\in N} v_i(d,t,\lambda,\alpha) .$$

That is, μ_N maximizes the Lagrangian function (3.7).

The essential idea in defining our general bargaining solution is that if the warranted claims for a set of rational threats can actually be achieved by an incentive-compatible mechanism, then this mechanism may be called an bargaining solution for the game. Some care is needed in formulating this idea precisely, to permit an existence theorem to be proven. The problem is that the warrant equations (5.4) are only known to be solvable if all $\lambda_i(t_i)$ are strictly positive, so that $\lambda\in\Lambda^0$. But the Kakutani [1941] fixed point theorem cannot be applied to the interior of a simplex. We solve this dilemma by allowing that some of our positive $\lambda_i(t_i)$ weights may be infinitesimal. In standard analysis, this is done by considering a sequence of vectors in Λ^0, some of whose components may converge to zero.

(This dilemma also arises in the case of complete information, where the resolution proposed by Shapley [1969] is not quite satisfactory. For Shapley's definition, if there is a dummy in a game then any feasible allocation will be an NTU value, with all nondummies having $\lambda_i = 0$. The definition developed below refines Shapley's definition in a way that rules out such perverse solutions without losing existence.)

We say that $\bar{\mu}_N$ is a <u>bargaining solution</u> (or an <u>NTU value</u>) for Γ iff $\bar{\mu}_N$ is an incentive-efficient mechanism and there exists a sequence $\{(\lambda^k,\alpha^k,\mu^k,\omega^k)\}_{k=1}^{\infty}$ such that

(5.5) $\alpha^k\in\underset{\sim}{A}$, $\mu^k\in M$, and $\lambda^k\in\Lambda^0$ (so all $\lambda_i^k(t_i) > 0$), $\forall k$;

(5.6) μ^k is a vector of rational threats for λ^k and α^k, $\forall k$;

(5.7) ω^k is warranted by λ^k,α^k, and μ^k, $\forall k$;

(5.8) $\limsup_{k\to\infty} \omega_i^k(t_i) \leq U_i(\bar{\mu}_N|t_i)$, $\forall i\in N$, $\forall t_i\in T_i$.

That is, a bargaining solution is an incentive-efficient mechanism such that there is a vector of warranted claims, supported by positive utility-weights and rational threats, in which no type's warranted claim exceeds the utility that it expects from the mechanism by more than an arbitrarily small amount.

We can now state our main existence and individual-rationality theorems.

Theorem 2. There exists at least one bargaining solution $\bar{\mu}_N$ for Γ.

Theorem 3. If $\bar{\mu}_N$ is a bargaining solution then

$$U_i(\bar{\mu}_N|t_i) \geq \underset{\mu_{N-i}\in M_{N-i}}{\text{minimum}} \ \underset{\mu_i\in M_{\{i\}}}{\text{maximum}} \ U_i((\mu_{N-i},\mu_i)|t_i) \qquad \forall i\in N, \ \forall t_i\in T_i.$$

Proofs are deferred to Section 7.

For any positive number δ, (5.5)-(5.8) imply that, for all sufficiently large k, $\omega_i^k(t_i) \leq U_i(\bar{\mu}_N|t_i) + \delta$ for every i and t_i, and so

$$\begin{aligned}
(5.9) \qquad & \sum_{t\in T} p(t) \max_{d\in D} \sum_{i\in N} v_i(d,t,\lambda^k,\alpha^k) \\
&= \sum_{t\in T} p(t)\, W_N(\mu^k,t,\lambda^k,\alpha^k) \\
&= \sum_{t\in T} p(t) \sum_{i\in N} \phi_i(W(\mu^k,t,\lambda^k,\alpha^k)) \\
&= \sum_{t\in T} \sum_{i\in N} \lambda_i^k(t_i)\, \omega_i^k(t_i) \\
&\leq \sum_{i\in N} \sum_{t_i\in T_i} \lambda_i^k(t_i)\, U_i(\bar{\mu}_N|t_i) + n\delta.
\end{aligned}$$

(Here the first equality holds because μ_N^k is a rational strategy for N with

respect to λ^k and α^k. The second equality is the Pareto-optimality of the Shapley value. The third equality follows from summing the warrant equations (5.3) over i and t_i. The final inequality follows from (5.8) and the fact that the $\lambda_i(t_i)$ sum to n, since $\lambda \in \Lambda^0$.) But $\bar{\mu}_N$ is incentive compatible; so by duality, if $\delta > 0$ then, for all sufficiently large k, $\bar{\mu}_N$ and α^k are respectively within $n\delta$ of the optimum in the primal and dual problems for λ^k.

The following theorem follows from (5.9), and lists some convenient necessary conditions for a bargaining solution. Notice that these conditions seem to be well-determined, in the sense that (5.10)-(5.13) can determine $\bar{\mu}_N$, α, μ, and ω, and (5.14) has one equation for each component in λ. This suggests a conjecture that the set of bargaining solutions might be generically finite.

Theorem 4. If $\bar{\mu}_N$ is a bargaining solution for Γ then there exist $(\lambda,\alpha,\mu,\omega)$ such that

(5.10) $\bar{\mu}_N$ is an optimal solution of the primal problem for λ;

(5.11) α is an optimal solution of the dual problem for λ;

(5.12) μ is a vector of rational threats for λ and α, and $\mu_N = \bar{\mu}_N$;

(5.13) ω is warranted by λ, α, and μ ;

(5.14) $\lambda_i(t_i) \geq 0$, $\omega_i(t_i) \leq U_i(\bar{\mu}_N | t_i)$, and

$\lambda_i(t_i)\,\omega_i(t_i) = \lambda_i(t_i)\,U_i(\bar{\mu}_N | t_i)$, $\quad \forall i \in N, \forall t_i \in T_i$;

(5.15) $(\lambda,\alpha) \neq (\underset{\sim}{0},\underset{\sim}{0})$.

(That is, λ may be any vector in the nonnegative orthant $\underset{i\in N}{\times} \mathbb{R}_+^{T_i}$, not necessarily in the simplex Λ. But, to avoid trivial solutions, λ and α cannot both be zero vectors.)

See Section 7 for the proof of this theorem.

6. Interpretation of the rational-threat criterion

Our rational-threat criterion (5.4) postulates that each coalition should seek to maximize the expected difference between the total virtual utility that its members would earn and the total virtual utility that the complementary coalition would earn, if both carried out their threats. The rational threats for coalitions other than N are not required to satisfy any equity or incentive-compatibility constraints. These aspects of our rational-threat criterion deserve some interpretive discussion.

In any bargaining situation, a coalition's threat normally has both defensive and offensive objectives. The defensive objective is to show that the coalition could maintain high payoffs for its members if the complementary coalition refused to cooperate. The offensive objective is to show that the complementary coalition's members would be hurt by such a breakdown in cooperation. Obviously, a threat that is strong both defensively and offensively would be the ideal; but the best defensive threat will generally not be the best for offensive purposes. Thus, a coalition may have to make some tradeoff between these two objectives. As observed by Harsanyi [1963], the Shapley value implicitly defines such a tradeoff, since it only depends on the difference $W_S - W_{N\setminus S}$. The defensive and offensive objectives are combined with this tradeoff in our rational-threats criterion.

This interpretation of the rational-threat criterion relies on our identifying W_S as the natural defensive objective function for coalition S. Once this identification is made, then the natural offensive objective function is $-W_{N\setminus S}$ (the opposite of the complement's defensive objective), and $W_S - W_{N\setminus S}$ is a natural combination of defensive and offensive objectives. But, in what sense is $W_S(\mu,t,\alpha,\lambda)$ an appropriate measure of the defensive strength of coalition S?

We can best understand the purely defensive aspect of threats by studying games in which each coalition can only influence its own members' payoffs, so that there are no offensive possibilities to consider. In the terminology of Shapley and Shubik [1973], these are games with orthogonal coalitions. That is, a game Γ has orthogonal coalitions iff,

$$(6.1) \qquad u_i((d_S,d_{N\setminus S}),t) = u_i((d_S,\hat{d}_{N\setminus S}),t)$$
$$\forall S\in CL,\ \forall i\in S,\ \forall d_S\in D_S,\ \forall d_{N\setminus S}\in D_{N\setminus S},\ \forall \hat{d}_{N\setminus S}\in D_{N\setminus S},\ \forall t\in T,$$

so that the threat of coalition $N\setminus S$ cannot affect the payoffs to members of S. Market games of pure exchange are examples of games with orthogonal coalitions.

In a game with orthogonal coalitions, suppose $i\in S$. Then, we can let $u_i(d_S,t)$ denote the utility payoff for i in state t if d_S in D_S is carried out. That is, $u_i(d_S,t) = u_i((d_S,d_{N\setminus S}),t)$ for any $d_{N\setminus S}$ in $D_{N\setminus S}$. (Recall $D_S \times D_{N\setminus S} \subseteq D$.) We similarly define $v_i(d_S,t,\lambda,\alpha)$ as $v_i((d_S,d_{N\setminus S}),t,\lambda,\alpha)$ for any $d_{N\setminus S}$. Then for any μ_S in M_S, the obvious generalizations of (2.6),(2.7) and (5.1) are:

$$U_i^*(\mu_S,r_i|t_i) = \sum_{t_{N-i}} p_i(t_{N-i}|t_i) \sum_{d_S} \mu_S(d_S|t_{S-i},r_i)\, u_i(d_S,t)$$

$$U_i(\mu_S|t_i) = U_i^*(\mu_S,t_i|t_i)$$

$$W_S(\mu_S,t,\lambda,\alpha) = \sum_{d_S} \mu_S(d_S|t_S) \sum_{j\in S} v_j(d_S,t,\lambda,\alpha).$$

Let us now consider what threats would be defensively optimal for a given player, say player 1, in a game with orthogonal coalitions. To be specific,

suppose that player 1 is acting as a coordinator or leader for all the coalitions to which he can belong. Suppose that, to maintain his leadership, player 1 must use a threat-plan that offers each type t_i of each player i at least some minimal expected utility $\omega_i(t_i)$. For any coalition $S \supseteq \{1\}$, let q_S denote the probability that S will be the coalition forming under player 1's leadership. Ordinarily, q_S would depend on how much player 1 offers the other players, but for simplicity let us suppose that q_N will be some fixed number close to one and all other q_S will be small positive numbers, for any threat-plan that gives all players at least their $\omega_i(t_i)$ payoffs. (Then q_S for $S \neq N$ may be thought of as a "trembling-hand" probability of the coalition S forming instead of N.)

In such a situation, if player 1's type is t_1, then he wants to choose his threat-plan $(\mu_S)_{S \supseteq \{1\}}$ in $\underset{S \supseteq \{1\}}{\times} M_S$ so as to maximize his expected utility $\sum_{S \supseteq \{1\}} q_S U_1(\mu_S | t_1)$ subject to the minimum-payoff constraints

$$(6.2) \qquad \sum_{S \supseteq \{1,i\}} q_S U_i(\mu_S | t_i) \geq \left(\sum_{S \supseteq \{1,i\}} q_S \right) \omega_i(t_i), \quad \forall i \in N-1, \ \forall t_i \in T_i,$$

and the incentive-compatibility constraints

$$(6.3) \qquad \sum_{S \supseteq \{1,i\}} q_S U_i(\mu_S | t_i) \geq \sum_{S \supseteq \{1,i\}} q_S U_i^*(\mu_S, r_i | t_i), \quad \forall i \in N, \ \forall t_i \in T_i, \ \forall r_i \in T_i.$$

This constraint (6.3) asserts that no player i should have any incentive to lie about his type when agreeing to follow player 1 in a coalition. We assume that 1 can negotiate separately with each other player i, so that i agrees without knowing which coalition $S \supseteq \{1,i\}$ will actually form.

To conceal his own type, player 1 must use a threat-plan which achieves some balance between the objectives of his various types. (See Myerson [1981]

for detailed discussion of this issue.) At the very least, however, player 1 should choose a threat-plan such that there is no other threat-plan satisfying (6.2) and (6.3) that gives higher expected utility to all types t_1 in T_1. For any such undominated threat-plan there must exist some vector λ such that $(\mu_S)_{S \supseteq \{1\}}$ maximizes

$$(6.4) \qquad \sum_{t_1 \in T_1} \lambda_1(t_1) \sum_{S \supseteq \{1\}} q_S \, U_1(\mu_S | t_1)$$

subject to the constraints (6.2) and (6.3).

So optimal defensive threats for player 1 should maximize (6.4) over $(\mu_S)_{S \supseteq \{1\}}$, subject to (6.2) and (6.3). The Lagrangian for this problem can be written as follows

$$(6.5) \qquad \sum_{t_1 \in T_1} \lambda_1(t_1) \sum_{S \supseteq \{1\}} q_S \, U_1(\mu_S | t_1)$$

$$+ \sum_{i \in N-1} \sum_{t_i \in T_i} \lambda_i(t_i) \sum_{S \supseteq \{1,i\}} q_S \, (U_i(\mu_S | t_i) - \omega_i(t_i))$$

$$+ \sum_{i \in N} \sum_{t_i \in T_i} \sum_{r_i \in T_i} \alpha_i(r_i | t_i) \sum_{S \supseteq \{1,i\}} q_S \, (U_i(\mu_S | t_i) - U_i^*(\mu_S, r_i | t_i))$$

$$= \sum_{S \supseteq \{1\}} q_S \left(\sum_{t \in T} p(t) \, W_S(\mu_S, t, \lambda, \alpha) - \sum_{i \in S-1} \sum_{t_i \in T_i} \lambda_i(t_i) \, \omega_i(t_i) \right).$$

(This equality follows straightforwardly from the definitions of U_i, W_S, and virtual utility.)

Thus, by (6.5), in any plan of optimal defensive threats for player 1, there must exist some λ and α such that every coalition is choosing a threat that maximizes the expected sum of its members' virtual utilities with respect to λ and α. The maximum value of 1's weighted objective function (6.4) is

equal to the expected sum of these virtual utilities for the coalition forming around player 1, minus terms in (6.5) that do not depend on the threat-plans. This is exactly the result that we wanted, since it shows that the sum of virtual utilities can be a valid measure of the defensive strength of a coalition.

In particular, suppose that q_N is almost one, and all other q_S are only infinitesimal probabilities. Then constraints (6.2) and (6.3) require that μ_N must be (almost) incentive compatible and must give all players their minimum payoffs (or at most infinitesimally less). Any other coalitional threat μ_S does not need to be either incentive compatible or equitable, as it is only a small component of a plan which is incentive compatible and equitable overall. Thus, our rational-threat criterion can be justified in situations where the coalitions commit themselves to their threats and gather type-reports from their members before the realized coalition structure is determined, provided that all players believe that the probability of the grand coalition forming is close to one. In such situations, the value of a threat for a coalition S ($S \neq N$) depends on what it can contribute to the required utility and incentive-compatibility of the overall plan. With appropriate shadow prices λ and α, expected virtual utility W_S measures this contribution. For $S \neq N$, the threat μ_S does not need to be either equitable or incentive compatible itself, because the members of S do not expect to carry out this threat when they agree to make it part of their threat-plans.

7. Proofs

First, we cite a basic lemma.

Lemma 1: Given any player i, α in $\underset{\sim}{A}$, λ in Λ^0, and h_i in $\mathbb{R}^{T_i}$, there is a unique vector ω_i in $\mathbb{R}^{T_i}$ satisfying

$$(7.1) \qquad \Big(\lambda_i(t_i) + \sum_{s_i} \alpha_i(s_i|t_i)\Big)\omega_i(t_i) - \sum_{s_i} \alpha_i(t_i|s_i)\omega_i(s_i) = h_i(t_i) \ , \quad \forall t_i \in T_i .$$

Furthermore, the solution ω_i to these linear equations is increasing in the vector h_i. (That is, if $h_i'(t_i) \geq h_i(t_i)$ $\forall t_i$, and ω_i' solves (7.1) for h_i' instead of h_i, then $\omega_i'(t_i) \geq \omega_i(t_i)$ $\forall t_i$.)

This result is proven as "Lemma 1" in Myerson [1981].

Lemma 2: Suppose that μ is a vector of rational threats with respect to λ and α, and ω is the vector of warranted claims for λ, α, and μ, where $\lambda \in \Lambda^0$ and $\alpha \in \underset{\sim}{A}$. Then

$$\omega_i(t_i) \geq \underset{\nu_{N-i} \in M_{N-i}}{\text{minimum}} \quad \underset{\nu_i \in M_{\{i\}}}{\text{maximum}} U_i(\nu_{N-i}, \nu_i | t_i).$$

for any player i and type t_i.

Proof: Let i be any fixed player. For any coalition $S \supseteq \{i\}$, let μ_i^S be a mechanism in $M_{\{i\}}$ such that

$$\mu_i^S \in \underset{\nu_i \in M_{\{i\}}}{\text{argmax}} \ U_i(\mu_{S-i}, \nu_i, \mu_{N \setminus S} | t_i)$$

for every t_i and T_i. That is, $\mu_i^S(d_i|t_i) > 0$ only if d_i would be a best response for player i if his type were t_i and the coalitions $S-i$ and $N \setminus S$ were expected to independently implement their threats from μ. Then let $\hat{\mu}$ in M be

defined so that

$$\hat{\mu}_S = (\mu_{S-i}, \mu_i^S) \quad \text{if } i \in S$$
$$\hat{\mu}_S = \mu_S \quad \text{if } i \notin S.$$

For the threat-vector $\hat{\mu}$, no coalition changes its threat when player i joins it. So $W_S(\hat{\mu},t,\lambda,\alpha)$ and $W_{S-i}(\hat{\mu},t,\lambda,a)$ differ only by the addition of i's expected virtual utility in state t when μ_{S-i}, μ_i^S, and $\mu_{N\setminus S}$ are carried out. Thus, for any t_i,

$$\sum_{t_{N-i}} p(t)\left(W_S(\hat{\mu},t,\lambda,\alpha) - W_{S-i}(\hat{\mu},t,\lambda,\alpha)\right)$$

$$= \left(\lambda_i(t_i) + \sum_{r_i} \alpha_i(r_i|t_i)\right) U_i(\mu_{S-i}, \mu_i^S, \mu_{N\setminus S}|t_i)$$

$$- \sum_{r_i} \alpha_i(t_i|r_i)\, U_i^*((\mu_{S-i}, \mu_i^S, \mu_{N\setminus S}), t_i|r_i).$$

Let $$\eta_i(t_i) = \sum_{S \supseteq \{i\}} \frac{(|S|-1)!(n-|S|)!}{n!} U_i(\mu_{S-i}, \mu_i^S, \mu_{N\setminus S}|t_i)$$

$$\eta_i^*(t_i|r_i) = \sum_{S \supseteq \{i\}} \frac{(|S|-1)!(n-|S|)!}{n!} U_i^*((\mu_{S-i}, \mu_i^S, \mu_{N\setminus S}), t_i|r_i).$$

Since $\mu_i^S(\cdot|r_i)$ is the best response for type r_i against $(\mu_{S-i}, \mu_{N\setminus S})$, for each S, it follows that $\eta_i(r_i) \geq \eta_i^*(t_i|r_i)$.

Consider now the following chain of inequalities

$$\left(\lambda_i(t_i) + \sum_{r_i} \alpha_i(r_i|t_i)\right)\eta_i(t_i) - \sum_{r_i} \alpha_i(t_i|r_i)\, \eta_i(r_i)$$

$$\leq \left(\lambda_i(t_i) + \sum_{r_i} \alpha_i(r_i|t_i)\right)\eta_i(t_i) - \sum_{r_i} \alpha_i(t_i|r_i)\, \eta_i^*(t_i|r_i)$$

$$= \sum_{t_{N-i}} p(t) \sum_{S \supseteq \{i\}} \frac{(|S|-1)!(n-|S|)!}{n!} \left(W_S(\hat{\mu},t,\lambda,\alpha) - W_{S-i}(\hat{\mu},t,\lambda,\alpha)\right)$$

$$= \sum_{t_{N-i}} p(t) \sum_{S \supseteq \{i\}} \frac{(|S|-1)!(n-|S|)!}{n!} \left(W_S((\mu_{-S},\hat{\mu}_S),t,\lambda,\alpha) - W_{N\setminus S}((\mu_{-S},\hat{\mu}_S),t,\lambda,\alpha)\right)$$

$$\leq \sum_{t_{N-i}} p(t) \sum_{S \supseteq \{i\}} \frac{(|S|-1)!(n-|S|)!}{n!} \left(W_S(\mu,t,\lambda,\alpha) - W_{N\setminus S}(\mu,t,\lambda,\alpha)\right)$$

$$= \left(\lambda_i(t_i) + \sum_{r_i} \alpha_i(r_i|t_i)\right)\omega_i(t_i) - \sum_{r_i} \alpha_i(t_i|r_i)\,\omega_i(r_i).$$

In this chain, the fourth line holds because $W_S(\hat{\mu},t,\lambda,\alpha)$ and $W_{N\setminus S}(\hat{\mu},t,\lambda,a)$ depend only on $\hat{\mu}_S$ and $\hat{\mu}_{N\setminus S} = \mu_{N\setminus S}$. Then the next inequality uses the fact that μ is a vector of rational threats, and the last equality uses the fact that ω is the vector of warranted claims.

Since the above chain of inequalities holds for all t_i, Lemma 1 implies that $\omega_i(t_i) \geq \eta_i(t_i)$ for all t_i. But $\eta_i(t_i)$ is an average of best-response payoffs for type t_i against a variety of mechanisms for N-i, and so $\eta_i(t_i)$ is not smaller than the right-hand side of the inequality in Lemma 2. Q.E.D.

Proof of Theorem 2

We begin with some definitions. For any k larger than $\sum_{i=1}^{n} |T_i|$, let

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

There exists a compact convex set $\underset{\sim}{A}^*$ such that $\underset{\sim}{A}^* \subseteq \underset{\sim}{A}$ and, for each λ in Λ there is some α in $\underset{\sim}{A}^*$ such that α is an optimal solution of the dual for λ. The proof of this fact is given in the proof of Theorem 6 in Myerson [1981].

Let $B = \underset{i,d,t}{\text{maximum}} |u_i(d,t)| + 1$

and let $X = \{\omega \in \underset{i \in N}{\times} \mathbb{R}^{T_i} \mid -B \leq \omega_i(t_i) \leq B, \forall i, \forall t_i\}$.

For each k greater than $\sum_{i \in N} |T_i|$, we define a correspondence $Z^k: M \times \underset{\sim}{A}^* \times X \times \Lambda^k \Rightarrow M \times \underset{\sim}{A}^* \times X \times \Lambda^k$ so that $(\hat{\mu},\hat{\alpha},\hat{\omega},\hat{\lambda}) \in Z^k(\mu,\alpha,\omega,\lambda)$ iff the following conditions are satisfied

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

(7.3) $\hat{\mu}_S \in \underset{\nu_S \in M_S}{\text{argmax}} \sum_{t \in T} p(t)\left(W_S((\mu_{-S},\nu_S)t,\lambda,\alpha) - W_{N \setminus S}((\mu_{-S},\nu_S)t,\lambda,\alpha)\right)$.

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

(7.5) $\hat{\omega}_i(t_i) = \max\{-B, \min\{B, \tilde{\omega}_i(t_i)\}\}$, $\forall i, \forall t_i$, where $\tilde{\omega}$ is the vector of claims warranted by λ, α, and μ;

(7.6) $\lambda_i(t_i) = 1/k$ for every t_i such that

$$\omega_i(t_i) - U_i(\mu_N | t_i) < \underset{j,s_j}{\text{maximum}} (\omega_j(s_j) - U_j(\mu_N | s_j)).$$

By the Kakutani fixed-point theorem, for each k there exists some $(\mu^k,\alpha^k,\omega^k,\lambda^k)$ such that

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

Since this sequence of fixed points is in a compact domain, there exists a convergent sequence, converging to some $(\bar{\mu},\bar{\alpha},\bar{\omega},\bar{\lambda})$ in $M \times \underset{\sim}{A}^* \times X \times \Lambda$. We will show that $\bar{\mu}_N$ is a bargaining solution.

By the fixed-point condition, each μ^k is a vector of rational threats for λ^k and α^k.

Let $\tilde{\omega}^k$ be the vector of claims warranted by λ^k, α^k, and μ^k. By Lemma 2, since $\tilde{\omega}^k$ is a vector of warranted claims supported by rational threats, $\tilde{\omega}^k_i(t_i) \geq -B$ for every i and t_i. Thus $\omega^k_i(t_i)$ can differ from

$\tilde{\omega}_i^k(t_i)$ only if $\tilde{\omega}_i^k(t_i) > B$, in which case $\omega_i^k(t_i) \leq \tilde{\omega}_i^k(t_i)$.

By summing the warrant equations, we get

$$\sum_{i\in N} \sum_{t_i \in T_i} \lambda_i^k(t_i)\, \tilde{\omega}_i^k(t_i) = \sum_{i\in N} \lambda_i^k(t_i)\, U_i(\mu_N^k | t_i).$$

For any i and t_i, if $\tilde{\omega}_i^k(t_i) < U_i(\mu_N^k | t_i)$ then, by (7.6) and (7.7), $\lambda_i^k(t_i) = 1/k$; thus

$$\text{(7.8)} \qquad \text{if } \liminf_{k\to\infty} \tilde{\omega}_i^k(t_i) < U_i(\bar{\mu}_N | t_i) \text{ then } \lim_{k\to\infty} \lambda_i^k(t_i) = 0.$$

Now, suppose that there were some j and r_j such that

$$\text{(7.9)} \qquad \limsup_{k\to\infty} \tilde{\omega}_j^k(r_j) > U_j(\bar{\mu}_N | r_j) = \lim_{k\to\infty} U_j(\mu_N^k | r_j).$$

Then (7.8) could be strengthened to:

$$\text{if } \liminf_{k\to\infty} \tilde{\omega}_i^k(t_i) \leq U_i(\bar{\mu}_N | t_i) \text{ then } \lim_{k\to\infty} \lambda_i^k(t_i) = 0.$$

Since each λ^k is in the simplex Λ, we could find some j and r_j satisfying (7.9) such that $\bar{\lambda}_j(r_j) > 0$. But then we would get

$$0 < \limsup_{k\to\infty} \lambda_j^k(r_j)(\tilde{\omega}_j^k(r_j) - U_j(\mu_N^k | r_j))$$

$$= \limsup_{k\to\infty} \sum_{(i,t_i)\neq(j,r_j)} \lambda_i^k(t_i)(U_i(\mu_N^k | t_i) - \tilde{\omega}_i^k(t_i)) \leq 0,$$

using (7.8) (and the fact that $\tilde{\omega}_i^k(t_i)$ does not diverge to $-\infty$ as $k\to\infty$, since it is bounded below by -B) to get the last inequality. But $0 < 0$ is impossible, so no (j,r_j) pair satisfying (7.9) can exist. That is, for every i and t_i,

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

So $\{(\lambda^k,\alpha^k,\mu^k,\tilde{\omega}^k)\}_{k=1}^{\infty}$ form a sequence verifying (5.5)-(5.8) for $\bar{\mu}_N$. Q.E.D.

Proof of Theorem 3

Theorem 3 follows immediately from Lemma 2 and the definition of a bargaining solution.

Proof of Theorem 4

Given the bargaining solution $\bar{\mu}_N$, let $\{(\lambda^k,\alpha^k,\mu^k,\omega^k)\}_{k=1}^{\infty}$ satisfy (5.5)-(5.8). Let $\hat{\lambda}^k$ and $\hat{\alpha}^k$ be defined by

$$\hat{\lambda}_i^k(t_i) = \lambda_i^k(t_i)/(|\lambda^k| + |\alpha^k|)$$

$$\hat{\alpha}_i^k(r_i|t_i) = \alpha_i^k(r_i|t_i)/(|\lambda^k| + |\alpha^k|)$$

where

$$|\lambda^k| + |\alpha^k| = \sum_{i\in N} \sum_{t_i\in T_i} (\lambda_i^k(t_i) + \sum_{r_i\in T_i} \alpha_i^k(r_i|t_i)) \geqslant n.$$

So for each k, $(\hat{\lambda}^k,\hat{\alpha}^k)$ lies in a unit simplex. By the linear homogeneity of all formulas concerned, μ^k is a vector of rational threats for $\hat{\lambda}^k$ and $\hat{\alpha}^k$, as well as for λ^k and α^k, and ω^k is warranted by $\hat{\lambda}^k$, $\hat{\alpha}^k$, and μ^k. By Lemma 2 and condition (5.8), each ω^k must lie within the compact set X defined in the proof of Theorem 2, and each μ^k is in the compact set M. So there must exist a subsequence $\{(\hat{\lambda}^k,\hat{\alpha}^k,\mu^k,\omega^k)\}_k$ that is convergent to some limit $(\lambda,\alpha,\mu,\omega)$.

The vectors λ and α cannot both be zero, because (λ,α) has a summation-norm of one. By continuity of the rational-threat and warranted-claim conditions, μ is a vector of rational threats for λ and α, and ω is warranted by λ, α, and μ. By (5.8), $\omega_i(t_i) \leqslant U_i(\bar{\mu}_N|t_i)$ for every i and t_i. From

(5.9) (dividing through by $|\lambda^k|+|\alpha^k|$, and letting $\delta \to 0$ as $k \to \infty$), we get

$$\sum_{t \in T} p(t) \max_{d \in D} \sum_{i \in N} v_i(d,t,\lambda,\alpha) \leq \sum_i \sum_{t_i} \lambda_i(t_i)\, U_i(\bar{\mu}_N | t_i),$$

and so by duality $\bar{\mu}_N$ and α are optimal solutions of the primal and dual for λ, respectively. Duality also implies that the above inequality must be an equality, which gives us the complementary slackness conditions in (5.14).

Thus we have all of the conditions in Theorem 4, except that letting $\mu_N = \lim_{k \to \infty} \mu_N^k$ does <u>not</u> imply $\mu_N = \bar{\mu}_N$. However, since $\bar{\mu}_N$ is an optimal solution of the primal for λ, it must also maximize the sum of the virtual utilities in every state. Thus, if we redefine μ_N as being equal to $\bar{\mu}_N$, we do not change $W_N(\mu,t,\lambda,\alpha)$ for any t, and so ω is still warranted by λ,α,μ. Q.E.D.

REFERENCES

D'Aspremont, C., and L.-A. Gerard-Varet [1979], "Incentives and Incomplete Information," Journal of Public Economics 11, 25-45.

Harsanyi, J. C. [1963], "A Simplifed 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.

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

Holmström, B. and R. B. Myerson [1981], "Efficient and Durable Decision Rules with Incomplete Information," mimeo, Northestern University. To appear in Econometrica.

Kakutani, S. [1941], "A Generalization of Brouwer's Fixed Point Theorem," Duke Journal of Mathematics 8, 457-459.

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

Myerson, R. B. [1981], "Mechanism Design by an Informed Principal," mimeo, Northestern University. To appear in Econometrica.

Myerson, R. B. [1982], "Two-Person Bargaining Problems with Incomplete Information," Northwestern University, mimeo. To appear in Econometrica.

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 Decision, 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.