Discussion Paper No. 288

THREAT EQUILIBRIA AND FAIR SETTLEMENTS IN COOPERATIVE GAMES

by

Roger B. Myerson*

July 1977

Abstract: The role of threats is studied in cooperative normal form games. A threats-game is constructed, in which every set of players selects a joint threat strategy and then a settlement function determines the final outcome. Threat equilibria and cooperative solutions are defined for any settlement function. Two axioms are introduced which determine a unique settlement function for games with transferable utility. This settlement function is closely related to the Shapley value, and has attractive Pareto-optimality and individual-rationality properties. A simple oligopoly problem is studied to illustrate these ideas.

*Graduate School of Management, Northwestern University, Evanston, IL 60201.

THREAT EQUILIBRIA AND FAIR SETTLEMENTS IN COOPERATIVE GAMES

by

Roger B. Myerson

1. Introduction

In cooperative situations involving many individuals, threats can play an important part in determining how the fruits of cooperation will be allocated to the participants. To study the role of threats in cooperation, we may conceptually divide the negotiation process into two stages. First comes the threats stage, when individuals and groups announce threats against each other. Then comes the settlement stage, in which an agreement is generated, allocating payoffs to individuals in proportion to their power (as expressed in the threats).

To describe how threats might be generated in an n-person cooperative game, Harsanyi [1] suggested constructing a larger noncooperative game in which there are 2^n-1 agents, each of whom must select a threat strategy for one possible coalition in the original game. The Shapley value [[illegible]] has an important role in Harsanyi's model, as what we will call a settlement function. In this paper we will see why it may be more natural to revise Harsanyi's model by using the partition function value [5] in place of the Shapley value.

2. The Basic Framework

We will describe our basic conflict situation by an n-person game in normal form. Formally, Γ is an n-person game in normal form if

(1) $$\Gamma = (N; D_1, \ldots, D_n; U_1, \ldots, U_n)$$

where $N = \{1,2,\ldots,n\}$, each D_i is a nonempty finite set, and each U_i is a real-valued function defined on the domain $D_1 \times D_2 \times \ldots \times D_n$. We interpret N as the set of players in the game, who are numbered from 1 to n. For each player i, D_i is the set of pure strategies which player i could choose if he played the game noncooperatively. And each U_i is the utility function for player i, so that $U_i(d_1,\ldots,d_n)$ would be the payoff to player i (measured in some vonNeumann-Morgenstern utility scale) if $(d_1,\ldots,d_n)$ were the combination of strategies chosen by the players.

If Γ is a cooperative game, then the players can form coalitions to choose their strategies jointly. For any set of players $S \subseteq N$, the set of joint strategies for S, which would be available to the members of S if they acted together as a coalition, is

(2) $$B_S = \times_{i \in S} D_i .$$

For example, $B_{\{1\}} = D_1$ and $B_{\{1,3\}} = D_1 \times D_3$.

Let CL be the set of all possible coalitions which could form among the players in N; that is:

(3) $$CL = \{S : S \subseteq N,\ S \neq \emptyset\} .$$

A threat for coalition S is a commitment by the members of S to carry out some joint strategy in B_S if the negotiations should break down during settlement stage in such a way that the members of S are in agreement with each other but with no other players. We will assume that a threat will be generated for every possible coalition in CL during the threats stage of the negotiations. That is, the set of all possible combinations of threats which may be generated is $\times_{S \in CL} B_S$.

Notice that we are allowing all 2^n-1 sets of players to plan threats against each other, not just some disjoint collection of coalitions. There is no inconsistency problem here, since the threats of intersecting coalitions are conditioned on mutually exclusive events. For example, suppose $N = \{1,2,3\}$. Then a threat for coalition $\{1,2\}$ is a conditional commitment to carry out some joint strategy $b_{\{1,2\}} = (d_1,d_2)$ if players 1 and 2 should find themselves in agreement against 3 after a breakdown of negotiations. On the other hand, a threat for coalition $\{1,3\}$ is a conditional commitment to carry out some joint strategy $b_{\{1,3\}} = (d_1',d_3')$ if 1 and 3 should find themselves in agreement against 2 after a breakdown of negotiations. Thus there is no reason to require that d_1 should equal d_1', since these are commitments for player 1 conditioned on mutually exclusive events.

Since the universal coalition N is itself a coalition, we are implicitly including the possibility of a threat for coalition N. One may object that the conditional plans of the universal coalition should not be called a "threat", since there is no one left to threaten. In a later section, we will assume that the threat for coalition N should be part of the final settlement which is actually carried out. Nevertheless, for uniformity of terminology, we will refer to the plans of all coalitions, including N, as "threats".

So far we have only discussed the first stage of negotiations, the threats stage. The outcome of the threats stage is the vector of coalition threats $b \in \times_{S \in CL} B_S$ selected by the 2^n-1 coalitions. But the significance of these threats is measured by their impact on the second stage of negotiations, the settlement stage.

At this point we shall leave the settlement stage of negotiations as something of a black box. One may imagine that an arbitrator will come along, look at all the threats which have been made, and then compute some fair payoff allocation to be the final binding settlement. But the only assumptions we require now are that the settlement stage of the negotiations will always generate some payoff allocation and that this outcome will depend on some or all of the coalitions' threats. Thus, the settlement stage of the negotiations may be modeled by some settlement function of the form

$$(4) \qquad f\colon \underset{S \in CL}{\times} B \to R^n$$

mapping vectors of threats into payoff allocation vectors. That is, if $f(\cdot)$ were the settlement function in a negotiation process, and if $b = (b_S)_{S \in CL}$ were the combination of threats chosen by the various coalitions, then $f_i(b)$ would be the player i's expected utility payoff in the final settlement.

3. Threat Equilibria and Cooperative Solutions

In this section we assume that a settlement function $f\colon \underset{S \in CL}{\times} B_S \to R^n$ has been specified for our game Γ. (We shall return to the problem of finding a reasonable settlement function in Section 4.) Given this settlement function, we can try to answer the question: how should the coalitions choose their threats?

We may assume that, for each $S \in CL$, there is some coalition agent responsible for selecting the threat $b_S \in B_S$ for S. Then the threats stage of the negotiation process begins to look like a normal form game, in which CL is the set of "players", and B_S is the set of pure strategies for coalition S. All that we need to complete the normal form structure is to specify a utility function for each coalition S.

The agent for coalition S does not expect his threat to be actually carried out, so he must judge his threat solely in terms of its impact on what the members of S will get in the final settlement. Thus a reasonable utility function for coalition S in the threats game is $U_S^f\colon \times_{S\in CL} B_S \to R$ where:

$$(5) \qquad U_S^f(b) = \sum_{i\in S} f_i(b) .$$

That is, the "payoff" to coalition S is the sum of the payoffs to the members of S in the final settlement allocation.

The threats-game for Γ, with respect to the settlement function f, is then defined to be:

$$(6) \qquad \Gamma^*(f) = (CL;(B_S)_{S\in CL};(U_S^f)_{S\in CL}).$$

$\Gamma^*(f)$ is a mathematical game in normal form, and so, by Nash [6], it must have an equilibrium in mixed strategies (that is, in which the coalitions may randomize their threat selections). We define a threat equilibrium for Γ to be an equilibrium for this threats-game $\Gamma^*(f)$.

In a threat equilibrium, every coalition's threat maximizes the total payoff which its members can expect from the final settlement, given the threats planned by the other coalitions. In this sense, the equilibrium threats are the rational threats for coalitions to use in negotiations. Thus, we can define a cooperative solution for Γ (with respect to f) to be any payoff allocating $x \in R^n$ such that $x = f(b)$ for some threat equilibrium b (or such that $x = E(f(\tilde{b}))$ for some mixed-strategy threat equilibrium $\tilde{b}$).

4. The Fair Settlement Function

In the last section, we saw that it is straightforward to develop a theory of cooperative solutions and rational threats for Γ, once a settlement function has been specified. So the crucial step in building a theory of cooperative games must be the construction of a settlement function. In this section, we will define two axioms which a reasonable settlement function might satisfy, and we will show the unique function satisfying both axioms.

We shall henceforth assume that utility is transferable between the players in Γ. That is, the payoffs to the players are assumed to be measured in units of some freely exchangeable commodity, like money. See [2] for some theoretical arguments to justify the important role which this transferrable utility assumption has played in game theory.

One basic requirement of a settlement function is that its allocation should always be feasible. With transferable utility, we can guarantee feasibility if $f(\cdot)$ satisfies the following equation:

$$(7) \qquad \sum_{i\in N} f_i(b) = \sum_{i\in N} U_i(b_N) \quad \text{for any} \quad b \in \mathop{\times}_{S\in CL} B_S .$$

(Here b_N is the N-component of the vector $b = (b_S)_{S\in CL}$. Notice that $b_N \in B_N = \mathop{\times}_{i\in N} D_i$, so $U_i(b_N)$ is well-defined.) The idea behind (7) is that the joint strategy chosen by the universal coalition N should be actually carried out in the settlement, but that it might be followed by some utility transfers or sidepayments between the players. We can think of these transfers as bribes which may be required to induce some smaller coalitions not to carry out their threats.

Axiom 1 expresses a stronger version of (7). The basic idea is that, if there are two groups of players who effectively ignore each other in all coalition plans, then the fair settlement should not require any coordination or transfers between these two groups. To develop this idea formally, we need some additional notation.

Suppose S and T are two disjoint sets of players, and suppose $b_S = (d_i)_{i\in S}$ and $b_T = (d_i)_{i\in T}$ are threats for S and T (in B_S and B_T respectively). Then we let $\langle b_S, b_T\rangle = (d_i)_{i\in S\cup T}$. That is, $\langle b_S, b_T\rangle$ is the threat for $S\cup T$ in which the S and T factions act as in their b_S and b_T threats separately.

For any set of players $T \subseteq N$ and any threats-combination $b = (b_S)_{S \in CL}$, we say that b is <u>T-decomposable</u> iff

$$(8) \qquad b_S = \langle b_{S\cap T}, b_{S\setminus T} \rangle, \text{ for every } S \in CL.$$

When threats are T-decomposable, there is no effective threat coordination between the members of T and those outside T. Axiom 1 suggests that, if the vector of threats happens to be T-decomposable, then there should be no utility transfers between the players in T and those outside T.

<u>Axiom 1</u>. For any $T \subseteq N$ and any $b \in \underset{S \in CL}{\times} B_S$,

if b is T-decomposable then

$$\sum_{i \in T} f_i(b) = \sum_{i \in T} U_i(b_N).$$

(Notice that $b_N = \langle b_T, b_{N\setminus T} \rangle$ if b is T-decomposable.

Notice also that b is always N-decomposable, so Axiom 1 implies (7).)

Consider the problem of a coalition S as it chooses its strategy in the threat-game. According to (5), the criterion to be maximized in the selection of $b_S \in B_S$ is $\sum_{i \in S} f_i(b)$. But a typical player $j \in S$ really wants to maximize $f_j(b)$. If these criteria are not strategically equivalent for the purposes of choosing b_S, then there may be some conflicting interests among the members of S when they choose their threat. To prevent such conflict of interests, a fair settlement function should be designed so that all members of coalition S would always gain or lose equally by a change in their collective S-threat b_S. A natural unit for comparing gains of different players is the transferrable unit of utility. So Axiom 2 suggests that the difference between what two players get in the final settlement should not be influenced by the threat of a coalition to which both belong.

Axiom 2. For any $S \in CL$, any $i \in S$, and any $j \in S$, $f_i(b) - f_j(b)$ does not depend on $b_S \in B_S$. That is, for any $b \in \underset{T \in CL}{\times} B_T$ and any $b'_S \in B_S$, $f_i(b) - f_j(b) = f_i(b_{-S}, b'_S) - f_j(b_{-S}, b'_S)$.
(Here (b_{-S}, b'_S) denotes the vector identical to b except that the S-component is changed to b'_S.)

Let PT be the set of all partitions of N; that is:

$$\text{(9)} \qquad PT = \{\{S_1, S_2, \ldots, S_k\} : S_i \cap S_j = \emptyset \text{ if } i \neq j, \bigcup_{i=1}^{k} S_i = N\}$$

If S is a coalition, we will let s be the number of players in S; and if Q is a partition, we will let q be the number of coalitions in Q.

Our main result is that Axioms 1 and 2 uniquely determine the settlement function $f(\cdot)$.

Theorem 1. A settlement function $f: \underset{S \in CL}{\times} B_S \to R^n$ satisfies Axioms 1 and 2 if and only if, for all $i \in N$:

$$f_i(b) = \sum_{j \in N} \sum_{Q \in PT} (-1)^{q-1}(q-1)! \left(\frac{1}{n} - \sum_{\substack{S \in Q \\ (i \notin S) \\ (j \notin S)}} \frac{1}{(q-1)(n-s)} \right) U_j(\langle b_T \rangle_{T \in Q})$$

($U_j(\langle b_T \rangle_{T \in Q})$ is the utility payoff which player j would get before transfers if all the coalitions in Q carried out their threats.)

5. Relationship to the Shapley Value

The formula in Theorem 1 can be interpreted in terms of partition function games. We define an embedded coalition to be an ordered pair (S,Q) such that Q is a partition of N and S is one of the coalitions in Q. Let ECL be the set of all embedded coalitions, so:

(10) $$ECL = \{(S,Q): Q \in PT, S \in Q\} .$$

Lucas and Thrall [3] defined a partition function game to be a vector w in R^{ECL} (so $w = (w_{S,Q})_{(S,Q)\in ECL}$). In a partition function game w, the (S,Q)-component $w_{S,Q}$ is interpreted as the wealth of transferrable utility which would be available to the members of coalition S if the players were aligned into the coalitions of partition Q.

Given a vector of coalition threats $b \in \times_{S\in CL} B_S$, let W(b) be the partition function game defined by:

(11) $$W_{S,Q}(b) = \sum_{i\in S} U_i(\langle b_T\rangle_{T\in Q}), \quad \forall (S,Q)\in ECL.$$

It can be shown that the formula in Theorem 1 is then equivalent to:

(12) $$f(b) = \Phi(W(b)),$$

where $\Phi: R^{ECL} \to R^N$ is the natural generalization of the Shapley value to partition function games, as defined in [5].

Harsanyi presents a bargaining model in [1], which differs from ours mainly in that he uses a settlement function based on the Shapley value for characteristic function games. To be precise, we may define $V: \times_{S\in CL} B_S \to R^{CL}$ by:

(13) $$V_S(b) = \sum_{i\in S} U_i(\langle b_S, b_{N\setminus S}\rangle), \quad \forall\, S \in CL.$$

Then Harsanyi's settlement function (for games with transferable utility) turns out to be

$$(14) \qquad \hat{f}(b) = \varphi(V(b)),$$

where φ is the Shapley value [8]. Harsanyi's settlement function satisfies our Axiom 2, but does not satisfy Axiom 1.

For two-person games (n=2) with transferable utility, the formulas of (12) and (14) coincide, and both generate the same cooperative solution theory as the Nash bargaining solution [7]. For $n \geq 3$, however, (12) and (14) are strictly different settlement functions.

6. Individual Rationality and Pareto-Optimality

Let f be as in Theorem 1, and suppose $b \in \underset{S \in CL}{\times} B_S$ is a threat-equilibrium for Γ with respect to f. If b is the vector of threats made in the first stage of the negotiations, then player i will expect payoff $f_i(b)$ from the final settlement. But can player i do better on his own?

We must allow that any player i always has the option to reject the settlement, and to drop out of the universal coalition. This action would leave player i alone against the players in the complementary coalition $N\backslash\{i\}$. The threats-vector b tells us what they would do; they would have to carry out their threat $b_{N\backslash\{i\}}$. Theorem 2 asserts that player i could not do better against this threat than he can do in the settlement. Thus, no player should want to reject the settlement if everyone else is willing to accept it. In this sense, the settlement in our cooperative solution is individually rational or stable.

Theorem 2. If b is a threat equilibrium with respect to f (the settlement function from Theorem 1), then for any $i \in N$ and any $d_i \in D_i$:

$$f_i(b) \geq U_i(b_{N\setminus\{i\}}, d_i).$$

Theorem 3 asserts that the settlement in our cooperative solution is also Pareto-optimal.

Theorem 3. For b and f as in Theorem 2,

$$\sum_{i=1}^{n} f_i(b) = \sum_{i=1}^{n} U_i(b_N) = \underset{b'_N \in B_N}{\text{maximum}} \sum_{i=1}^{n} U_i(b'_N).$$

7. Example

Consider a three-person game in which:

$$N = \{1,2,3\};$$

$$D_1 = \{0,1\};\ D_2 = \{2,3\} = D_3;$$

$$U_i(d_1,d_2,d_3) = d_i \cdot (8 - d_1 - d_2 - d_3).$$

We can interpret this game as an oligopoly problem, where player 1 is a small firm and players 2 and 3 are large firms, all producing for the same market. The small firm can produce either no output or one unit, and each of the large firms can produce either 2 or 3 units of output. Then the market price depends on total output in such a way that profit-per-unit equals eight minus total output. Each firm's profit equals its own output times the profit-per-unit.

The set of coalitions playing in the threats game $\Gamma^*(f)$ is:

$$CL = \{\{1\},\{2\},\{3\},\{1,2\},\{1,3\},\{2,3\},\{1,2,3\}\}.$$

The coalition strategy sets are as follows:

$B_{\{1\}} = \{(0),(1)\}$; $B_{\{2\}} = \{(2),(3)\} = B_{\{3\}}$;

$B_{\{1,2\}} = \{(0,2),(1,2),(0,3),(1,3)\} = B_{\{1,3\}}$;

$B_{\{2,3\}} = \{(2,2),(2,3),(3,2),(3,3)\}$; and

$B_{\{1,2,3\}} = \{(0,2,2),(1,2,2),(0,3,2),(0,2,3),(1,2,2),(1,3,2),(1,2,3),(0,3,3),(1,3,3)\}$

When n=3, our formula for f in Theorem 1 turns out to give:

$$
\begin{aligned}
(15)\qquad f_1(b) = {} & \tfrac{1}{3}(U_1(b_{\{1,2,3\}}) + U_2(b_{\{1,2,3\}}) + U_3(b_{\{1,2,3\}})) + \\
& + \tfrac{1}{6}(U_1(b_{\{1,2\}},b_{\{3\}}) + U_2(b_{\{1,2\}},b_{\{3\}})) - \tfrac{1}{3}U_3(b_{\{1,2\}},b_{\{3\}}) \\
& + \tfrac{1}{6}(U_1(b_{\{1,3\}},b_{\{2\}}) + U_3(b_{\{1,3\}},b_{\{2\}})) - \tfrac{1}{3}U_2(b_{\{1,3\}},b_{\{2\}}) \\
& + \tfrac{2}{3}(U_1(b_{\{1\}},b_{\{2,3\}}) - \tfrac{1}{3}(U_2(b_{\{1\}},b_{\{2,3\}}) + U_3(b_{\{1\}},b_{\{2,3\}})) \\
& + \tfrac{1}{6}U_2(b_{\{1\}},b_{\{2\}},b_{\{3\}}) + \tfrac{1}{6}U_3(b_{\{1\}},b_{\{2\}},b_{\{3\}}) \\
& - \tfrac{1}{3}U_1(b_{\{1\}},b_{\{2\}},b_{\{3\}}),
\end{aligned}
$$

along with similar equations for $f_2(b)$ and $f_3(b)$.

Computing the utility functions for the threats game $\Gamma^*(f)$ as in (5), we can then analyze $\Gamma^*(f)$ to find a threat equilibrium. It can be shown that there is a unique equilibrium for $\Gamma^*(f)$, which is described in the following table.

S	b_S
{1}	(1)
{2}	(3)
{3}	(3)
{1,2}	(1,3)
{1,3}	(1,3)
{2,3}	(2,2)
{1,2,3}	(0,2,2)

<u>Table 1.</u> The threat equilibrium

Thus, in threat equilibrium, all coalitions are threatening to produce as much as they can, except for $\{2,3\}$ and $\{1,2,3\}$. The strategy $b_N = (0,2,2)$ is easy to understand, since it maximizes total payoff to the three players ($\sum_{i=1}^{3} U_1(0,2,2) = 16$), and is therefore the only Pareto-optimal joint strategy (when utility is transferable).

To understand the threat of coalition $\{2,3\}$, suppose that the coalition were considering changing to another threat, say $b'_{\{2,3\}} = (3,3)$. Such a change (increasing the threatened production level for $\{2,3\}$ to 6) would give

$$U_1(b_{\{1\}},b'_{\{2,3\}}) = 1 < 3 = U_1(b_{\{1\}},b_{\{2,3\}}),$$

and

$$U_2(b_{\{1\}},b'_{\{2,3\}}) + U_2(b_{\{1\}},b'_{\{2,3\}}) = 6$$

$$< U_2(b_{\{1\}},b_{\{2,3\}}) + U_3(b_{\{1\}},b_{\{2,3\}}) = 12.$$

Decreasing $U_1(b_{\{1\}},b_{\{2,3\}})$ tends to improve the bargaining position of 2 and 3, but decreasing $U_2(b_{\{1\}},b_{\{2,3\}}) + U_3(b_{\{1\}},b_{\{2,3\}})$ tends to hurt their bargaining position. The net effect of these two changes must be determined by the settlement function, and it turns out that

$$f(b) = (3\tfrac{1}{3},\ 6\tfrac{1}{3},\ 6\tfrac{1}{3}),$$

while

$$f(b_{-\{2,3\}},b'_{\{2,3\}}) = (4,6,6).$$

Thus, the net effect of changing from $b_{\{2,3\}}$ to $b'_{\{2,3\}}$ would be to hurt the bargaining position of players 2 and 3.

Similar arguments can be given to verify that no coalition can benefit by changing from its threat in Table 1. Thus this is indeed a threat equilibrium, and the cooperative solution is $f(b) = (3\frac{1}{3},\ 6\frac{1}{3},\ 6\frac{1}{3})$.

If we used the Harsanyi-Shapley settlement function $\hat{f}$ (recall (14)), we would find that b from Table 1 would still be an equilibrium for the threats-game $\Gamma^*(\hat{f})$. But $\hat{f}(b) = (2\frac{2}{3}, 6\frac{2}{3}, 6\frac{2}{3})$. Notice that

$$\hat{f}_1(b) = 2\frac{2}{3} < 3 = U_1(b_{\{1\}}, b_{\{2,3\}}).$$

Thus the cooperative solution generated by the Harsanyi-Shapley settlement function does not satisfy the individual rationality property of Theorem 2.

8. Proofs

PROOF OF THEOREM 1.

Throughout this proof, we shall assume that f satisfies the formula in the Theorem, which is equivalent to (12). We must first show that this f satisfies Axioms 1 and 2.

Axiom 1 follows from Corollary 1 in [5]. Suppose that b is T-decomposable. Let $T^* = \{T, N\backslash T\}$. Then the partition function game W(b) (recall (11)) satisfies:

$$\begin{aligned} W_{S,Q}(b) &= \sum_{i\in S} U_i(\langle b_R\rangle_{R\in Q}) \\ &= \sum_{i\in S} U_i(\langle b_R\rangle_{R\in Q\wedge T^*}) \\ &= \sum_{\substack{\hat{S}\in Q\wedge T^* \\ \hat{S}\subseteq S}} W_{\hat{S},Q\wedge T^*}(b), \quad \forall (S,Q)\in ECL . \end{aligned}$$

This equation implies that W(b) is T*-decomposable as defined in [5]. Then Corollary 1 of [5] implies

$$\sum_{i\in T} \Phi_i(W(b)) = W_{T,T^*}(b) .$$

Translating back, using (11) and (12), gives us

$$\sum_{i\in T} f_i(b) = \sum_{i\in T} U_i(\langle b_T, b_{N\backslash T}\rangle) = \sum_{i\in T} U_i(b_N)$$

(using T-decomposability of b for the second equality). This proves that f satisfies Axiom 1.

Let $a_{i,k,Q} = (-1)^{q-1}(q-1)!\left(\frac{1}{n} - \sum_{\substack{S'\in Q\\(i\notin S')\\(k\notin S')}} \frac{1}{(q-1)(n-s')}\right)$.

Now, suppose S is a given coalition and $\{i,j\} \subseteq S$. Observe that:

(16) $$f_i(b) - f_j(b) = \sum_{k\in N} \sum_{Q\in PT} (a_{i,k,Q} - a_{j,k,Q})\, U_k(\langle b_T\rangle_{T\in Q})\ ;$$

But $U_k(\langle b_T\rangle_{T\in Q})$ depends on b_S only if $S\in Q$, and $a_{i,k,Q} = a_{j,k,Q}$ if $S \in Q$. So the terms in (16) which depend on b_S all have zero coefficients, which proves Axiom 2.

Now suppose that f' satisfies Axioms 1 and 2. We must show that $f' = f$. To do so, we will need some graph-theoretical concepts developed in [4].

A <u>graph</u> is a set of links between pairs of players. We will denote the <u>link</u> between i and j by i:j. (Since these are <u>undirected</u> links, i:j=j:i.)

For any coalition S and graph g, let S/g be the partition of S into components which are connected by g within S. (That is, two members of S are grouped together in S/g iff they can be connected by a path in g which stays within S.) N/g is thus the partition of N into the connected components of g.

Given $b \in \times_{S\in CL} B_S$, we define $b/g \in \times_{S\in CL} B_S$ so that:

$$(b/g)_S = \langle b_T\rangle_{T\in S/g}, \quad \forall S \in CL.$$

We shall prove that

(17) $$f'(b/g) = f(b/g), \quad \text{for every graph } g.$$

This will imply that $f(b) = f(b')$, because $b = b/\bar{g}$, where $\bar{g}$ is the complete graph linking all pairs of players.

Suppose that (17) does not hold. Then we can find a graph g such that $f'(g) \neq f(g)$ but $f'(g') = f(g')$ for every graph g' with fewer links than g.

For any link i:j in g, consider the graph $g_{ij} = g\backslash\{i:j\}$ in which this one link is removed. Observe that $S/g = S/g_{ij}$ unless i and j both belong to S. Thus, the only components of b/g_{ij} which differ from b/g are the components for coalitions containing both i and j. So, using Axiom 1 for f and f′, and using the fact that g_{ij} has fewer links than g, we get:

$$f'_i(b/g) - f'_j(b/g) = f'_i(b/g_{ij}) - f'_j(b/g_{ij})$$

$$= f_i(b/g_{ij}) - f_j(b/g_{ij}) = f_i(b/g) - f_j(b/g).$$

So for any two players i and j connected in N by g, we have

$$f'_i(b/g) - f_i(b/g) = f'_j(b/g) - f_j(b/g).$$

But for any connected component $S \in N/g$, b/g is S-decomposable, so

$$\sum_{i\in S} f'_i(b/g) = \sum_{i\in S} U_i((b/g)_N) = \sum_{i\in S} f_i(b/g) .$$

These two results imply that $f'_i(b/g) = f_i(b/g)$ for all i, contrary to the way that g was constructed. This proves (17) and completes the proof of the theorem.

PROOF OF THEOREM 2.

Given the threat equilibrium b, suppose (contrary to the theorem) that $f_i(b) < U_i(b_{N\backslash\{i\}},d_i)$. Then let $b' \in \underset{S\in CL}{\times} B_S$ satisfy:

$$b'_S = \begin{cases} \langle b_{S\backslash\{i\}},d_i\rangle , & \text{if } i \in S, \\ b_S , & \text{if } i \notin S. \end{cases}$$

Since b′ is {i}-decomposable, Axiom 1 implies:

$$f_i(b') = U_i(b_{N\backslash\{i\}},d_i).$$

Observe that each term in the formula for f_i (as given in Theorem 1) depends on the threat of only one coalition containing i. Furthermore b and b' differ only in components corresponding to coalitions which contain i. Thus:

$$f_i(b') - f_i(b) = \sum_{\substack{S \in CL \\ (i \in S)}} (f_i(b_{-S}, b'_S) - f_i(b))$$

So for at least one $S \in CL$ containing i, we must have

$$f_i(b_{-S}, b'_S) > f_i(b).$$

Then, by Axiom 2,

$$\sum_{j \in S} f_j(b_{-S}, b'_S) > \sum_{j \in S} f_j(b) .$$

By (5), this contradicts the fact that b is a threat equilibrium, so the hypothetical d_i could not exist and the theorem is proved.

(Note: with more involved notation, this proof can be extended to prove Theorem 2 for mixed-strategy equilibria as well.)

PROOF OF THEOREM 3.

The first equation follows from Axiom 1. Then, since b is an equilibrium, for any other $b'_N \in B_N$:

$$\sum_{i \in N} f_i(b) \geq \sum_{i \in N} f_i(b_{-N}, b'_N) = \sum_{i \in N} U_i(b'_N).$$

References

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

[2] Kalai, E. and R.B. Myerson (1977). Values of Games without Sidepayments. Discussion Paper 267, Center for Mathematical Studies in Economics and Management Science, Northwestern University.

[3] Lucas, W.F. and R.M. Thrall (1963). n-Person Games in Partition Function Form. Naval Research Logistics Quarterly 10, 281-298.

[4] Myerson, R.B. (1977). Graphs and Cooperation in Games. Forthcoming in Mathematics of Operations Research.

[5] Myerson, R.B. (1977). Values of Games in Partition Function Form. Forthcoming in International Journal of Game Theory.

[6] Nash, J.F. (1951). Non-Cooperative Games. Annals of Mathematics 54, 286-295.

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

[8] Shapley, L.S. (1953). A Value for n-Person Games. Contributions to the Theory of Games, H.W. Kuhn and A.W. Tucker, eds., Princeton University Press, Princeton, N.J., 307-317.

we will define two axioms which a reasonable settlement function might satisfy, and we will show the unique function satisfying both axioms.

We shall henceforth assume that utility is transferable between the players in Γ. That is, the payoffs to the players are assumed to be measured in units of some freely exchangeable commodity, like money. See [2] for some theoretical arguments to justify the important role which this transferrable utility assumption has played in game theory.

One basic requirement of a settlement function is that its allocation should always be feasible. With transferable utility, we can guarantee feasibility if $f(\cdot)$ satisfies the following equation:

$$(7) \qquad \sum_{i \in N} f_i(b) = \sum_{i \in N} U_i(b_N) \quad \text{for any } b \in \underset{S \in CL}{\times} B_S .$$

(Here b_N is the N-component of the vector $b = (b_S)_{S \in CL}$. Notice that $b_N \in B_N = \underset{i \in N}{\times} D_i$, so $U_i(b_N)$ is well-defined.) The idea behind (7) is that the joint strategy chosen by the universal coalition N should be actually carried out in the settlement, but that it might be followed by some utility transfers or sidepayments between the players. We can think of these transfers as bribes which may be required to induce some smaller coalitions not to carry out their threats.

Axiom 1 expresses a stronger version of (7). The basic idea is that, if there are two groups of players who effectively ignore each other in all coalition plans, then the fair settlement should not require any coordination or transfers between these two groups. To develop this idea formally, we need some additional notation.

Suppose S and T are two disjoint sets of players, and suppose $b_S = (d_i)_{i \in S}$ and $b_T = (d_i)_{i \in T}$ are threats for S and T (in B_S and B_T respectively). Then we let $\langle b_S, b_T \rangle = (d_i)_{i \in S \cup T}$. That is, $\langle b_S, b_T \rangle$ is the threat for $S \cup T$ in which