GAMES AND ECONOMIC BEHAVIOR 4, 565–581 (1992)

Coalition-Proof Nash Equilibria and the Core in Three-Player Games

BENNY MOLDOVANU*

Department of Economics, University of Bonn, Adenauerallee 24-42, 5300 Bonn 1, Federal Republic of Germany

Received September 11, 1990

We study 3-person noncooperative games of coalition formation where the underlying situation is represented by a game in coalitional form without side payments. We look at coalition-proof Nash equilibria and we show that if the underlying game is balanced (in the sense of Scarf), then, except for indifferences, the grand coalition forms, and the payoff is in the core. If the underlying game has an empty core, then only a two-player coalition can form, and the payoff to its members is given by the respective coordinates of a unique "outside-options" vector. If the underlying game is not balanced but has a nonempty core, then either one of the two mentioned cases may hold. *Journal of Economic Literature* Classification Numbers: 022, 020.

1. INTRODUCTION

Explicit models of coalition formation cannot take into account, without becoming very intricate, all aspects of group communication and joint action. Instead of incorporating these aspects in the extensive form we use a solution concept that supplements the Nash equilibrium concept with some additional "cooperative" intuition.

We study a noncooperative game of coalition formation and payoff division, based on a nontransferable utility (NTU) three-person game in

* I owe special thanks to Reinhard Selten and Avner Shaked for their advice, and to Eyal Winter with whom I cooperated on several related topics. I also thank Ken Binmore, Jean-Francois Mertens, Georg Nöldeke, Bezalel Peleg, Motty Perry, Phil Reny, and Asher Wolinsky for helpful discussions. The comments of an associate editor and a referee greatly improved the quality of the exposition. Financial support from Deutsche Forschungsgemeinschaft, SFB 303, is gratefully acknowledged.

0899-8256/92 $5.00

coalitional form. To each coalition we associate a set of feasible utility allocations, and use an explicit set of rules for the interaction between the agents. The bargaining procedure is based on a method due to Selten (1981).

The solution concept is the coalition-proof Nash equilibrium (CPNE) due to Bernheim *et al.* (1987). Unlike the Nash equilibrium concept, this more refined concept takes into account joint deviations of coalitions. However, only self-enforcing deviations are considered to be credible threats. A deviation by a coalition is self-enforcing if no subcoalition has an incentive to initiate a new deviation. The CPNE is formally a refinement of the Nash equilibrium. However, its underlying intuition is close to "cooperative" reasoning since the detailed mechanism that describes group interaction is left unmodeled.

We look at subgame-perfect coalition-proof Nash equilibria in stationary strategies. Quite surprisingly, we find that the concept of balancedness for NTU games (see Scarf (1967)) plays the main role in the analysis. Recall that balancedness is sufficient but not necessary for the nonemptiness of the core of NTU games.

We prove the following for the noncooperative game: When the three-player game in coalitional form is balanced then, except for indifferences, the grand coalition forms, and the payoff is in the core. When the underlying game has an empty core (and hence is not balanced) then only a two-player coalition will form, and the payoff to its members is given by the respective coordinates of a unique "outside-options" vector. When the underlying game is not balanced but has a nonempty core, either the grand coalition forms with payoff in the core, or a two-player coalition forms with payoff according to the "outside-options" vector.

The intuition behind the "outside-options" vector goes back to Harsanyi who argued that a particular payoff vector "will represent the equilibrium outcome of a bargaining among the n-players only if no pair of players has any incentive to redistribute their payoffs between them, as long as the other players' payoffs are kept constant" (Harsanyi, 1977, p. 196).

The paper is organized as follows: In Section 2 we introduce games in coalitional form, the core, and the concept of balanced games. In Section 3 we describe an explicit bargaining procedure based on games in coalitional form and we introduce coalition-proof Nash equilibria. In Section 4 we characterize coalition formation and payoff division under CPNE in three-player games. In Section 5 we discuss the results.

2. Games in Coalitional Form

Let $N = \{1, 2, \ldots, n\}$ be a set of players. A coalition S is a nonempty subset of N. A payoff vector for N is a function x: $N \to \mathbb{R}$.

Notation. We denote by x^S the restriction of x to members of S. The restriction of $\mathbb{R}^S$ to vectors with nonnegative coordinates is denoted by $\mathbb{R}^S_+$. The zero vector in $\mathbb{R}^S$ is denoted by 0^S. For $x, y \in \mathbb{R}^S$ we write $x \geq y$ if $x^i \geq y^i$ for all $i \in S$. Let K be a subset of $\mathbb{R}^S_+$. Then int K denotes the interior of K relative to $\mathbb{R}^S_+$ and ∂K denotes the set $K \backslash \text{int } K$.

DEFINITION 1. A *nontransferable utility (NTU) game in coalitional form* is a pair (N, V), where V is a function that assigns to each coalition S in N a set $V(S) \subseteq \mathbb{R}^S$ such that:

$$V(S) \text{ is a non-empty, closed, and bounded subset of } \mathbb{R}^S_+; \quad (2.1)$$

$$\forall i \in N,\ V(i) = 0^i; \quad (2.2)$$

$$\text{If } y \in \mathbb{R}^S_+,\ x \in V(S) \text{ and } x \geq y \text{ then } y \in V(S); \quad (2.3)$$

$$\text{If } x, y \in \partial V(S) \text{ and } x \geq y \text{ then } x = y. \quad (2.4)$$

Condition 2.2 is a normalization. Condition 2.3 ensures that utility is freely disposable. A set $V(S)$ satisfying 2.3 is said to be *comprehensive*. Condition 2.4 requires that the Pareto-frontier of $V(S)$ coincide with the strong Pareto-frontier, and it implies that utility cannot be transferred at a rate of zero or infinity. A set $V(S)$ satisfying 2.4 is said to be *nonleveled*.

Denote by $A \times B$ the Cartesian product of the sets A and B. An NTU game (N, V) is *superadditive* if the following holds:

$$\forall S, T \subset N \text{ with } S \cap T = \emptyset, \qquad V(S) \times V(T) \subseteq V(S \cup T) \quad (2.5)$$

A game with transferable utility (TU) in coalitional form is a pair (N, v), where v is a function that assigns to each coalition S in N a real nonnegative number $v(S)$.

We next define two concepts that will play an important role in our analysis:

DEFINITION 2. Let (N, V) be an NTU game, and let $x \in V(N)$. x can be *improved upon* if there exists a coalition S and a vector $y^S \in V(S)$ with $y^i > x^i$ for all $i \in S$.

The *core* of (N, V), $C(N, V)$, is the set of all $x \in V(N)$ that cannot be improved upon.

DEFINITION 3. A collection B of coalitions of N is called *balanced* iff the system of equations

$$\sum_{S: j \in S} \lambda_S = 1 \qquad \text{for all } j \in N \quad (2.6)$$

has a nonnegative solution with $\lambda_S = 0$ for $S \notin B$.

An NTU game (N, V) is said to be *balanced* if and only if the following statement holds for any balanced collection B:

$$\text{If } x \in \mathbb{R}^N \text{ and } x^S \in V(S) \text{ for all } S \in B, \text{ then } x \in V(N). \qquad (2.7)$$

A fundamental theorem due to Scarf (1967) states that the core of a balanced NTU game is nonempty. The converse is not true; i.e., there are unbalanced NTU games with a nonempty core. For TU games balancedness is equivalent to the existence of a core. We note also that games arising from exchange markets are balanced.

3. The Bargaining Procedure and Coalition-Proof Nash Equilibria

We now describe a bargaining procedure based on a game (N, V): A player $i \in N$ has the first initiative. An initiator may shift the initiative to another player, or he may make a proposal. A proposal consists of a coalition S, a payoff vector $x^S \in V(S)$, and a responder who must be a player of S. The responder can reject or accept the proposal. If the responder rejects, then he becomes the new initiator. If the responder accepts there are two possibilities: 1. Coalition S forms and the game ends if the responder was the last player in S needed to accept the proposal. The members of S are paid according to x^S, the other players receive zero payoffs. 2. Otherwise the responder must select the next responder to the existing proposal.

An infinite play results in zero payoffs to all players.

Whenever our results hold independently of the identity of the first player with the initiative we omit the dependence of the game form on this parameter, and we denote the sequential bargaining game by $\Gamma(N, V)$. For a formal description of this game we refer the reader to Selten (1981), where a version is studied in which the underlying situation is described by a TU game. For a TU game (N, v) the bargaining procedure has the same structure as above. The only difference is that a proposed payoff vector x^S must satisfy $x(S) \leq v(S)$, where $x(S)$ denotes the sum $\sum_{i \in S} x^i$.

We note that a general version of this game would allow the consecutive formation of several coalitions. However, we concentrate on three-player games, and it is clear that in such games only one essential coalition can form—a two-player coalition or the grand coalition. Hence it makes sense to look at the simplified version described above.

We restrict our attention to subgame-perfect equilibria in stationary strategies. A stationary behavioral strategy for a player in $\Gamma(N, V)$ assigns to each decision node of that player a probability distribution over the set of actions available at that node. A player uses the same mixture of possible actions whenever he acts as proposer. The action of a responder depends only on the existing proposal and on the set of players that have proposed or accepted this proposal. Nash and subgame-perfect Nash equilibria are defined in the usual way. Note that Nash equilibria in stationary strategies are stable also against nonstationary deviations. Again, for a more formal treatment of these matters we refer the reader to Selten (1981). We note that without the stationarity assumption one obtains results of the Folk-Theorem type even if time discounting is introduced (see Chatterjee *et al.* (1990)).

Let (N, V) be an NTU game, and let $\Gamma(N, V)$ be the associated noncooperative game. Let $\sigma = (\sigma^1, \sigma^2, \ldots, \sigma^n)$ be a stationary strategy profile for $\Gamma(N, V)$. Denote by q^i the expected payoff of player i, given the strategy σ and that i has the initiative. This is well defined because of the stationarity assumption. Let $q = (q^1, q^2, \ldots, q^n)$ denote the vector of expectations for all players in N. Note that q *is not* necessarily feasible for the grand coalition.

These considerations apply unchanged for the TU case. Selten calls q the "*demand vector*" of σ, and proves the following useful theorem:

THEOREM (Selten, 1981). *Let (N, v) be an n-person TU game, and let $\Gamma(N, v)$ be the associated bargaining game. Assume that a stationary subgame-perfect Nash equilibrium σ is played, and let q be the demand vector of σ.*

If coalition S forms in a subgame starting with an initiator's decision node, then $v(S)$ is divided according to q^S. The demand vector q has the following properties:

$$(1)\ \forall S \subseteq N, q(S) \geq v(S) \tag{3.1}$$

$$(2)\ \forall i \in N, \exists S \subseteq N \text{ such that } i \in S \text{ and } q(S) = v(S) \tag{3.2}$$

Conversely, for any vector q satisfying conditions 3.1 and 3.2 there exists a pure stationary equilibrium strategy σ such that q is the demand vector of σ.

For the NTU case we prove the following lemma (the proof is analogous to Selten's proof of the previous Theorem, and it uses conditions 2.3 and 2.4):

LEMMA 1. *Let (N, V) be an n-person NTU game, and let $\Gamma(N, V)$ be the associated bargaining game. Assume that a stationary subgame-perfect Nash equilibrium σ is played, and let q be the demand vector of σ.*

If coalition S forms in a subgame starting with an initiator's decision *node, then $V(S)$ is divided according to q^S. The demand vector q has the following properties:*

$$(1)\ \forall S \subseteq N,\ q^S \notin int\ V(S) \tag{3.3}$$

$$(2)\ \forall i \in N,\ \exists S \subseteq N \text{ such that } i \in S \text{ and } q^S \in V(S) \tag{3.4}$$

Conversely, for any vector q satisfying conditions 3.3 and 3.4 there exists a pure stationary equilibrium strategy σ such that q is the demand vector of σ.

Note that the set of subgame-perfect equilibrium payoffs may be quite large even when attention is restricted to equilibria in stationary strategies. Example 2 in Section 5 displays a "counterintuitive" result related to this phenomenon.

Bernheim *et al.* (1987) proposed the concept of "coalition-proof Nash equilibrium" (CPNE) for situations where communication is possible but binding commitment is not. The concept requires stability against deviations of coalitions, but internal consistency requires that only self-enforcing deviations be regarded as credible threats. A deviation by a coalition is self-enforcing if no subset of this coalition has the incentive to deviate yet again, while taking as given the strategies of the nondeviating players. We recall that an earlier concept, the *strong Nash equilibrium* due to Aumann (1959), requires stability against deviations of all conceivable coalitions, but the deviations are not in any way restricted. The idea behind the concept of CPNE is closely related to that of "renegotiation-proofness". We now proceed to the formal definition of coalition-proof Nash equilibria. We first need some preparations:

DEFINITION 4. An n-person *game in normal, or strategic form* consists of a finite set of players, $N = \{1, 2, \ldots, n\}$; an n-tuple of nonempty strategy sets, $(\Delta^1, \Delta^2, \ldots, \Delta^n)$; and an n-tuple $(g^1, g^2, \ldots, g^n)$ of payoff functions $g^i\colon \Delta^N \to \mathbb{R}$, where Δ^S denotes the Cartesian product of Δ^i over $i \in S$.

Notation. Let G be an n-person game in normal form, and let S be a coalition in N. Let $-S$ denote the complement of S in N. For $\tau \in \Delta^N$ let G^S_τ denote the game induced on S by the actions τ^{-S} for coalition $-S$; i.e., $G^S_\tau = (S, \{h^i\}_{i\in S}, \{\Delta^i\}_{i\in S})$ where $h^i\colon \Delta^S \to \mathbb{R}$ is given by $h^i(\sigma^S) = g^i(\sigma^S, \tau^{-S})$ for all $i \in S$ and $\sigma^S \in \Delta^S$.

DEFINITION 5. In a single player game G, $\tau_* \in \Delta^N$ is a *coalition-proof Nash equilibrium* (CPNE) if and only if τ_* maximizes $g^1(\tau)$.

Let $n > 1$ and assume that CPNE have been defined for all games with fewer than n players. Then

(a) For any game G with n players $\tau_* \in \Delta^N$ is *self-enforcing* if, for all coalitions $S \subsetneqq N$, τ_*^S is a CPNE in the game G_τ^S.

(b) For any game G with n players, $\tau_* \in \Delta^N$ is a CPNE if it is self-enforcing and if there does not exist another self-enforcing strategy $\tau \in \Delta^N$ such that $g^i(\tau) > g^i(\tau_*)$ for all $i \in N$.

We note that the concept of CPNE is a less stringent refinement of Nash equilibrium than the strong Nash equilibrium (indeed all strong equilibrium are CPNE), and despite this CPNE may not exist even for simple games.

4. Three-Player Games

Three-player games offer the first interesting case for the study of coalition formation and payoff division.

In the sequel an equilibrium will mean a subgame-perfect Nash equilibrium. We study equilibria that are also coalition-proof. We are now ready to prove:

Proposition A. *Let (N, V) be a three-player balanced game and let $\Gamma(N, V)$ be the associated bargaining game. Then*

1. *Let σ be a CPNE in stationary strategies. Then q, the demand vector of σ, belongs to the core of (N, V).*
2. *For each vector x in the core of (N, V) there exists a CPNE in pure stationary strategies with payoff x.*

Proof. 1. By Lemma 1,

$$\forall S,\ S \subseteq N\ q^S \notin \text{int } V(S). \tag{4.1}$$

To prove that $q \in C(N, V)$ it suffices to show that $q \in V(N)$. The proof is by contradiction: if $q \notin V(N)$ then there are self-enforcing deviations, hence we obtain a contradiction to the assumption that σ is a CPNE.

Assume therefore that

$$q = (q^1, q^2, q^3) \notin V(N). \tag{4.2}$$

We first show that q must have some special properties—here we use the fact that the game is balanced. If σ is played then the grand coalition cannot form, since if it does it must divide the payoff according to q (by Lemma 1), and this is impossible by 4.2. Certainly $q \neq 0^N$, since the zero vector is in $V(N)$ by the definition of NTU games. Therefore a two-player coalition can form if σ is played. Assume without loss of generality that player 1 has the initiative, and that at a certain end node the coalition $\{1, 2\}$ forms. By Lemma 1 we know that the payoff is $(q^1, q^2, 0)$, and that

$$(q^1, q^2) \in \partial V(1, 2). \tag{4.3}$$

Again by Lemma 1, we must have a coalition S with $3 \in S$ and $q^S \in V(S)$. By assumption 4.2 $q \notin V(N)$, hence $S \neq N$. If $S = \{3\}$ then $q^3 = 0$. The game (N, V) is balanced and therefore also superadditive. With 4.3 this implies that $q \in V(N)$, which contradicts $q^3 = 0$. Thus

$$q^3 > 0 \qquad \text{and} \qquad q^3 \notin V(3). \tag{4.4}$$

Coalition S must be a two-person coalition, and we assume without loss of generality that $S = \{2, 3\}$. With 4.1 this yields

$$(q^2, q^3) \in \partial V(2, 3). \tag{4.5}$$

If it were true that $(q^1, q^3) \in V(1, 3)$ then 4.3, 4.5, and the balancedness of (N, V) imply that $q \in V(N)$, which contradicts $(q^1, q^3) \in V(1, 3)$. (Note that the collection $(\{1, 2\}, \{2, 3\}, \{1, 3\})$ is balanced.) Hence it must be that

$$(q^1, q^3) \notin V(1, 3). \tag{4.6}$$

By 4.2, 4.5, and the superadditivity of (N, V) it must be that $q^1 > 0$. By Lemma 1, and by the definition of the demand vector q, the only possible coalition at an end node of any subgame starting with player 1 as initiator is the coalition $\{1, 2\}$, and the payoff is $(q^1, q^2, 0)$.

We now describe a self-enforcing joint deviation of players 2 and 3. First we choose $\varepsilon > 0$ such that

$$q^3 - \varepsilon > 0 \qquad (\text{so that } (q^3 - \varepsilon) \notin V(3)) \tag{4.7}$$

$$(q^1, q^3 - \varepsilon) \notin V(1, 3) \tag{4.8}$$

$$(q^1, q^2, q^3 - \varepsilon) \notin V(N). \tag{4.9}$$

Such a choice is possible because $V(3)$, $V(2, 3)$, $V(1, 2, 3)$ are closed sets and the distances between q^3 and $V(3)$, between (q^1, q^3) and $V(1, 3)$, and between (q^1, q^2, q^3) and $V(N)$ are positive. Let y^2 be defined by

$$(y^2, q^3 - \varepsilon) \in \partial V(2, 3). \tag{4.10}$$

If ε is sufficiently small then y^2 will be close to q^2 (see 4.5), and hence it will also be true that

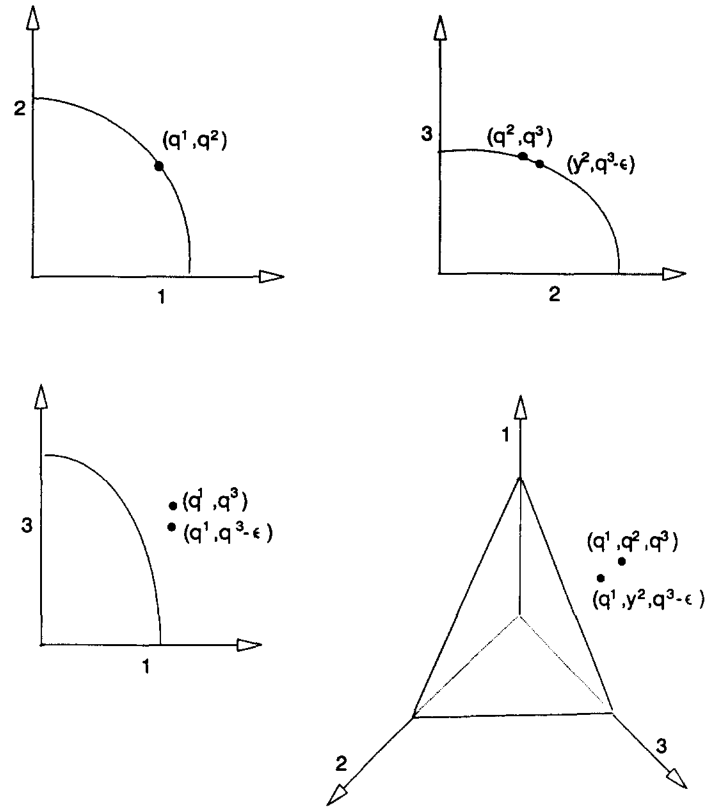

FIGURE 1

$$(q^1, y^2, q^3 - \varepsilon) \notin V(N). \tag{4.11}$$

Nonlevelness and comprehensiveness of $V(2, 3)$ (see 2.3 and 2.4) guarantee that

$$y^2 > q^2. \tag{4.12}$$

(See Fig. 1.)

The joint deviation of players *2* and *3* is as follows: Player *2* proposes the coalition $\{2, 3\}$ with the division $(y^2, q^3 - \varepsilon)$, and accepts only offers that yield him at least y^2. Player *3* proposes the coalition $\{2, 3\}$ with the division $(y^2, q^3 - \varepsilon)$ and accepts only offers that yield him at least $q^3 - \varepsilon$.

If player *1* adheres to his strategy, he never accepts proposals that yield him less than q^1, since by rejecting such a proposal he becomes initiator and is guaranteed q^1. Hence, if players *2* and *3* deviate as above, then this deviation constitutes a Pareto-optimal Nash equilibrium in their induced game and they can both benefit by it. The reader may confirm this with the help of 4.3, 4.7, 4.8, 4.10, 4.11, 4.12. By Definition 5, a deviation that constitutes a Pareto-optimal Nash equilibrium in an induced two-player game is self-enforcing.

The strategy profile σ is a CPNE; hence the assumption that q, the demand vector of σ, does not belong to $V(N)$ (4.2) has led to a contradiction. Hence $q \in V(N)$ and, by 4.1, q belongs to $C(N, V)$.

2. Let $x \in C(N, V)$. Consider the following strategy profile: Each player proposes the coalition N with division x (note that $x \in V(N)$), and a responder. A responder accepts a proposal if all players who have not yet accepted the standing proposal (including himself) get at least their respective payoff in x. If he is not the last responder to such a proposal he designates the next responder from the set of players who have not yet accepted. Other proposals are rejected.

Assume that an initiator i deviates and proposes coalition S with $y^S \in V(S)$, and $y^i > x^i$. With $x \in C(N, V)$ and nonlevelness this implies that $y^j < x^j$ for a player j in S. This player will reject the proposal, hence i cannot benefit from this deviation. By backward induction it is easy to check that the actions of a player are optimal whenever he has to respond, and therefore we have described a subgame-perfect Nash equilibrium in pure, stationary strategies. This is also a strong Nash equilibrium, hence it is also a CPNE. Q.E.D.

Remark. The fact that the grand coalition does not necessarily form at all possible endpoints is due to possible indifferences. A responder j may be indifferent between accepting and rejecting a proposal if in both cases he is guaranteed the same payoff. Assume that each responder faced with such an indifference accepts the proposal (this can be interpreted as a secondary preference for shorter plays). Then each initiator is indifferent between the proposal according to σ, and the proposal (N, q) that yields the formation of the grand coalition with a payoff in the core of the game. By backward induction the proposal (N, q) will be accepted by all responders. The tie-breaking rule in favor of the grand coalition can be interpreted as "nondiscrimination": if the inclusion of a player in the final coalition *does not* affect the payoff of other players in that coalition, then this player should not be excluded. The tie-breaking problem is a "boundary" phenomenon and it does not appear if q is in the relative interior of the core.

The next example demonstrates that balancedness is necessary for the first part of the previous proposition. We describe a nonbalanced, super-additive game with a nonempty core.

EXAMPLE 1. Let $N = \{1, 2, 3\}$, define a game (N, V) as follows:

$$V(i) = 0^i, \qquad \forall i \in N. \tag{4.14}$$

$$V(2,3) = \{(x^2, x^3) \in \mathbb{R}_+^{\{2,3\}} | x^2 + x^3 \leq 2\}. \tag{4.15}$$

$$V(1,2) = \{(x^1, x^2) | (x^1, x^2) \in \text{conv}[(0,0), (0,2), (1,1), (\tfrac{7}{6}, 0)]\}. \tag{4.16}$$

$$V(1,3) = \{(x^1, x^3) | (x^1, x^3) \in \text{conv}[(0,0), (0,2), (1,1), (\tfrac{7}{6}, 0)]\}. \tag{4.17}$$

$$V(N) = \left\{(x^1, x^2, x^3) \in \mathbb{R}_+^N | x^1 + \frac{x^2}{3} + \frac{x^3}{3} \leq 1.5\right\}. \tag{4.18}$$

Let σ^1 be the following strategy combination for the game $\Gamma(N, V)$: Player i, $1 \leq i \leq 3$, proposes coalition $\{i, j\}$ together with the division $(1, 1)$, where $j = i + 1 \pmod 3$. He accepts only those offers that yield him and all receivers that have not yet accepted at least 1.

Let σ^2 be the following strategy combination: Each player proposes the grand coalition with division $z = (z^1, z^2, z^3) = (0.5, 1.5, 1.5)$, and accepts only offers that yield him and those responders that have not yet accepted the respective payoff in z. The reader may verify that both σ^1 and σ^2 are CPNE. Under σ^2 the grand coalition forms immediately and the payoff z is in the core of the game. Under σ^1 only two-player coalitions can form with payoff $(1, 1)$ to each, and zero to the other player. Observe that the demand vector of σ^1, namely $y = (1, 1, 1)$, has the property that $(y^i, y^j) \in \partial V(i, j)$ for all $i, j \in N$, $i \neq j$, but $y \notin V(N)$. This is possible because the game (N, V) is not balanced.

We next consider games with an empty core. By Scarf's Theorem these games are not balanced. We first prove a lemma that has also some independent interest.

LEMMA 2. *Let (N, V) be a three player super-additive game with an empty core. Then there exists a unique vector $y = (y^1, y^2, y^3)$ such that:*

$$y \notin V(N), \tag{4.19}$$

$$(y^i, y^j) \in \partial V(i,j) \qquad \textit{for all } i,j \in N, i \neq j. \tag{4.20}$$

(See Fig. 2.)

Proof. Let $\partial V(i,j)$ be described by the graph of a function $u^j = f_{ij}(u^i)$, for $i, j \in N$, $i < j$. By the conditions in the definition of NTU games these functions are continuous and monotonically decreasing.

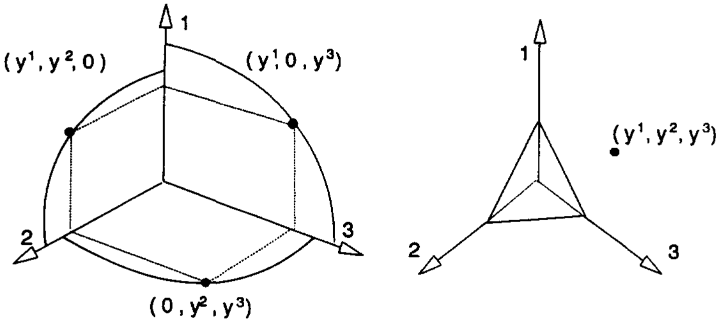

FIGURE 2

Uniqueness of y: Assume there exists a vector $z \neq y$ such that $z \notin V(N)$ and $(z^i, z^j) \in \partial V(i, j)$ for all $i, j \in N$, $i \neq j$. Without loss of generality assume that $z^1 > y^1$. The monotonicity of the functions f_{ij} and the properties of z imply that $z^2 < y^2$ and $z^3 < y^3$. This contradicts $(z^2, z^3) \in \partial V(2, 3)$.

Existence of y: It suffices to prove existence of a vector y with property 4.20. The emptiness of the core implies that such a vector also has property 4.19. The existence of y is proved using a constructive fixed point argument: y will be determined by the intersection of a specially constructed path with the boundary of the feasible set belonging to a two-player coalition.

We first construct the end point of the path. Consider the vectors $(0, f_{13}(0) \in \partial V(1, 3)$ and $(0, f_{23}(0) \in \partial V(2, 3)$. These are the best feasible outcomes for player 3 in coalitions $\{1, 3\}$ and $\{2, 3\}$, respectively. Assume without loss of generality that $f_{13}(0) \leq f_{23}(0)$. The set $\partial V(2, 3)$ is comprehensive, hence $(0, f_{13}(0) \in V(2, 3)$ and $f_{23}^{-1}(f_{13}(0))$ is well defined. Consider the vector (x^1, x^2, x^3), where

$$x^3 = f_{13}(0), \qquad x^2 = f_{23}^{-1}(f_{13}(0)), \qquad x^1 = f_{13}^{-1}(f_{13}(0)) = 0. \quad (4.21)$$

The definition of x implies that

$$(x^1, x^3) \in \partial V(1, 3); \quad (4.22)$$

$$(x^2, x^3) \in \partial V(2, 3). \quad (4.23)$$

By superadditivity, 4.23, and 4.21 we obtain

$$x \in V(N). \quad (4.24)$$

If it were true that $(x^1, x^2) \notin V(1, 2)$ then 4.24, 4.23, and 4.22 imply that x belongs to the core of (N, V), but this is impossible because the core is empty. Hence it must be that

$$(x^1, x^2) \in V(1, 2). \tag{4.25}$$

We now describe a path in $\mathbb{R}_+^{\{1,2\}}$ that ends at (x^1, x^2): For $0 \leq t \leq x^3$, let

$$\omega(t) = (\omega^1(t), \omega^2(t)) = (f_{13}^{-1}(t), f_{23}^{-1}(t)). \tag{4.26}$$

By 4.26 the following statement holds:

$$\forall t,\ 0 \leq t \leq x^3: \qquad (\omega^1(t), t) \in \partial V(1, 3) \text{ and } (\omega^2(t), t) \in \partial V(2, 3). \tag{4.27}$$

We now look at the starting point $(\omega^1(0), \omega^2(0))$. We claim that

$$(\omega^1(0), \omega^2(0)) \notin V(1, 2). \tag{4.28}$$

Assume for the moment that the statement in 4.28 is true. The path $\omega(t)$ is monotonically decreasing in both coordinates. Hence, by 4.25 and 4.28, this path must intersect the boundary $\partial V(1, 2)$ for $t = t^*$ with $0 \leq t^* \leq x^3$. It holds that

$$(\omega^1(t^*), \omega^2(t^*)) \in \partial V(1, 2). \tag{4.29}$$

By 4.27 and 4.29 we obtain that $y = (\omega^1(t^*), \omega^2(t^*), t^*)$ is the desired vector.

It remains to prove 4.28. The proof is by contradiction: if 4.28 does not hold then we find a vector in the core of (N, V), hence we obtain a contradiction to the assumption that the core is empty. Assume therefore that $(\omega^1(0), \omega^2(0)) \in V(1, 2)$. We may increase the coordinates of this vector until $(z^1, z^2) \in \partial V(1, 2)$ is reached. By superadditivity, $(z^1, z^2, 0) \in V(N)$. By 4.27 and the construction of (z^1, z^2) we obtain that $(z^1, z^2, 0) \in C(N, V)$. This is a contradiction, and 4.28 must be true. Q.E.D.

The vector described in Lemma 2 is called the *outside-options vector*.

If the core is empty, and if the grand coalition forms, there will always be objections to the proposed outcome. We observe that in this case the set of strong equilibria is empty, because there are always joint deviations that benefit some coalition. This is closely related to the emptiness of the core.

PROPOSITION B. *Let (N, V) be a three-player superadditive game with empty core, and let $\Gamma(N, V)$ be the associated bargaining game.*

Let $y = (y^1, y^2, y^3)$ be the unique vector such that $(y^i, y^j) \in \partial V(i,j)$ for all $i, j \in N$ with $i \neq j$. Then

(1) *The set of CPNE is not empty.*

(2) *Let σ be a CPNE in stationary strategies. In any subgame starting with an initiator's decision node only a two-player coalition can form. If $S = \{i, j\}$ forms then the payoffs are always (y^i, y^j) for the members of S, and zero for the other player.*

Proof. 1. Let σ^1 be the following strategy profile: Player i, $1 \leq i \leq 3$, proposes the coalition $S_i = \{i, j\}$ together with payoff y^{S_i}, where $j = i + 1 \pmod 3$. He accepts only those offers that yield him and all players that have not yet accepted at least the respective payoff in y. It is easy to check that this describes an equilibrium.

It is easy to verify that, given the strategy of any player i, the strategies of the other two players constitute a Pareto-undominated Nash equilibrium in the game induced on those players by the strategy of player i. Hence σ^1 is a CPNE in the game $\Gamma(N, V)$. Note that this equilibrium is very simple, using only pure stationary strategies.

2. Assume without loss of generality that the subgame starts with player 1 as the initiator. We first show that q, the demand vector of σ, must be equal to our outside-options vector y. In particular σ is an equilibrium, therefore q must satisfy conditions 3.3 and 3.4 in Lemma 1. Condition 3.3 and the emptiness of the core imply that $q \notin V(N)$. By condition 3.4, there exists a coalition $S \subseteq N$, with $1 \in S$ and $q^S \in V(S)$. It must be the case that $|S| \leq 2$, where $|S|$ denotes the cardinality of the set S. Assume first that $|S| = 2$, and that, without loss of generality, $S = \{1, 2\}$. Then, by condition 3.3, we obtain $(q^1, q^2) \in \partial V(1, 2)$. Similarly, for player 3 it must hold that, say, $(q^2, q^3) \in \partial V(2, 3)$. By condition 3.3 we know that $(q^1, q^3) \notin \operatorname{int} V(1, 3)$. If $(q^1, q^3) \in \partial V(1, 3)$, then q has the desired form. Otherwise one can find, using the construction of Proposition A, a self-enforcing deviation of players 2 and 3. This is a contradiction because σ is a CPNE. Similarly, one can check the result for the case $|S| = 1$ (then $q^1 = 0$).

We have proved that the demand vector of any CPNE must be y. Because $y \notin V(N)$ we know by the first part of Lemma 1 that the grand coalition cannot form in a CPNE. Only two-player coalitions can form, and, again by the first part of Lemma 1, the payoff division must be according to y. Q.E.D.

The next result concludes the characterization of CPNE in stationary strategies for the noncooperative bargaining based on three-player superadditive games.

PROPOSITION C. *Let (N, V) be a three-player, superadditive, and non-*

balanced game with nonempty core, let $\Gamma(N, V)$ be the associated bargaining game, and let σ be a CPNE in stationary strategies. Then either $q = (q^1, q^2, q^3)$, the demand vector of σ, is an element of $C(N, V)$, or it is the outside options vector.

Proof. If $q \in V(N)$, then we obtain $q \in C(N, V)$ by condition 3.3 in Lemma 1. Otherwise we find, using the arguments of Proposition A, that the respective projections of q must be on the boundary of at least two of the sets representing the feasible allocations for two-player coalitions (see 4.3, 4.5, and Fig. 1). If the respective projections of q are also on the boundary of the feasible allocations set for the remaining two-player coalition, then q is the outside options vector. Otherwise we find, using the construction of Proposition A, a self-enforcing deviation that benefits a pair of players. This is a contradiction to the assumption that σ is a CPNE. Q.E.D.

5. Concluding Remarks

We have illustrated strong relations between outcomes in coalition-proof Nash equilibria of a sequential bargaining game and the fine structure of the underlying cooperative game in coalitional form.

It may be argued that it is inappropriate to study coalition formation using a concept that already embodies the fact that coalitions may form. However, the kind of study attempted here—a combination of cooperative and noncooperative analysis—is in its infancy, and we believe that using equilibrium concepts that "go already some of the way" may simplify the models while still focusing on the main issues.

The next example shows that subgame-perfect Nash equilibria (that are not CPNE) may lead, in our model, to results that contradict our intuition:

Example 2. Consider a market with two buyers, B_1 and B_2, and a seller A. The seller owns an indivisible object and his reservation price for the object is normalized to zero. The buyers have reservation prices of 1 and 2, respectively. This defines the following TU game:

$$v(B_1) = v(B_2) = v(A) = 0 \tag{5.1}$$

$$v(B_1, A) = 1; v(B_2, A) = v(B_1, B_2, A) = 2. \tag{5.2}$$

Consider the following strategies for the bargaining game:

B_1 proposes the coalition $\{B_1, A\}$ with division (0.5, 0.5) and accepts only proposals where he is offered at least 0.5.

B_2 proposes the coalition $\{B_2, A\}$ with division (1.5, 0.5) and accepts only proposals where he is offered at least 1.5.

A proposes the coalition $\{B_2, A\}$ with division (1.5, 0.5) and accepts only proposals where he is offered at least 0.5.

This profile is an equilibrium (irrespective of who has the first initiative), and trade always takes place at price of 0.5. The associated demand vector is $(q^{B_1}, q^{B_2}, q^{A}) = (0.5, 1.5, 0.5)$, and this is not feasible. By proposition A, in coalition-proof Nash equilibria this "paradox" disappears, and trade can take place only at prices of at least 1.

The outside-options vector for three-player games reminds one of *bargaining aspirations* for NTU games (Bennett and Zame, 1988). In general however, the philosophy of the CPNE concept is very different from that of aspirations. Harsanyi's intuition about pairwise renegotiation-proofness (see our Introduction) is captured by the *stable bargaining equilibria* due to Moldovanu (1990). However, these equilibria always exist only in NTU games that generalize two-sided markets.

Note that the outside-options vector is, in a sense, a limiting case of the core. Consider for example a TU, superadditive, three-player game with an empty core. Let the value of $v(N)$ continuously increase. At a certain point the core appears and it contains a unique vector—the outside-options vector. After that point the core grows larger and larger.

Binmore (1985) studies three-player games where each pair of players controls the division of a different "cake," but only one of the cakes can be divided. His "telephone bargaining model" is similar to the procedure we used here, and its subgame-perfect equilibria may have the paradoxical nature exhibited in Example 2. In Binmore's "market bargaining model" players announce sequentially real numbers that represent the utility they require if agreement is to be reached. The results are remarkably similar to ours: either a cake is divided according to the outside-options vector, or one cake is relatively much bigger than the remaining two—this yields a kind of "core"—and this cake is divided according to a vector in the core. A full comparison with our results is not possible because of the absence of a cake for the grand coalition.

Let (N, V) be an n-player NTU game with a nonempty core. The game $\Gamma(N, V)$ has subgame-perfect equilibria that are also strong, and hence coalition-proof. It is more difficult to characterize the payoffs in CPNE. Even the existence of CPNE is not clear if the core of (N, V) is empty. As in our proofs, it appears that one must actually go over the balanced collections for n-person games.

Bernheim *et al.* (1987) have also defined the notion of *perfect coalition-proof Nash equilibrium* for extensive-form games with a *finite* number of nested proper subgames. Because of its inductive nature, it is not at all clear how to generalize this concept for games with an infinite number of nested subgames. For some possible approaches see Asheim (1990) and Bernheim and Ray (1989).

References

ASHEIM, G. B. (1990). *Extending Renegotiation-Proofness to Infinite Horizon Games*. Discussion paper, The Norwegian School of Economics and Business Administration.

AUMANN, R. (1959). "Acceptable Points in General Cooperative *n*-Person Games," in *Contributions to the Theory of Games IV*. Princeton, NJ: Princeton Univ. Press.

BENNETT, E., AND ZAME, W. R. (1988). Bargaining in Cooperative Games, *Int. J. Game Theory* **17**, 279–301.

BERNHEIM, B. D., PELEG, B., AND WHINSTON, M. D. (1987). "Coalition-Proof Nash Equilibria. I. Concepts," *J. Econ. Theory* **42**, 1–12.

BERNHEIM, B. D., AND RAY, D. (1989). "Collective Dynamic Consistency in Repeated Games, *Games Econ. Behav.* **1**, 295–327.

BINMORE, K. G. (1985). "Bargaining and Coalitions," in *Game-Theoretic Models of Bargaining* (Alvin Roth, Ed.). Cambridge: Cambridge Univ. Press.

CHATTERJEE, K., DUTTA, B., RAY, D., AND SENGUPTA, K. (1990). "A Noncooperative Theory of Coalitional Bargaining," unpublished paper.

HARSANYI, J. C. (1977). *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge: Cambridge Univ. Press.

MOLDOVANU, B. (1990). Stable Bargained Equilibria for Assignment Games without Side Payments," *Int. J. Game Theory* **19**, 171–190.

SCARF, H. (1967). "The Core of an *N*-Person Game," *Econometrica* **37**, 50–69.

SELTEN, R. (1981). "A Non-cooperative Model of Characteristic Function Bargaining," in *Essays in Game Theory and Mathematical Economics in honor of O. Morgenstern"* (V. Böhm and H. Nachtkamp, Eds.). Mannheim: Bibliographisches Institut.