Available online at www.sciencedirect.com
ScienceDirect
Journal of Mathematical Economics 42 (2006) 771–793
JOURNAL OF Mathematical ECONOMICS
www.elsevier.com/locate/jmateco

Implementation of the recursive core for partition function form games

Chen-Ying Huang[a], Tomas Sjöström[b,*]

[a] *Department of Economics, National Taiwan University, Taipei, Taiwan*
[b] *Department of Economics, Rutgers University, New Brunswick, NJ 08901, USA*

Received 6 January 2004; received in revised form 7 April 2005; accepted 30 August 2005
Available online 24 July 2006

Abstract

In partition function form games, the *recursive core* (r-core) is implemented by a modified version of Perry and Reny's [Perry, M., Reny, P., 1994. A non-cooperative view of coalition formation and the core. Econometrica 62, 795-817] non-cooperative game. Specifically, every stationary subgame perfect Nash equilibrium (SSPNE) outcome is an r-core outcome. With the additional assumption of total r-balancedness, every r-core outcome is an SSPNE outcome.

JEL classification: C71; C72

Keywords: Cooperative game theory; Partition function form games; Consistency; Non-cooperative implementation; Recursive core

1. Introduction

The core is an important solution concept with intuitive appeal. It is an appropriate solution for situations where players have unhampered ability to sign binding agreements. Two lines of research have recently been prominent. The first line of research extends the core concept to situations with externalities across coalitions. The α- and r-cores are based on a *partition function* instead of a characteristic function. Important new solution concepts have been proposed by Ray and Vohra (1997), Koczy (2003). The second line of research concerns non-cooperative

* Corresponding author. Fax: +1 732 9327416.
E-mail address: tsjostrom@economics.rutgers.edu (T. Sjöström).

0304-4068/$ – see front matter
doi:10.1016/j.jmateco.2005.08.009

implementation of the core for characteristic function form games. Important contributions to this literature include Kalai et al. (1979), Chatterjee et al. (1993), Moldovanu and Winter (1994, 1995), Serrano and Vohra (1997). Non-cooperative games of coalition formation in the presence of externalities have been studied by Bloch (1995), Yi (1997), Ray and Vohra (1997).

In non-cooperative games of coalition formation, the set of equilibrium payoffs often depends on the fine details of the bargaining protocol. Changing the order in which players make and accept proposals may change the distribution of bargaining power. This happens, for example, in the very natural game of Chatterjee et al. (1993). To address this problem, Moldovanu and Winter (1995) look at outcomes that are equilibrium outcomes for *any* order of moves. This approach is interesting, but differs from the usual concept of implementation. Serrano and Vohra (1997) implement the core correspondence in economic environments using a two-stage game. In the first stage, each player proposes an outcome. In the second stage, a player may propose a coalition, which forms if all members accept. It is not clear how to generalize this game to the general partition function form games studied in our paper. At the very least, more stages would have to be added in order to address the key issue of how outsiders react when a coalition is formed.

Perry and Reny (1994) introduce a continuous-time game with no fixed protocol for making offers. Their game seems to capture the spirit of free competition underlying the core. The outcome does not depend on some arbitrarily specified order of moves. In their original paper, Perry and Reny did not allow externalities across coalitions. However, in their game outsiders can react to the formation of a coalition, so it seems well suited to handle externalities. After a coalition has "left" the Perry and Reny game, the remaining players enter a subgame which has the same structure as the original game. Since the original game implements the core, the subgame naturally implements the core of the "reduced game" consisting of the players who did not leave. Accordingly, when the original players leave, they will expect that the remaining players behave in a way which is consistent with the core of the reduced game. The self-similar structure of the Perry and Reny game suggests a connection with the *recursive core* (briefly, *r-core*). The recursive core is a solution concept for partition function form games which is *consistent* in the sense that the worth of a coalition is calculated by recursively computing the cores of the "left reduced societies". The connection between the recursive core and Perry and Reny's game when there are only three players was pointed out by Huang and Sjöström (2003). The current paper shows that Perry and Reny's game, suitably modified to allow for externalities across coalitions, provides a non-cooperative implementation of the r-core, with any number of players. Perry and Reny's game needs to be modified mainly because, in the presence of externalities, bargaining cannot be directly over payoffs. In our game, players instead bargain over *sharing rules*. We believe these results shed some light on how to define the core based on partition functions. Specifically, it suggests that the recursive core is a natural generalization of the core to games with externalities.

The rest of the paper is organized as follows. Section 2 contains the definition of the r-core. Section 3 provides an example which illuminates the issues that arise when the core is extended to allow for externalities. Section 4 describes the modified Perry and Reny (1994) game. We prove our two main theorems in Section 5. Section 6 discusses the assumption of total r-balancedness.

2. The recursive core

Let $N = \{1, 2, \ldots, n\}$ be the set of players. A coalition is a non-empty subset of N. A coalition structure is a partition of N. A transferable utility game in partition function form is denoted $\langle N, P\rangle$, where P is the partition function. The partition function form is the natural way to model externalities across coalitions. For any coalition structure $\mathcal{P}_N$ and for any coalition S in $\mathcal{P}_N$,

$P(S|\mathcal{P}_N) > 0$ is the value (or worth) of S when players partition themselves according to $\mathcal{P}_N$.[1] Thus, the worth of S can depend on the coalitional structure formed by the players in $N\backslash S$. For any payoff vector $x \equiv (x_i)_{i \in N}$ and any coalition $S \subseteq N$, let $x(S) \equiv \sum_{i \in S} x_i$. A payoff vector $x \in \mathbb{R}^n$ is *feasible under the partition* $\mathcal{P}_N$ if for every coalition S in $\mathcal{P}_N$, $x(S) = P(S|\mathcal{P}_N)$. If $S \in \mathcal{P}_N$, then we have $\mathcal{P}_N = (S, A_1, A_2, \ldots, A_k)$ for some coalitions $A_1, A_2, \ldots, A_k$. Notice that $\mathcal{P}_{N\backslash S} \equiv (A_1, A_2, \ldots, A_k)$ is a partition of $N\backslash S$. With a slight abuse of notation, we write $\mathcal{P}_N = (S, \mathcal{P}_{N\backslash S})$ and $P(S|S, \mathcal{P}_{N\backslash S}) \equiv P(S|\mathcal{P}_N)$.

Now we can define the recursive core (r-core). Consider any coalition $S \subseteq N$, and suppose the players in $N\backslash S$ have partitioned themselves into $\mathcal{P}_{N\backslash S}$. The r-core for coalition S given $\mathcal{P}_{N\backslash S}$, denoted $C(S|S, \mathcal{P}_{N\backslash S})$, is defined as follows. For a single-player society $S = \{i\}$, we define $C(\{i\}|\{i\}, \mathcal{P}_{N\backslash\{i\}})$ to be the set of payoff vectors feasible under $(\{i\}, \mathcal{P}_{N\backslash\{i\}})$. Proceeding recursively, suppose the r-core has been defined for all coalitions with at least $s-1$ members, and all partitions on players other than these $s-1$ members. Now suppose coalition S has s members, and other players partition themselves according to $\mathcal{P}_{N\backslash S}$. For any coalition $T \subseteq S$, define

$$V(T|S, \mathcal{P}_{N\backslash S}) \equiv \begin{cases} P(S|S, \mathcal{P}_{N\backslash S}) & \text{if } T = S \\ \min\{x(T) : x \in C(S\backslash T|S\backslash T, T, \mathcal{P}_{N\backslash S})\} & \text{if } T \neq S \end{cases}$$

Now, $x \in C(S|S, \mathcal{P}_{N\backslash S})$ if and only if there exists some partition $\mathcal{P}_S$ of S such that x is feasible under the partition $(\mathcal{P}_S, \mathcal{P}_{N\backslash S})$, and $x(T) \geq V(T|S, \mathcal{P}_{N\backslash S})$ for each coalition $T \subseteq S$. This completes the definition.

The r-core predicts how S will partition itself given that $N\backslash S$ is partitioned according to $\mathcal{P}_{N\backslash S}$, but the prediction may not be unique. Let $\mathcal{P}(S|S, \mathcal{P}_{N\backslash S})$ denote the set of all r-core partitions of S given $\mathcal{P}_{N\backslash S}$. That is, $\mathcal{P}_S \in \mathcal{P}(S|S, \mathcal{P}_{N\backslash S})$ if and only if there is $x \in C(S|S, \mathcal{P}_{N\backslash S})$ which is feasible under the partition $(\mathcal{P}_S, \mathcal{P}_{N\backslash S})$. We will show below that if $\langle N, P\rangle$ is strictly superadditive, then the r-core makes the unique prediction that S stays together, i.e., $\mathcal{P}(S|S, \mathcal{P}_{N\backslash S})$ contains only (S).

The final step of the recursive definition occurs when $S = N$. Although the method is the same in each step, it is useful to simplify the notation at the last step. At the last step the worth of $T \subseteq N$ will be denoted $V(T)$ instead of $V(T|N)$. Similarly, we write $C(N)$ instead of $C(N|N)$, and $\mathcal{P}(N)$ instead of $\mathcal{P}(N|N)$. By a slight abuse of terminology, we refer to $C(N)$ as the r-core of $\langle N, P\rangle$. Notice that $x \in C(N)$ if and only if there is some partition $\mathcal{P}_N$ of N such that x is feasible under the partition $\mathcal{P}_N$, and $x(T) \geq V(T)$ for each coalition $T \subseteq N$. This simplified notation should not cause any confusion.

Intuitively, the recursive core may be justified as follows. In a stable outcome, each coalition must get at least " what it is worth." However, in partition function form games the payoff to coalition $T \subseteq N$ depends on the behavior of the players in $N\backslash T$. Therefore, in order to give coalition T "what it is worth," we must predict what the players in $N\backslash T$ will do if the coalition T forms. The α-core applies *maximal pessimism*: the players in T think that what is worst for them will happen, regardless of the *incentives* of the members of $N\backslash T$ to hurt the players in T. The recursive core instead relies on a notion of *consistency*: the players in T believe that every subcoalition of $N\backslash T$ will insist on getting " what it is worth" (rather than trying to punish the members of T). Therefore, the players in T use the recursive core as a solution concept to make a prediction for the "reduced society" $N\backslash T$. The set of payoff vectors that can occur when the

[1] Allowing $P(S|\mathcal{P}_N) \leq 0$ would not change the results, but it would make the exposition slightly more awkward.

players in $N\backslash T$ behave according to the solution concept is $C(N\backslash T|N\backslash T, T)$. Accordingly, we define the worth of coalition T to be

$$V(T) \equiv \min\{x(T) : x \in C(N\backslash T|N\backslash T, T)\}$$

and x is in the r-core of $\langle N, P\rangle$ if and only if $x(T) \geq V(T)$ for all $T \subseteq N$. This recursive logic is less pessimistic than the α-core logic and accordingly computes a higher worth for any coalition. Thus, the r-core is always a subset of the α-core.

Each characteristic function form corresponds to a "trivial" partition function form, where the payoff to any coalition S is independent of the behavior of the players in $N\backslash S$. If the characteristic function form game is totally balanced, then the r-core for the corresponding partition function form game is non-empty and equals the α-core, and also the core of the original characteristic function form game (because there is only one possible way to define the worth of a coalition). If the characteristic function form game is not totally balanced, then the r-core may be empty, because the recursive construction requires all subgames to have non-empty cores. In the Perry and Reny game, this corresponds to the requirement of subgame perfection (i.e., an equilibrium must exist in every subgame).

Peleg (1986), Tadenuma (1992) characterize the core for characteristic function form games using a consistency axiom. In their definition of "reduced game", a set of "remaining players" may cooperate with a set of "leaving players". With externalities, we need to specify the behavior of those "remaining players" who are *not* cooperating. It seems natural to assume that they partition themselves according to the solution concept. We leave for future work a characterization of the r-core along these lines.

Just like classical cooperative game theory, we assume agreements are fully binding. To avoid any misunderstanding, suppose contracts are *legally* binding and will be enforced by a court of law. Since a coalition can form under a legally binding contract, they need not worry about destabilizing deviations *within* the coalition. Therefore, a coalition can always insist on getting at least "what it is worth".

The r-core is non-empty if and only if $C(N)$ is a non-empty set. Of course, it is easy to construct examples where the r-core is not non-empty. Even if the r-core is non-empty, the grand coalition may not form. Consider the following example. Let $N = \{1, 2\}$. The grand coalition is worth $V(N) \equiv P(N|N) = 1$. Each player $i \in \{1, 2\}$ on his own is worth 2, i.e., $V(\{i\}) = P(\{i\}|\{1\}, \{2\}) = 2$. The r-core is a singleton, $C(N) = (2, 2)$, and the r-core partition structure for N is $\mathcal{P}(N) = (\{1\}, \{2\})$. Here the grand coalition breaks apart by mutual agreement. This type of situation will complicate the non-cooperative implementation of the r-core. To simplify the analysis, we would like to be assured that each coalition S prefers to stick together rather than break apart, i.e., $\mathcal{P}(S|S, \mathcal{P}_{N\backslash S}) = (S)$. Thus, we introduce the following definition.

Definition 1. The game $\langle N, P\rangle$ is totally r-balanced if and only if for any S and $\mathcal{P}_{N\backslash S}$, $P(S|S, \mathcal{P}_{N\backslash S}) = (S)$.

We will discuss this assumption in Section 6. This property is not easy to check as it stands. However, it turns out that strictly superadditive games are totally r-balanced, provided the r-core is non-empty. Formally, $\langle N, P\rangle$ is *strictly superadditive* if for any two disjoint coalitions S and T and any coalitional structure $\mathcal{P}_{N\backslash(S\cup T)}$ on the remaining players

$$P(S|S, T, \mathcal{P}_{N\backslash(S\cup T)}) + P(T|S, T, \mathcal{P}_{N\backslash(S\cup T)}) < P(S \cup T|S \cup T, \mathcal{P}_{N\backslash(S\cup T)}) \tag{1}$$

For example, symmetric Bertrand competition with differentiated commodities (Deneckere and Davidson, 1985) satisfies (1). In strictly superadditive games, a coalition will maximize its joint payoff by staying together.

Proposition 1. *If the game $\langle N, P\rangle$ is strictly superadditive and the r-core is non-empty, then $\langle N, P\rangle$ is totally r-balanced.*

Proof. By definition, $V(S|S, \mathcal{P}_{N\setminus S}) = P(S|S, \mathcal{P}_{N\setminus S})$. Strict superadditivity implies that if S breaks up into several coalitions, the sum of the payoffs of the members of S will be strictly lower than $P(S|S, \mathcal{P}_{N\setminus S})$. But if $x \in C(S|S, \mathcal{P}_{N\setminus S})$ then $x(S) \geq V(S|S, \mathcal{P}_{N\setminus S})$, which implies that S must stay together. □

3. A three-player example

We illustrate some aspects of the recursive core and its implementation in a symmetric three-player example. Suppose $N = \{1, 2, 3\}$, and for $i, j, k \in N$ distinct

$$P(\{i\}|(\{i\},\{j\},\{k\})) = 0, \qquad P(\{i\}|(\{i\},\{j,k\})) = a, \qquad P(\{j,k\}|(\{i\},\{j,k\})) = b,$$
$$P(N|N) = 1$$

To calculate his own worth, player i must predict what players j and k would do if player i "leaves the game", i.e., refuses to cooperate with j and k. If players j and k form a coalition, the resulting structure is $(\{i\}, \{j, k\})$, in which case players j and k share b while player i gets a. If players j and k break apart and induce $(\{i\}, \{j\}, \{k\})$, all three players receive zero. A reasonable prediction is that players j and k will stick together if $b > 0$, but break apart if $b < 0$. This is indeed what will happen in subgame perfect equilibrium of the Perry and Reny game. Accordingly, if player i "leaves the game" as a singleton coalition, he expects to earn $V(\{i\}) = 0$ if $b < 0$ and $V(\{i\}) = a$ if $b > 0$.[2] If a two-player coalition $\{j, k\}$ "leaves the game," the resulting coalition structure is $(\{i\}, \{j, k\})$, so $V(\{j, k\}) = b$. Perry and Reny's game does not allow any more moves by players j and k after they have "left," so coalition $\{j, k\}$ cannot break apart and it must get b. (The intuitive justification is that agreements to form a coalition are legally binding.) Finally, $V(N) = 1$.

There are three cases:

- *Case 1*. If $b < 0$ then the recursive core $C(N)$ consists of all payoff vectors such that $x_1 + x_2 + x_3 = 1$ and $x_i \geq 0$ for all i. Notice that $C(N) \neq \emptyset$. We predict that the grand coalition forms even if $a + b > 1$. The structure $(\{i\}, \{j, k\})$ is not a possible r-core structure when $b < 0$ because under this structure, feasibility implies that players j and k get b in sum. At least one of them would get strictly less than the worth of a singleton coalition (zero), which is not possible. In the Perry and Reny game, a player would rather leave on his own than take a negative payoff.
- *Case 2*. If $b > 0$ and $a + b < 1$, then $C(N)$ consists of all payoff vectors such that $x_1 + x_2 + x_3 = 1$, $x_i \geq a$ for all i, and $x_i + x_j \geq b$ for all i, j distinct. This implies that $C(N) \neq \emptyset$ as long as $a \leq 1/3$ and $b \leq 2/3$. If $C(N) \neq \emptyset$ then the prediction is that the grand coalition forms.
- *Case 3*. If $b > 0$ and $a + b > 1$, then $C(N)$ consists of all payoff vectors such that $x_i \geq a$ for all i, $x_i + x_j \geq b$ for all i, j distinct, and there is i such that $x_i = a$ and $x_j + x_k = b$ for j ,k

[2] If $b = 0$, then player i is assumed to have pessimistic expectations, $V(\{i\}) = \min\{0, a\}$.

distinct and different from i. This implies that $C(N) \neq \emptyset$ only if $a = b/2$, in which case two players will form a coalition and share b, leaving the third player to stand alone and get a.

These r-core predictions seem intuitively plausible. In contrast, according to the α-core logic, player i must fear that if he refuses to cooperate, the other two players will induce whatever coalition structure is the worst for player i. Therefore, player i is worth only $\min\{0, a\}$, regardless of b. However, these fears may be unfounded. If $a < 0$ and $b < 0$, why would players j and k form a coalition just to hurt player i? If $a > 0$ and $b > 0$, why would players j and k break up just to hurt player i? The recursive core rules out such incredible threats.

An important aspect of Perry and Reny's game is that it does not specify a fixed order of moves. If $a > 0$ and $b > 0$, then each player will insist on at least a, because he thinks that he can leave, and then the other two will react by forming a coalition. On the other hand, suppose there is an exogenously given order of moves (say 1, 2, 3), as assumed for example by Maskin (2003). Then the following might be an equilibrium if a is large: player 1 starts by leaving, then players 2 and 3 merge. Even if player $j \in \{2, 3\}$ gets less than a, when he moves player 1 has already left, and so player j may be unable to get a. In contrast, in the Perry and Reny game, player $j \in \{2, 3\}$ can always preempt player 1 by leaving first. A fixed order of moves would yield predictions even in cases when the r-core is empty; however, it would not capture the spirit of free competition underlying the core.

A refinement of the α-core could be obtained by recursively defining stable coalition structures, starting at the finest and building toward coarser structures.[3] Suppose we postulate that $(\{i\}, \{j\}, \{k\})$ is always a stable partition, and then define $(\{i\}, \{j, k\})$ to be stable if and only if players j, k prefer to stay together rather than induce $(\{i\}, \{j\}, \{k\})$. Suppose, finally, we postulate that if player i defects from the grand coalition, he fears the worst of all *stable* partitions. This refined α-core seems more compelling than the original α-core. If $a < 0$ and $b < 0$, it would yield the " correct" prediction that player i on his own is worth 0. However, suppose $a > 0$ and $b > 0$. According to the refined α-core logic, player i fears that by defecting from the grand coalition, he may trigger the stable structure $(\{i\}, \{j\}, \{k\})$ and get zero. Here he is too pessimistic, because if player i leaves then players j and k actually prefer to form a coalition (and this is what they will do in any equilibrium of the Perry and Reny game). In general, if we start by postulating that the finest coalition structure is always stable, and then recursively define stability for coarser structures, and finally postulate that a deviator fears the worst of all stable coalition structures, then a deviator may fear structures that are " too fine" to be truly credible. The refined α-core would be biased in favor of "splitting up" with not enough "re-merging." It would not correspond to equilibria of the Perry and Reny game, because that game has no such bias: players j and k will not fall apart if they prefer to merge.

We end this section by considering the question of internal instability of coalitions. Although the Perry and Reny game rules it out by assumption, one can imagine scenarios where a coalition falls apart after it has "left." There are in fact two separate issues. First, if (contrary to our assumption) legally binding contracts *cannot* be signed, then a coalition member may worry about other members leaving the coalition, thereby hurting him. The fear of such defections may prevent the coalition from forming in the first place. Second, even if legally binding contracts *can* be signed, one can imagine coalitions breaking apart by *mutual agreement* if this is beneficial to *all* its members. A judge may be unwilling to stop the break-up of a coali-

3 This discussion was prompted by the comments of a referee.

tion if its members unanimously agree to "tear up the contract".[4] We address these issues in turn.

To be specific, suppose $0 < b \leq (1 - a)/2$ and $a < 0$ (which is case 2 above). Consider the payoff vector x where $a \leq x_1 < 0$, $b \leq x_2 < b - x_1$, and $x_3 = 1 - x_1 - x_2 \geq b$. Such x exists (for instance, take $x = (a, b, 1 - a - b)$). Since $x_1 + x_2 < b$, we have $x \notin C(N)$. If legally binding contracts are available, then x will be rejected by players 1 and 2, because they can sign a binding agreement which guarantees them b. However, suppose the players *cannot* sign binding agreements. Player 1 certainly cannot improve on x on his own, because if he refuses to cooperate, then players 2 and 3 will form a coalition and share $b > 0$, and player 1 would only get $a \leq x_1$. Suppose instead player 1 proposes coalition $\{1, 2\}$, and he offers $x_2 + \varepsilon$ to player 2, where $\varepsilon > 0$. Player 1 would get $b - (x_2 + \varepsilon) < 0$. For ε small enough, $b - (x_2 + \varepsilon) > x_1$ so both players 1 and 2 would be better off forming coalition $\{1, 2\}$ than accepting x. However, without a binding contract, player 2 may suspect that player 1 plans to break up $\{1, 2\}$ and induce the finest coalition structure $(\{1\}, \{2\}, \{3\})$. This would give each player zero, which is better for player 1 than staying in coalition $\{1, 2\}$ (since $b - (x_2 + \varepsilon) < 0$). Fearing this internal instability of $\{1, 2\}$, player 2 may reject player 1's proposal, preferring to get $x_2 > 0$. Thus, x may be viable if binding contracts are not available.[5] In fact, x is an equilibrium binding agreement for the grand coalition (Ray and Vohra (1997)).[6] By allowing legally binding contracts, we avoid having to deal with these issues.

Now consider the issue of break-up by mutual agreement. Suppose $b < 0$, and suppose a two-player coalition $\{i, j\}$ forms, sharing b equally. If instead they manage to break apart and induce $(\{1\}, \{2\}, \{3\})$, they are *both* made better off. A coalition S is said to be *not credible* if there is a partition $\mathcal{P}_S$ of S such that

$$\sum_{T \in \mathcal{P}_S} V(T) > V(S)$$

Thus, if $b < 0$ then two-player coalitions are not credible. One might argue that it is not realistic to assume that non-credible coalitions stick together. However, this problem is moot, because if a coalition is not credible, then the corresponding core constraint is anyway redundant (see Ray, 1989). Notice that if $b < 0$, then $x_i \geq 0$ for all i implies that $x_j + x_k \geq b$ for all j, k distinct.

Suppose we modify the example by setting

$$P(\{1\}|(\{i\}, \{j\}, \{k\})) = P(\{2\}|(\{i\}, \{j\}, \{k\})) = c > 0$$

Otherwise, the partition function is as before. Suppose $a < 0$, $c < b < 2c$, and $a + b < 1$. Now $b > c$ implies that if player $i \in \{1, 2\}$ leaves, then the other two will form a coalition. Therefore

$$V(\{i\}) = P(\{i\}|(\{i\}, \{j, k\})) = a$$

for $i \in \{1, 2\}$. But $b < 2c$ means that if player 3 leaves, the other two will split up. Therefore

$$V(\{3\}) = P(\{3\}|(\{i\}, \{j\}, \{k\})) = 0$$

[4] Of course, there are many real-world situations where even unanimous consent is not enough to break a coalition. An example would be certain laws of marriage.

[5] A potential problem with this argument is the hypothesis that player 1 can induce the finest coalition structure $(\{1\}, \{2\}, \{3\})$ by defecting from $\{1, 2\}$. Recall that player 1 cannot block x on his own, because he fears that the other two players will merge. But then, why does he think he can induce $(\{1\}, \{2\}, \{3\})$ by defecting from $\{1, 2\}$? The argument seems to put too much emphasis on the breaking apart of coalitions and too little on the possibility of re-merging.

[6] No internally stable EBA for the coalition $\{1, 2\}$ can block x, since $x_2 \geq b$. More generally, if $0 < b \leq (1 - a)/2$ and $a < 0$, then payoff vector x is an EBA for the grand coalition if $x_i \geq a$ for each $i \in N$, $x_1 + x_2 + x_3 = 1$ and *either* there are at least two distinct players $j, k \in N$ such that $x_j, x_k \geq b$, *or* $x_i + x_j \geq b$ for all distinct $i, j \in N$.

For any two-player coalition, $V(\{j, k\}) = b$, and $V(N) = 1$. Thus, the recursive core $C(N)$ consists of all payoff vectors such that $x_1 + x_2 + x_3 = 1,\ x_1 \geq a, x_2 \geq a, x_3 \geq 0$ and $x_i + x_j \geq b$ for all $i, j \in N$ distinct. Intuitively, one might argue as above that perhaps $\{1, 2\}$ is unlikely to stick together, because if they split up and induce $(\{1\}, \{2\}, \{3\})$, they get $c > b/2$ each. However, they cannot be assured that $(\{1\}, \{2\}, \{3\})$ would be the final outcome. Each player $i \in \{1, 2\}$ fears that if he stands alone the other two will merge. Coalition $\{1, 2\}$ is credible in the sense that

$$V(\{1, 2\}) = b > 2a = V(\{1\}) + V(\{2\})$$

Accordingly, it is not unrealistic to assume that $\{1, 2\}$ sticks together. It is important to notice that the recursive method computes the value of a coalition based on the whole strategic situation, it does not automatically declare the finest partition stable.

4. The non-cooperative game

We will show that a modified Perry and Reny (1994) extensive form game yields a natural non-cooperative implementation of the r-core. The modification is required in order to allow for externalities across coalitions. First, we will informally describe the non-cooperative game, following Section 2 of Perry and Reny (1994).

4.1. Informal description

The game starts at $t = 0$ and time is continuous. At any point in time, a player can either: (1) make a proposal; (2) accept the current proposal; (3) stay quiet or (4) leave.[7] A proposal $((w_i)_{i\in S}, S)$ by any player who has not left consists of a *division rule* $(w_i)_{i\in S}$ and a coalition S. Here $w_i \in \mathbb{R}_+$ represents i's share of the worth of S. We require $\sum_{i\in S} w_i = 1$ and $w_i \geq 0$ for all $i \in S$. In Perry and Reny (1994), a player proposes a *payoff vector*, but this would not work here because the members of S do not know what payoffs are feasible until all other coalitions have formed. (In a partition function form game the coalitional structure determines the value of S.) A division rule does work, because it implies a feasible distribution of payoffs within S for *every* possible coalitional structure that might form.[8]

When a proposal is made, it is *effective* as long as no new proposal is made. Once a new proposal is made, the previous proposal is no longer effective. To avoid the simultaneous proposals of distinct proposals, when this happens it is ruled that no new proposal is effective. So there is at most one effective proposal at any point in time.

If an effective proposal $((w_i)_{i\in S}, S)$ is accepted by all members in S, then the proposal becomes *binding*, and S is a *binding coalition*. If any player in binding coalition S chooses to leave, then all players in S must leave at the same time. At any point in time, there might be several binding proposals among the players who have not left. If a new proposal contains any player in any binding coalition, then it must contain *all* players in that binding coalition. This reflects the idea that annulment of a binding agreement has to be approved by every member in it. To avoid the problem where a player is involved in two different binding coalitions, if a new proposal $((w_i)_{i\in S}, S)$ becomes binding, then any old binding proposal that involves members of S

[7] At the very beginning of the game when $t = 0$, players can only choose either (1) or (3).

[8] We could generalize to allow a proposal pertaining to a coalition S to specify a complete contingent plan regarding how to divide S's value under all possible coalitional structures $N\backslash S$ may form. Allowing this does not change the result of the paper.

is annulled. Hence, at any point in time, all binding coalitions are disjoint. Players consume only after all have left. Notice that when all players leave, the coalitional structure is uniquely defined. Thus, the partition function implies a well-defined payoff for each coalition. Players share the payoff of the coalition to which they belong according to the binding proposals they have signed.

We now more formally define the rules and the equilibrium concept. Again, the description closely follows Perry and Reny (1994).

4.2. Histories

A proposal specifies a division rule $(w_i)_{i\in S}$ and a coalition S. Thus, the set of feasible proposals is

$$P \equiv \{((w_i)_{i\in S}, S) : S \subseteq N, \sum_{i\in S} w_i = 1 \quad \text{and}\, w_i \geq 0 \quad \text{for all}\, i \in S\}$$

Denote by a the choice to accept the current effective proposal, q the choice to be quiet and l the choice to leave. A history for player i up to time $t > 0$ is a function h_i such that

$$h_i : [0, t) \to P \cup \{a, q, l\}$$

If h_i is a history for player i up to time t, and $t' < t$, then let $h_i|t'$ denote the history for player i up to time t' which is implied by h_i (i.e., $h_i|t'$ is the truncation of h_i at time t').

Since players can only leave once and for all, $h_i^{-1}(l)$ is either empty or a singleton. We follow Perry and Reny (1994) by assuming that $h_i^{-1}(P \cup \{a\})$ is a finite set. At $t = 0$, since nothing has happened, $h_i(0) = \emptyset$. For convenience, denote a history up to time t by the n-tuple of functions $h \equiv (h_1, h_2, \ldots, h_n)$. For $t' < t$, let $h|t'$ denote the truncation of h at t'. Let $H(t)$ denote the set of all histories up to time t and $H \equiv \cup_{t=0}^{\infty} H(t)$ the set of all histories.

Let $p(h)$ denote the current effective proposal according to h. To make it well-defined, if according to h, either no proposal has been made, multiple distinct proposals are simultaneously made, the current effective proposal has become binding or some member in a binding coalition which is involved in the current effective proposal has exercised to leave, then $p(h) = \emptyset$.[9]

Let $\tau(h)$ for $h \in H(t)$ denote the amount of time that has passed up to time t since $p(h)$ was proposed. Whenever $p(h) = \emptyset$, $\tau(h)$ measures the time that has passed since the previous effective proposal becomes binding. When there is never any effective proposal, $\tau(h)$ measures the time that has passed since time 0.

Let $N(h) \subseteq N$ denote the set of players who have not left and $A(h) \subseteq N(h)$ the set of players who have accepted $p(h)$. Whenever $p(h) = \emptyset$, $A(h) = \emptyset$.

Player i is said to have accepted the current effective proposal $p(h)$ for $h \in H(t)$ if $p(h)$ is made at time $\bar{t} < t$ and $h_i(t') = a$ for some $t' \in (\bar{t}, t)$. If everyone involved in $p(h)$ has accepted it, then $p(h)$ is binding. The coalitions in binding proposals are called binding coalitions.

Let $\Pi(h)$ denote the set of binding proposals among the players in $N(h)$. Since there are externalities across coalitions, we also need to keep track of those binding coalitions that have left. Let $K(h)$ denote the coalitional structure formed by those players that have left, i.e., $N \backslash N(h)$,

[9] When some member in a binding coalition which is involved in the current effective proposal has exercised to leave, we need to reset $p(h)$ to an empty set to avoid the following from happening. Suppose players 1, 2 and 3 remain and the current proposal pertains to them. Suppose player 1 and 2 have accepted the current proposal and 3 has not. Suppose coalition {3} is binding. If 3 exercises to leave, then by the rules in Perry and Reny, if $p(h)$ is not reset to an empty set, players 1 and 2 cannot do anything further and they have to stay in the game forever.

according to h. Perry and Reny (1994) do not keep track of $K(h)$ since for characteristic function form games the remaining players' values are not affected by the coalitional structure of the players who left. In our setting, the coalitional structure of the players who left does affect the values of the remaining players.

4.3. Payoffs

If a proposal $((w_i)_{i\in S}, S)$ becomes binding and player $i \in S$ leaves, then in contrast to Perry and Reny (1994) he cannot consume immediately (because the worth of S depends on the final coalitional structure $\mathcal{P}_N$). All he can guarantee himself by leaving is that coalition S will be part of the final coalitional structure. When all players have left, so that $N(h) = \emptyset$, a structure $\mathcal{P}_N$ of binding coalitions has formed. Every coalition in $\mathcal{P}_N$ distributes its value according to its binding division rule. Thus, if $S \in \mathcal{P}_N$ and its division rule is $(w_i)_{i\in S}$, then player $i \in S$ gets $w_i P(S|\mathcal{P}_N)$. If there is a player who never leaves the game, then every player $i \in N$ gets $-\infty$.[10]

4.4. Strategies

A strategy for any player is a function which maps every possible history to an action. Hence, a strategy for any player $i \in N$, denoted by f_i is

$$f_i : H \to P \cup \{a, q, l\}$$

Denote the n-tuple of strategies by $f \equiv (f_1, f_2, \ldots, f_n)$. We impose several restrictions on the strategies.

- (S0) For $h \in H(0)$, $f_i(h) \in P \cup \{q\}$. That is, at the very beginning of the game, players can only make a proposal or be quiet.
- (S1) If $i \in N(h)$ has accepted the current effective proposal $p(h)$, then $f_i(h) = q$. That is, before the current effective proposal becomes binding, an accepting player can only be quiet. If $i \in N\backslash N(h)$, then $f_i(h) = q$. Thus, a leaving player can only be quiet.
- (S2) If $f_i(h) = ((w_i)_{i\in S'}, S')$, then for any binding coalition S, either $S \cap S' = \emptyset$ or $S \subseteq S'$. Moreover, $S' \subseteq N(h)$. That is, if a new proposal contains some players in a binding coalition, it has to include all of them. Any new proposal can only contain players who have not left.
- (S3) If a player i is not a member of any binding coalition, then $f_i(h) \neq l$. That is, a player can only leave if he belongs to a binding coalition.
- (S4) For all i and all $\bar{t} > 0$ and for all $h \in H(\bar{t})$ and $t \in [0, \bar{t})$, there exists an $\varepsilon > 0$ such that $f_i(h|\tau) = q$ for all $\tau \geq 0$ and $\tau \in (t - \varepsilon, t + \varepsilon)\backslash\{t\}$. That is, there are two open intervals $(t - \varepsilon, t)$ and $(t, t + \varepsilon)$ in which player i is quiet. This assumption makes sure that players always have enough time to respond. (See Section 3, and the last paragraph on page 806 in Perry and Reny (1994), for a discussion.)

[10] One can relax this rather strong assumption. For instance, if some binding coalitions have left while others remain in the game forever, it might be argued that although the remaining players might get the worst possible payoffs, any leaving coalition should at least get its value in the coalitional structure where all the leaving coalitions have formed and the remaining players form into the worst possible coalitional structure for this coalition in consideration. Allowing this does not change the result.

Lastly, denote player i's payoff induced by the strategy tuple f after h by $u_i(f|h)$. Let F_i denote the set of strategies for i which satisfy (S0) to (S4).

4.5. Equilibrium concept

The equilibrium concept is stationary subgame perfect Nash equilibrium (SSPNE). By definition, a strategy profile $\hat{f} \equiv (\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_n)$ is an SSPNE if (E1) and (E2) are satisfied:

(E1) Perfection: for all $i \in N$, $h \in H$ and $f_i \in F_i$

$$u_i(\hat{f}|h) \geq u_i((f_1, \hat{f}_2, \ldots, \hat{f}_n)|h)$$

(E2) Stationarity: for $h, h' \in H$, if

$$(p(h), \tau(h), N(h), A(h), \Pi(h), K(h)) = (p(h'), \tau(h'), N(h'), A(h'), \Pi(h), K(h')),$$

then $\hat{f}(h) = \hat{f}(h')$

Notice that we have one more state variable $K(h)$ than Perry and Reny because of externalities across coalitions.

5. Results

We will prove two theorems, corresponding to Theorems 1 and 2 in Perry and Reny (1994). The first theorem states that every SSPNE outcome of the extensive form game is in the r-core of the partition function game $\langle N, P\rangle$. The second theorem needs some qualification. We will show that for any totally r-balanced game $\langle N, P\rangle$, every r-core outcome can be supported as an SSPNE outcome of the extensive form game.

The proof of Theorem 1 is more complicated than the corresponding proof in Perry and Reny (1994). This is because in our model the value of a coalition is not given by a characteristic function, but instead has to be *derived* recursively from the partition function. Thus, we first need to show that if there exists an SSPNE of the extensive form game, then the value of each coalition is well defined. After having done this, we show the existence of an r-core outcome which corresponds to this SSPNE outcome.

The following proposition is used to prove Theorem 1.

Proposition 2. *Suppose an SSPNE $\hat{f}$ exists. Take any $S \subseteq N$ and any partition $\mathcal{P}_{N\backslash S}$ on $N\backslash S$. Let x denote the subgame equilibrium outcome induced by $\hat{f}$ where players in $N\backslash S$ have left according to $\mathcal{P}_{N\backslash S}$, i.e. in the subgame where the states are*

$$(p(h), \tau(h), N(h), A(h), \Pi(h), K(h)) = (\emptyset, 0, S, \emptyset, \emptyset, \mathcal{P}_{N\backslash S})$$

The following is true:

(a) $C(S|S, \mathcal{P}_{N\backslash S}) \neq \emptyset$.
(b) There must exist $y \in C(S|S, \mathcal{P}_{N\backslash S})$ such that $x_i = y_i$ for all $i \in S$.
(c) The coalitional structure induced by $\hat{f}$ must be $(\mathcal{P}_S, \mathcal{P}_{N\backslash S})$ where $\mathcal{P}_S \in \mathcal{P}(S|S, P_{N\backslash S})$.

Proof. We will proceed by induction. Suppose $|S| = 1$. For any $S = \{i\}$ and any $\mathcal{P}_{N\backslash\{i\}}$, $C(\{i\}|\{i\}, \mathcal{P}_{N\backslash\{i\}}) \neq \emptyset$ since by definition, it consists of every payoff vector where i gets $P(\{i\}|\{i\}, \mathcal{P}_{N\backslash\{i\}})$ and every coalition $S' \in \mathcal{P}_{N\backslash\{i\}}$ gets $P(S'|\{i\}, \mathcal{P}_{N\backslash\{i\}})$.

In the subgame where all players in $N\backslash\{i\}$ have left and formed the coalitional structure $\mathcal{P}_{N\backslash\{i\}}$, player i must first propose $((1), \{i\})$ and then leave according to $\hat{f}_i$. For otherwise, i is staying in the game forever and getting $-\infty$, which cannot be an equilibrium strategy. Parts (b) and (c) follow immediately.

Thus, we have established that parts (a), (b) and (c) hold for any S such that $|S| = 1$ and any partition $\mathcal{P}_{N\backslash S}$ on $N\backslash S$.

To continue the induction, suppose that for any S such that $|S| \leq k-1 < n$ and any partition $\mathcal{P}_{N\backslash S}$ on $N\backslash S$, parts (a), (b) and (c) hold.

Suppose $|S| = k$. For any partition $\mathcal{P}_{N\backslash S}$ on $N\backslash S$, to show part (a), we first need to make sure that $V(T|S, \mathcal{P}_{N\backslash S})$ is well defined for all $T \subseteq S$. By definition, $V(S|S, \mathcal{P}_{N\backslash S}) = P(S|S, \mathcal{P}_{N\backslash S})$. And $V(T|S, \mathcal{P}_{N\backslash S})$ is also well defined because $C(S\backslash T|S\backslash T, T, \mathcal{P}_{N\backslash S}) \neq \emptyset$ by the induction hypothesis (because $|S\backslash T| \leq k-1$).

Suppose by contradiction that $C(S|S, \mathcal{P}_{N\backslash S}) = \emptyset$. At time t, let h be any history such that

$$(p(h), \tau(h), N(h), A(h), \Pi(h), K(h)) = (\emptyset, 0, S, \emptyset, \emptyset, \mathcal{P}_{N\backslash S})$$

Consider the continuation equilibrium outcome x induced by $\hat{f}$. Since $C(S|S, \mathcal{P}_{N\backslash S}) = \emptyset$, there must exist a coalition $S' \subseteq S$ such that $x(S') < V(S'|S, \mathcal{P}_{N\backslash S})$. Without loss of generality, suppose $S' = \{1, 2, \ldots, s'\}$ where $s' = |S'|$. Let

$$y_i = x_i + \frac{V(S'|S, \mathcal{P}_{N\backslash S}) - x(S')}{|S'|} \quad \text{for all } i \in S'$$

and define

$$w_i = \frac{y_i}{V(S'|S, \mathcal{P}_{N\backslash S})} \quad \text{for all } i \in S'$$

Consider any time $t' > t$, any history $h' \in H(t')$ such that $h'|t = h$ and (i) $p(h') = ((w_i)_{i\in S'}, S')$, (ii) $N(h') = S$, (iii) $A(h') = \{1, 2, \ldots, s'-1\}$, (iv) $\Pi(h') = \emptyset$, (v) $K(h') = \mathcal{P}_{N\backslash S}$, and (vi) $\hat{f}_i(h') = q$ for all $i \in N$. Thus, at time t' player s' is the only member of $S' = \{1, \ldots, s'\}$ who has not yet accepted the proposal $((w_i)_{i\in S'}, S')$.

Claim 1. According to $\hat{f}$, player s' will accept $((w_i)_{i\in S'}, S')$ before any new proposal is made and thus $((w_i)_{i\in S'}, S')$ will become binding.

Proof of claim. A feasible action for player s' is to accept the current effective proposal $((w_i)_{i\in S'}, S')$ and leave before anything happens. In this case the whole coalition S' leaves by the rules of the game. The resulting coalitional structure will be $(\mathcal{P}_{S\backslash S'}, S', \mathcal{P}_{N\backslash S})$ for some r-core coalitional structure $\mathcal{P}_{S\backslash S'} \in \mathcal{P}(S\backslash S'|S\backslash S', S', \mathcal{P}_{N\backslash S})$, according to the induction hypothesis. When S' is calculating its value $V(S'|S, \mathcal{P}_{N\backslash S})$, it expects the worst possible r-core coalitional structure and $\mathcal{P}_{S\backslash S'}$ is not necessarily the worst possible one, so

$$V(S'|S, \mathcal{P}_{N\backslash S}) \leq P(S'|\mathcal{P}_{S\backslash S'}, S', \mathcal{P}_{N\backslash S})$$

(If $S' = S$, then $V(S'|S, \mathcal{P}_{N\backslash S}) = P(S'|S', \mathcal{P}_{N\backslash S})$.) Player s' gets

$$w_{s'} P(S'|\mathcal{P}_{S\backslash S'}, S', \mathcal{P}_{N\backslash S}) \geq y_{s'} > x_{s'}$$

Suppose, according to $\hat{f}$, player s' never accepts any proposal ever. Then player s' gets $-\infty$, which contradicts the assumption that $\hat{f}$ is an SSPNE, since we have just shown that by accepting and leaving he can do better. Suppose instead that, according to $\hat{f}$, some new proposal $((w_i)_{i\in S''}, S'')$ is made before player s' accepts $((w_i)_{i\in S'}, S')$. In the continuation equilibrium, s'

must get at least $w_{s'}P(S'|\mathcal{P}_{S\setminus S'}, S', \mathcal{P}_{N\setminus S}) > x_{s'}$ (we have shown that he has a feasible action which gives him $w_{s'}P(S'|\mathcal{P}_{S\setminus S'}, S', \mathcal{P}_{N\setminus S})$). By stationarity, this means whenever any history h' yields the states

$$(p(h'), \tau(h'), N(h'), A(h'), \Pi(h'), K(h')) = (((w_i)_{i\in S''}, S''), 0, S, \emptyset, \emptyset, \mathcal{P}_{N\setminus S})$$

player s' gets strictly more than $x_{s'}$. But then player s' could have proposed $((w_i)_{i\in S''}, S'')$ at time close enough to t. This is in contradiction to x being a continuation equilibrium outcome. Hence according to $\hat{f}$, player s' will accept $((w_i)_{i\in S'}, S')$ before any new proposal is made and thus $((w_i)_{i\in S'}, S')$ will become binding. This proves the claim.

Next consider at any time $t'' > t$, any history $h'' \in H(t'')$ such that $h''|t = h$ and (i) $p(h'') = ((w_i)_{i\in S'}, S')$, (ii) $N(h'') = S$, (iii) $A(h'') = \{1, 2, \dots, s' - 2\}$, (iv) $\Pi(h'') = \emptyset$, (v) $K(h'') = \mathcal{P}_{N\setminus S}$ and (vi) $\hat{f}_i(h'') = q$ for all $i \in N$.

Claim 2. According to $\hat{f}$, players $s' - 1$ and s' will accept $((w_i)_{i\in S'}, S')$ before any new proposal is made and $((w_i)_{i\in S'}, S')$ will become binding.

Proof of claim. The argument is virtually the same as the proof of the previous claim. A feasible action for player $s' - 1$ is to accept the proposal, and if he does, then the previous claim applies so s' will accept as well. Thus player $s' - 1$ can guarantee himself the payoff of

$$w_{s'-1}P(S'|\mathcal{P}_{S\setminus S'}, S', \mathcal{P}_{N\setminus S}) \geq y_{s'-1} > x_{s'-1}$$

by accepting and leaving after s' accepts. As in the previous claim, suppose to the contrary that either according to $\hat{f}$ player $s' - 1$ never accepts anything, or a new proposal is made before player $s' - 1$ accepts $((w_i)_{i\in S'}, S')$. In the first possibility, player $s' - 1$ gets $-\infty$, which is a contradiction of the fact that $\hat{f}$ is an SSPNE. In the second possibility, by an analogous argument, when the new proposal is made, player $s' - 1$ must get strictly more than $x_{s'-1}$. But he could have made that proposal at time close enough to t. This proves the claim.

Proceeding stepwise just as in these two claims, we can finally establish the following claim. Let $t''' > t$, and consider a history $h''' \in H(t''')$ such that $h'''|t = h$, and suppose: (i) $p(h''') = ((w_i)_{i\in S'}, S')$, (ii) $N(h''') = S$, (iii) $A(h''') = \{1\}$, (iv) $\Pi(h''') = \emptyset$, (v) $K(h''') = \mathcal{P}_{N\setminus S}$ and (vi) $\hat{f}_i(h''') = q$ for all $i \in N$. Then, according to $\hat{f}$, players 2 to s' will accept $((w_i)_{i\in S'}, S')$ before any new proposal is made and $((w_i)_{i\in S'}, S')$ will become binding.

However, this claim contradicts the hypothesis that x is a continuation equilibrium outcome induced by $\hat{f}$, because player 1 can always propose $((w_i)_{i\in S'}, S')$ and subsequently accept it at time close enough to t. By this deviation, he gets at least $y_1 > x_1$. Thus $C(S|S, \mathcal{P}_{N\setminus S}) \neq \emptyset$ and part (a) is proved.

For part (b), suppose x is the continuation equilibrium outcome induced by $\hat{f}$ where players in $N\setminus S$ have left according to $\mathcal{P}_{N\setminus S}$, i.e., in the subgame where the states are

$$(p(h), \tau(h), N(h), A(h), \Pi(h), K(h)) = (\emptyset, 0, S, \emptyset, \emptyset, \mathcal{P}_{N\setminus S})$$

If there does not exist a $y \in C(S|S, \mathcal{P}_{N\setminus S})$ such that

$$x_i = y_i \quad \text{for all } i \in S$$

then there must exist a coalition $S' \subseteq S$ such that $x(S') < V(S'|S, \mathcal{P}_{N\setminus S})$. By the same argument in the proof of part (a), this leads to a contradiction. Part (c) follows from part (b) immediately. □

Proposition 2 directly implies the following theorem, which states that every SSPNE outcome is in the r-core of $\langle N, P\rangle$. This is directly analogous to Theorem 1 in Perry and Reny (1994).

Theorem 1. *Suppose an SSPNE $\hat{f}$ induces an equilibrium outcome x. Then $x \in C(N)$, and the coalitional structure induced by $\hat{f}$ belongs to the r-core structure $\mathcal{P}(N)$.*

Now consider the converse of Theorem 1. For any r-core outcome, we need strategies that support it as an SSPNE. It turns out that strategies similar to those used by Perry and Reny (1994) in the proof of their Theorem 2 work here as well. The main modification is due to the fact that players must propose division rules instead of payoff vectors. If we assume total r-balancedness, then analogues of Lemmas 1 and 2 in Perry and Reny (1994) can be readily proved, and from this we obtain a converse to Theorem 1.

If the r-core exists, then it exists for any reduced society S given any $\mathcal{P}_{N\setminus S}$. Thus, for any $S \subset N$ and $\mathcal{P}_{N\setminus S}$ we can select an r-core payoff vector from $C(S|S, \mathcal{P}_{N\setminus S})$. In the following, we denote this payoff vector by $x(S|\mathcal{P}_{N\setminus S})$, where

$$x(S|\mathcal{P}_{N\setminus S}) \in C(S|S, \mathcal{P}_{N\setminus S})$$

For the grand coalition, we choose

$$x(N|\mathcal{P}_{N\setminus N}) \in C(N)$$

Note that when the game is totally r-balanced, for any reduced society S given any $\mathcal{P}_{N\setminus S}$, $\mathcal{P}(S|S, \mathcal{P}_{N\setminus S}) = (S)$.

Perry and Reny (1994) construct two continuation equilibrium payoff vectors: one for the continuing equilibrium when the current proposal is rejected, the other for the continuation equilibrium when the current proposal is accepted. For any history h where

$$\Pi(h) = \{((w_i^1)_{i\in S^1}, S^1), ((w_i^2)_{i\in S^2}, S^2), \ldots, ((w_i^m)_{i\in S^m}, S^m)\}$$

define

$$z_i(h) = \begin{cases} w_i^k \sum_{j\in S^k} x_j(N(h)|K(h)) & \text{if } i \in S^k \text{ for some } k \in \{1, 2, \ldots, m\} \\ x_i(N(h)|K(h)) & \text{if } i \in N(h)\setminus(S^1 \cup S^2 \cup \ldots \cup S^m) \end{cases}$$

If the current effective proposal is $p(h) = ((w_i)_{i\in S}, S)$, without loss of generality we can assume that there exists an integer r such that in the current proposal $p(h)$, the coalition S contains all the coalitions S^k where $k \le r$.[11] On the other hand, S is disjoint from all the coalitions S^k where $k \ge r+1$. Thus when the current proposal becomes binding, the resulting new set of binding proposals becomes

$$\hat{\Pi}(h) = \{((w_i)_{i\in S}, S), ((w_i^{r+1})_{i\in S^{r+1}}, S^{r+1}), \ldots, ((w_i^m)_{i\in S^m}, S^m)\}$$

Define

$$\hat{z}_i(h) = \begin{cases} w_i \sum_{j\in S} x_j(N(h)|K(h)) & \text{if } i \in S, \\ w_i^k \sum_{j\in S^k} x_j(N(h)|K(h)) & \text{if } i \in S^k \text{ where } k \in \{r+1, \ldots, m\} \\ x_i(N(h)|K(h)) & \text{if } i \in N(h)\setminus(S \cup S^{r+1} \cup \ldots \cup S^m) \end{cases}$$

By construction, $z_i(h)$ will be player i's payoff in the continuation equilibrium if the current effective proposal gets rejected, while $\hat{z}_i(h)$ will be his payoff if it gets accepted. When $p(h) = \emptyset$, $z_i(h) = \hat{z}_i(h)$.

[11] Recall that if a new proposal contains any player in any binding coalition, then it must contain all players in that binding coalition.

The equilibrium strategies are defined as follows. For every $t \geq 0$, $h \in H(t)$ and $i \in N(h)$, if $\tau(h)$ is not a positive integer then $f_i(h) = q$. If $\tau(h)$ is a positive integer, then player i behaves as follows:

(a) if $p(h) = ((w_i)_{i\in S}, S) \neq \emptyset$, and $\hat{z}_j(h) \geq z_j(h)$ for all $j \in S\backslash A(h)$, then

$$f_i(h) = \begin{cases} a & \text{if } i \in S\backslash A(h) \\ q & \text{otherwise} \end{cases}$$

(b) otherwise

$$f_i(h) = \begin{cases} l & \text{if } i \in N(h) \quad \text{and } \Pi(h) = \{(\cdot, N(h))\} \\ \left(\left(\dfrac{z_j(h)}{\sum_{l\in N(h)} x_l(N(h)|K(h))}\right)_{j\in N(h)}, N(h)\right) & \text{if } i \in N(h)\backslash A(h) \quad \text{and } \Pi(h) \neq \{(\cdot, N(h))\} \\ q & \text{if } i \in A(h) \end{cases}$$

Here $\Pi(h) = \{(\cdot, N(h))\}$ means that the binding coalition is $N(h)$[12] and $\Pi(h) \neq \{(\cdot, N(h))\}$ means that the binding coalition is not $N(h)$.

These strategies depend only on the state variable. On the equilibrium path, all players propose

$$\left(\left(\frac{x_i(N|\mathcal{P}_{N\backslash N})}{\sum_{j\in N} x_j(N|\mathcal{P}_{N\backslash N})}\right)_{i\in N}, N\right)$$

at time 1, accept at time 2, and leave at time 3. The equilibrium outcome is $x(N|\mathcal{P}_{N\backslash N})$ where every $i \in N$ gets $x_i(N|\mathcal{P}_{N\backslash N})$. For any history h where $K(h)$ has left and nothing else has happened, in the continuing subgame, on the equilibrium path every $i \in N(h)$ gets $x_i(N(h)|K(h))$.

With these strategies, we can easily prove two lemmas analogous to Lemmas 1 and 2 in Perry and Reny (1994). The proof of Lemma 1 shows how total r-balancedness is used.

Lemma 1. *Assume total r-balancedness. For any $h \in H$, if $p(h) = ((w_i)_{i\in S}, S)$ and*

$$\Pi(h) = \{((w_i^1)_{i\in S^1}, S^1), ((w_i^2)_{i\in S^2}, S^2), \ldots, ((w_i^m)_{i\in S^m}, S^m)\}$$

then,[13]

(a) *for any $k \in \{1, 2, \ldots, m\}$and $i \in S^k$*

$$z_i(h) \geq w_i^k \sum_{j\in S^k} x_j(N(h)\backslash S^k|S^k, K(h))$$

(b) *for any $k \in \{1, 2, \ldots, m\}$*

$$\sum_{i\in S^k} z_i(h) = \sum_{i\in S^k} x_i(N(h)|K(h))$$

(c) $$\sum_{i\in S} z_i(h) = \sum_{i\in S} \hat{z}_i(h) = \sum_{i\in S} x_i(N(h)|K(h))$$

[12] In other words, there exists a division rule $(w_i)_{i\in N(h)}$ such that $\Pi(h) = \{((w_i)_{i\in N(h)}, N(h))\}$.

[13] Recall that there exists an integer r such that the current proposal S contains all the coalitions S^k where $k \leq r$ and is disjoint from all the coalitions S^k where $k \geq r + 1$.

Proof.

(a) By definition

$$z_i(h) = w_i^k \sum_{j\in S^k} x_j(N(h)|K(h))$$

Since $x(N(h)|K(h))$ belongs to the r-core for $N(h)$ given $K(h)$

$$\sum_{j\in S^k} x_j(N(h)|K(h)) \geq V(S^k|N(h), K(h)) = P(S^k|S^k, N(h)\backslash S^k, K(h))$$

The equality follows because by total r-balancedness, when S^k breaks off, the remaining players in $N(h)\backslash S^k$ stay together in the r-core. Now $x(N(h)\backslash S^k|S^k, K(h))$ belongs to the r-core for $N(h)\backslash S^k$ given $(S^k, K(h))$, and by total r-balancedness players in $N(h)\backslash S^k$ stay together. So

$$\sum_{j\in S^k} x_j(N(h)\backslash S^k|S^k, K(h)) = P(S^k|S^k, N(h)\backslash S^k, K(h))$$

Hence

$$z_i(h) \geq w_i^k \sum_{j\in S^k} x_j(N(h)\backslash S^k|S^k, K(h))$$

(b) By definition

$$\sum_{i\in S^k} z_i(h) = \sum_{i\in S^k} w_i^k \left(\sum_{j\in S^k} x_j(N(h)|K(h))\right) = \sum_{i\in S^k} x_i(N(h)|K(h))$$

since $\sum_{i\in S^k} w_i^k = 1$.

(c) By definition

$$\sum_{i\in S} z_i(h) = \sum_{k\in\{1,\ldots,r\}} \left(\sum_{i\in S^k} z_i(h)\right) + \sum_{i\in S\backslash(S^1\cup\ldots\cup S^r)} x_i(N(h)|K(h)) = \sum_{i\in S} x_i(N(h)|K(h))$$

where the second equality follows because of (b). By definition

$$\sum_{i\in S} \hat{z}_i(h) = \sum_{i\in S} w_i \left(\sum_{j\in S} x_j(N(h)|K(h))\right) = \sum_{i\in S} x_i(N(h)|K(h))$$

because $\sum_{i\in S} w_i = 1$. □

Lemma 2. *Assume total r-balancedness. For any $t \geq 0$, $h \in H(t)$ and $A(h) = \emptyset$, the outcome generated by the strategies $f \equiv (f_1, \ldots, f_n)$ after h is such that player i gets $z_i(h)$ for all $i \in N(h)$.*

Proof. Let t_1 be the smallest time at least as large as t such that $t_1 - \tau(h)$ is an integer. According to the strategies f, if $t \neq t_1$, then all players are quiet in the interval $[t, t_1)$. Denote the history generated by $h_1 \in H(t_1)$. There are four possible cases depending on what players will do at t_1 according to the strategies f.

Case 1. If $f_i(h_1) = l$, then it must be the case that $\Pi(h_1) = \{(\cdot, N(h_1))\}$. This implies that there exists a division rule $(w_i)_{i\in N(h_1)}$ such that

$$\Pi(h_1) = \{((w_i)_{i\in N(h_1)}, N(h_1))\}$$

Thus player i gets

$$w_i \sum_{j\in N(h_1)} x_j(N(h_1)|K(h_1))$$

for all $i \in N(h_1)$ according to f. If $t \neq t_1$, then since everyone is quiet in $[t, t_1)$, $\Pi(h) = \Pi(h_1)$, $K(h) = K(h_1)$, and $N(h) = N(h_1)$. This implies that

$$z_i(h) = w_i \sum_{j\in N(h)} x_j(N(h)|K(h)) = w_i \sum_{j\in N(h_1)} x_j(N(h_1)|K(h_1))$$

Hence player i gets $z_i(h)$.[14]

Case 2. If $f_i(h_1) = a$ for some $i \in N(h_1)\backslash A(h_1)$ and

$$p(h_1) = ((w_i)_{i\in N(h_1)}, N(h_1))$$

since everyone is quiet in $[t, t_1)$ and thus $A(h_1) = A(h) = \emptyset$, then it implies $\hat{z}_i(h_1) \geq z_i(h_1)$ for all $i \in N(h_1)$ by the construction of the strategies f. So the current proposal becomes binding at t_1. By Lemma 1 (c)

$$\sum_{i\in N(h_1)} z_i(h_1) = \sum_{i\in N(h_1)} \hat{z}_i(h_1)$$

hence $\hat{z}_i(h_1) = z_i(h_1)$ for all $i \in N(h_1)$. Moreover, $N(h_1) = N(h)$ and $z_i(h_1) = z_i(h)$ for all $i \in N(h)$ because everyone is quiet in $[t, t_1)$.

By the equilibrium strategies, everyone is quiet in $(t_1, t_1 + 1)$. Let $t_2 = t_1 + 1$ and denote the history generated by $h_2 \in H(t_2)$. Since everyone is quiet in (t_1, t_2), $N(h_1) = N(h_2)$ and $K(h_1) = K(h_2)$. At t_2, since $\Pi(h_2) = \{(\cdot, N(h_2))\}$, so all players leave. Thus player $i \in N(h_2)$ gets

$$w_i \sum_{j\in N(h_2)} x_j(N(h_2)|K(h_2)) = w_i \sum_{j\in N(h_1)} x_j(N(h_1)|K(h_1)) = \hat{z}_i(h_1) = z_i(h_1) = z_i(h)$$

Case 3. If

$$f_i(h_1) = \left(\left(\frac{z_j(h_1)}{\sum_{l\in N(h_1)} x_l(N(h_1)|K(h_1))}\right)_{j\in N(h_1)}, N(h_1)\right)$$

then everyone is quiet between $(t_1, t_1 + 1)$. Let $t_2 = t_1 + 1$ and denote the history generated by $h_2 \in H(t_2)$. By definition for all $i \in N(h_1)$

$$\hat{z}_i(h_2) = \frac{z_i(h_1)}{\sum_{j\in N(h_1)} x_j(N(h_1)|K(h_1))} \sum_{l\in N(h_1)} x_l(N(h_2)|K(h_2)) = z_i(h_1) = z_i(h_2)$$

The second equality follows because no one leaves in $[t_1, t_2)$, so $N(h_1) = N(h_2)$ and $K(h_1) = K(h_2)$. The third equality follows because no new proposal binds in $[t_1, t_2)$, so $z_i(h_1) = z_i(h_2)$.

[14] If $t = t_1$ then it is obvious that $h = h_1$ and therefore $z_i(h) = z_i(h_1)$.

Thus all players in $N(h_1)$ accept the proposal at time t_2. According to the equilibrium strategies, everyone is quiet in $(t_2, t_2 + 1)$. Let $t_3 = t_2 + 1$ and denote the history generated by $h_3 \in H(t_3)$. Since no one leaves in $[t_2, t_3)$, so $N(h_2) = N(h_3)$. Hence all players in $N(h_3)$ leave at time t_3. Thus player $i \in N(h_3)$ gets

$$\frac{z_i(h_1)}{\sum_{j\in N(h_1)} x_j(N(h_1)|K(h_1))} \sum_{l\in N(h_3)} x_l(N(h_3)|K(h_3)) = z_i(h_1) = z_i(h)$$

The first equality follows because no one leaves in $[t_1, t_3)$ so $N(h_1) = N(h_3)$ and $K(h_1) = K(h_3)$. The second equality follows because nothing happens in $[t, t_1)$.

Case 4. If $f_i(h_1) = a$ for some $i \in S\backslash A(h_1)$ and $p(h_1) = ((w_i)_{i\in S}, S)$ where $S \neq N(h_1)$, since everyone is quiet in $[t, t_1)$ and thus $A(h_1) = A(h) = \emptyset$, then it implies $\hat{z}_i(h_1) \geq z_i(h_1)$ for all $i \in S$ by the construction of the strategies f. So the current proposal becomes binding at t_1. By Lemma 1 (c)

$$\sum_{i\in S} z_i(h_1) = \sum_{i\in S} \hat{z}_i(h_1)$$

hence $\hat{z}_i(h_1) = z_i(h_1)$ for all $i \in S$. Denote

$$\Pi(h_1) = \{((w_i^1)_{i\in S^1}, S^1), ((w_i^2)_{i\in S^2}, S^2), \ldots, ((w_i^m)_{i\in S^m}, S^m)\}$$

and without loss of generality, assume that S contains all the coalitions S^k where $k \leq r$ and is disjoint from all the coalitions S^k where $k \geq r + 1$. Then for all $i \in N(h_1)\backslash(S \cup S^{r+1} \cup \ldots \cup S^m)$

$$\hat{z}_i(h_1) = z_i(h_1) = x_i(N(h_1)|K(h_1))$$

and for all $i \in S^k$ where $k \geq r + 1$

$$\hat{z}_i(h_1) = z_i(h_1) = w_i^k \sum_{j\in S^k} x_j(N(h_1)|K(h_1))$$

Thus $\hat{z}_i(h_1) = z_i(h_1)$ for all $i \in N(h_1)$. Note that all players are quiet in $(t_1, t_1 + 1)$. Let $t_2 = t_1 + 1$ and denote the history generated by $h_2 \in H(t_2)$. Since no one leaves in $[t_1, t_2)$, so $N(h_1) = N(h_2)$.

Since $S \neq N(h_2)$, at time t_2, for all $i \in N(h_2)$,

$$f_i(h_2) = \left(\left(\frac{z_j(h_2)}{\sum_{l\in N(h_2)} x_l(N(h_2)|K(h_2))}\right)_{j\in N(h_2)}, N(h_2)\right)$$

Everyone is quiet in $(t_2, t_2 + 1)$. Let $t_3 = t_2 + 1$ and denote the history generated by $h_3 \in H(t_3)$. By definition for all $i \in N(h_2)$

$$\hat{z}_i(h_3) = \frac{z_i(h_2)}{\sum_{j\in N(h_2)} x_j(N(h_2)|K(h_2))} \sum_{l\in N(h_2)} x_l(N(h_3)|K(h_3)) = z_i(h_2) = z_i(h_3)$$

The second equality follows because no one leaves in $[t_2, t_3)$, so $N(h_2) = N(h_3)$ and $K(h_2) = K(h_3)$. The third equality follows because no new offer binds in $[t_2, t_3)$. Thus all players in $N(h_2)$ accept the proposal at time t_3. Everyone is quiet in $(t_3, t_3 + 1)$. Let $t_4 = t_3 + 1$ and denote the history generated by $h_4 \in H(t_4)$.

Since no one leaves in $[t_3, t_4)$, so $N(h_3) = N(h_4)$ and $K(h_3) = K(h_4)$. At t_4, since $\Pi(h_4) = \{(\cdot, N(h_4))\}$, so all players leave. Thus player $i \in N(h_4)$ gets

$$\frac{z_i(h_2)}{\sum_{j\in N(h_2)} x_j(N(h_2)|K(h_2))} \sum_{l\in N(h_4)} x_l(N(h_4)|K(h_4)) = z_i(h_2)$$

However $z_i(h_2) = \hat{z}_i(h_1)$ because at time t_1, $p(h_1)$ binds. Combining with $\hat{z}_i(h_1) = z_i(h_1)$, thus $z_i(h_2) = z_i(h_1) = z_i(h)$. The last equality follows because nothing happens in the interval $[t, t_1)$.[15] Thus player $i \in N(h)$ gets $z_i(h)$ in the continuing equilibrium. □

The following result is the partial converse to Theorem 1.

Theorem 2. *If $\langle N, P\rangle$ is totally r-balanced, then any payoff vector in the r-core can be supported as an SSPNE outcome.*

Proof. We want to show that the strategy f does constitute an equilibrium. Therefore, we want to show that any player i has no profitable deviation after any history, given all others are playing according to the equilibrium. For any $t \geq 0, h \in H(t)$ and $i \in N(h)\backslash A(h)$, if $f_j(h) = l$ for some $j \in N(h)\backslash\{i\}$, then according to the equilibrium strategies, it must be the case that $\Pi(h) = \{(\cdot, N(h))\}$. Hence player i has to leave anyway. So $f_i(h) = l$ is clearly optimal. Thus we only need to show that for any $t \geq 0$, $h \in H(t)$ and $i \in N(h)\backslash A(h)$, if $f_j(h) \neq l$ for all $j \in N(h)\backslash\{i\}$, then using the equilibrium strategy f_i given all others are using their corresponding equilibrium strategies f_{-i} is optimal for player i.

Consider another strategy f'_i for player i. Notice that f'_i and f_{-i} generate a unique continuation path h' subsequent to h. Since i gets at least $z_i(h) > -\infty$ by following f_i, he does no better if he never leaves according to f'_i. Thus suppose i leaves at time $t' \geq t$ with the proposal $((w'_j)_{j\in S'}, S')$. Notice that since others are following the equilibrium strategies f_{-i}, everyone must leave the game ultimately. The coalitional structure thus formed, denoted by $\mathcal{P}'_N$, must satisfy:

$$S \in \mathcal{P}'_N \quad \text{for all } S \in K(h) \quad \text{and } S' \in \mathcal{P}'_N$$

That is, it must respect the coalitions that have left. Thus player i obtains the payoff of $w'_i P(S'|\mathcal{P}'_N)$. Since the proposal $((w'_j)_{j\in S'}, S')$ must be binding before the coalition S' can leave, thus

$$((w'_j)_{j\in S'}, S') \in \Pi(h'|t')$$

By Lemma 1

$$z_i(h'|t') \geq w'_i \sum_{j\in S'} x_j(N(h'|t')\backslash S'|S', K(h'|t')) = w'_i P(S'|\mathcal{P}'_N)$$

The equality follows because after S' has left, since others are playing according to the equilibrium strategies, they must stay together and form a grand coalition of their own. Thus

$$\mathcal{P}'_N = \{N(h'|t')\backslash S'\} \cup \{S'\} \cup K(h'|t')$$

Moreover, the payoff vector

$$x(N(h'|t')\backslash S'|S', K(h'|t'))$$

is in the core

$$C(N(h'|t')\backslash S'|N(h'|t')\backslash S', S', K(h'|t'))$$

By total r-balancedness

$$\mathcal{P}(N(h'|t')\backslash S'|N(h'|t')\backslash S', S', K(h'|t')) = (N(h'|t')\backslash S')$$

[15] Hence $N(h) = N(h_1)$ and $K(h) = K(h_1)$.

hence the sum of payoffs for players in S' is simply

$$\sum_{j\in S'} x_j(N(h'|t')\backslash S'|S', K(h'|t')) = P(S'|\mathcal{P}'_N)$$

There are three exhaustive cases.

- Case A: $A(h) = \emptyset$.
- Case B: $p(h) = ((w_j)_{j\in T}, T)$ and either $\hat{z}_i(h) \le z_i(h)$ or $\hat{z}_j(h) < z_j(h)$ for some $j \in T\backslash A(h)$.
- Case C: $p(h) = ((w_j)_{j\in T}, T)$, $\hat{z}_i(h) > z_i(h)$ and $\hat{z}_j(h) \ge z_j(h)$ for all $j \in T\backslash A(h)$.

Notice that case B covers the instances where $i \notin T$ since then $\hat{z}_i(h) = z_i(h)$. Case A covers the instances where $p(h) = \emptyset$. As in Perry and Reny (1994), the argument applying to cases A and B have a common component, thus we treat them together until it is necessary to separate them.

Cases A and B. By Lemma 2, in case A, player i will get $z_i(h)$ by using the equilibrium strategy f_i. In case B, by the equilibrium strategies, either player i does not belong to T, he is the only player who has not accepted the current proposal $p(h)$ and will accept it or at least a player will reject the proposal by making a new proposal involving all the players $N(h)$. In any case, player i gets $z_i(h)$. Thus suppose to the contrary that by following f'_i, player i made a profitable deviation. Thus $w'_i P(S'|\mathcal{P}'_N) > z_i(h)$.

Since $z_i(h'|t') \ge w'_i P(S'|\mathcal{P}'_N) > z_i(h)$, it must be the case that $t' > t$. This is because $h'|t = h$. Hence, let

$$t^* = \inf\{\hat{t} \in [t, t']|z_i(h'|\hat{t}) > z_i(h)\}$$

It follows that $z_i(h) \ge z_i(h'|t^*)$. To see this, note if $t^* = t$, then since $h'|t = h$, the weak inequality is certainly true. If $t^* > t$ and suppose to the contrary that $z_i(h) < z_i(h'|t^*)$, then by (S4), there exists an $\varepsilon > 0$ small enough so that $t^* - \varepsilon > t$ and nothing happens in $[t^* - \varepsilon, t^*)$. Hence $z_i(h'|t^* - \frac{\varepsilon}{2}) = z_i(h'|t^*) > z_i(h)$. But then t^* is not the infimum. Hence $z_i(h) \ge z_i(h'|t^*)$. This implies $t^* \ne t'$ because $z_i(h'|t') > z_i(h)$. Thus $t' > t^*$.

Because $z_i(h) \ge z_i(h'|t^*)$ and t^* is the infimum, there must exist a sequence of positive numbers $\{\varepsilon_n\}$ where $\lim_{n\to\infty} \varepsilon_n = 0$ and $z_i(h'|t^* + \varepsilon_n) > z_i(h) \ge z_i(h'|t^*)$ for every ε_n. By (S4), there must exist an n^* large enough such that nothing happens in $(t^*, t^* + \varepsilon_{n^*})$. Thus something must happen at time t^* for otherwise it cannot be the case that $z_i(h'|t^* + \varepsilon_{n^*}) > z_i(h'|t^*)$. Thus either the current proposal $p(h'|t^*)$ which contains player i becomes binding at t^* or someone leaves at t^*. The latter cannot happen because according to f'_i, player i leaves at $t' > t^*$. Players other than i cannot leave at t^* either for otherwise, since they are playing according to the equilibrium strategies f_{-i}, when one leaves, all must leave, contradicting that player i leaves at $t' > t^*$. Therefore, the current proposal $p(h'|t^*)$ becomes binding at t^*. Let $p(h'|t^*) = ((w^*_j)_{j\in S}, S)$. Notice that $i \in S$ as argued.

Since the current proposal $p(h'|t^*)$ becomes binding at t^* and nothing happens in $(t^*, t^* + \varepsilon_{n^*})$, this implies that $\hat{z}_i(h'|t^*) = z_i(h'|t^* + \varepsilon_{n^*})$. Because $z_i(h'|t^* + \varepsilon_{n^*}) > z_i(h'|t^*)$, it follows that $\hat{z}_i(h'|t^*) > z_i(h'|t^*)$. By part (c) of Lemma 1

$$\sum_{j\in S} z_j(h'|t^*) = \sum_{j\in S} \hat{z}_j(h'|t^*)$$

Because i is in S, this implies that there exists a player k in S such that $\hat{z}_k(h'|t^*) < z_k(h'|t^*)$. We now separate the discussion for cases A and B.

Case A. Since $p(h'|t^*)$ becomes binding at t^* and at time t no one has accepted any proposal because $A(h) = \emptyset$, thus player k must accept $p(h'|t^*)$ at some point of time $t_k \in [t, t^*]$. At that time, since k is playing according to the equilibrium strategy, so $\hat{z}_k(h'|t_k) \geq z_k(h'|t_k)$. Note that $\Pi(h'|t_k) = \Pi(h'|t^*)$, $N(h'|t_k) = N(h'|t^*)$ and $K(h'|t_k) = K(h'|t^*)$ since $p(h'|t_k) = p(h'|t^*)$. Hence $\hat{z}_k(h'|t_k) = \hat{z}_k(h'|t^*)$ and $z_k(h'|t_k) = z_k(h'|t^*)$. This implies that $\hat{z}_k(h'|t^*) \geq z_k(h'|t^*)$, yielding a contradiction. Hence there is no profitable deviation for case A.

Case B. Note that $p(h) = ((w_j)_{j\in T}, T)$ will not bind. This is because if there exists some $j \neq i$ where $j \in T\backslash A(h)$ such that $\hat{z}_j(h) < z_j(h)$, then he will not accept the proposal and will make another proposal pertaining to $N(h)$ at the next integer time if no one has done so. If there exists no $j \neq i$ where $j \in T\backslash A(h)$ such that $\hat{z}_j(h) < z_j(h)$, then either $\hat{z}_i(h) = z_i(h)$ or $\hat{z}_i(h) < z_i(h)$. When $\hat{z}_i(h) = z_i(h)$, according to the equilibrium strategies, all $j \neq i$, $j \in T\backslash A(h)$ will accept the proposal at the next integer time. Thus if player i accepts as well, the proposal will bind. However, once it binds, say at time t'', then $p(h'|t'') = \emptyset$. Since we have shown in Case A that no profitable deviation is possible, player i's optimal strategy is to follow the equilibrium strategy f_i from t'' on. This implies i's payoff will be $\hat{z}_i(h) = z_i(h)$ by using f_i'. Since $\hat{z}_i(h) = w_i' P(S'|\mathcal{P}_N')$. This is in contradiction to $w_i' P(S'|\mathcal{P}_N') > z_i(h)$. When $\hat{z}_i(h) < z_i(h)$, there are two possibilities. Either there exists an $j \neq i$ where $j \in T\backslash A(h)$ or $\{i\} = T\backslash A(h)$. In the first situation when there exists a $j \neq i$, $j \in T\backslash A(h)$, then according to player j's equilibrium strategy, he will not accept the proposal[16] and will make another proposal pertaining to $N(h)$ if no on has done so. In the second situation where $\{i\} = T\backslash A(h)$, player i will not accept the proposal. For if he did, say at time t''', then $p(h'|t''') = \emptyset$. Again, since we have shown in Case A that no profitable deviation is possible from t''' on, player i gets $\hat{z}_i(h) < z_i(h)$, yielding a contradiction. Therefore in all possible situations, $p(h) = ((w_j)_{j\in T}, T)$ will not bind.

Since $p(h'|t^*)$ becomes binding at t^* and $p(h) = ((w_j)_{j\in T}, T)$ will not bind, $p(h'|t^*)$ must be proposed at time t or later but before t^*. Hence player k must accept $p(h'|t^*)$ at some point of time $t_k \in [t, t^*]$. Now apply exactly the same logic in case A to get a contradiction. Thus there is no profitable deviation for case B.

Case C. We will show that player i has no profitable deviation. If player i plays according to the equilibrium strategy f_i, since all others are also playing the equilibrium strategies, his payoff is $\hat{z}_i(h) > z_i(h)$. If instead player i deviates to another strategy f_i', there are two possibilities.

In the first possibility $p(h)$ becomes binding. This implies player i accepts $p(h)$ at some time t''. Since all other players accept $p(h)$ by the equilibrium strategies at the next integer time, say t''', this implies $p(h)$ becomes binding at $\max\{t'', t'''\}$. Therefore $p(h'|\max\{t'', t'''\}) = \emptyset$. By the argument in case A, it is optimal for player i to follow the equilibrium strategy from time $\max\{t'', t'''\}$ on. Hence player i's payoff from using f_i' is at most $\hat{z}_i(h)$.

In the second possibility $p(h)$ does not become binding. This implies either player i makes another proposal at some time t'' or i leaves before accepting. In the first situation, since others are playing according to the equilibrium strategies, if the next integer time arrives before than or at t'', all others accept $p(h)$ at the next integer time except player i. If the next integer time arrives after t'', this new proposal is made before anyone has accepted it. Both imply $p(h'|t'') = \emptyset$. By the argument in case A, it is optimal for player i to follow the equilibrium strategy from time t'' on. Hence player i's payoff from using f_i' is at most $z_i(h)$. In the second situation where i leaves before accepting, it must be the case that player i is in a binding coalition S^k. After S^k leaves, all players still play according to the equilibrium strategies. Thus $x(N(h)\backslash S^k|S^k, K(h))$ is expected

[16] Because according to the equilibrium strategy, player j will not accept $p(h)$ since $\hat{z}_i(h) < z_i(h)$.

in the continuing equilibrium. Hence player i's payoff is

$$w_i^k \sum_{j\in S^k} x_j(N(h)\backslash S^k|S^k, K(h)) \le z_i(h)$$

by part (a) of Lemma 1.

Thus there is no profitable deviation for case C. □

6. Discussion

We note that total r-balancedness indeed guarantees the existence of the r-core. In fact, it makes a precise prediction that for any S and $\mathcal{P}_{N\backslash S}$, $\mathcal{P}(S|S, \mathcal{P}_{N\backslash S}) = (S)$. When there is no externality across coalitions, it naturally reduces to the standard notion of total balancedness. In addition, total r-balancedness plays an important role in the proof of Theorem 2. More precisely, the assumption of total r-balancedness is convenient for the following reasons.

First, suppose S_1 has signed a binding agreement. After this history, the relevant partition function game is no longer $\langle N, P\rangle$. Since S_1 cannot break apart, we should treat S_1 as a " composite player" and consider the "derived" partition function consistent with the fact that S_1 stays together. According to Theorem 1, after S_1 forms, in the continuation equilibrium some r-core outcome of the " derived" partition function game must occur. The non-emptyness of the r-core for this "derived" partition function game is not guaranteed, however. Indeed, the non-emptyness of the r-core in the original game only implies that when every player is *on its own*, the r-core for any reduced society is non-empty. For totally r-balanced games, since the grand coalition for any reduced society always forms, treating S_1 as a" composite player" actually reduces the number of the inequalities to check for the r-core to be non-empty. For instance, suppose $S_1 = \{1, 2\}$ and $N = \{1, 2, 3, 4\}$. When we treat S_1 as a " composite player," for the grand coalition N to form in the " derived" partition function form game, we do not need to worry, for instance, whether the sum of payoffs of players 1 and 3 is greater than their worth. Therefore, if the r-core is non-empty in the original game, then it must be non-empty in this " derived" partition function game. The grand coalition is always the resulting r-core structure. But if the game is not totally r-balanced, then existence may be a problem. Example 1 demonstrates this.

Example 1. Suppose $N = \{1, 2, 3, 4\}$. Suppose

$$P(\{1,2,4\}|(\{1,2,4\},\{3\})) = P(\{1,2,3\}|(\{1,2,3\},\{4\})) = P(\{3,4\}|(\{1,2\},\{3,4\})) = 2,$$

$$P(\{1\}|(\{1\},\{2\},\{3,4\})) = P(\{2\}|(\{1\},\{2\},\{3,4\})) = P(\{3,4\}|(\{1\},\{2\},\{3,4\})) = 1$$

and $P(S|\mathcal{P}_N) = 0$ in all other cases. For this game

$$V(\{1,2,3\}) = V(\{1,2,4\}) = 2, \quad V(\{1\}) = V(\{2\}) = V(\{3,4\}) = 1$$

and the values for all other coalitions are zero. The r-core is non-empty and the unique r-core structure for the society N is $(\{1\}, \{2\}, \{3, 4\})$. Now subgame perfection requires that a continuation equilibrium exists after any possible history. So suppose coalition $\{1, 2\}$ has formed. Treating $\{1, 2\}$ as a composite player, there exists no outcome in the r-core of the " derived" partition function form game. This is because now the value of $\{3, 4\}$ becomes 2, since players 1 and 2 cannot break apart. Combined with the fact that the values of $\{1, 2, 3\}$ and $\{1, 2, 4\}$ are 2, there is no way to give each coalition its worth. Therefore, after $\{1, 2\}$ has formed, no continuation equilibrium exists.

Second, suppose the r-core structure is not the grand coalition. After some coalitions in the r-core structure have formed, since we have to treat these coalitions as " composite players," it is not guaranteed that the r-core in the "derived" partition function form game agrees with the original r-core. Example 2 demonstrates this. (For the same reason as argued above, when the game is totally r-balanced, the problem does not occur.)

Example 2. Suppose $N = \{1, 2, 3, 4\}$. Suppose

$$P(\{1,2\}|(\{1,2\},\{3\},\{4\})) = 2, \quad P(\{1\}|(\{1\},\{2\},\{3,4\})) = P(\{2\}|(\{1\},\{2\},\{3,4\}))$$
$$= P(\{3,4\}|(\{1\},\{2\},\{3,4\})) = P(\{3\}|(\{1,2\},\{3\},\{4\})) = P(\{4\}|(\{1,2\},\{3\},\{4\})) = 1,$$
$$P(\{3,4\}|(\{1,2\},\{3,4\})) = 3$$

and $P(S|\mathcal{P}_N) = 0$ in all other cases. For this game

$$V(\{3,4\}) = V(\{1\}) = V(\{2\}) = V(\{3\}) = V(\{4\}) = 1$$

and the values for all other coalitions are zero. The r-core is non-empty and the unique r-core structure for the society N is $(\{1, 2\}, \{3\}, \{4\})$. However, suppose $\{1, 2\}$ has formed. Treating $\{1, 2\}$ as a composite player, the unique r-core structure in the "derived" partition function form game is $(\{1, 2\}, \{3, 4\})$. This is because now the value of $\{3, 4\}$ becomes 3 since players 1 and 2 cannot break apart.

Acknowledgements

The comments of two anonymous referees have greatly improved the exposition. We also wish to thank seminar participants at the 2003 European Meeting of the Econometric Society for helpful comments. Financial support from the NSC Grant No. 90-2415-H-002-029 is gratefully acknowledged.

References

Bloch, F., 1995. Endogenous structures of association in oligopolies. Rand Journal of Economics 26, 537–556.

Chatterjee, K., Dutta, B., Ray, D., Sengupta, K., 1993. A non-cooperative theory of coalitional bargaining. Review of Economic Studies 60, 463–477.

Deneckere, R., Davidson, C., 1985. Incentives to form coalitions with Bertrand competition. Rand Journal of Economics 16, 473–486.

Huang, C.-Y., Sjöström, T., 2003. Consistent solutions for cooperative games with externalities. Games and Economic Behavior 43, 196–213.

Kalai, E., Postlewaite, A., Roberts, J., 1979. A group incentive compatible mechanism yielding core allocations. Journal of Economic Theory 20, 13–22.

Koczy, L., 2003. The Core in Normal Form Games. University of Maastricht, Mimeo.

Maskin, E., 2003. Bargaining, Coalitions and Externalities, Mimeo, Princeton.

Moldovanu, B., Winter, E., 1994. Core implementation and increasing returns to scale for cooperation. Journal of Mathematical Economics 23, 533–548.

Moldovanu, B., Winter, E., 1995. Order independent equilibria. Games and Economic Behavior 9, 21–34.

Peleg, B., 1986. On the reduced game property and its converse. International Journal of Game Theory 15, 187–200.

Perry, M., Reny, P., 1994. A non-cooperative view of coalition formation and the core. Econometrica 62, 795–817.

Ray, D., 1989. Credible coalitions and the core. International Journal of Game Theory 18, 185–187.

Ray, D., Vohra, R., 1997. Equilibrium binding agreements. Journal of Economic Theory 73, 30–78.

Serrano, R., Vohra, R., 1997. Non-cooperative implementation of the core. Social Choice and Welfare 14, 513–525.

Tadenuma, K., 1992. Reduced games, consistency, and the core. International Journal of Game Theory 20, 325–334.

Yi, S.-S., 1997. Stable coalition structures with externalities. Games and Economic Behavior 20, 201–237.