Games **2010**, *1*, 66-88; doi:10.3390/g1020066

OPEN ACCESS

games

ISSN 2073-4336

www.mdpi.com/journal/games

Article

The Recursive Core for Non-Superadditive Games

Chen-Ying Huang [1], **Tomas Sjöström** [2,⋆]

[1] Department of Economics, National Taiwan University, 21 Hsu-Chow Road, Taipei, Taiwan

[2] Department of Economics, Rutgers University, New Brunswick, N.J. 08901, USA

⋆ Author to whom correspondence should be addressed; E-Mail: tsjostrom@economics.rutgers.edu; Tel.: +1-732-932-7362; Fax: +1-732-932-7416.

Received: 23 January 2010; in revised form: 2 April 2010 / Accepted: 8 April 2010 / Published: 15 April 2010

Abstract: We study the recursive core introduced in Huang and Sjöström [8]. In general partition function form games, the recursive core coalition structure may be either coarser or finer than the one that maximizes the social surplus. Moreover, the recursive core structure is typically different from the one predicted by the α-core. We fully implement the recursive core for general games, including non-superadditive games where the grand coalition does not form in equilibrium. We do not put any restrictions, such as stationarity, on strategies.

Keywords: coalition formation; non-cooperative implementation; partition function; recursive core

1. Introduction

Characteristic functions are used to study games without externalities across coalitions. If coalitions have unhampered ability to sign binding agreements, then a natural solution concept for games in characteristic function form is the *core*. However, in many applications there are externalities across coalitions, and the characteristic function is replaced by a *partition function*. Several generalizations of the core have been proposed for partition function form games. The well-known α-theory is based on incredible threats: the members of any coalition S assume that if S forms, then the outsiders will try to hurt the members of S as much as they can, without regard to their own payoffs. Huang and Sjöström [8] introduced an alternative theory, called the r-theory, which rules out incredible threats. According to the r-theory, the members of any coalition S assume that if S forms, then the outsiders in $N\backslash S$ behave in

their own best interest. This yields a *recursive* solution concept called the *recursive core* or, more briefly, *r-core*. According to this recursive solution concept, the reaction of the outsiders must be consistent with the notion of the core itself, applied to the reduced society $N\backslash S$.

In Huang and Sjöström [8], we studied the recursive core for partition function form games that are derived from *normal form games*. In such games, the grand coalition consisting of all agents will maximize the total social surplus; no finer coalition structure can do better in the aggregate. For such games, since binding agreements and side-payments are assumed to be possible, if the recursive core is non-empty then it must predict that the grand coalition forms. In Huang and Sjöström [9], we implemented the recursive core for partition function form games which are not necessarily derived from normal form games. However, assumptions were made to again ensure that the recursive core (if non-empty) predicts that the grand coalition will form. Moreover, this outcome was efficient, in the sense of maximizing the social surplus. The α-core (if non-empty) also predicts the grand coalition forms in this case. Therefore, in our previous work, Huang and Sjöström [8,9], equilibrium coalition formation was fairly trivial, and the value of using the recursive core to make predictions about coalition formation was obscured.

The classical literature on cooperative games often took for granted that the grand coalition would be efficient (*i.e.*, maximize total social surplus). This was justified by the reasoning that the grand coalition could always replicate what smaller coalitions could do separately - so a merger would not reduce the available surplus. However, Aumann and Dreze [1] pointed out that this argument is not always valid: "the very act of "acting together" may be difficult, costly, or illegal, or the players may, for various "personal" reasons, not wish to do so" (Aumann and Dreze [1], p. 233). Forming a large coalition may even be illegal (e.g., due to anti-trust laws). Aumann and Dreze [1] argued in favor of incorporating costs of forming coalitions into the characteristic function; the value of an "illegal" coalition might even be minus infinity. In many applications, it is indeed natural to allow for the possibility that forming a large coalition is costly. In such cases, even computing an efficient coalition structure can be a difficult task if the number of players is large. Computer scientists have developed algorithms for this (see Sandholm et. al. [17]). We are however not interested in computational but conceptual issues. Specifically, we will not take for granted that the coalition structure must be efficient. Indeed, we will find in Section 3 that the recursive core structure often does not maximizes the total surplus.

In Section 3 we will discuss an example drawn from politics, where forming certain coalitions may be costly. The players are three parties, left, middle and right; any two (or all three) can form a coalition government. We would like a theory to predict which government forms. The political economy literature frequently assumes that a left-wing party may suffer a cost from cooperating with, or even talking to, a right-wing party (see Bandyopadhyay *et al.* [2]). The cost may be a loss of support from its base, if it is seen as "selling out" by compromising its ideological position. Moreover, it may not be possible to completely specify in advance the policies a coalition government would implement; these will arise out of some future bargaining process. But the implemented policies may be expected to differ greatly from the "ideal outcome" of the left party, causing it to suffer large "compromise costs". If times are tough, any member of a governing coalition may have to share responsibility for austere economic policies; even a left-middle coalition government may have to cut government spending. Thus, the left party may suffer significant costs from joining a middle-left coalition, and even greater costs from cooperating with the right-wing party. A grand coalition government involving all three parties may, due to compromise

costs, generate significantly smaller surplus than a middle-right coalition would. We believe the best way to formally model this kind of situation is by a partition function where forming a large coalition may be costly, hence a large coalition may not be able to replicate what smaller coalitions can do. Formally, we do not want to assume superadditivity.

In strictly superadditive games with non-empty recursive cores, the grand coalition must form. But motivated by examples such as the one in Section 3 , we want to study more general partition functions. Thus, in this article we will not assume superadditivity. In non-superadditive games the grand coalition need not form, and the coalition structure predicted by the recursive core may be either coarser or finer than the efficient coalition structure which maximizes social surplus. Moreover, the recursive core coalition structure is typically different from the one predicted by the α-core. Indeed, we will argue that the recursive core prediction is more intuitively appealing. This paper, therefore, has two motivations. First, we want to argue that the recursive core predicts intuitively appealing equilibrium coalition structures for non-superadditive games.[1] Second, we want to find a non-cooperative implementation of the recursive core for general partition function form game, including non-superadditive ones, which does not make any restrictions on strategies. For example, we do not want to assume stationarity.

The characteristic function approach assumes there are no externalities across coalitions, so the value of a coalition can be defined without specifying the behavior of outsiders. Accordingly, to non-cooperatively implement the core for characteristic function form games, one need not worry too much about how outsiders react to the formation of a coalition.[2] However, when externalities exist, the reaction of outsiders becomes crucial. As it happens, the literature on games without externalities contains an extensive form game, Perry and Reny [15], which does allow outsiders to react to the formation of a coalition in a reasonable way. Specifically, if a group of players signs a binding contract to form a coalition, then the group becomes a "composite player" which cannot break apart, but can join other players in a larger coalition. Except for that, the game continues with the same rules as before. In subgame perfect equilibrium, the reaction of outsiders will, by definition, be consistent with their own best interests. This is precisely the idea behind the recursive core. In Huang and Sjöström [9], we showed that a slightly modified version of Perry and Reny's [15] game implements the recursive core for strictly superadditive partition function form games.[3] However, unlike in the current article, it was necessary to restrict attention to stationary strategies.

It turns out that a game of the Perry-Reny type cannot implement the recursive core if superadditivity is violated. In the Perry-Reny game, coalitions form sequentially. After a coalition S has formed, this coalition will behave as a "composite player". This causes two problems if the game is not superadditive. The first problem is that if coalition S forms in order to block a proposed allocation, then there may

[1] Laboratory experiments could be used to test these predictions. For games without externalities, Yan and Friedman [20] found that considerations of "fairness" frequently cause deviations from core outcomes. However, fairness issues are orthogonal to the distinction between the α-core and the recursive core. We believe experiments will show that, after filtering out fairness considerations which equally affect both concepts, the recursive core has more predictive power than the α-core.

[2] For implementation of the core correspondence for cooperative games without externalities, see Kalai, Postlewaite and Roberts [10], Chatterjee *et al.* [4], Moldovanu and Winter [13,14], Serrano and Vohra [18] and Perry and Reny [15].

[3] The modification is required because with externalities, a proposal to form a coalition cannot specify a distribution of payoffs (because the final payoffs cannot be determined until outsiders have reacted). Instead, a proposal must specify an allocation of *shares* of the coalition's final payoff, whatever it turns out to be.

not exist any continuation equilibrium of the ensuing subgame (even if the recursive core of the original game where S has not formed is non-empty). Therefore, implementation in subgame perfect equilibrium fails. Intuitively, the formation of coalition S changes the incentives of the remaining players, and the recursive core of the original game (without the "composite player") cannot predict the outcome with a composite player (see Example 1 in Huang and Sjöström [9]). The second problem is that without superadditivity, the recursive core may predict that the grand coalition does not form. Instead, several smaller coalitions must form in equilibrium. But again, after some coalition S has formed in the Perry-Reny game (where coalitions form sequentially), the recursive core structure of the "derived" partition function form game (where S cannot break apart) may not be consistent with the recursive core of the original game. Therefore, even if the continuation equilibrium of the non-cooperative game implements the recursive core of the "derived" game, this may not guarantee implementation of the recursive core of the original game (see Example 2 in Huang and Sjöström [9]). Intuitively, if coalitions form sequentially as in the Perry-Reny game, one must ensure that after some coalitions have formed, the remaining coalitions form according to the recursive core prediction of the original game. In general, this cannot be achieved, because each coalition which forms changes the incentives for subsequent coalitions to form.

The above discussion suggests that to fully implement the recursive core for non-superadditive games with externalities, the non-cooperative game should have three properties. The first and most basic property is "self-similarity": if a blocking coalition forms, then the outsiders play a game which is a scaled-down version of the original game. This gives the blocking coalition an expected payoff which is consistent with the recursive core. Second, if a coalition forms in order to block an allocation, then in order to avoid the first problem discussed in the previous paragraph, its members cannot be allowed to join a subsequent, larger coalition. This ensures that the continuation equilibrium implements the recursive core among the remaining players. Third, in equilibrium all coalitions should form simultaneously, in order to avoid the second problem discussed in the previous paragraph. In this paper, we will show that a game satisfying these three properties fully implements the recursive core. Thus, we achieve full non-cooperative implementation of the recursive core for general games, including games where the grand coalition does not form in equilibrium.

Our starting point is Serrano and Vohra's [18] non-cooperative game. Their game has two stages: a coalition formation stage and a blocking stage. In the coalition formation stage, all players simultaneously propose an allocation and a permutation of the player set. If all players propose the same allocation, then this allocation becomes the "status quo", and the game proceeds to the blocking stage. In the blocking stage, a proposer is selected according to the composition of the proposed permutations by all players in the coalition formation stage. The proposer may name a blocking coalition (which must include himself). If he does, all other members of the blocking coalition are asked to respond. If they all say yes, then the blocking coalition forms. Serrano and Vohra [18] showed that their game fully implements the core correspondence. Indeed, if any player is not happy with the status quo, he can attain the right to propose a blocking coalition by unilaterally changing his announced permutation in the coalition formation stage. This elegant construction, due to Thomson [19], does not rely on any fixed protocol of proposals and responses, but instead allows unrestricted coalitional blocking. Thus, it captures the spirit of free and unhampered competition which underlies the core.

For our purposes, we need to modify one aspect of the Serrano and Vohra [18] game. They assumed outsiders cannot react at all to the formation of a coalition. This was not an important assumption for them, since they did not allow externalities across coalitions. With externalities, the reactions of outsiders become crucial. According to the recursive core, the outsiders will react in a way which is consistent with the notion of the core itself. Therefore, we modify the Serrano-Vohra [18] game so that it satisfies self-similarity: if a blocking coalition forms, the outsiders will play a scaled-down version of the original game among themselves. A lack of superadditivity is not a problem for the self-similar Serrano-Vohra [18] game. First, once a blocking coalition has formed, we will not allow its members to join a larger coalition (only the remaining active players will play a scaled-down version of the game), which avoids the first problem mentioned above. Second, the game starts with all players simultaneously proposing a way to distribute payoffs and a coalition structure, which in the non-superadditive case may be a non-trivial partitioning of N. Since all coalitions form simultaneously, the second problem mentioned above is avoided.

The recursive core does not always exist. Koczy [11,12] proposed "optimistic" and "pessimistic" recursive cores which mitigate the non-existence problem. We will discuss existence and Koczy's work further in Section 2. The spirit of Koczy's work is the same as ours: extend the definition of the core to allow for externalities across coalitions, and then find a non-cooperative game to implement this solution concept. In particular, we follow the traditional theory of the core by assuming every coalition has unhampered ability to sign a binding agreement to block the status quo. In a recent paper, Zheng [21] takes a fresh new look at this issue. In his theory, if a coalition S signs a binding agreement to cooperate, the outsiders in $N\backslash S$ will react according to the core defined on the reduced society $N\backslash S$; this part is just as in our theory. What is new in his approach is that a coalition which tries to sign a binding agreement does not necessarily succeed: with some probability the outsiders may coordinate a preemptive response. Thus, attempting to form a coalition is a risky activity - it may be challenged and preempted. Zheng's version of the core allows interesting existence results. In an externality problem where each player can either "pollute" or "not pollute", Zheng's core is always nonempty and the outcome is efficient.

A different strand of literature takes a different approach. It does not directly try to generalize the classic core notion to games with externalities. Instead, it incorporates natural coalition formation dynamics into new solution concepts. The most well known such concept is Ray and Vohra's [16] Equilibrium Binding Agreement (EBA). Ray and Vohra [16] originally assumed coalitions break apart but do not re-merge. Diamantoudi and Xue's [5] extended EBA (EEBA) allows coalitions to do both. Funaki and Yamato [7] develop a related solution concept called sequentially stable coalition structures. In their theory, coalition structures form sequentially. If there is a sequence of step-by step changes, where in each step the involved coalition members are better off with the final outcome if they proceed with the change, then the starting coalition structure is dominated by the final coalition structure. A coalition structure which dominates all other coalition structures is said to be sequentially stable. Like Diamantoudi and Xue [5], Funaki and Yamato [7] allow coalitions to merge as well as break apart. In a common pool game (previously studied in Funaki and Yamato [6]), the efficient grand coalition can be sequentially stable.

2. Definitions

Let $N = \{1, 2, ..., n\}$ be the set of players. A coalition is a non-empty subset of N. A coalition structure $\mathcal{P}_N$ is a partition of N. A transferable utility game in partition function form is denoted $< N, P >$, where P is the partition function. The partition function is the natural way to model externalities across coalitions. For any coalition structure $\mathcal{P}_N$ and any coalition $S \in \mathcal{P}_N$, let $P(S \mid \mathcal{P}_N)$ denote the value (or worth) of S when players partition themselves according to $\mathcal{P}_N$. Thus, the worth of S can depend on the coalitional structure formed by the "outside" players in $N \backslash S$ (the set of players who belong to N but not to S). To simplify the exposition, we assume

$$P(S \mid \mathcal{P}_N) > 0 \tag{1}$$

for all $\mathcal{P}_N$ and all $S \in \mathcal{P}_N$. For any payoff vector $x \equiv (x_i)_{i \in N}$ and any coalition $S \subseteq N$, define the sum of payoffs across players in S as

$$x(S) \equiv \sum\nolimits_{i \in S} x_i.$$

Given a coalition structure $\mathcal{P}_N$, a payoff vector $x \in \mathbb{R}^n$ is said to be *efficient under the partition* $\mathcal{P}_N$ if for every $S \in \mathcal{P}_N$, $x(S) = P(S \mid \mathcal{P}_N)$.

If $S \in \mathcal{P}_N$, then we have $\mathcal{P}_N = (S, A_1, A_2, ..., A_k)$ for some coalitions $A_1, A_2, ..., A_k$. Notice that $\mathcal{P}_{N \backslash S} \equiv (A_1, A_2, ..., A_k)$ 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 $P(S \mid S, \mathcal{P}_{N \backslash S}) \equiv P(S \mid \mathcal{P}_N)$. A partition $\mathcal{P}_S$ of a coalition S is said to be *non-trivial* if $\mathcal{P}_S \neq (S)$, *i.e.*, if it contains at least two non-empty subcoalitions.

The recursive core (r-core) is a solution concept for partition function form games (Huang and Sjöström [8,9]). To calculate the value of coalition S in a partition function form game, we need to predict how the outsiders (*i.e.*, the players in $N \backslash S$) will react if S forms. The recursive core theory makes the following recursive prediction: the outsiders will behave in a way which is consistent with the recursive core of the natural "reduced game". In the reduced game, it is taken as given that S has formed, because agreements to form a coalition are binding. But the outsiders still have to decide which coalitions to form. Recursively applying the idea of the core, we assume each coalition of outsiders will insist on getting "what it is worth".

Before giving the formal definition, let us briefly discuss the notation. Suppose a partition $\mathcal{P}_N$ is given, and coalition $S \in \mathcal{P}_N$ is part of this coalition structure. Thus, we can write $\mathcal{P}_N = \left(S, \mathcal{P}_{N \backslash S}\right)$. If the members of S think the "outsiders" will partition themselves according to $\mathcal{P}_{N \backslash S}$, what would the members of S themselves want to do? For $S \in \mathcal{P}_N$, we would like to make a prediction of payoff vectors that, in some sense, would be acceptable to the members of S if the outsiders partition themselves according to $\mathcal{P}_{N \backslash S}$. We will let $C\left(S \mid \mathcal{P}_N\right)$ denote the set of such "acceptable" payoff vectors.[4] Notice that we are not requiring that S itself sticks together. Instead, $C\left(S \mid \mathcal{P}_N\right) = C\left(S \mid S, \mathcal{P}_{N \backslash S}\right)$ denotes the payoff vectors that would be acceptable to members of $S \in \mathcal{P}_N$, allowing for the possibility that S itself may break apart, but assuming the outsiders stick to the partition $\mathcal{P}_{N \backslash S}$. Thus, notation $C\left(S \mid \mathcal{P}_N\right)$ means S is singled out for special consideration, and treated differently than other coalitions in $\mathcal{P}_N$.

[4] Payoff vectors are points in $\mathbb{R}^n$. Suppose S is a strict subset on N, and so has fewer than n members. The members of S care only about their own payoffs, not the payoffs of outsiders. Thus, when asking whether a payoff vector $x \in \mathbb{R}^n$ is acceptable to the members of S, the payoffs x assigns to "outsiders" is a matter of indifference to the members of S.

We now give the formal definition. The recursive core for coalition $S \subseteq N$, given a partition $\mathcal{P}_{N\backslash S}$ of the outsiders, is denoted $C\left(S \mid S, \mathcal{P}_{N\backslash S}\right)$. It is defined as follows. For a single-player society $S = \{i\}$, we define $C\left(\{i\} \mid \{i\}, \mathcal{P}_{N\backslash\{i\}}\right)$ to be the set of payoff vectors that are efficient under $\left(\{i\}, \mathcal{P}_{N\backslash\{i\}}\right)$. Proceeding recursively, suppose the recursive core has been defined for all coalitions with at most $s-1$ 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 the value (or worth) of T given $\mathcal{P}_{N\backslash S}$ as

$$V\left(T \mid S, \mathcal{P}_{N\backslash S}\right) \equiv \begin{cases} P(S \mid S, \mathcal{P}_{N\backslash S}) & \text{if } T = S \\ \min\left\{x(T) : x \in C\left(S\backslash T \mid S\backslash T, T, \mathcal{P}_{N\backslash S}\right)\right\} & \text{if } T \neq S \end{cases}. \tag{2}$$

Note that when $T \neq S$, to calculate the value (or worth) of T we assume the outsiders in $S\backslash T$ will play according to the recursive core of the reduced society given T, while the players in $N\backslash S$ have arranged themselves according to $\mathcal{P}_{N\backslash S}$. This is where we recursively apply the concept of the core. Now, $x \in C\left(S \mid S, \mathcal{P}_{N\backslash S}\right)$ if and only if there exists some partition $\mathcal{P}_S$ of S such that x is efficient under the partition $\left(\mathcal{P}_S, \mathcal{P}_{N\backslash S}\right)$, and $x(T) \geq V\left(T \mid S, \mathcal{P}_{N\backslash S}\right)$ for each coalition $T \subseteq S$. This completes the definition.

Notice that according to the definition of the recursive core, to any $x \in C(S \mid S, \mathcal{P}_{N\backslash S})$ there corresponds a partition $\mathcal{P}_S$ of S. Such $\mathcal{P}_S$ is referred to as a *recursive core partition* of S given $\mathcal{P}_{N\backslash S}$. This partition may not be unique. Let $\mathcal{P}(S \mid S, \mathcal{P}_{N\backslash S})$ denote the set of all recursive core partitions of S given $\mathcal{P}_{N\backslash S}$. That is, $\mathcal{P}_S \in \mathcal{P}(S \mid S, \mathcal{P}_{N\backslash S})$ if and only if there is an $x \in C(S \mid S, \mathcal{P}_{N\backslash S})$ which is efficient under the partition $(\mathcal{P}_S, \mathcal{P}_{N\backslash S})$. Thus, the recursive 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. Notice that since we make no assumption of superadditivity, the members of coalition S may well prefer to break apart by mutual agreement, in which case $(S) \notin \mathcal{P}(S \mid S, \mathcal{P}_{N\backslash S})$.

We have defined $C(S \mid S, \mathcal{P}_{N\backslash S})$ for any $S \subseteq N$ and any partition $\mathcal{P}_{N\backslash S}$ of $N\backslash S$. However, we are naturally interested in the case $S = N$. If $S = N$, then $N\backslash S = \emptyset$ so we obtain

$$C(S \mid S, \mathcal{P}_{N\backslash S}) = C(N \mid N)$$

In fact, it is useful to simplify the notation even further by writing $C(N)$ instead of $C(N \mid N)$. Similarly, we write $\mathcal{P}(N)$ instead of $\mathcal{P}(N \mid N)$ and $V(T)$ instead of $V(T \mid N)$. By a slight abuse of terminology, $C(N)$ is *the recursive core of* $< N, P >$, $\mathcal{P}(N)$ is the set of *recursive core partitions*, and $V(T)$ is the *worth of coalition* $T \subseteq N$. Notice that $x \in C(N)$ if and only if there is some partition $\mathcal{P}_N$ of N such that x is efficient under the partition $\mathcal{P}_N$, and $x(T) \geq V(T)$ for each coalition $T \subseteq N$.

We are interested in games where the recursive core exists and is non-empty, *i.e.*, $C(N) \neq \emptyset$. We will argue in Section 3 that when the recursive core is non-empty, it makes intuitively reasonable predictions.

For the recursive core of $< N, P >$ to be non-empty, it must also be non-empty for every reduced society. Otherwise, it is not possible to make consistent predictions about what will happen when coalitions break up. Formally, since we recursively apply the concept of the core from smaller societies up, if $C(N) \neq \emptyset$ then $C(S \mid S, \mathcal{P}_{N\backslash S}) \neq \emptyset$ for any nonempty $S \subseteq N$ and any partition $\mathcal{P}_{N\backslash S}$ of $N\backslash S$. However, the requirement that the recursive core must be non-empty for all reduced societies is quite strong, and may cause non-existence. Huang and Sjöström (pp. 202-203) [8] discussed the following

example, suggested by D. Ray. Suppose $N = \{1, 2, 3, 4\}$, and consider coalition $S = \{2, 3, 4\}$. Suppose the recursive method just described has been used to calculate the worth of coalition S, when player 1 stands alone, as well as the worth of all the subcoalitions of S. Suppose the calculated worths are as follows: $V(S \mid S, \{1\}) = 6$, and $V(\{j, k\} \mid S,\ \{1\}) = 5$ for any two-player subcoalition $\{j, k\} \subset S$, and $V(\{j\} \mid S,\ \{1\}) = 0$ for any singleton $j \in S$. In this example, if player 1 stands alone it is efficient for S to form and share 6. There is no way for S to break up and get more than 6. But there is no way to allocate 6 among players 2,3 and 4 so as to satisfy every two-player subcoalition of S, because each two-player subcoalition $\{j, k\}$ demands 5. This implies that the recursive core of the reduced society $N\backslash\{1\} = S$ is empty, $C(S \mid S, \{1\}) = \emptyset$, so we cannot predict what the other players would do if player 1 refuses to cooperate with them. Therefore, we cannot use the recursive method to calculate the worth $V(\{1\})$ of player 1. Therefore, the recursive core $C(N)$ does not exist. However, suppose it happens that $P(\{1\} \mid \{1\}, \mathcal{P}_S) = 5$ for any partition $\mathcal{P}_S$ of S. Even if we cannot predict what the other players will do if player 1 stands alone, *whatever* the other players do player 1 will get 5. It seems reasonable then that he should be worth 5; yet $V(\{1\})$ cannot be calculated by our recursive method.

Huang and Sjöström [8] left it as a topic for future research to develop solution concepts which would not require non-empty cores of all reduced societies. Koczy [11,12] in fact develops several such solution concepts. When a coalition cannot make any prediction about the behavior of the outsiders, they might become pessimistic and expect the worst (as in the α-core); or they might be optimistic and expect the best. Thus, Koczy [11,12] obtains a "pessimistic" and an "optimistic" recursive core. In the example of the previous paragraph, both versions would assign the worth 5 to player 1. This is the major difference between the recursive core in Huang and Sjöström [8,9] and the one in Koczy [11,12]. There is another difference: in Koczy's theory, a coalition S blocks by proposing a partition on S, while in our theory, S blocks by signing a binding agreement to cooperate. This however seems to be a rather technical point, and it is hard to argue that one assumption dominates the other.

In this article, we will maintain the original definition of Huang and Sjöström [8,9]. Thus, the core must be non-empty for every reduced society. Huang and Sjöström [8] noted that for a non-cooperative implementation, the requirement of non-empty cores of reduced societies corresponds to the requirement that an equilibrium exists in every subgame. In the particular game we develop below, if the recursive core is empty for some subgame, there would be no continuation equilibrium for that subgame. Thus, although the requirement seems strong, it does seem to have a natural interpretation in the non-cooperative setting. Koczy [12] (Theorem 8) implements the "pessimistic" recursive core in stationary, order independent equilibria of a modified version of Bloch's [3] game. He does not assume superadditivity, but he needs an extra assumption: the core is non-empty for all reduced societies. Our new implementation result will not require any assumptions about stationarity or order independence. In addition, our notion of "recursive core" is somewhat different from Koczy's [12], as explained in the previous paragraph.

We want to emphasize that $C(N) \neq \emptyset$ does not mean, in general, that the grand coalition must form. The recursive core may predict that the grand coalition breaks apart, leading to some finer partition. For example, suppose $N = \{1, 2\}$, $V(N) \equiv P(N \mid N) = 1$, and $V(\{i\}) = P(\{i\} \mid \{1\}, \{2\}) = 2$ for each $i \in \{1, 2\}$. In this example, the grand coalition $\{1, 2\}$ cannot be expected to form; it would produce a total surplus of only 1, but if it breaks apart then the players each get 2. Obviously, the recursive core partition structure for N is $\mathcal{P}(N) = (\{1\}, \{2\})$. That is, singleton coalitions form. The payoffs are also

uniquely determined, since obviously neither player can get anything else than 2. Indeed, the recursive core is $C(N) = (2, 2)$. Huang and Sjöström [9] ruled out this kind of example by focusing on totally r-balanced games, where (by definition) the recursive core prediction is that no coalition ever breaks up. By definition, the game $< N, P >$ is *totally r-balanced* if for any nonempty $S \subseteq N$ and any partition $\mathcal{P}_{N\setminus S}$ of the players not in S, it holds that $\mathcal{P}(S \mid S, \mathcal{P}_{N\setminus S}) = (S)$.

In the current paper, we will consider equilibrium coalition formation in games which are *not* totally r-balanced. Since total r-balancedness is an unfamiliar concept, we will clarify its connection to a more familiar idea, the concept of superadditivity. By definition, the game $< N, P >$ is *strictly superadditive* if for any two disjoint nonempty coalitions S and T and any coalitional structure $\mathcal{P}_{N\setminus(S\cup T)}$ on the remaining players,

$$P(S \mid S, T, \mathcal{P}_{N\setminus(S\cup T)}) + P(T \mid S, T, \mathcal{P}_{N\setminus(S\cup T)}) < P(S \cup T \mid S \cup T, \mathcal{P}_{N\setminus(S\cup T)}). \tag{3}$$

The connection between total r-balancedness and superadditivity is shown in the following proposition.

Proposition 1 If the game $< N, P >$ is strictly superadditive and the recursive core is non-empty, then $< N, P >$ is totally r-balanced.

Proof. The proof is by induction on the number of agents in S, denoted $|S|$. The induction hypothesis is the following.

Hypothesis $H(s)$. For any coalition S such that $1 \leq |S| \leq s$, the following two conditions both hold: (C1) For any partition $\mathcal{P}_{N\setminus S}$ of the complement of S,

$$\mathcal{P}(S \mid S, \mathcal{P}_{N\setminus S}) = (S). \tag{C1}$$

(C2) For any partition $\mathcal{P}_{N\setminus S}$ of the complement of S, and any *non-trivial* partition $\mathcal{P}_S$ of S,

$$\sum_{T_i \in \mathcal{P}_S} P(T_i \mid \mathcal{P}_S, \mathcal{P}_{N\setminus S}) < P(S \mid S, \mathcal{P}_{N\setminus S}). \tag{C2}$$

Condition C1 states that, according to the recursive core, coalition S will not break apart. Condition C2 states that if S did break apart, and the ex-members partitioned themselves according to $\mathcal{P}_S$, their total payoff would be strictly less than what they get by sticking together in S. Thus, by sticking together, the members of S maximize their total surplus.

A singleton coalition cannot possibly break apart, so $H(1)$ holds trivially. For the inductive step, we need to show that for any $s \geq 2$, if $H(s-1)$ holds then $H(s)$ holds as well.

Assume the induction hypothesis $H(s-1)$, and consider S such that $|S| = s$. We need to verify conditions C1 and C2.

By hypothesis $C(N) \neq \emptyset$, so $C(S \mid S, \mathcal{P}_{N\setminus S}) \neq \emptyset$ for any S and any $\mathcal{P}_{N\setminus S}$. Consider any nontrivial partition $\mathcal{P}_S$ of S. We claim $\mathcal{P}_S \notin \mathcal{P}(S \mid S, \mathcal{P}_{N\setminus S})$. That is, we claim it is not a recursive core structure for S to break apart and partition themselves according to $\mathcal{P}_S$. There are two cases, which we consider separately: either there are at least three non-empty coalitions in $\mathcal{P}_S$ or there are only two.

Case 1: Suppose there are at least three nonempty coalitions in $\mathcal{P}_S$. Then we can write $\mathcal{P}_S = (T_1, T_2, ..., T_k)$, where $k \geq 3$. If S breaks apart by formation of a subcoalition $T_j \subset S$, then we can apply the induction hypothesis to $S\backslash T_j$. By C1,

$$\mathcal{P}(S\backslash T_j \mid S\backslash T_j, T_j, \mathcal{P}_{N\backslash S}) = (S\backslash T_j).$$

That is, if T_j leaves, the remaining players in $S\backslash T_j$ will stick together. By C2,

$$\sum_{T_i \in \mathcal{P}_S\backslash\{T_j\}} P(T_i \mid \mathcal{P}_S\backslash\{T_j\}, T_j, \mathcal{P}_{N\backslash S}) =$$
$$\sum_{T_i \in \mathcal{P}_S\backslash\{T_j\}} P(T_i \mid \mathcal{P}_S, \mathcal{P}_{N\backslash S}) < P(S\backslash T_j \mid S\backslash T_j, T_j, \mathcal{P}_{N\backslash S}). \quad (4)$$

Suppose $y \in C(S \mid S, \mathcal{P}_{N\backslash S})$. By definition, this means the players in S can partition themselves in such a way that y is an efficient payoff vector, and y gives each subcoalition of S at least its "worth". By the induction hypothesis, if S breaks apart by formation of subcoalition $S\backslash T_j \subset S$, then the remaining players in T_j will stick together, and the resulting coalition structure will be $(S\backslash T_j, T_j, \mathcal{P}_{N\backslash S})$. Since $y \in C(S \mid S, \mathcal{P}_{N\backslash S})$, it must hold that

$$y(S\backslash T_j) \geq V(S\backslash T_j \mid S, \mathcal{P}_{N\backslash S}) = P(S\backslash T_j \mid S\backslash T_j, T_j, \mathcal{P}_{N\backslash S}). \quad (5)$$

Combining (4) and (5) and summing over all T_j in $\mathcal{P}_S$, we get

$$y(S) > \sum_{T_i \in \mathcal{P}_S} P(T_i \mid \mathcal{P}_S, \mathcal{P}_{N\backslash S}). \quad (6)$$

Since y is an efficient payoff vector for S, the inequality (6) means $\mathcal{P}_S$ is not a recursive core structure: $\mathcal{P}_S \notin \mathcal{P}(S \mid S, \mathcal{P}_{N\backslash S})$.

Case 2: Suppose there are only two nonempty coalitions in $\mathcal{P}_S$. Then we can write $\mathcal{P}_S = (T_1, T_2)$. By strict superadditivity,

$$P(T_1 \mid \mathcal{P}_S, \mathcal{P}_{N\backslash S}) + P(T_2 \mid \mathcal{P}_S, \mathcal{P}_{N\backslash S}) < P(S \mid S, \mathcal{P}_{N\backslash S}). \quad (7)$$

Again, this implies that $\{T_1, T_2\}$ is not a recursive core structure: $(T_1, T_2) \notin \mathcal{P}(S \mid S, \mathcal{P}_{N\backslash S})$, because S can do better.

Thus, for any nontrivial partition $\mathcal{P}_S$ of S, we have established our claim that $\mathcal{P}_S \notin \mathcal{P}(S \mid S, \mathcal{P}_{N\backslash S})$. But $\mathcal{P}(S \mid S, \mathcal{P}_{N\backslash S}) \neq \emptyset$ since $C(S \mid S, \mathcal{P}_{N\backslash S}) \neq \emptyset$. The only possibility is that S sticks together, that is, $\mathcal{P}(S \mid S, \mathcal{P}_{N\backslash S}) = (S)$. So condition C1 holds.

It remains to verify C2, for any nontrivial partition $\mathcal{P}_S$ of S. Since $\mathcal{P}(S \mid S, \mathcal{P}_{N\backslash S}) = (S)$, it holds that

$$y(S) = P(S \mid S, \mathcal{P}_{N\backslash S}).$$

Hence in case 1, the inequality (6) implies

$$P(S \mid S, \mathcal{P}_{N\backslash S}) = y(S) > \sum_{T_i \in \mathcal{P}_S} P(T_i \mid \mathcal{P}_S, \mathcal{P}_{N\backslash S}).$$

Hence, C2 is verified. In case 2, C2 follows from (7).

Thus, we have verified that for any S such that $|S| = s$, conditions C1 and C2 hold. This means $H(s)$ holds, and the induction can be continued until we reach $S = N$. Condition C1 therefore holds for any S of any size, and for any $\mathcal{P}_{N\setminus S}$. This proves the proposition. □

The recursive argument in the proof of Proposition 1 guarantees that the grand coalition cannot increase its surplus by breaking apart. In this case, as in Huang and Sjöström [9], there is a strong incentive for the grand coalition to form. Now, however, we want to study coalition formation in non-superadditive games, where coalitions might prefer to break up by mutual agreement. Recall that $C(N) \neq \emptyset$ does not require the grand coalition to form; the recursive core partition structure may be finer. An example was discussed above (namely, $V(\{1,2\}) = 1$, $V(\{i\}) = 2$ for each $i \in \{1,2\}$). Indeed, the recursive core is frequently non-empty for games which are not superadditive. We claim the recursive core gives interesting insights into equilibrium coalition formation in such situations.

3. Examples

In this section, we will argue that the recursive core makes intuitively appealing predictions for games that are not superadditive. Consider an example with three players, L, M and R, interpreted as three political parties, Left, Middle and Right. No party has a majority of the seats in parliament. Any combination of two or more parties can form a coalition government. If no coalition of size two or more forms, then a *caretaker government* takes over, and each player's payoff is normalized to zero:

$$P(\{i\} \mid \{i\},\{j\},\{k\}) \equiv 0, \text{ for } i \in \{L,M,R\}.$$

The value of a two-party coalition government containing players i and j is

$$V(\{i,j\}) \equiv P(\{i,j\} \mid \{i,j\},\{k\})$$

and the "outsider" gets $O(k) \equiv P(\{k\} \mid \{i,j\},\{k\})$. The value of a singleton, by (2), is

$$V(\{i\}) \equiv \begin{cases} 0 & \text{if} \quad P(\{j,k\} \mid \{j,k\},\{i\}) < 0\,, \quad j,k \neq i \\ \min\{0, O(i)\} & \text{if} \quad P(\{j,k\} \mid \{j,k\},\{i\}) = 0\,, \quad j,k \neq i \\ O(i) & \text{if} \quad P(\{j,k\} \mid \{j,k\},\{i\}) > 0\,, \quad j,k \neq i \end{cases}$$

The value of a grand coalition government containing all three parties is

$$V(\{L,M,R\}) \equiv P(\{L,M,R\} \mid \{L,M,R\})$$

The following is easy to check.

Claim 2 *(i)* $(\{i,j\},\{k\})$ *is a recursive core structure if and only if*

$$V(\{i,j\}) \geq \max\{V(\{i\}), V(\{i,k\}) - O(k)\} + \max\{V(\{j\}), V(\{j,k\}) - O(k)\} \tag{8}$$

and

$$O(k) \geq V(\{k\}) \tag{9}$$

and

$$V(\{L,M,R\}) \leq V(\{i,j\}) + O(k). \tag{10}$$

(ii) $(\{L\},\{M\},\{R\})$ *is a recursive core structure if and only if* $V(S) \leq 0$ *for all* $S \subseteq \{L,M,R\}$.
(iii) $(\{L,M,R\})$ *is a recursive core structure if and only if* V *is balanced.*

To interpret part (i) intuitively, consider coalition structure $(\{i,j\},\{k\})$, where two coalition partners i and j share $V(\{i,j\})$, and the outside party k gets $O(k)$. Player i will stay with player j as long as his payoff satisfies $x_i \geq \max\{V(\{i\}), V(\{i,k\}) - O(k)\}$. Indeed, party i's options are to either leave the government, or to entice party k to form a two-party government. In the former case, party i expects to get $V(\{i\})$, in the latter case $V(\{i,k\}) - O(k)$ (since party k must get at least $O(k)$ to be willing to join the government). Similarly, player j will not defect from the coalition as long as $x_j \geq \max\{V(\{j\}), V(\{j,k\}) - O(k)\}$. The inequality (8) says that coalition $\{i,j\}$ generates enough surplus to simultaneously prevent both coalition partners from defecting. Next, the inequality (9) ensures that the outside player k gets at least his worth. Finally, (10) says that there is no gain from forming a grand coalition government. Part (ii) is obvious: the caretaker government is a reasonable outcome if and only if no coalition can generate a positive value. Part (iii) is standard.

The grand coalition uniquely maximizes the total social surplus if and only if

$$V(\{L,M,R\}) > \max\{0, V(\{M,R\}) + O(L), V(\{L,M\}) + O(R), V(\{L,R\}) + O(M)\}$$

In this case, the claim implies that only the grand coalition can be a recursive core structure. Outside of this case, however, the recursive core structure can be different from the social surplus maximizing structure. To be specific, suppose the $\{L,M\}$ coalition uniquely maximizes the social surplus:

$$V(\{L,M\}) + O(R) > \max\{0, V(\{M,R\}) + O(L), V(\{L,R\}) + O(M), V(\{L,M,R\})\}. \quad (11)$$

Now suppose the government has to deal with a financial crisis. The right wingers would prefer if a left-middle government deals with it, but the left-middle coalition would suffer disutility from it. In this case, it is unreasonable to expect a left-middle coalition to form, even if it is social surplus maximizing. (The inequality (11) holds because the social surplus on the left hand side includes $O(R)$, the large benefit the right-wingers enjoy from shifting responsibility to the left-middle government). Which coalition would actually form depends on the parameters.

Example 1: Suppose

$$V(\{L,M,R\}) = V(\{M,R\}) = -10, V(\{L,M\}) = -20, V(\{L,R\}) = 10,$$
$$O(R) = 40, O(L) = 7, O(M) = -14.$$

In this case, the only coalition government which can create a positive surplus for itself is $\{L,R\}$, so $V(M) = -14$, $V(R) = V(L) = 0$. Part (i) of the claim can be verified for two-party coalitions $\{M,R\}$ and $\{L,R\}$, but not for $\{L,M\}$, and neither (ii) nor (iii) holds. Thus, in example 1, there are two recursive core structures: $(\{M,R\},\{L\})$ and $(\{L,R\},\{M\})$. Any two-party coalition *except* the surplus maximizing one can form. It is intuitively plausible that the $\{L,R\}$ coalition can be an equilibrium outcome, given these payoffs (the outside party M would suffer but this is of no concern to L and R). It is more interesting that the $\{M,R\}$ coalition can form, even though it generates a negative surplus for itself. Why doesn't the middle-right coalition fall apart and trigger the caretaker government? Neither party wants to take the first step. Player M thinks that if he breaks up with R, the result is *not* the caretaker government, instead party R would join a left-right (surplus producing) government. This would give M a large negative payoff, $O(M) = -14$. Player M's fear of this outcome makes him

unwilling to defect from the middle-right government. Specifically, M is willing to stay with R as long as M's payoff is at least -14. So R can get $V(\{M,R)) - (-14) = 4$, and in this case, he is happy to stay with M. (R's best outside option is to govern with L, but L requires at least 7 to be made no worse off, and that leaves only $V(\{L,R)\} - 7 = 3$ for R.) Party M cannot compete for party L's favors, since the middle-left government is only worth -20. The right wing party has a strong outside option, namely to join with the left-wingers, but this allows it to extract a large surplus from the middle party, which holds the middle-right coalition together.

Notice that if our theory had allowed only the splitting up of coalitions, with no "re-merging", the middle-right coalition could not be stable in example 1. If re-merging were not allowed, then M could freely break-up with R, knowing that R could never re-merge with L. The fact that re-merging *is* allowed contributes to the appeal of the recursive core. After some reflection, it seems intuitively plausible that the right wing party could form an equilibrium coalition with *either one* of the other two parties.

Example 2: Suppose

$$V(\{L,M,R\}) = 40, V(\{M,R\}) = 0, V(\{L,M\}) = -50, V(\{L,R\}) = 0,$$
$$O(R) = 100, O(L) = 0, O(M) = 0.$$

In this case, the only coalition government that can generate a positive surplus for itself is the grand coalition. This implies $V(\{M\}) = V(\{R\}) = V(\{L\}) = 0$. We can check that only part (iii) of the claim holds. Thus, in this example, the unique recursive core structure is for the grand coalition to form, and this is also the only intuitively plausible outcome. The coalition structure $(\{L,M\},\{R\})$ maximizes the social surplus but is implausible. The right wing party benefits greatly from the formation of a left-middle government, and the criterion of social surplus maximization takes this into account. However, one cannot expect that the left-middle government will form simply to benefit the right wingers.

Example 3: Suppose

$$V(\{L,M,R\}) = -5, V(\{M,R\}) = -1, V(\{L,M\}) = -50, V(\{L,R\}) = -50,$$
$$O(R) = 100, O(L) = -1, O(M) = -1.$$

In this case, the "financial crisis" is so severe that no coalition government can create positive surplus for itself. This implies $V(\{M\}) = V(\{R\}) = V(\{L\}) = 0$. We can check that only part (ii) of the claim holds. Thus, in this example, the unique recursive core structure is $(\{L\},\{M\},\{R\})$. Since no coalition has anything to gain by forming a government, the caretaker government is the only intuitively plausible outcome.

In all three examples, the coalition structures that are predicted by the recursive core seem intuitively plausible. It is, therefore, worth mentioning that the recursive core can be quite different from the α-core. In example 3, according to the α-theory, $V^\alpha(\{M\}) = -1$, $V^\alpha(\{R\}) = 0$, and $V^\alpha(\{L\}) = -1$. Therefore, the α-core admits the middle-right government. Consider the structure $(\{M,R\},\{L\})$. According to the α-theory, M pessimistically expects that if the middle-right government breaks up, then a left-right government will form instead. So M is satisfied with the payoff of -1, and this holds the middle-right coalition together. However, player M should expect that if he *refuses to participate in any coalition government*, then the other two parties will definitely *not* form a government: L and R have no reason to form a coalition of worth -50 just to "punish" M. Therefore, the α-core does not make much

sense. (The caretaker government is also consistent with the α-core in example 3, but the left-middle coalition is not, so it is not that the α-core is biased toward surplus maximization).

4. The self-similar Serrano-Vohra game

In this section, we define a non-cooperative game Γ which fully implements the recursive core of $< N, P >$ in subgame perfect equilibria. We make no assumptions about superadditivity, so the equilibrium coalition structure might be a non-trivial partition of N. The game Γ is a self-similar version of Serrano and Vohra's [18] two-stage game, which they used to implement the core correspondence in economic environments without externalities. "Fines" are imposed if the players cannot agree with each other, which is naturally possible in economic environments. More generally, we consider a transferable utility game. Utility can be freely transferred across agents using some commodity (Shubik's u-money) which can be freely disposed of. A more complicated game could be constructed which does not rely on free disposal, but for the sake of transparency, we mimic the Serrano-Vohra [18] construction.

We assume externalities exist, so the key issue is how outsiders react when a coalition is formed. In Serrano and Vohra's [18] original game, outsiders have no chance to react. We need to make the Serrano-Vohra [18] game self-similar: after a coalition has formed, the outsiders go on to play a scaled-down version of the original game. Subgame perfection rules out the kind of incredible threats that define the α- and β-cores.

Since the game may continue after a coalition has formed, we need to distinguish between *inactive players*, who have joined a coalition and cannot make any more moves, and *active players*. Let A denote the set of active players and let $I = N \backslash A$ denote the set of inactive players. The inactive players in I must have formed some coalition structure $\mathcal{P}_I$, a partition of I. The members of each coalition $S \in \mathcal{P}_I$ must also have agreed on how to divide their coalition's payoff. Specifically, they have agreed to share their coalition's final payoff according to some weights $w^S = (w_i^S)_{i \in S}$, where

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

Let $w^I = (w^S)_{S \in \mathcal{P}_I}$ be the list of weights for all the coalitions in $\mathcal{P}_I$. In game Γ, coalitions may block outcomes, and several blocking coalitions may form in sequence. As will become clear below, it is useful to keep track of the *last blocking coalition* that has formed so far, say $L \in \mathcal{P}_I$. We will refer to $(\mathcal{P}_I, w^I, L)$ as the *state*. The state summarizes the history of the inactive players, and is an expositional device only. There are typically several ways to reach the same state. In particular, the sequence in which coalitions have formed is not completely specified by the state: only the *last* blocking coalition is specified. We keep track of the "state" $(\mathcal{P}_I, w^I, L)$ only to simplify the exposition. Nothing prevents the players from conditioning their actions on the full history of the game. We make no a priori restriction to stationary or Markov strategies.

For the sake of exposition, all subgames of Γ that start at the same state will be considered simultaneously. The set of all subgames that start at state $(\mathcal{P}_I, w^I, L)$ will be denoted by $\Gamma(\mathcal{P}_I, w^I, L)$. If no coalition has formed, then no player is inactive, and this state is denoted $(\emptyset, \emptyset, \emptyset)$. The game Γ itself starts at $(\emptyset, \emptyset, \emptyset)$.

Consider any state $(\mathcal{P}_I, w^I, L)$ such that $I \neq N$. Each subgame in $\Gamma(\mathcal{P}_I, w^I, L)$ has two stages, the coalition formation stage and the blocking stage.

A. Coalition formation stage. The subgame starts with the coalition formation stage. Each active player $i \in A$ simultaneously proposes a triplet $(\mathcal{P}_A, w^A, \pi_A)$.[5] Here $\mathcal{P}_A$ is the coalition structure that the active players will form (a partition of A), and $w^A = (w_i^S)_{i\in S, S\in\mathcal{P}_A}$ is the list of weights according to which each coalition in $\mathcal{P}_A$ should distribute its joint payoffs. The weights must satisfy, for each $S \in \mathcal{P}_A$,

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

Thus, according to the proposal, w_i^S is player i's share of the payoff of coalition S which he belongs to. Finally, π_A is a permutation of the player set A. We compose the permutations proposed by all players in A according to some pre-specified order and denote it by π_A^*. The key point is that by unilaterally changing the permutation he proposes, each player $i \in A$ can become the first player according to the composition π_A^*.

If in the coalition formation stage all active players propose the same $\mathcal{P}_A$ and w^A, then their common proposal $(\mathcal{P}_A, w^A)$ is designated the *status quo* and the game proceeds to the blocking stage. If it is *not* the case that all active players propose the same $\mathcal{P}_A$ and w^A, then the game ends, the last player according to π_A^* gets $-\varepsilon$ and all other players get zero. The coalition structure can be arbitrary in this case. Feasibility is guaranteed by "free disposal of u-money" (since all coalitional values are positive by (1)).

B. Blocking stage. The first player according to π_A^* is the proposer. The proposer can either (i) say Pass, or (ii) propose a blocking coalition $S \subseteq A$ of which he is a member, along with weights $w^S = (w_i^S)_{i\in S}$ for sharing its payoffs, where

$$\sum\nolimits_{i\in S} w_i^S = 1.$$

There are two possibilities.

(i) If the proposer says Pass, then the game ends. The status quo $(\mathcal{P}_A, w^A)$ is implemented. This makes sense because the active players all announced $(\mathcal{P}_A, w^A)$ in the coalition formation stage, and the proposer has passed on his right to name a blocking coalition. Moreover, the state records that inactive players partition themselves according to $\mathcal{P}_I$. So the final coalition structure on N is $(\mathcal{P}_I, \mathcal{P}_A)$. Each coalition distributes its payoffs according to the agreed-upon weights (w^I, w^A).

(ii) If the proposer proposes a blocking coalition S, then all members of S (except the proposer himself) must respond by saying Yes or No sequentially (according to some pre-specified order). If any member says No, then the coalition S is not formed, and just as in (i) the game ends with the status quo implemented. But if all members say Yes, then the blocking coalition S has formed (this always happens if S consists of only the proposer, since then there is nobody who can say No). When S forms, all members of S become inactive, so the state changes from $(\mathcal{P}_I, w^I, L)$ to $(\mathcal{P}_{I'}, w^{I'}, L')$. The new set of inactive players is $I' = I \cup S$, the new set of active players is $A' = N\backslash I' = A\backslash S$, and the most recently formed blocking coalition is $L' = S$. The players in I cannot rearrange themselves, but the new coalition S has formed, so the coalition structure on I' is $\mathcal{P}_{I'} = (\mathcal{P}_I, S)$. The new set of weights $w^{I'}$ is the concatenation of w^I and w^S, denoted $w^{I'} = (w^I, w^S)$. Now there are two possible cases. If there are no active players left, $A' = A\backslash S = \emptyset$, then the game ends. The final coalition structure on

[5] Strictly speaking, player i's proposal should be indexed by i, but this index is omitted to simplify notation.

N is $(\mathcal{P}_I, S)$. The players distribute their payoffs according to the agreed-upon weights (w^I, w^S). If $A' = A\backslash S \neq \emptyset$, then the game continues to a subgame in $\Gamma(\mathcal{P}_{I'}, w^{I'}, L') = \Gamma((\mathcal{P}_I, S), (w^I, w^S), S)$. This subgame is played out with a coalition formation stage and a blocking stage, according to the rules we have described. Notice that only the active players in $A' = A\backslash S$ can participate in this subgame. The inactive players in $I' = I \cup S$ are on the sidelines looking on, having already formed their coalitions.

5. Results

We prove two results in this section. First we show that every payoff vector in $C(N)$ is a subgame perfect Nash equilibrium (SPNE) outcome of the game Γ. Later, we will prove the converse: every SPNE outcome in the game Γ is a payoff vector in $C(N)$.

To implement $x \in C(N)$, with the corresponding recursive core coalition structure $\mathcal{P}_N$, we want all players to propose $(\mathcal{P}_N, w^N)$ at the very first coalition formation stage of Γ, where w^N are weights that give us the payoff vector x. The blocking stage gives each coalition a chance to block the allocation, reflecting the spirit behind the core, if it does not get at least what it is worth. Thus if a coalition $T \subseteq N$ forms in the blocking stage, then we want it to get exactly its value $V(T)$. The definition of $V(T)$ implies that the players in $N\backslash T$ should behave in such a way that the outcome is a point in the recursive core $C(N\backslash T \mid N\backslash T, T)$ which is worst for T. This can be done, because the subgame played out among the players in $N\backslash T$ is simply a smaller version of Γ. In the subgame, any subcoalition $S \subset N\backslash T$ will have a chance to block. If blocking coalition S forms, then we want it to get exactly its value $V(S \mid N\backslash T, T)$. Hence, we want the remaining players in $N\backslash(T \cup S)$ to behave in such a way that the outcome is a point in the recursive core $C(N\backslash(T \cup S) \mid N\backslash(T \cup S), S, T)$ which is worst for S, the most recent blocking coalition. And so on, recursively. Notice that behavior will depend on the last blocking coalition which has formed at any stage. This last blocking coalition L is part of the "state" $(\mathcal{P}_I, w^I, L)$.

In the proof of Proposition 3, strategies will be constructed where history matters *only* through the state $(\mathcal{P}_I, w^I, L)$. For any $(\mathcal{P}_I, w^I, L)$, whenever the players find themselves in this state, they will behave the same way. Indeed, we may abuse terminology by referring to $\Gamma(\mathcal{P}_I, w^I, L)$ as a subgame, although it is understood to be the set of all subgames starting at the state $(\mathcal{P}_I, w^I, L)$. This is harmless because behavior will, by construction, be the same in any subgame in $\Gamma(\mathcal{P}_I, w^I, L)$. Similarly, we will talk about the equilibrium continuation payoff vector $x(\mathcal{P}_I, w^I, L)$ at the state $(\mathcal{P}_I, w^I, L)$. Because behavior will be the same in any two subgames in $\Gamma(\mathcal{P}_I, w^I, L)$, the continuation payoff will also be the same.

Proposition 3 *Every payoff vector in $C(N)$ is an SPNE payoff vector in the game Γ.*

Proof. Suppose $x \in C(N)$, and let $\mathcal{P}_N$ be the corresponding recursive core coalition structure. We need to show that x is an equilibrium payoff vector for Γ. Recall that $C(N) \neq \varnothing$ implies $C(S \mid S, \mathcal{P}_{N\backslash S}) \neq \emptyset$ for any $S \subseteq N$ and any $\mathcal{P}_{N\backslash S}$.

Let

$$w^N = (w_i^S)_{i \in S, S \in \mathcal{P}_N}$$

be the list of weights according to which each coalition in $\mathcal{P}_N$ should distribute its joint payoffs in order to reach the payoff vector x. That is, if $i \in S \in \mathcal{P}_N$ then

$$x_i = w_i^S P(S \mid \mathcal{P}_N).$$

Before constructing the equilibrium strategy to support x as an equilibrium outcome, we first specify what the equilibrium outcome will be in any proper subgame $\Gamma(\mathcal{P}_I, w^I, L)$ where $(\mathcal{P}_I, w^I, L) \neq (\emptyset, \emptyset, \emptyset)$, $A = N \backslash I$ is the set of active players and the inactive players in I have formed coalition structure $\mathcal{P}_I$. We want the continuation equilibrium payoff vector to be a point in $C(A \mid A, \mathcal{P}_I)$. Specifically, we will select a payoff vector $x(\mathcal{P}_I, w^I, L)$ in $C(A \mid A, \mathcal{P}_I)$ which is the worst possible for the last blocking coalition that has formed so far, denoted $L \in \mathcal{P}_I$. Thus, let $x(\mathcal{P}_I, w^I, L)$ be a payoff vector which has the following three properties: (i)

$$x(\mathcal{P}_I, w^I, L) \in C(A \mid A, \mathcal{P}_I), \tag{12}$$

with the corresponding recursive core structure $\mathcal{P}(\mathcal{P}_I, w^I, L) \in \mathcal{P}(A \mid A, \mathcal{P}_I)$;

(ii)

$$\sum_{i \in L} x_i(\mathcal{P}_I, w^I, L) = \min \left\{ \sum_{i \in L} y_i : \ y \in C(A \mid A, \mathcal{P}_I) \right\} = V(L \mid (A \cup L), \mathcal{P}_I) \tag{13}$$

(iii) if $i \in S \in \mathcal{P}_I$ then

$$x_i(\mathcal{P}_I, w^I, L) = w_i^S P(S \mid \mathcal{P}(\mathcal{P}_I, w^I, L), \mathcal{P}_I) \tag{14}$$

where w_i^S is the weight for i which was agreed upon when coalition S formed.

Notice that (13) implies that the last blocking coalition L is punished as much as possible without violating the constraints of the recursive core. That is, $\sum_{i \in L} y_i$ is minimized in $C(A \mid A, \mathcal{P}_I)$, and this minimized sum is independent of w^I. Intuitively, the second equality in (13) holds because the minimized sum that L can get occurs when the remaining players A play out the worst recursive core for L. This is exactly the value of L, when L is contemplating whether to deviate in the reduced society $A \cup L$.

Let

$$w^A(\mathcal{P}_I, w^I, L) = (w_i^S(\mathcal{P}_I, w^I, L))_{i \in S, S \in \mathcal{P}(\mathcal{P}_I, w^I, L)}$$

be the list of weights according to which each coalition in $\mathcal{P}(\mathcal{P}_I, w^I, L)$ should distribute its joint payoffs in order to reach the payoff vector $x(\mathcal{P}_I, w^I, L)$. That is, if $i \in S \in \mathcal{P}(\mathcal{P}_I, w^I, L)$ then

$$w_i^S(\mathcal{P}_I, w^I, L) = \frac{x_i(\mathcal{P}_I, w^I, L)}{P(S \mid \mathcal{P}(\mathcal{P}_I, w^I, L), \mathcal{P}_I)} \tag{15}$$

This implies that

$$x_i(\mathcal{P}_I, w^I, L) = w_i^S(\mathcal{P}_I, w^I, L) P(S \mid \mathcal{P}(\mathcal{P}_I, w^I, L), \mathcal{P}_I) \tag{16}$$

Equation (15) is well-defined because $P(S \mid \mathcal{P}(\mathcal{P}_I, w^I, L), \mathcal{P}_I) > 0$.

We can now construct the equilibrium strategy to support x as an equilibrium outcome.

(E1) At the coalition formation stage, if the state is $(\emptyset, \emptyset, \emptyset)$, all players start by proposing $(\mathcal{P}_N, w^N, \pi_N)$ where π_N is an arbitrary permutation of N. If the state is $(\mathcal{P}_I, w^I, L) \neq (\emptyset, \emptyset, \emptyset)$ and $A = N \backslash I$ is the set of active players, then all players start by proposing $(\mathcal{P}(\mathcal{P}_I, w^I, L), w^A(\mathcal{P}_I, w^I, L), \pi_A)$, where π_A is an arbitrary permutation of A.

(E2) At the blocking stage when the state is $(\mathcal{P}_I, w^I, L)$, by the game rules, active players $A = N \backslash I$ must have agreed on some common proposal $(\mathcal{P}_A, w^A, \pi_A)$. According the common proposal (which

becomes the status quo), the final coalition structure will be $(\mathcal{P}_I, \mathcal{P}_A)$, and each active player $i \in A$ will get payoff

$$y_i \equiv w_i^S P(S \mid \mathcal{P}_A, \mathcal{P}_I),$$

where w_i^S is his share of payoffs of the coalition $S \in \mathcal{P}_A$ to which he belongs.

Without loss of generality, call the first player according to the composition of the permutations proposed by A player 1. Suppose player 1 proposes to form a blocking coalition $S \subseteq A$ of which he is a member, along with weights $w^S = (w_i^S)_{i \in S}$ for sharing its payoffs. If S forms in the blocking stage, the remaining players $A \backslash S$ play out the recursive core which is the worst for S. The continuation equilibrium payoff will be a point

$$x((\mathcal{P}_I, S), (w^I, w^S), S) \in C(A \backslash S \mid A \backslash S, S, \mathcal{P}_I).$$

By (13), the total surplus for coalition $S \subseteq A$ will be

$$\sum_{i \in S} x_i((\mathcal{P}_I, S), (w^I, w^S), S) = V(S \mid A, \mathcal{P}_I).$$

After paying every other member in S his status quo payoff, the surplus available to player 1 is

$$V(S \mid A, \mathcal{P}_I) - \sum_{i \in S, i \neq 1} y_i.$$

Pick any S^* such that

$$S^* \in \arg \max_{S \subseteq A, 1 \in S} \left(V(S \mid A, \mathcal{P}_I) - \sum_{i \in S, i \neq 1} y_i \right).$$

If

$$V(S^* \mid A, \mathcal{P}_I) - \sum_{i \in S^*, i \neq 1} y_i \leq y_1,$$

then let player 1 say Pass when he proposes. Otherwise, let player 1 propose to form S^* and share the payoffs according to the weights

$$w_i^{S^*} = \frac{y_i}{V(S^* \mid A, \mathcal{P}_I)}$$

for each $i \in S^*,\ i \neq 1$. Each such player i's payoff will be exactly y_i, since the total surplus for S^* is $V(S^* \mid A, \mathcal{P}_I)$. Finally, we need to specify the behavior of the responding players. When any player i is called to respond to a proposal from player 1, let him say Yes if he is given weakly more than his status quo payoff y_i and No otherwise. In other words, if S is proposed by player 1, together with the weights $w^S = (w_j^S)_{j \in S}$, and player i is asked to respond (where $1 \neq i \in S$), then player i says Yes if and only if $w_i^S V(S \mid A, \mathcal{P}_I) \geq y_i$.

Note that every subgame in $\Gamma(\mathcal{P}_I, w^I, L)$ has a coalition formation stage and a blocking stage. In the coalition formation stage, the equilibrium strategies specify that each active player $i \in A = N \backslash I$ proposes $(\mathcal{P}(\mathcal{P}_I, w^I, L), w^A(\mathcal{P}_I, w^I, L), \pi_A)$, where π_A is an arbitrary permutation of A. This implies that $(\mathcal{P}(\mathcal{P}_I, w^I, L), w^A(\mathcal{P}_I, w^I, L))$ becomes the status quo in the subsequent blocking stage. The equilibrium strategies also specify that the proposer says Pass in this case (we will show this in the next paragraph). Therefore, if $\Gamma(\mathcal{P}_I, w^I, L)$ is reached, then the continuation equilibrium payoff vector will be $x(\mathcal{P}_I, w^I, L)$. In particular, in $\Gamma(\emptyset, \emptyset, \emptyset)$, every player $i \in N$ proposes $(\mathcal{P}_N, w^N, \pi_N)$ and the proposer says Pass in the blocking stage. Hence x is the equilibrium outcome in $\Gamma(\emptyset, \emptyset, \emptyset)$.

To see why the proposer says Pass on the equilibrium path, suppose all players follow their equilibrium strategies in the coalition formation stage of $\Gamma(\mathcal{P}_I, w^I, L)$. In the blocking stage, the status quo gives payoff $x_i(\mathcal{P}_I, w^I, L)$ to each player $i \in A = N\backslash I$. Some player, say player $1 \in A$, is the proposer (as determined by the composition of the permutations in π_A). Player 1 can either say Pass, or propose a blocking coalition S such that $1 \in S \subseteq A$, with weights w^S. If player 1 says Pass (or if he proposes a coalition which is rejected by some member), then the status quo is implemented. If player 1 proposes a blocking coalition S which is accepted by all members, then S forms. The players in S become inactive, so the new set of inactive players is $I \cup S$ and the new set of active players is $A\backslash S$. The coalition structure of the new set of inactive players is $(\mathcal{P}_I, S)$. Play then moves to a subgame in $\Gamma((\mathcal{P}_I, S), (w^I, w^S), S)$, because S is the last coalition to have formed. As argued above, the continuation equilibrium payoff will be

$$x((\mathcal{P}_I, S), (w^I, w^S), S) \in C(A\backslash S \mid A\backslash S, \mathcal{P}_I, S).$$

For player $i \in S$, $i \neq 1$, the equilibrium strategy specifies that he accepts player 1's proposal if and only if he is offered at least his status quo payoff (every responder, by saying No, can induce the status quo). Therefore, when player $1 \in A$ makes a proposal to form blocking coalition S in the blocking stage of $\Gamma(\mathcal{P}_I, w^I, L)$, the most he can hope to get for himself is

$$\sum_{i \in S} x_i((\mathcal{P}_I, S), (w^I, w^S), S) - \sum_{\substack{i \in S \\ i \neq 1}} x_i(\mathcal{P}_I, w^I, L) \tag{17}$$

But we claim that this expression is no bigger than player 1's status quo payoff $x_1(\mathcal{P}_I, w^I, L)$, so proposing a blocking coalition S doesn't pay for player 1. To prove this claim, we need to show that

$$\sum_{i \in S} x_i(\mathcal{P}_I, w^I, L) \geq \sum_{i \in S} x_i((\mathcal{P}_I, S), (w^I, w^S), S) \tag{18}$$

for all S such that $1 \in S \subseteq A$. However, $x(\mathcal{P}_I, w^I, L) \in C(A \mid A, \mathcal{P}_I)$. Using the definition of the recursive core and (13) we get

$$\sum_{i \in S} x_i(\mathcal{P}_I, w^I, L) \geq V(S \mid A, \mathcal{P}_I) = \sum_{i \in S} x_i((\mathcal{P}_I, S), (w^I, w^S), S) \tag{19}$$

Therefore, (18) holds for all S such that $1 \in S \subseteq A$, so there is no opportunity for profitable blocking. The equilibrium strategy specifies that player 1 says Pass, which maximizes his expected payoff.

In particular, consider the blocking stage of $\Gamma(\emptyset, \emptyset, \emptyset)$ with status quo payoff vector $x \in C(N)$. If player 1 proposes a blocking coalition S with weights w^S which is accepted by all members, then S forms. The set of inactive players becomes S and the set of active players becomes $N\backslash S$. The coalition structure of the new set of inactive players is (S). Play then moves to a subgame in $\Gamma((S), w^S, S)$. As argued above, the continuation equilibrium payoff will be

$$x((S), w^S, S) \in C(N\backslash S \mid N\backslash S, S).$$

Equation (13) implies that this outcome is worse for S than any other payoff vector in $C(N\backslash S \mid N\backslash S, S)$. By definition of the recursive core,

$$\sum_{i \in S} x_i \geq V(S) = \sum_{i \in S} x_i((S), w^S, S)$$

That is, (18) holds for status quo payoff vector x. So player 1 maximizes his payoff by saying Pass at the blocking stage of $\Gamma(\emptyset, \emptyset, \emptyset)$.

We can now verify that the strategies in (E1) and (E2) constitute a subgame perfect Nash equilibrium. First, suppose we have reached a subgame $\Gamma(\mathcal{P}_I, w^I, L)$ such that $A = N\backslash I$ contains only one active player, say $A = \{1\}$. In this situation, player 1's strategy specifies that he forms the singleton coalition $\{1\}$, takes a share $w_1 = 1$ of the payoff of his coalition, and passes in the blocking stage. This is obviously a continuation equilibrium.

Next, we make an induction hypothesis: suppose for any A such that $|A| \leq a - 1$, whenever we reach a subgame $\Gamma(\mathcal{P}_I, w^I, L)$ where $A = N\backslash I$ is the set of active players, then the strategies we have described constitute a continuation equilibrium for $\Gamma(\mathcal{P}_I, w^I, L)$. Now suppose we reach a subgame $\Gamma(\mathcal{P}_I, w^I, L)$ where the set of active players is $A = N\backslash I$ and A has $|A| = a$ members, $a \geq 2$. We need to show that the strategies we have described constitute a continuation equilibrium for $\Gamma(\mathcal{P}_I, w^I, L)$. By construction, in the blocking stage each responder is using a best response (he says Yes if and only if the continuation equilibrium payoff would be no less than his status quo payoff). Each proposer is also using a best response: say Pass if no profitable blocking is possible, and propose the most profitable blocking coalition otherwise. Now consider the coalition formation stage of $\Gamma(\mathcal{P}_I, w^I, L)$. Since $|A| \geq 2$, deviating by proposing anything other than $(\mathcal{P}(\mathcal{P}_I, w^I, L), w^A)$ results in either the payoff of 0 or $-\varepsilon$. The equilibrium payoff for any active player $i \in A$ is $x_i(\mathcal{P}_I, w^I, L)$, which is strictly positive because $x(\mathcal{P}_I, w^I, L) \in C(A \mid A, \mathcal{P}_I)$ and (1) holds. Hence, all active players prefer to announce $(\mathcal{P}(\mathcal{P}_I, w^I, L), w^A)$. Given these announcements, since (19) holds for any blocking coalition $S \subseteq A$,. in the blocking stage the proposer will say Pass (as argued above). This implies that every active player i will get $x_i(\mathcal{P}_I, w^I, L)$ if the blocking stage is reached, no matter who is the proposer. So it does not matter to any active player which permutation he proposes. Therefore, at the coalition formation stage, there is nothing to gain from deviating on the announced permutation, either. □

We can now prove the converse of Proposition 3. Notice that we will not restrict attention to stationary or Markov strategies that only depend on history via the "state". Any kind of history dependence is allowed.

Proposition 4 *Every SPNE payoff vector in the game* Γ *is a payoff vector in* $C(N)$.

Proof. Let f denote any SPNE strategy profile in the game Γ. The proof is by induction. The induction hypothesis is the following.

Hypothesis $H(k)$. For any coalition $A \subseteq N$ such that $1 \leq |A| \leq k$, any partition $\mathcal{P}_I$ on $I = N\backslash A$, and any $L \in \mathcal{P}_I$, the following two conditions both hold:
(C1) For any subgame in $\Gamma(\mathcal{P}_I, w^I, L)$, if x is the payoff vector induced by f in this subgame, then $x \in C(A \mid A, \mathcal{P}_I)$.
(C2) For any subgame in $\Gamma(\mathcal{P}_I, w^I, L)$, if $(\mathcal{P}_A, \mathcal{P}_I)$ is the coalitional structure induced by f in this subgame, then $\mathcal{P}_A \in \mathcal{P}(A \mid A, \mathcal{P}_I)$.

First, consider the case $k = 1$. If $1 \leq |A| \leq k = 1$ then $A = \{i\}$ for some $i \in N$. Let $I = N\backslash\{i\}$. For any $\mathcal{P}_I$, $C(\{i\} \mid \{i\}, \mathcal{P}_I)$ contains every payoff vector where i gets $P(\{i\} \mid \{i\}, \mathcal{P}_{N\backslash\{i\}})$ and every coalition $S \in \mathcal{P}_I$ gets $P(S \mid \{i\}, \mathcal{P}_I)$. Since player i is the only active player, his only option is to propose coalition $\{i\}$ and take the share $w_i = 1$. This becomes the status quo. In the blocking stage, the

rules do not allow player i to propose anything else than the status quo (another alternative is that i can say Pass), so the game must end with the status quo. Hence, player i must get $x_i = P(\{i\} \mid \{i\}, \mathcal{P}_I)$ and every coalition $S \in \mathcal{P}_I$ gets $P(S \mid \{i\}, \mathcal{P}_I)$. So $x \in C(\{i\} \mid \{i\}, \mathcal{P}_I)$ and the coalition structure induced on $A = \{i\}$ is, trivially, $\mathcal{P}_A = \{\{i\}\}$. This proves $H(k)$ for $k = 1$.

To continue the induction, fix k such that $2 \leq k \leq n$ and suppose the induction hypothesis $H(k-1)$ holds. We need to prove $H(k)$.

Consider any coalition $A \subseteq N$ such that$|A| = k$, any partition $\mathcal{P}_I$ on $I = N \backslash A$, any $L \in \mathcal{P}_I$, and any subgame in $\Gamma(\mathcal{P}_I, w^I, L)$. Let x denote the payoff vector, and $(\mathcal{P}_A, \mathcal{P}_I)$ the coalitional structure, induced by f in this subgame. We need to verify (C1) and (C2).

We first observe that $V(T \mid A, \mathcal{P}_I)$ is well defined for all $T \subseteq A$. Indeed, by definition

$$V(A \mid A, \mathcal{P}_I) = P(A \mid A, \mathcal{P}_I).$$

And $V(T \mid A, \mathcal{P}_I)$ is also well defined for $T \subseteq A$, $T \neq A$ because $C(A \backslash T \mid A \backslash T, T, \mathcal{P}_I) \neq \emptyset$ by the induction hypothesis (because $|A \backslash T| \leq k-1$).

At the coalition formation stage of any subgame in $\Gamma(\mathcal{P}_I, w^I, L)$, every active player must propose the same coalition structure and the same weights. For otherwise, since there are at least two active players, the last player according to the composition of permutations proposed by all active players would get $-\varepsilon$. Changing the permutation he proposes would make his payoff 0. Hence there would exist a profitable deviation.

Suppose $x \notin C(A \mid A, \mathcal{P}_I)$. We will derive a contradiction. By definition of $C(A \mid A, \mathcal{P}_I)$ there must exist a coalition $S \subseteq A$ such that

$$\sum_{i \in S} x_i < V(S \mid A, \mathcal{P}_I).$$

Without loss of generality, suppose $S = \{1, 2, ..., s\}$. In the coalition formation stage, given the permutations announced by the other active players, there is for each active player a permutation he can announce that will make him the proposer. Specifically, player 1 can make sure that he is the proposer by announcing an appropriate permutation. At the blocking stage, player 1 can propose to form S and give each member $i \in S$ an amount y_i so that

$$y_i = x_i + s^{-1} \left[V(S \mid A, \mathcal{P}_I) - \sum_{i \in S} x_i \right] > x_i.$$

Accordingly, for all $i \in S$, define

$$w_i = \frac{y_i}{V(S \mid A, \mathcal{P}_I)}.$$

When player 1 makes this proposal, without loss of generality, suppose players $2, 3, ..., s$ will be called to respond in this order. We want to show that they will all say Yes. There are two cases.

Case I: $S = A$. In this case, the game ends immediately when blocking coalition S forms, by the rules of the game. When player s is called to respond, if all other responders have said Yes, then if player s also says Yes, S successfully forms. Player s gets

$$w_s P(A \mid A, \mathcal{P}_I) = y_s > x_s,$$

because $P(A \mid A, \mathcal{P}_I) = V(A \mid A, \mathcal{P}_I)$. Hence, player s must say Yes (saying No would only result in the status quo payoff x_s). When player $s-1$ is called to respond, if all previous responders have said Yes, then if player $s-1$ also says Yes, he expects player s to say Yes after him. By the same logic as for player s, player $s-1$ will say Yes. Inductively, all responders will say Yes. So, after player 1 makes his proposal, the blocking coalition $S = A$ forms and the game ends. Player 1 gets

$$w_1 P(A \mid A, \mathcal{P}_I) = y_1 > x_1.$$

So player 1 has a profitable deviation, contradicting the assumption that f is an SPNE.

Case II: $S \subseteq A$ and $S \neq A$. In this case, if the blocking coalition S forms, the set of active players become $A' = A \backslash S$. When player s is called to respond, if all other responders have said Yes, then if player s says Yes, S successfully forms. By the induction hypothesis, the continuation equilibrium payoff vector z will satisfy

$$z \in C(A \backslash S \mid A \backslash S, S, \mathcal{P}_I). \tag{20}$$

Player s will get $w_s z(S)$. Since $V(S \mid A, \mathcal{P}_I)$ is the worst possible joint payoff for S in $C(A \backslash S \mid A \backslash S, S, \mathcal{P}_I)$, it must be the case that $z(S) \geq V(S \mid A, \mathcal{P}_I)$. Therefore,

$$w_s z(S) \geq w_s V(S \mid A, \mathcal{P}_I) = y_s > x_s.$$

So s will say Yes. When player $s-1$ is called to respond, if all previous responders have said Yes, then if he says Yes, he expects player s to say Yes after him. By the same logic as for player s, player $s-1$ will say Yes. Inductively, all responders will say Yes. So, after player 1 makes his proposal, the blocking coalition S forms. Player 1 will get

$$w_1 z(S) \geq w_1 V(S \mid A, \mathcal{P}_I) = y_1 > x_1,$$

so player 1 has a profitable deviation, contradicting the assumption that f is an SPNE.

Since both case I and case II lead to a contradiction, we conclude that $x \in C(A \mid A, \mathcal{P}_I)$. Thus, (C1) holds. Clearly, the coalitional structure $(\mathcal{P}_A, \mathcal{P}_I)$ induced by f in the subgame $\Gamma(\mathcal{P}_I, w^I, L)$ must satisfy $\mathcal{P}_A \in \mathcal{P}(A \mid A, \mathcal{P}_I)$, so (C2) holds as well. Therefore, we have proven $H(k)$.

The induction proves that $H(n)$ holds. From (C1), any SPNE payoff vector x induced by f in the game itself must satisfy $x \in C(N)$. □

Acknowledgements

We thank two referees for very useful comments, and Nadide Banu Olcay for expert research assistance.

References

1. Aumann R.J.; Dreze J. Cooperative Games with Coalition Structures. *Int. J. Game Theory* **1974**, *3*, 217-237.
2. Bandyopadhyay, S.; Chatterjee, K.; Sjöström, T. Pre-Electoral Coalitions and Post-Election Bargaining. Unpublished manuscript, Rutgers University, New Brunswick, NJ, USA, 2009.

3. Bloch, F. Sequential Formation of Coalitions in Games with Externalities and Fixed Payoff Division, *Games Econ. Behav.* **1996**, *14*, 90-123.
4. Chatterjee, K.; Dutta, B.; Ray, D.; Sengupta, K. A Noncooperative Theory of Coalitional Bargaining. *Rev. Econ. Stud.* **1993,** *60*, 463-477.
5. Diamantoudi, E.; Xue, L. Coalitions, Agreements and Efficiency. *J. Econ. Theory* **2007**, *136*, 105-125.
6. Funaki, Y.; Yamato, T. The Core of an Economy with a Common Pool Resource: A Partition Function Form Approach. *Int. J. Game Theory* **1999,** *28*, 157-171.
7. Funaki, Y.; Yamato, T. Sequentially Stable Coalition Structures. Unpublished manuscript, Tilburg University, Tilburg, The Netherlands, 2009.
8. Huang C.-Y.; Sjöström, T. Consistent Solutions for Cooperative Games with Externalities. *Games Econ. Behav.* **2003,** *43*, 196-213.
9. Huang C.-Y.; Sjöström, T. Implementation of the Recursive Core for Partition Function Form Games. *J. Math. Econ.* **2006,** *42*, 771-793.
10. Kalai, E.; Postlewaite, A.; Roberts, J. A Group Incentive Compatible Mechanism Yielding Core Allocations. *J. Econ. Theory* **1979,** *20*, 13-22.
11. Kóczy, L.Á. A Recursive Core for Partition Function Form Games. *Theory Decis.* **2007,** *63*, 41-51.
12. Kóczy, L.Á. Sequential Coalition Formation and the Core in the Presence of Externalities. *Games Econ. Behav.* **2009**, *66*, 559-565
13. Moldovanu, B.; Winter, E. Core Implementation and Increasing Returns to Scale for Cooperation. *J. Math. Econ.* **1994,** *23*, 533-548.
14. Moldovanu, B.; Winter, E. Order Independent Equilibria. *Games Econ. Behav.* **1995,** *9*, 21-34.
15. Perry, M.; Reny, P. A Noncooperative View of Coalition Formation and the Core. *Econometrica* **1994,** *62*, 795-817.
16. Ray, D.; Vohra, R. Equilibrium Binding Agreements. *J. Econ. Theory* **1997,** *73*, 30-78.
17. Sandholm, T.; Larson, K.; Andersson, M.; Shehory, O.; Tohmé, F. Coalition Structure Generation with Worst Case Guarantees. *Artificial Intelligence* **1999**, *111*, 209-238.
18. Serrano, R.; Vohra, R. Noncooperative Implementation of the Core. *Soc. Choice Welfare* **1997,** *14*, 513-525.
19. Thomson, W. Divide and Permute. *Games Econ. Behav.* **2005,** *52*, 186-200.
20. Yan, H.; Friedman, D. Testing the Counterexample of the Core. Unpublished manuscript, University of California, Santa Cruz, CA, USA, 2009.
21. Zheng, C. A Coase Theorem Based on a New Concept of the Core. Unpublished manuscript, Iowa State University, Ames, IO, USA, 2009.