The Harsanyi paradox and the "right to talk" in bargaining among coalitions*

Juan J. Vidal-Puga†

Universidade de Vigo

December 15, 2005

Abstract

I describe a new coalitional value from a non-cooperative point of view, assuming coalitions are formed for the purpose of bargaining. The idea is that all the players have the same chances to make proposals. This means that players maintain their own "right to talk" when joining a coalition. The resulting value coincides with the weighted Shapley value in the game between coalitions, with weights given by the size of the coalitions. I apply this value to an intriguing example presented by Krasa, Temimi and Yannelis (Journal of Mathematical Economics, 2003) and show that the Harsanyi paradox (forming a coalition may be disadvantageous) disappears. These results throw certain doubts on the reasonability of the Carrier axiom as presented by Hart and Kurz (Econometrica, 1983).

JEL codes: C71, C78

*Latest version at `http://webs.uvigo.es/vidalpuga/`. Financial support by the Spanish Ministerio de Ciencia y Tecnología and FEDER through grant BEC2002-04102-C02-01 and Xunta de Galicia through grant PGIDIT03PXIC30002PN is gratefully acknowledged.

†Facultade de Ciencias Sociais. Campus A Xunqueira. 36005 Pontevedra. Spain. Phone: +34 986 802014. Fax: +34 986 812401. Email: `vidalpuga@uvigo.es`

Keywords: weighted Shapley value, non-cooperative bargaining, coalition structures, Harsanyi paradox

1 Introduction

Many economic situations can be modelled as a set of agents or players with independent interests who may benefit from cooperation. Moreover, it is not infrequent that these agents have partitioned themselves into coalitions (such as unions, cartels, or syndicates) for the purpose of bargaining.

Assuming that cooperation is carried out, the question is how to share the benefit between the coalitions and between the members inside each coalition, i.e. which "value" best represents the expectation of each individual. The economic theory has addressed this problem from two different points of view. One of them is axiomatic. The other is non-cooperative.

The axiomatic point of view focuses on finding allocations which satisfy "fair" (or at least "reasonable") properties, such as efficiency (the final outcome must be efficient), symmetry (players with the same characteristics must receive the same), etc. The non-cooperative point of view leads to the study of the allocations which arise in a given non-cooperative environment. In this paper, I follow a non-cooperative approach.

Taking an axiomatic point of view, Owen (1977) presented a value for transfer utility games with coalition structure. Another axiomatic characterization was provided by Hart and Kurz (1983).

Owen assumed that this structure was exogenously given. Hart and Kurz (1983) reinterpreted the Owen value assuming that players form coalitions in order to improve their bargaining power.

Under both approaches, the main idea is that the coalitions play among themselves as individual agents in a game between coalitions, and the surplus obtained by each coalition is distributed among its members.

Recently, the Owen value has been non-cooperatively supported by Vidal-Puga and Bergantiños (2003) and Vidal-Puga (2005). In these papers, the players play a non-cooperative mechanism[1] in two stages: in the first stage, the players inside a coalition bargain among themselves the strategy to follow in the second stage, where bargaining takes place among coalitions.

In Vidal-Puga (2005) I generalize a previous mechanism of Hart and Mas-Colell (1996). In Hart and Mas-Colell's model, a player is randomly chosen in order to propose a payoff. If this proposal is not accepted by all the other players, the mechanism is played again under the same conditions with probability $\rho \in [0, 1)$. With probability $1 - \rho$, the proposer leaves the game and the mechanism is repeated with the rest of the players.

In Vidal-Puga (2005), this procedure is played in two stages. First, agreements are negotiated within coalitions and then through delegates among coalitions. In the first stage, a player is randomly chosen out of each coalition and proposes a payoff. Each proposal is voted by the rest of the members of its own coalition. If one of them rejects the proposal, the mechanism is either played again under the same conditions (probability ρ), or the proposer leaves the game and the mechanism is repeated with the rest of the players (probability $1 - \rho$). If there is no rejection, the proposal of one of the coalitions is randomly chosen. If this proposal is not accepted by all other coalitions, the mechanism is played again under the same conditions (probability ρ), or the entire proposing coalition leaves the game and the mechanism is repeated with the rest of the players (probability $1 - \rho$).

This mechanism in two stages implements the Owen value in a non-restrictive class of games (Vidal-Puga (2005)). Notice that each coalition is acting as a single unit in the second stage. The entire proposing coalition leaves the game when the proposal made by one of its members is rejected by the other players.

Frequently, it is interpreted that players form coalition structures in order to improve their bargaining strength (Hart and Kurz (1983)). However, as Harsanyi (1977, p. 203) points out, the bargaining strength does not improve

[1] To avoid ambiguities with cooperative games, I use the term *non-cooperative mechanism*, or simply *mechanism*, rather than non-cooperative game.

in general. An individual can be worse off bargaining as a member of a coalition than bargaining alone. Formally stated, the Harsanyi paradox[2] is as follows: Consider a simple n-person unanimity game in which n players can share a pie of size 1 as long as all of them agree on the division. Under a symmetric assumption, each player will typically expect to get a share of the pie of size $1/n$. Assume now two players decide to join forces and act as one single player. Harsanyi claims that this situation is equivalent to a symmetric $(n-1)$-person unanimity game and thus each player's expectation should be a pie of size $1/(n-1)$. Hence, by joining forces, the two players have moved from a joint expectation of $2/n$ to an expectation of just $1/(n-1)$. Of course the same result holds if more than two players decide to act as one player (except in the trivial case in which *all* n players participate in this agreement).

This paradox seems somehow problematic. It implies that cooperation can be harmful in bargaining environments. Chae and Heidhues (2004, p. 47) provide the following explanation: By merging in a coalition structure, players reduce their multiple "rights to talk" to a single right in the game between coalitions, hence improving the position of the outsiders.

The meaning of "rights to talk" is not clear from an axiomatic viewpoint (see for example Chae and Moulin (2004)). However, it has a clear meaning in the mechanism in Vidal-Puga (2005). The right to talk is simply the right to make a proposal. This right is dispelled as the size of the coalition increases. For example in the n-person unanimity game where two players act as one unit, the proposal comes from one of the members of the joined coalition with a probability $1/(n-1)$, whereas when no coalition is formed the proposal would come from one of them with probability $2/n$.

In this paper, I study the effect that provides to maintain the "rights to talk" of the players inside a coalition. I modify the mechanism in Vidal-Puga (2005) so that players maintain their "rights to talk". Hence, the coalitions with more members have more chances to make proposals. In the previous

[2] Harsanyi calls it the joint-bargaining paradox.

example, this means that the proposal from a member of the joined coalition will come with a probability $1/n$, as if he were acting alone. However, the coalitions still bargain as single units: The entire proposing coalition leaves the game when its proposal is rejected.

As a consequence of this modification, the resulting equilibrium payoff coincides with the weighted Shapley value (Shapley, 1953a) in the game between coalitions, with weights given by the size of the coalition.

Moreover, the final outcome in unanimity games is not affected: The equilibrium payoffs would be the same irrespective of the coalition structure (see Proposition 4.5). However, this is not true in general games. Krasa, Temimi and Yannelis (2003) recently presented an intriguing example with three players in which the benefit of joining a coalition critically depend on informational asymmetries. More specifically, when information is complete, players 1 and 2 find it advantageous to bargain as one unit. However, when players 1 and 2 lack certain information that is only available to the outside party, they are better off bargaining separately (even though in either case they are in a weaker position than before). With the proposed modification, the final outcome seems much more intuitive: Players 1 and 2 are still in a weaker position when information is not complete; however, bargaining as one unit is *always* advantageous, and the benefit of joining forces in the differential information case is *exactly the same* as when information is complete.

The new proposed mechanism is still a generalization of the mechanism of Hart and Mas-Colell (1996), in the sense that they coincide when the coalition structure is trivial (i.e. all the coalitions are singletons, or there exists a unique coalition).

The notation used throughout the paper and some previous results are presented in Section 2. The new coalitional value is presented in Section 3. The formal mechanism and the main result are presented in Section 4. In Section 5 the example presented by Krasa, Temimi and Yannelis (2003) is analyzed. Finally, Section 6 is devoted to a brief discussion.

2 Preliminaries

Let $U = \{1, 2, ...\}$ be a (may be infinite) set of potential *players*. A *non-transferable utility game*, or *NTU game*, is a pair (N, V) where $N \subseteq U$ is finite and V is a correspondence which assigns to each $S \subseteq N$, $S \neq \emptyset$ a nonempty, closed, convex and bounded-above subset $V(S) \subset \mathbb{R}^S$ representing all the possible payoffs that the members of S can obtain for themselves when playing cooperatively. For $S \subset N$, I maintain the notation V when referring to the application V restricted to S as player set. For simplicity, I denote $V(i)$ instead of $V(\{i\})$, $S \cup i$ instead of $S \cup \{i\}$ and $N \backslash i$ instead of $N \backslash \{i\}$. I denote the set of NTU games as NTU.

For each $i \in N$, let $r_i := \max \{x : x \in V(i)\}$.

When
$$V(S) = \left\{ x \in \mathbb{R}^S : \sum_{i \in S} x_i \leq v(S) \right\}$$
for some $v : 2^N \to \mathbb{R}$ with $v(\emptyset) = 0$, I say that (N, V) is a *transferable utility game* (or *TU game*) and I represent it as (N, v). As before, I maintain the notation v when referring to the application v restricted to 2^S.

A TU game is *superadditive* if it satisfies $v(S) + v(T) \leq v(S \cup T)$ for all $S, T \subset N$ with $S, T \neq \emptyset$, $S \cap T = \emptyset$. A TU game is *convex* if it satisfies $v(T \cup i) - v(T) \leq v(S \cup i) - v(S)$ for all $i \in N$ and $T \subset S \subseteq N \backslash i$. If the previous inequalities are strict, the TU game is *strictly superadditive* and *strictly convex*, respectively. All (strictly) convex TU games are (strictly) superadditive. A *unanimity game* is a TU game satisfying $v(N) = 1$ and $v(S) = 0$ otherwise. All unanimity games are convex.

When $V(S) = \{r^S\}$ for all $S \subsetneq N$, where $r_i^S = r_i$ for all $i \in S$, and $r^N \in V(N)$, I say that (N, V) is a *pure bargaining problem*.

Unanimity games are both TU games and pure bargaining problems.

Given $N \subseteq U$ finite, I call *coalition structure* over N a partition of the player set, i.e. $\mathcal{C} = \{C_1, C_2, ..., C_m\} \subset 2^N$ is a coalition structure if it satisfies $\bigcup_{C_q \in \mathcal{C}} C_q = N$ and $C_q \cap C_r = \emptyset$ when $q \neq r$. I also assume $C_q \neq \emptyset$ for all q. A coalition structure $\mathcal{C}$ over N is *trivial* if either $\mathcal{C} = \{\{i\}\}_{i \in N}$ or $\mathcal{C} = \{N\}$.

For any $S \subset N$, I denote the restriction of $\mathcal{C}$ to the players in S as $\mathcal{C}_S$ (notice that this implies that $\mathcal{C}_S$ may have less or the same number of coalitions as $\mathcal{C}$). Given a TU game (N, v) and a coalition structure $\mathcal{C} = \{C_1, C_2, ..., C_m\}$ over N, the *game between coalitions* is the TU game $(M, v/\mathcal{C})$ where $M = \{1, 2, ..., m\}$ and $v/\mathcal{C}(Q) = v\left(\bigcup_{q \in Q} C_q\right)$ for all $Q \subseteq M$.

I denote an NTU game (N, V) with coalition structure $\mathcal{C}$ over N as $(N, V, \mathcal{C})$. I denote the set of NTU games with coalition structure as $CNTU$.

Given G is a subset of NTU or $CNTU$, a *value* in G is a correspondence ψ which assigns to each $(N, V) \in G$ or $(N, V, \mathcal{C}) \in G$ a vector $\psi^N(V) \in \mathbb{R}^N$. With a slight abuse of notation, I say that $\psi^N(V)$ is the value of (N, V), and each $\psi_i^N(V)$ is the value of i. A value ψ is *efficient* if $\psi^N(V)$ belongs to the upper boundary of $V(N)$ for all (N, V). For any TU game (N, v), this condition is equivalent to say $\sum_{i \in N} \psi_i^N(v) = v(N)$.

Two well-known efficient values in TU games and in bargaining problems are respectively the *Shapley value* (Shapley (1953b)) and the *Nash solution* (Nash (1950)). I denote the Shapley value of the TU game (N, v) as $\varphi^N(v) \in \mathbb{R}^N$. A simple inductive method to compute the Shapley value of (N, v) is as follows: $\varphi_i^{\{i\}}(v) = r_i$ for all $i \in N$. Assume we know $\varphi^T(v) \in \mathbb{R}^T$ for all $T \subsetneq S$. Then,

$$\varphi_i^S(v) = \frac{1}{|S|}\left[v(S) - v(S \backslash i) + \sum_{j \in S \backslash i} \varphi_i^{S \backslash j}(v)\right]$$

or equivalently, by efficiency,

$$\varphi_i^S(v) = \frac{1}{|S|}\left[v(S) + \sum_{j \in S \backslash i}\left(\varphi_i^{S \backslash j}(v) - \varphi_j^{S \backslash i}(v)\right)\right] \tag{1}$$

for all $i \in S$.

In NTU games that are both TU games and pure bargaining problems, the Shapley value and the Nash solution coincide. In unanimity games, $\varphi_i^N(v) = 1/|N|$ for all $i \in N$.

A non symmetric generalization of $\varphi^N(v)$ is the *weighted Shapley value* (Shapley (1953a), Kalai and Samet (1987, 1988)). I denote the weighted

Shapley value of the TU game (N,v) as $\varphi^{\omega N}(v) \in \mathbb{R}^N$, where $\omega \in \mathbb{R}^N_{++}$ is a vector of weights. Pérez-Castrillo and Wettstein (2001, Lemma 1) proved that the weighted Shapley value can be inductively computed as follows: $\varphi_i^{\omega\{i\}}(v) = r_i$ for all $i \in N$. Assume we know $\varphi^{\omega T}(v) \in \mathbb{R}^T$ for all $T \subsetneq S$. Then,

$$\varphi_i^{\omega S}(v) = \frac{1}{\sum_{j \in S} \omega_j} \left[\omega_i v(S) - \omega_i v(S \backslash i) + \sum_{j \in S \backslash i} \omega_j \varphi_i^{\omega S \backslash j}(v) \right] \tag{2}$$

or equivalently, by efficiency,

$$\varphi_i^{\omega S}(v) = \frac{1}{\sum_{j \in S} \omega_j} \left[\omega_i v(S) + \sum_{j \in S \backslash i} \left(\omega_j \varphi_i^{\omega S \backslash j}(v) - \omega_i \varphi_j^{\omega S \backslash i}(v) \right) \right] \tag{3}$$

for all $i \in N$.

As it becomes clear from the previous formulas, when $\omega_i = \omega_j$ for all i, j, the weighted Shapley value coincides with the Shapley value. Moreover, when $\omega, \omega' \in \mathbb{R}^N_{++}$ are two weight vectors such that there exists $\alpha > 0$ with $\omega_i = \alpha \omega'_i$ for all $i \in N$, then $\varphi^{\omega N}(v) = \varphi^{\omega' N}(v)$.

The weight vector breaks the symmetric treatment of players in a TU game, but they should not be interpreted as a measure of bargaining power. In particular, Owen (1968) presented a simple example in which one of the players was worse-off when his weight increased. See, for example, Haeringer (2000, Section 4).

However, for convex games, a higher weight never implies a lower weighted Shapley value (Haeringer (2000, Section 4)).

I now focus on TU games with coalition structure. Fix $\mathcal{C} = \{C_1, ..., C_m\}$ and $M = \{1, ..., m\}$. Owen (1977) proposed an efficient value based on Shapley's which takes into account the coalition structure. I call this value the *Owen coalitional value*, or simply the *Owen value*. The Owen value is defined as follows: Given $C_q \in \mathcal{C}$, the *reduced TU game* (C_q, v_q) is defined as $v_q(T) := \varphi_q^M \left(v / \mathcal{C}_{N \backslash (C_q \backslash T)} \right)$ for all $T \subseteq C_q$. Thus, each $v_q(T)$ is the Shapley value of the coalition T in the game between coalitions assuming that the members of $C_q \backslash T$ are out. The Owen value is then defined as

$$\phi_i^N(v) := \varphi_i^{C_q}(v_q)$$

for all $i \in C_q$.

The interpretation of this definition is as follows: Players in C_q should divide $\varphi_q^M(v/\mathcal{C})$, which is their value in the game between coalitions. In order to compute the contribution of each player, a new game is defined, where the worth of a (sub)coalition $T \subseteq C_q$ is the value that T would get should the other players in C_q are not present and T plays the role of $q \in M$ in the game between coalitions.

Obviously, when the coalition structure is trivial, the Owen value coincides with the Shapley value.

Levy and McLean (1989) studied the *weighted coalitional value with intracoalitional symmetry*. This value is defined as follows (Levy and LcLean (1989, Proposition C(2))). Given a vector of weights $\omega \in \mathbb{R}_{++}^M$ for the coalitions and $C_q \in \mathcal{C}$, the *weighted reduced TU game* $\left(C_q, v_q^\omega\right)$ is defined as $v_q^\omega(T) := \varphi_q^{\omega M}\left(v/\mathcal{C}_{N\backslash(C_q\backslash T)}\right)$ for all $T \subseteq C_q$. The weighted coalitional value with intracoalitional symmetry, with weights given by ω, is defined as

$$\phi_i^{\omega N}(v) := \varphi_i^{C_q}\left(v_q^\omega\right)$$

for all $i \in N$.

The interpretation of this definition is as before. However, in this case the coalitions are not treated symmetrically in the game between coalitions.

When $\mathcal{C} = \{\{i\}\}_{i \in N}$, this value coincides with the weighted Shapley value. When $\mathcal{C} = \{N\}$, it coincides with the Shapley value. When $\omega_q = \omega_r$ for all q, r, it coincides with the Owen value.

When there is no ambiguity, I write φ^N, ϕ^N, $\varphi^{\omega N}$, $\phi^{\omega N}$ instead of $\varphi^N(v)$, $\phi^N(v)$, $\varphi^{\omega N}(v)$, $\phi^{\omega N}(v)$.

I now define formally the Harsanyi paradox. Given $C_q, C_r \in \mathcal{C}$, I define the coalition structure $\mathcal{C}^{q+r}$ as $(\mathcal{C} \backslash \{C_q, C_r\}) \cup \{C_q \cup C_r\}$. This means that the coalition structure $\mathcal{C}^{q+r}$ arises from $\mathcal{C}$ when coalitions C_q, C_r join forces and act as a single coalition $C_q \cup C_r$. Let ψ be a value defined on $G \subseteq CNTU$. Just in this case, I write $\psi^N(\mathcal{C})$ and $\psi^N(\mathcal{C}^{q+r})$ when the coalition structure is given by $\mathcal{C}$ and $\mathcal{C}^{q+r}$, respectively. I say that ψ is *joint-monotonic* in G if

$$\sum_{i \in C_q \cup C_r} \psi_i^N(\mathcal{C}) \leq \sum_{i \in C_q \cup C_r} \psi_i^N\left(\mathcal{C}^{q+r}\right)$$

for all $(N, V, \mathcal{C}) \in G$ and all $C_q, C_r \in \mathcal{C}$. I say that a value yields the *Harsanyi paradox* if it is not joint-monotonic in unanimity games. It is well-known that the Owen value is not joint-monotonic in unanimity games. The Shapley value is joint-monotonic in all TU games, but this is because φ does not take into account the coalition structure[3].

When a value is not joint-monotonic, the members of a coalition can be better off acting alone than acting as a single unit that tries to improve its members' aggregate payoff (*cf.* the explanation given by Harsanyi (1977, p. 204-205)).

3 A new coalitional value

One feature of the Owen value is that the aggregate value received by each coalition depends only on the game between coalitions $v/\mathcal{C}$. In fact, this is one of the properties that Owen (1977, Axiom A3) uses to characterize ϕ. Hart and Kurz (1983, p.1051) consider that this property "is the most difficult to accept", and propose an alternative characterization without it.

An important consequence of this property, together with symmetry, is that two coalitions that affect the game between coalitions in a symmetric way will receive the same aggregate payoff. Levy and McLean (1989, p.235) claim that this intercoalitional symmetry may not be a reasonable requirement for a value. A classical example (Kalai and Samet (1987)) is the case where coalitions represent groups of different size. In these cases it seems reasonable to assign a size-depending weight to each coalition. A natural way to proceed is to give each coalition a weight proportional to its size (see

[3] For the same reason, the Nash solution is joint-monotonic in pure bargaining problems.

Kalai and Samet (1987, Section 7) for additional arguments supporting this particular choice).

An obvious candidate is the Levy-McLean value $\phi^{\omega N}$ with weights ω given by $\omega_q = |C_q|$ for each $C_q \in \mathcal{C}$. However, we should be cautious with the definition of the weighted reduced TU game $\left(C_q, v_q^{\omega}\right)$. Remark that $v_q^{\omega}(T) = \varphi_q^{\omega M}\left(v/\mathcal{C}_{N\setminus(C_q\setminus T)}\right)$ is interpreted as the value that T would get should $C_q\setminus T$ be not present and T play the role of $q \in M$. However, *coalition* T *has size* $|T| \leq |C_q|$ *in* $\mathcal{C}_{N\setminus(C_q\setminus T)}$. Under the above interpretation, if players in $C_q\setminus T$ are not present, there is no reason to assume that the weight of $q \in M$ remains unchanged.

In order to take into account the *real* size of the subcoalitions, let $\lambda \in \mathbb{R}^M_{++}$ be the weight system given by $\lambda_q = |T|$ and $\lambda_r = |C_r|$ otherwise. A new reduced TU game $\left(C_q, v_q^{*N}\right)$ is defined as

$$v_q^{*N}(T) := \varphi_q^{\lambda M}\left(v/\mathcal{C}_{N\setminus(C_q\setminus T)}\right)$$

for all $T \subseteq C_q$. Thus, each $v_q^{*N}(T)$ is the weighted Shapley value, with weights given by the size of each coalition, of coalition T in the game between coalitions assuming that the members of $C_q\setminus T$ are out.

Remark 3.1 *Another possible interpretation of the worth of* T *is that players in* $C_q\setminus T$ *are present, but they do not take part in the negotiation with the other coalitions. Hence, coalition* T *maintains its weight in the game between coalitions. In this case, the weighted reduced TU game* $\left(C_q, v_q^{\omega}\right)$ *makes sense.*

I denote the resulting coalitional value as ζ. The formal definition is as follows:

Definition 3.1 *Given a TU game with coalition structure* $(N, v, \mathcal{C})$*, the value* ζ *is defined as*

$$\zeta_i^N(v) := \varphi_i^{C_q}\left(v_q^{*N}\right)$$

for all $i \in C_q \in \mathcal{C}$.

As usual, I write ζ^N instead of $\zeta^N(v)$.

Remark that the TU game $\left(C_q, v_q^{*N}\right)$ is not a weighted reduced TU game $\left(C, v_q^{\omega}\right)$, and the value ζ is not a weighted coalitional value with intracoalitional symmetry. The weights λ that appear in the definition of $v_q^{*N}(T)$ depend on T, whereas in the definition of v_q the weights are the same for each possible T.

From now on, I consider the normalized version of λ for each $S \subseteq N$, that is

$$\lambda_q^S = \frac{\left|C'_q\right|}{|S|}$$

for all $S \subseteq N$ and all $C'_q \in \mathcal{C}_S$.

In the next proposition I describe an inductive formula to compute ζ.

Proposition 3.1 *Let $(N, v, \mathcal{C})$ be a TU game with coalition structure. Then, ζ can be defined inductively as follows: $\zeta_i^{\{i\}} = r_i$ for all $i \in N$. Assume we know $\zeta^T \in \mathbb{R}^T$ for all $T \subsetneq S$. Then, $\zeta_i^S =$*

$$\frac{1}{\left|C'_q\right|}\left[\lambda_q^S v(S) + \sum_{j \in C'_q \setminus i}\left(\zeta_i^{S \setminus j} - \zeta_j^{S \setminus i}\right) + \sum_{C'_r \in \mathcal{C}_S \setminus C'_q}\left(\lambda_r^S \sum_{j \in C'_q} \zeta_j^{S \setminus C'_r} - \lambda_q^S \sum_{j \in C'_r} \zeta_j^{S \setminus C'_q}\right)\right]$$

for all $i \in C'_q \in \mathcal{C}_S$.

Proof. The result is clear for $\zeta^{\{i\}}$. I prove the result for $(S, v, \mathcal{C}_S)$. Let $M' = \left\{q : C'_q \in \mathcal{C}_S\right\}$ and $m' = |M'|$.

Claim 3.1 *Given $i, j \in C'_q \in \mathcal{C}_S$, $\varphi_i^{C'_q \setminus j}\left(v_q^{*S}\right) = \varphi_i^{C'_q \setminus j}\left(v_q^{*S \setminus j}\right)$.*

It is enough to prove that $v_q^{*S}(T) = v_q^{*S \setminus j}(T)$ for all $T \subseteq C'_q \setminus j$. Given $T \subseteq C'_q \setminus j$,

$$\begin{aligned} v_q^{*S}(T) &= \varphi_q^{\lambda^{S \setminus \left(C'_q \setminus T\right)} M'}\left(v / \mathcal{C}_{S \setminus \left(C'_q \setminus T\right)}\right) \\ &= \varphi_q^{\lambda^{(S \setminus j) \setminus \left(\left(C'_q \setminus j\right) \setminus T\right)} M'}\left(v / \mathcal{C}_{(S \setminus j) \setminus \left(\left(C'_q \setminus j\right) \setminus T\right)}\right) = v_q^{*S \setminus j}(T). \end{aligned}$$

Claim 3.2 *Given $q, r \in M'$, $\varphi_q^{\lambda^S M' \setminus r}\left(v / \mathcal{C}_S\right) = v_q^{*S \setminus C'_r}\left(C'_q\right)$.*

The weights λ_q^S are proportional to the weights $\lambda_q^{S\backslash C_r'}$ for all $q \in M'\backslash r$. Hence,

$$\varphi_q^{\lambda^S M'\backslash r}\left(v/\mathcal{C}_S\right) = \varphi_q^{\lambda^{S\backslash C_r'} M'\backslash r}\left(v/\mathcal{C}_S\right).$$

Moreover, $v/\mathcal{C}_S\left(Q\right) = v/\mathcal{C}_{S\backslash C_r'}\left(Q\right)$ for all $Q \subseteq M'\backslash r$. Hence,

$$\varphi_q^{\lambda^{S\backslash C_r'} M'\backslash r}\left(v/\mathcal{C}_S\right) = \varphi_q^{\lambda^{S\backslash C_r'} M'\backslash r}\left(v/\mathcal{C}_{S\backslash C_r'}\right) = v_q^{*S\backslash C_r'}\left(C_q'\right).$$

Claim 3.3 *Given* $q, r \in M'$, $v_q^{*S\backslash C_r'}\left(C_q'\right) = \sum_{j\in C_q'} \zeta_j^{S\backslash C_r'}$.

By definition,

$$\sum_{j\in C_q'} \zeta_j^{S\backslash C_r'} = \sum_{j\in C_q'} \varphi_j^{C_q'}\left(v_q^{*S\backslash C_r'}\right) = v_q^{*S\backslash C_r'}\left(C_q'\right).$$

I now use the claims to prove the result. Given $i \in C_q' \in \mathcal{C}_S$,

$$\begin{aligned}
\zeta_i^S &= \varphi_i^{C_q'}\left(v_q^{*S}\right) \overset{(1)}{=} \frac{1}{\left|C_q'\right|}\left[v_q^{*S}\left(C_q'\right) + \sum_{j\in C_q'\backslash i}\left(\varphi_i^{C_q'\backslash j}\left(v_q^{*S}\right) - \varphi_j^{C_q'\backslash i}\left(v_q^{*S}\right)\right)\right] \\
&\overset{\text{(Claim 3.1)}}{=} \frac{1}{\left|C_q'\right|}\left[v_q^{*S}\left(C_q'\right) + \sum_{j\in C_q'\backslash i}\left(\zeta_i^{S\backslash j} - \zeta_j^{S\backslash i}\right)\right]. \qquad (4)
\end{aligned}$$

Taking into account that $\sum_{r\in M'} \lambda_r^S = 1$, $v_q^{*S}\left(C_q'\right) =$

$$\begin{aligned}
&\varphi_q^{\lambda^S M'}\left(v/\mathcal{C}_S\right) \overset{(3)}{=} \lambda_q^S v/\mathcal{C}_S\left(M'\right) + \sum_{r\in M'\backslash q}\left(\lambda_r^S \varphi_q^{\lambda^S M'\backslash r}\left(v/\mathcal{C}_S\right) - \lambda_q^S \varphi_r^{\lambda^S M'\backslash q}\left(v/\mathcal{C}_S\right)\right) \\
&\overset{\text{(Claim 3.2)}}{=} \lambda_q^S v\left(S\right) + \sum_{r\in M'\backslash q}\left(\lambda_r^S v_q^{*S\backslash C_r'}\left(C_q'\right) - \lambda_q^S v_r^{*S\backslash C_q'}\left(C_r'\right)\right) \\
&\overset{\text{(Claim 3.3)}}{=} \lambda_q^S v\left(S\right) + \sum_{C_r'\in\mathcal{C}_S\backslash C_q'}\left(\lambda_r^S \sum_{j\in C_q'} \zeta_j^{S\backslash C_r'} - \lambda_q^S \sum_{j\in C_r'} \zeta_j^{S\backslash C_q'}\right). \qquad (5)
\end{aligned}$$

The result comes from combining (4) and (5). ■

From now on, fix $(N, v, \mathcal{C})$.

Corollary 3.1 *For any* $S \subseteq N$ *and* $C_q' \in \mathcal{C}_S$,

$$\sum_{i\in C_q'} \zeta_i^S = \lambda_q^S v\left(S\right) + \sum_{C_r'\in\mathcal{C}_S\backslash C_q'}\left(\lambda_r^S \sum_{j\in C_q'} \zeta_j^{S\backslash C_r'} - \lambda_q^S \sum_{j\in C_r'} \zeta_j^{S\backslash C_q'}\right).$$

Proof. Under Proposition 3.1,

$$\begin{aligned}\sum_{i\in C'_q}\zeta_i^S &= \lambda_q^S v(S)+\frac{1}{|C'_q|}\sum_{i\in C'_q}\sum_{j\in C'_q\backslash i}\left(\zeta_i^{S\backslash j}-\zeta_j^{S\backslash i}\right)\\ &\quad+\sum_{C'_r\in\mathcal{C}_S\backslash C'_q}\left(\lambda_r^S\sum_{j\in C'_q}\zeta_j^{S\backslash C'_r}-\lambda_q^S\sum_{j\in C'_r}\zeta_j^{S\backslash C'_q}\right)\end{aligned}$$

but

$$\sum_{i\in C'_q}\sum_{j\in C'_q\backslash i}\left(\zeta_i^{S\backslash j}-\zeta_j^{S\backslash i}\right)=\sum_{i,j\in C'_q}\left(\zeta_i^{S\backslash j}-\zeta_j^{S\backslash i}\right)=\sum_{i,j\in C'_q}\zeta_i^{S\backslash j}-\sum_{i,j\in C'_q}\zeta_j^{S\backslash i}=0$$

and hence the result holds. ■

The next corollary states that each coalition gets its weighted Shapley value of the game between coalitions, with weights given by their size.

Corollary 3.2 *For any $S\subseteq N$ and $C'_q\in\mathcal{C}_S$,*

$$\sum_{i\in C'_q}\zeta_i^S=\varphi_q^{\lambda^S M'}(v/\mathcal{C}_S)$$

where $M'=\{r:C'_r\in\mathcal{C}_S\}$.

Proof. I proceed by induction on $m'=|M'|$, the number of coalitions in $\mathcal{C}_S$. For $m'=1$, the efficiency of ζ and φ^{λ^S} makes the result. Assume the result is true for coalition structures of size less than m'. Under Claim 3.2 and Claim 3.3, $\sum_{i\in C'_q}\zeta_i^{S\backslash C'_r}=\varphi_q^{\lambda^S M'\backslash r}(v/\mathcal{C}_S)$. Under Corollary 3.1,

$$\begin{aligned}\sum_{i\in C'_q}\zeta_i^S &= \lambda_q^S v/\mathcal{C}(M')+\sum_{r\in M'\backslash q}\left(\lambda_r^S\varphi_q^{\lambda^S M'\backslash r}(v/\mathcal{C}_S)-\lambda_q^S\varphi_r^{\lambda^S M'\backslash q}(v/\mathcal{C}_S)\right)\\ &\overset{(3)}{=}\varphi_q^{\lambda^S M'}(v/\mathcal{C}_S).\end{aligned}$$

■

I now prove that with this new value the Harsanyi paradox disappears.

Proposition 3.2 *The value ζ is joint-monotonic in convex games.*

Proof. I proceed by induction on m, the size of $\mathcal{C}$. For $m = 2$, the result is trivial. Assume the result is true for coalition structures of size $m-1$. Let $C_q, C_r \in \mathcal{C}$. I assume wlog $q = m-1$ and $r = m$. Let $\mathcal{C}^* = \left\{C_1^*, C_2^*, ..., C_{m-1}^*\right\}$ where $C_p^* = C_p$ for all $p < m-1$ and $C_{m-1}^* = C_{m-1} \cup C_m$. Let $M^* = \{1, 2, ..., m-1\}$, and let $\omega \in \mathbb{R}^M, \omega^* \in \mathbb{R}^{M^*}$ be defined as $\omega_p = \omega_p^* = \frac{|C_p|}{|N|}$ for all $p < m-1$, $\omega_{m-1} = \frac{|C_{m-1}|}{|N|}$, $\omega_m = \frac{|C_m|}{|N|}$ and $\omega_{m-1}^* = \omega_{m-1} + \omega_m$. Under Corollary 3.2, it is enough to prove that

$$\varphi_{m-1}^{\omega M}(v/\mathcal{C}) + \varphi_m^{\omega M}(v/\mathcal{C}) \overset{?}{\leq} \varphi_{m-1}^{\omega^* M^*}(v/\mathcal{C}^*).$$

For simplicity, I denote $u = v/\mathcal{C}$ and $u^* = v/\mathcal{C}^*$.

Under (2),

$$\begin{aligned}
\varphi_{m-1}^{\omega M}(u) + \varphi_m^{\omega M}(u) &= \omega_{m-1}u(M) - \omega_{m-1}u(M\backslash(m-1)) \\
&\quad + \sum_{p \in M\backslash(m-1)} \omega_p \varphi_{m-1}^{\omega M\backslash p}(u) \\
&\quad + \omega_m u(M) - \omega_m u(M\backslash m) + \sum_{p \in M\backslash m} \omega_p \varphi_m^{\omega M\backslash p}(u) \\
&= \omega_{m-1}u(M) - \omega_{m-1}u(M\backslash(m-1)) \\
&\quad + \omega_m u(M) - \omega_m u(M\backslash m) \\
&\quad + \omega_m \varphi_{m-1}^{\omega M\backslash m}(u) + \omega_{m-1}\varphi_m^{\omega M\backslash(m-1)}(u) \\
&\quad + \sum_{p<m-1} \omega_p \left(\varphi_{m-1}^{\omega M\backslash p}(u) + \varphi_m^{\omega M\backslash p}(u)\right)
\end{aligned}$$

and

$$\begin{aligned}
\varphi_{m-1}^{\omega^* M^*}(u^*) &= \omega_{m-1}^* u^*(M^*) - \omega_{m-1}^* u^*(M^*\backslash(m-1)) + \sum_{p<m-1} \omega_p^* \varphi_{m-1}^{\omega^* M^*\backslash p}(u^*) \\
&= (\omega_{m-1} + \omega_m)u(M) - (\omega_{m-1} + \omega_m)u(M\backslash\{m-1, m\}) \\
&\quad + \sum_{p<m-1} \omega_p^* \varphi_{m-1}^{\omega^* M^*\backslash p}(u^*).
\end{aligned}$$

Under the induction hypothesis, we have $\varphi_{m-1}^{\omega M\backslash p}(u) + \varphi_m^{\omega M\backslash p}(u) \leq \varphi_{m-1}^{\omega M^*\backslash p}(u^*)$

for all $p < m-1$. Hence, it is enough to prove,

$$\begin{aligned}
&\omega_{m-1} u(M) - \omega_{m-1} u(M \backslash (m-1)) + \omega_m u(M) \\
&-\omega_m u(M \backslash m) + \omega_m \varphi_{m-1}^{\omega M \backslash m}(u) + \omega_{m-1} \varphi_m^{\omega M \backslash (m-1)}(u) \\
&\overset{?}{\leq} (\omega_{m-1} + \omega_m) u(M) - (\omega_{m-1} + \omega_m) u(M \backslash \{m-1, m\}).
\end{aligned}$$

Simplifying and rearranging terms,

$$\begin{aligned}
&\omega_{m-1} \left[u(M \backslash (m-1)) - u(M \backslash \{m-1, m\}) - \varphi_m^{\omega M \backslash (m-1)}(u) \right] \\
&+\omega_m \left[u(M \backslash m) - u(M \backslash \{m-1, m\}) - \varphi_{m-1}^{\omega M \backslash m}(u) \right]
\end{aligned}$$

must be nonnegative. In fact, both terms are. We check it for the second one (the first is analogous):

$$\varphi_{m-1}^{\omega M \backslash m}(u) \overset{?}{\leq} u(M \backslash m) - u(M \backslash \{m-1, m\}).$$

It is well-known (Kalai and Samet (1987, Theorem 1)) that the weighted Shapley value is a weighted average of marginal contributions. Since (N, v) is convex, the TU game $(M \backslash m, u)$ is convex too. This implies that the maximal marginal contribution of $m-1$ in $(M \backslash m, u)$ is $u(M \backslash m) - u(M \backslash \{m-1, m\})$. Hence we conclude the result. ■

Proposition 3.2 does not hold in general for nonconvex games, as the next example shows:

Example 3.1 *Let* $N = \{1,2,3,4,5\}$ *and* v *be defined as* $v(\{1\}) = v(\{2\}) = v(T) = 0$, $v(\{1,2\}) = v(\{1,2\} \cup T) = 360$ *and* $v(\{1\} \cup T) = v(\{2\} \cup T) = 180$ *for all* $T \subseteq \{3,4,5\}$, $T \neq \emptyset$. *This TU game is superadditive but not convex. Consider the coalition structure* $\mathcal{C} = \{\{1\},\{2\},\{3,4\},\{5\}\}$, *i.e. players* 3 *and* 4 *form coalition. Then,*

$$\zeta^N = (147, 147, 12, 12, 42).$$

Consider now the coalition structure $\mathcal{C}^* = \{\{1\},\{2\},\{3,4,5\}\}$, *i.e. player* 5 *joins forces with coalition* $\{3,4\}$. *Then,*

$$\zeta^N = (153, 153, 18, 18, 18).$$

I now present a technical property that will be used in the next section. This property has the flavor of the balanced contributions property of Myerson's (1980), and it is also satisfied by the Owen value (Calvo, Lasaga and Winter (1996), Bergantiños and Vidal-Puga (2005)):

Proposition 3.3 *For all $S \subseteq N$ and $i \in C'_q \in \mathcal{C}_S$,*

$$\sum_{j \in C'_q \backslash i} \left(\zeta_i^S - \zeta_i^{S \backslash j} \right) = \sum_{j \in C'_q \backslash i} \left(\zeta_j^S - \zeta_j^{S \backslash i} \right).$$

Proof. Under Proposition 3.1 and Corollary 3.1,

$$\zeta_i^S = \frac{1}{\left| C'_q \right|} \left[\sum_{j \in C'_q} \zeta_j^S + \sum_{j \in C'_q \backslash i} \left(\zeta_i^{S \backslash j} - \zeta_j^{S \backslash i} \right) \right].$$

We have then

$$\begin{aligned} \left| C'_q \right| \zeta_i^S &= \sum_{j \in C'_q} \zeta_j^S + \sum_{j \in C'_q \backslash i} \left(\zeta_i^{S \backslash j} - \zeta_j^{S \backslash i} \right) \\ &= \sum_{j \in C'_q \backslash i} \zeta_j^S + \zeta_i^S + \sum_{j \in C'_q \backslash i} \left(\zeta_i^{S \backslash j} - \zeta_j^{S \backslash i} \right) \\ &= \sum_{j \in C'_q \backslash i} \left(\zeta_j^S - \zeta_i^S \right) + \left| C'_q \right| \zeta_i^S + \sum_{j \in C'_q \backslash i} \left(\zeta_i^{S \backslash j} - \zeta_j^{S \backslash i} \right) \end{aligned}$$

and hence

$$\sum_{j \in C'_q \backslash i} \left(\zeta_i^S - \zeta_j^S \right) = \sum_{j \in C'_q \backslash i} \left(\zeta_i^{S \backslash j} - \zeta_j^{S \backslash i} \right)$$

from where the result is easily deduced. ■

The next proposition states that, in strictly convex TU games, the aggregate payoff in a coalition is higher than when one of its members leaves and gets his autarky payoff.

Proposition 3.4 *For strictly convex TU games,*

$$\sum_{j \in C'_q \backslash i} \zeta_j^{S \backslash i} + r_i < \sum_{j \in C'_q} \zeta_j^S$$

for all $i \in C'_q \in \mathcal{C}_S$, $S \neq \{i\}$.

Proof. Let $M' = \{r : C'_r \in \mathcal{C}_S\}$. Since the game is strictly convex, $(M', v/\mathcal{C}_S)$ is also strictly convex and thus strictly superadditive. Assume first $C'_q = \{i\}$ (hence $\sum_{j \in C'_q \backslash i} \zeta_j^{S\backslash i} = 0$). Under Corollary 3.2, it is enough to prove $r_i < \varphi_q^{\lambda^S M'}(v/\mathcal{C}_S)$, which is straightforward given the strict superadditivity of $(M', v/\mathcal{C}_S)$ and the fact that $\varphi_q^{\lambda^S}$ is a weighted average of marginal contributions.

Assume now $C'_q \neq \{i\}$. Under Corollary 3.2, it is enough to prove

$$\varphi_q^{\lambda^{S\backslash i} M'}\left(v/\mathcal{C}_{S\backslash i}\right) + r_i \overset{?}{<} \varphi_q^{\lambda^S M'}\left(v/\mathcal{C}_S\right).$$

It is straightforward to check that $\lambda_r^{S\backslash i} = \frac{|S|}{|S|-1}\lambda_r^S$ for all $r \in M'\backslash q$, whereas $\lambda_q^{S\backslash i} = \frac{\left|C'_q\right|-1}{\left|C'_q\right|}\frac{|S|}{|S|-1}\lambda_q^S$. Hence, when weights change from λ^S to $\lambda^{S\backslash i}$, coalition q reduces its relative weight in the game between coalitions. Since $\left(M', v/\mathcal{C}_{S\backslash i}\right)$ is strictly convex, $\varphi_q^{\lambda^{S\backslash i} M'}\left(v/\mathcal{C}_{S\backslash i}\right) \leq \varphi_q^{\lambda^S M'}\left(v/\mathcal{C}_{S\backslash i}\right)$.

Hence, it is enough to prove

$$\varphi_q^{\lambda^S M'}\left(v/\mathcal{C}_{S\backslash i}\right) + r_i \overset{?}{<} \varphi_q^{\lambda^S M'}\left(v/\mathcal{C}_S\right).$$

Consider the following TU games on M':

$$u_q(Q) = \begin{cases} 0 & \text{if } q \notin Q \\ r_i & \text{if } q \in Q \end{cases}$$

and $v'(Q) = v/\mathcal{C}_{S\backslash i}(Q) + u_q(Q)$ for all $Q \subseteq M'$.

Under strict superadditivity, $v'(Q) = v/\mathcal{C}_S(Q)$ if $q \notin Q$ and $v'(Q) < v/\mathcal{C}_S(Q)$ if $q \in Q$. It is well-known from Kalai and Samet (1985) that the weighted Shapley value is monotonic. Thus $\varphi_q^{\lambda^S M'}(v') < \varphi_q^{\lambda^S M'}(v/\mathcal{C}_S)$.

Since the weighted Shapley value satisfies additivity $\varphi_q^{\lambda^S M'}\left(v/\mathcal{C}_{S\backslash i}\right) + \varphi_q^{\lambda^S M'}(u_q) = \varphi_q^{\lambda^S M'}(v')$. Moreover, $\varphi_q^{\lambda^S M'}(u_q) = r_i$. Hence,

$$\varphi_q^{\lambda^S M'}\left(v/\mathcal{C}_{S\backslash i}\right) + r_i = \varphi_q^{\lambda^S M'}(v') < \varphi_q^{\lambda^S M'}\left(v/\mathcal{C}_S\right).$$

■

4 The noncooperative mechanism

In this section I describe the non-cooperative mechanism. This mechanism arises as a modification of the mechanism presented in Vidal-Puga (2005).

Even though the model is defined for NTU games, I focus on TU games and bargaining problems.

Fix $(N, V, \mathcal{C}) \in CNTU$. For each $S \subseteq N$, I denote by Γ_S the set of applications $\gamma : \mathcal{C}_S \to S$ satisfying $\gamma\left(C'_q\right) \in C'_q$ for each $C'_q \in \mathcal{C}_S$. For simplicity, I denote $\gamma_q := \gamma\left(C'_q\right)$.

The coalitional non-cooperative mechanism associated with $(N, V, \mathcal{C})$ and $\rho \in [0, 1)$ is defined as follows:

> In each round there is a set $S \subseteq N$ of active players. In the first round, $S = N$. Each round has one or two stages. In the first stage, a *proposer* is randomly chosen from each coalition. Namely, a function $\gamma \in \Gamma_S$ is randomly chosen, being each γ equally likely to be chosen. The coalitions play sequentially (say, for example, in the order $(C'_1, C'_2, ..., C'_{m'})$) in the following way: γ_1 proposes a feasible payoff, i.e. a vector in $V(S)$. The members of $C'_1 \backslash \gamma_1$ are then asked in some prespecified order to accept or reject the proposal. If one of them rejects the proposal, then we move to the next round where the set of active players is S with probability ρ and $S \backslash \gamma_1$ with probability $1 - \rho$. In the latter case, player γ_1 gets r_{γ_1}. If all the players accept the proposal, we move on to the next coalition, C'_2. Then, players of C'_2 proceed to repeat the process under the same conditions, and so on. If all the proposals are accepted in each coalition, the proposers are called *representatives*. We denote the proposal of γ_q as $a\left(S, \gamma_q\right) \in V(S)$.
>
> In the second stage, a proposal is randomly chosen. *The probability of* $a(S, \gamma_r)$ *being chosen is* λ_r^S, *i.e. proportional to the size of the coalition that supports it.* Assume $a\left(S, \gamma_q\right)$ is chosen.

We call player γ_q the *representative-proposer*, or simply *RP*. If all the members of $S \backslash C'_q$ accept $a\left(S, \gamma_q\right)$– they are asked in some prespecified order – then the game ends with these payoffs. If it is rejected by at least one member of $S \backslash C'_q$, then we move to the next round where, with probability ρ, the set of active players is again S and, with probability $1-\rho$, the entire coalition C'_q *drops out* and the set of active players becomes $S \backslash C'_q$. In the latter case each $i \in C'_q$ gets r_i.

Clearly, given any set of strategies, this mechanism finishes in a finite number of rounds with probability 1.

This mechanism coincides with the mechanism in Vidal-Puga (2005) except that the probability of a coalition to be chosen is proportional of its size[4]. With this modification, when there is no rejection *each player has the same probability to be chosen RP.* Hence, players do not loose their "right to talk" when joining a coalition.

The mechanism also generalizes Hart and Mas-Colell's (1996) for trivial coalition structures. For $\mathcal{C} = \{N\}$, the second stage is trivial, since there is a single representative and a single proposal. Moreover, the first stage coincides with Hart and Mas-Colell's mechanism. For $\mathcal{C} = \{\{i\}\}_{i \in N}$, the first stage is trivial. Each player states a proposal, and in the second stage a proposal is randomly selected with equal probability and voted by the rest of the players/coalitions.

As usual, I consider stationary subgame perfect equilibria. In this context, an equilibrium is stationary if the players' strategies depend only on the set of active players. They do not depend, however, on the previous history or the number of played rounds.

Before studying the general stationary subgame perfect equilibria, it is worthy to analyze a particular example.

[4] In Vidal-Puga (2005) each coalition is chosen with the same probability.

Example 4.1 *Let $N = \{1, 2, 3\}$ and v be defined as $v(\{1, 2\}) = v(N) = 1$ and $v(S) = 0$ otherwise. Consider the coalition structure $\mathcal{C} = \{\{1\}, \{2, 3\}\}$, i.e. players 2 and 3 form coalition.*

When there are two active players, the mechanism coincides with the mechanism given by Hart and Mas-Colell, and thus the expected final payoffs are $\zeta^{\{1,2\}} = \left(\frac{1}{2}, \frac{1}{2}\right)$ and $\zeta^{\{1,3\}} = \zeta^{\{2,3\}} = (0, 0)$.

Assume now the set of active players is N. For simplicity, assume $\rho = 0$. Then, player 1 would propose $a(N, 1) = (1, 0, 0)$, i.e. he offers the other players their respective continuation payoff after rejection in the second stage. The proposals given by player 2 and player 3 are subtler, because they would not *propose to each other their continuation payoff after rejection in the first stage. Instead, they propose to each other a value that,* averaging with player 1's proposal, *results in their respective continuation payoffs after rejection. In particular, player 3 would propose $a(N, 3) = \left(0, \frac{3}{4}, \frac{1}{4}\right)$, because (taking into account that player 1 would be the RP in the second stage with probability $\frac{1}{3}$) player 2's expected final payoff after rejection is $\frac{1}{3}0 + \frac{2}{3}\frac{3}{4} = \frac{1}{2}$. Analogously, player 2 would propose $a(N, 2) = (0, 1, 0)$.*

Once these proposals are accepted in the first stage, in the second stage the proposals are either $a(N, 1)$ and $a(N, 2)$ (probability $\frac{1}{2}$), or $a(N, 1)$ and $a(N, 3)$ (probability $\frac{1}{2}$). In the second stage, the final proposal will be $a(N, 1)$ with probability $\frac{1}{3}$, and either $a(N, 2)$ or $a(N, 3)$ with probability $\frac{2}{3}$. On average, the expected final payoff is:

$$\frac{1}{3}a(N, 1) + \frac{2}{3}\left(\frac{1}{2}a(N, 2) + \frac{1}{2}a(N, 3)\right) = \left(\frac{1}{3}, \frac{7}{12}, \frac{1}{12}\right) = \zeta^N.$$

I now analyze the general stationary subgame perfect equilibria. Let S denote the set of active players. Given a set of stationary strategies, I denote by $a(S, i)^\gamma \in V(S)$ the payoff proposed by $i \in C'_q \in \mathcal{C}_S$ when the set of proposers is determined by some $\gamma \in \Gamma_S$ with $\gamma_q = i$. Thus, for a given $\gamma \in \Gamma_S$,

$$a(S)^\gamma := \sum_{C'_q \in \mathcal{C}_S} \lambda_q^S a\left(S, \gamma_q\right)^\gamma \in V(S) \tag{6}$$

is the expected final payoff when all the proposals are accepted and γ determines the set of proposers (or representatives).

I denote

$$a(S) := \sum_{\gamma \in \Gamma_S} \frac{1}{|\Gamma_S|} a(S)^{\gamma} \in V(S) \tag{7}$$

as the expected final payoff when all the proposals are accepted.

Given $i \in C'_q \in \mathcal{C}_S$, let $\Gamma_{S,i}$ be the subset of functions $\gamma \in \Gamma_S$ such that $\gamma_q = i$. Notice that $|\Gamma_S| = |\Gamma_{S,i}| \left|C'_q\right|$ for all $i \in C'_q \in \mathcal{C}_S$.

Let

$$a(S,i) := \sum_{\gamma \in \Gamma_{S,i}} \frac{1}{|\Gamma_{S,i}|} a(S,i)^{\gamma} \tag{8}$$

be the expected payoff proposed by $i \in C'_q \in \mathcal{C}_S$ when he is a proposer.

The next proposition states that the probability that the final proposal comes from a particular player (when all the proposals are accepted) is equal for all the players, i.e. they maintain their respective "rights to talk".

Proposition 4.1 *for all $S \subseteq N$,*

$$a(S) = \sum_{i \in S} \frac{1}{|S|} a(S,i).$$

Proof. Given $S \subseteq N$,

$$a(S) \overset{(7)}{=} \sum_{\gamma \in \Gamma_S} \frac{1}{|\Gamma_S|} a(S)^{\gamma} \overset{(6)}{=} \sum_{\gamma \in \Gamma_S} \frac{1}{|\Gamma_S|} \sum_{C'_q \in \mathcal{C}_S} \lambda_q^S a\left(S, \gamma_q\right)^{\gamma}$$

$$= \sum_{C'_q \in \mathcal{C}_S} \lambda_q^S \sum_{\gamma \in \Gamma_S} \frac{1}{|\Gamma_S|} a\left(S, \gamma_q\right)^{\gamma} = \sum_{C'_q \in \mathcal{C}_S} \lambda_q^S \sum_{i \in C'_q} \frac{1}{\left|C'_q\right|} \sum_{\gamma \in \Gamma_{S,i}} \frac{1}{|\Gamma_{S,i}|} a\left(S, \gamma_q\right)^{\gamma}.$$

Since $a\left(S, \gamma_q\right)^{\gamma} = a(S,i)^{\gamma}$ for all $i \in C'_q$, $\gamma \in \Gamma_{S,i}$,

$$a(S) = \sum_{C'_q \in \mathcal{C}_S} \lambda_q^S \sum_{i \in C'_q} \frac{1}{\left|C'_q\right|} \sum_{\gamma \in \Gamma_{S,i}} \frac{1}{|\Gamma_{S,i}|} a(S,i)^{\gamma}.$$

Under (8),

$$a(S) = \sum_{C'_q \in \mathcal{C}_S} \lambda_q^S \sum_{i \in C'_q} \frac{1}{\left|C'_q\right|} a(S,i) = \sum_{C'_q \in \mathcal{C}_S} \frac{1}{|S|} \sum_{i \in C'_q} a(S,i) = \sum_{i \in S} \frac{1}{|S|} a(S,i).$$

■

Proposition 4.2 *Assume a set proposals* $\left(a\left(S,i\right)^{\gamma}_{i\in S,\gamma\in\Gamma_{S,i}}\right)_{S\subseteq N}$ *satisfies the following three conditions for all* $S\subseteq N$*:*

P-1 $a_j\left(S,i\right)^{\gamma}=\rho a_j\left(S\right)+\left(1-\rho\right)a_j\left(S\backslash C'_q\right)$ *for all* $i\in C'_q\in\mathcal{C}_S$*,* $\gamma\in\Gamma_{S,i}$ *and* $j\in S\backslash C'_q$*;*

P-2 $a_j\left(S\right)^{\gamma}=\rho a_j\left(S\right)+\left(1-\rho\right)a_j\left(S\backslash i\right)$ *for all* $i\in C'_q\in\mathcal{C}_S$*,* $\gamma\in\Gamma_{S,i}$ *and* $j\in C'_q\backslash i$*;*

P-3 $\sum_{j\in S}a_j\left(S,i\right)^{\gamma}=v\left(S\right)$ *for all* $i\in S$ *and* $\gamma\in\Gamma_{S,i}$*.*

Then, $a\left(S\right)=\zeta^S$ *for all* $S\subseteq N$*.*

Proof. By P-3,

$$\sum_{i\in S}a_i\left(S\right)=v\left(S\right). \tag{9}$$

Fix $i\in C'_q\in\mathcal{C}_S$. From (6) it is readily checked that, for any $j\in C'_q\backslash i$, $\gamma\in\Gamma_{S,i}$:

$$a_j\left(S,i\right)^{\gamma}=\frac{1}{\lambda^S_q}a_j\left(S\right)^{\gamma}-\sum_{C'_r\in\mathcal{C}_S\backslash C'_q}\frac{\lambda^S_r}{\lambda^S_q}a_j\left(S,\gamma_r\right)^{\gamma}.$$

Under P-1 and P-2, $a_j\left(S,i\right)^{\gamma}=$

$$\begin{aligned}&\frac{1}{\lambda^S_q}\left[\rho a_j(S)+(1-\rho)a_j(S\backslash i)\right]-\sum_{C'_r\in\mathcal{C}_S\backslash C'_q}\frac{\lambda^S_r}{\lambda^S_q}\left[\rho a_j(S)+(1-\rho)a_j(S\backslash C'_r)\right]\\ =\;&\rho a_j(S)+(1-\rho)\left[\frac{1}{\lambda^S_q}a_j(S\backslash i)-\sum_{C'_r\in\mathcal{C}_S\backslash C'_q}\frac{\lambda^S_r}{\lambda^S_q}a_j(S\backslash C'_r)\right].\end{aligned} \tag{10}$$

Under Proposition 4.1 and (8),

$$\begin{aligned}&\left|S\right|a_i\left(S\right)\overset{\text{(Proposition 4.1)}}{=}\sum_{j\in S}a_i\left(S,j\right)\overset{(8)}{=}\sum_{j\in S}\sum_{\gamma\in\Gamma_{S,j}}\frac{1}{\left|\Gamma_{S,j}\right|}a_i\left(S,j\right)^{\gamma}\\ =\;&\sum_{\gamma\in\Gamma_{S,i}}\frac{1}{\left|\Gamma_{S,i}\right|}a_i\left(S,i\right)^{\gamma}+\sum_{j\in C'_q\backslash i}\sum_{\gamma\in\Gamma_{S,j}}\frac{1}{\left|\Gamma_{S,j}\right|}a_i\left(S,j\right)^{\gamma}+\sum_{j\in S\backslash C'_q}\sum_{\gamma\in\Gamma_{S,j}}\frac{1}{\left|\Gamma_{S,j}\right|}a_i\left(S,j\right)^{\gamma}.\end{aligned}$$

I study the three terms one by one. For the first term:

$$
\begin{aligned}
&\sum_{\gamma\in\Gamma_{S,i}}\frac{1}{|\Gamma_{S,i}|}a_i(S,i)^\gamma \overset{\text{(P-3)}}{=} v(S)-\sum_{\gamma\in\Gamma_{S,i}}\frac{1}{|\Gamma_{S,i}|}\sum_{j\in S\backslash i}a_j(S,i)^\gamma\\
=\;& v(S)-\sum_{\gamma\in\Gamma_{S,i}}\frac{1}{|\Gamma_{S,i}|}\sum_{C'_r\in\mathcal{C}_S\backslash C'_q}\sum_{j\in C'_r}a_j(S,i)^\gamma-\sum_{\gamma\in\Gamma_{S,i}}\frac{1}{|\Gamma_{S,i}|}\sum_{j\in C'_q\backslash i}a_j(S,i)^\gamma\\
\overset{\text{(P-1)-(10)}}{=}\;& v(S)-\sum_{C'_r\in\mathcal{C}_S\backslash C'_q}\sum_{j\in C'_r}\left[\rho a_j(S)+(1-\rho)a_j\left(S\backslash C'_q\right)\right]\\
&-\sum_{j\in C'_q\backslash i}\left[\rho a_j(S)+(1-\rho)\left[\frac{1}{\lambda_q^S}a_j(S\backslash i)-\sum_{C'_r\in\mathcal{C}_S\backslash C'_q}\frac{\lambda_r^S}{\lambda_q^S}a_j(S\backslash C'_r)\right]\right]
\end{aligned}
$$

under (9), $\sum_{j\in S\backslash i}\rho a_j(S)=\rho(v(S)-a_i(S))$ and thus

$$
\begin{aligned}
=\;& v(S)-\rho(v(S)-a_i(S))-(1-\rho)\sum_{C'_r\in\mathcal{C}_S\backslash C'_q}\sum_{j\in C'_r}a_j\left(S\backslash C'_q\right)\\
&-(1-\rho)\sum_{j\in C'_q\backslash i}\left[\frac{1}{\lambda_q^S}a_j(S\backslash i)-\sum_{C'_r\in\mathcal{C}_S\backslash C'_q}\frac{\lambda_r^S}{\lambda_q^S}a_j(S\backslash C'_r)\right].
\end{aligned}
$$

For the second term:

$$
\begin{aligned}
&\sum_{j\in C'_q\backslash i}\sum_{\gamma\in\Gamma_{S,j}}\frac{1}{|\Gamma_{S,j}|}a_i(S,j)^\gamma \overset{(10)}{=}\\
&\sum_{j\in C'_q\backslash i}\left[\rho a_i(S)+(1-\rho)\left[\frac{1}{\lambda_q^S}a_i(S\backslash j)-\sum_{C'_r\in\mathcal{C}_S\backslash C'_q}\frac{\lambda_r^S}{\lambda_q^S}a_i(S\backslash C'_r)\right]\right]\\
=\;& \rho\left(\left|C'_q\right|-1\right)a_i(S)\\
&+(1-\rho)\left[\sum_{j\in C'_q\backslash i}\frac{1}{\lambda_q^S}a_i(S\backslash j)-\left(\left|C'_q\right|-1\right)\sum_{C'_r\in\mathcal{C}_S\backslash C'_q}\frac{\lambda_r^S}{\lambda_q^S}a_i(S\backslash C'_r)\right].
\end{aligned}
$$

For the third term:

$$
\begin{aligned}
&\sum_{j\in S\backslash C'_q}\sum_{\gamma\in\Gamma_{S,j}}\frac{1}{|\Gamma_{S,j}|}a_i(S,j)^\gamma \overset{\text{(P-1)}}{=} \sum_{C'_r\in\mathcal{C}_S\backslash C'_q}\sum_{j\in C'_r}\left[\rho a_i(S)+(1-\rho)a_i(S\backslash C'_r)\right]\\
=\;& \rho\left(|S|-\left|C'_q\right|\right)a_i(S)+(1-\rho)\sum_{C'_r\in\mathcal{C}_S\backslash C'_q}|C'_r|\,a_i(S\backslash C'_r).
\end{aligned}
$$

Hence, adding terms, $|S|\, a_i(S) =$

$$
\begin{aligned}
& v(S) - \rho v(S) - (1-\rho) \sum_{C'_r \in \mathcal{C}_S \backslash C'_q} \sum_{j \in C'_r} a_j \left(S \backslash C'_q\right) \\
& - \sum_{j \in C'_q \backslash i} (1-\rho) \left[\frac{1}{\lambda_q^S} a_j (S \backslash i) - \sum_{C'_r \in \mathcal{C}_S \backslash C'_q} \frac{\lambda_r^S}{\lambda_q^S} a_j (S \backslash C'_r) \right] \\
& + (1-\rho) \left[\sum_{j \in C'_q \backslash i} \frac{1}{\lambda_q^S} a_i (S \backslash j) + \sum_{C'_r \in \mathcal{C}_S \backslash C'_q} \frac{\lambda_r^S}{\lambda_q^S} a_i (S \backslash C'_r) \right] \\
& + \rho |S|\, a_i(S).
\end{aligned}
$$

Rearranging terms and dividing by $1-\rho$, $|S|\, a_i(S) =$

$$
\begin{aligned}
= \; & v(S) + \sum_{j \in C'_q \backslash i} \frac{1}{\lambda_q^S} \left(a_i (S \backslash j) - a_j (S \backslash i)\right) \\
& + \sum_{C'_r \in \mathcal{C}_S \backslash C'_q} \left(\frac{\lambda_r^S}{\lambda_q^S} \sum_{j \in C'_q} a_j (S \backslash C'_r) - \sum_{j \in C'_r} a_j \left(S \backslash C'_q\right) \right).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
a_i(S) \; = \; & \frac{\lambda_q^S}{\left|C'_q\right|} v(S) + \sum_{j \in C'_q \backslash i} \frac{1}{\left|C'_q\right|} \left(a_i (S \backslash j) - a_j (S \backslash i)\right) \\
& + \sum_{C'_r \in \mathcal{C}_S \backslash C'_q} \left(\sum_{j \in C'_q} \frac{\lambda_r^S}{\left|C'_q\right|} a_j (S \backslash C'_r) - \sum_{j \in C'_r} \frac{\lambda_q^S}{\left|C'_q\right|} a_j \left(S \backslash C'_q\right) \right).
\end{aligned}
$$

Under Proposition 3.1, $a(S) = \zeta^S$ is easily deduced following a standard induction argument. ■

Proposition 4.3 *A set of proposals* $\left(a(S,i)^{\gamma}_{i \in S, \gamma \in \Gamma_{S,i}}\right)_{S \subseteq N}$ *can be supported as a stationary subgame perfect equilibrium for strictly convex games if and only if they satisfy P-1, P-2 and P-3.*

Proof. The only nonstraightforward step is to verify that proposers cannot prefer being rejected, i.e. $\rho a_i(S) + (1-\rho) r_i < a_i(S)^{\gamma}$ for all $i \in S$ and

$\gamma \in \Gamma_{S,i}$. Under Proposition 4.2, $a(S) = \zeta^S$ for all S. Hence,

$$a_i(S,i)^\gamma \overset{\text{(P-3)}}{=} v(S) - \sum_{C'_r \in \mathcal{C}_S \setminus C'_q} \sum_{j \in C'_r} a_j(S,i)^\gamma - \sum_{j \in C'_q \setminus i} a_j(S,i)^\gamma$$

$$\begin{aligned}
&\overset{\text{(P-1)-(10)}}{=} v(S) - \sum_{C'_r \in \mathcal{C}_S \setminus C'_q} \sum_{j \in C'_r} \left[\rho \zeta_j^S + (1-\rho)\zeta_j^{S \setminus C'_q}\right] \\
&\quad - \sum_{j \in C'_q \setminus i} \left[\rho \zeta_j^S + (1-\rho)\left[\frac{1}{\lambda_q^S}\zeta_j^{S \setminus i} - \sum_{C'_r \in \mathcal{C}_S \setminus C'_q} \frac{\lambda_r^S}{\lambda_q^S}\zeta_j^{S \setminus C'_r}\right]\right] \\
&= v(S) - \rho \sum_{j \in S \setminus i} \zeta_j^S - (1-\rho)v(S) + (1-\rho) \\
&\quad \left(v(S) - \sum_{C'_r \in \mathcal{C}_S \setminus C'_q} \sum_{j \in C'_r} \zeta_j^{S \setminus C'_q} - \sum_{j \in C'_q \setminus i} \frac{1}{\lambda_q^S}\zeta_j^{S \setminus i} + \sum_{j \in C'_q \setminus i} \sum_{C'_r \in \mathcal{C}_S \setminus C'_q} \frac{\lambda_r^S}{\lambda_q^S}\zeta_j^{S \setminus C'_r}\right) \\
&= \rho \zeta_i^S + (1-\rho) \\
&\quad \left(v(S) + \sum_{C'_r \in \mathcal{C}_S \setminus C'_q}\left(\sum_{j \in C'_q} \frac{\lambda_r^S}{\lambda_q^S}\zeta_j^{S \setminus C'_r} - \sum_{j \in C'_r} \zeta_j^{S \setminus C'_q}\right) - \sum_{C'_r \in \mathcal{C}_S \setminus C'_q} \frac{\lambda_r^S}{\lambda_q^S}\zeta_i^{S \setminus C'_r} - \sum_{j \in C'_q \setminus i} \frac{1}{\lambda_q^S}\zeta_j^{S \setminus i}\right) \\
&\overset{\text{(Proposition 3.1)}}{=} \rho \zeta_i^S + (1-\rho)\left(\frac{\left|C'_q\right|}{\lambda_q^S}\zeta_i^S - \frac{1}{\lambda_q^S}\sum_{j \in C'_q \setminus i} \zeta_i^{S \setminus j} - \sum_{C'_r \in \mathcal{C}_S \setminus C'_q} \frac{\lambda_r^S}{\lambda_q^S}\zeta_i^{S \setminus C'_r}\right). \qquad (11)
\end{aligned}$$

Hence

$$\begin{aligned}
a_i(S)^\gamma &\overset{(6)}{=} \lambda_q^S a_i(S,i)^\gamma + \sum_{C'_r \in \mathcal{C}_S \setminus C'_q} \lambda_r^S a_i(S,\gamma_r)^\gamma \\
&\overset{\text{(P-1)}}{=} \lambda_q^S a_i(S,i)^\gamma + \sum_{C'_r \in \mathcal{C}_S \setminus C'_q} \lambda_r^S \left(\rho \zeta_i^S + (1-\rho)\zeta_i^{S \setminus C'_r}\right) \\
&\overset{(11)}{=} \rho \zeta_i^S + (1-\rho)\left(\left|C'_q\right|\zeta_i^S - \sum_{j \in C'_q \setminus i} \zeta_i^{S \setminus j}\right).
\end{aligned}$$

Thus, it is enough to prove

$$r_i \overset{?}{<} \left|C'_q\right|\zeta_i^S - \sum_{j \in C'_q \setminus i} \zeta_i^{S \setminus j}.$$

Under Proposition 3.3 and Proposition 3.4, I have

$$r_i \overset{\text{(Proposition 3.4)}}{<} \sum_{j\in C'_q} \zeta_j^S - \sum_{j\in C'_q\backslash i} \zeta_j^{S\backslash i} \overset{\text{(Proposition 3.3)}}{=} \left|C'_q\right| \zeta_i^S - \sum_{j\in C'_q\backslash i} \zeta_i^{S\backslash j}.$$

■

Proposition 4.4 *There always exists a stationary subgame perfect equilibrium for strictly convex games.*

Proof. Under Proposition 4.3, it is enough to prove that there exits a set of proposals satisfying P-1, P-2 and P-3. I define $a_i(\{i\}, i) = r_i$ for all $i \in S$. Assume I have defined $a(T, i)$ for all $i \in T \subsetneq S$. I define:

$$a_j(S,i)^\gamma = \rho a_j(S) + (1-\rho)\, a_j\left(S\backslash C'_q\right)$$

for all $i \in C'_q \in \mathcal{C}_S$, $\gamma \in \Gamma_{S,i}$ and $j \in S\backslash C'_q$;

$$a_j(S,i)^\gamma = \rho a_j(S) + (1-\rho)\left[\frac{1}{\lambda_q^S} a_j(S\backslash i) - \sum_{C'_r\in\mathcal{C}_S\backslash C'_q} \frac{\lambda_r^S}{\lambda_q^S} a_j(S\backslash C'_r)\right]$$

for all $i \in C'_q \in \mathcal{C}_S$, $\gamma \in \Gamma_{S,i}$ and $j \in C'_q\backslash i$; and

$$a_i(S,i)^\gamma = v(S) - \sum_{j\in S\backslash i} a_j(S,i)$$

for all $i \in S$ and $\gamma \in \Gamma_{S,i}$.

It is straightforward to check that these proposals satisfy P-1, P-2 and P-3. ■

Theorem 4.1 *There exists a unique stationary subgame perfect equilibrium payoff in strictly convex games, which equals ζ^N.*

Proof. It is an immediate consequence of Propositions 4.2, 4.3 and 4.4. ■

In general, the mechanism does not implement ζ for nonconvex games. Take $\rho = 0$. Take the TU game given in Example 3.1 with coalition structure

$\{\{1\},\{2\},\{3,4,5\}\}$. Assume the only equilibrium payoff is ζ^S for all $S \subsetneq N$. Some of these values are given in the following table:

S	ζ^S
$\{1,2\}$	$(180,180)$
$\{1,2,4,5\}$	$(150,150,30,30)$
$\{1,3,4,5\}$	$(45,45,45,45)$
$\{2,3,4,5\}$	$(45,45,45,45)$

I compute the equilibrium payoff when $S = N$. In the second stage of the mechanism, coalitions $\{1\}$ and $\{2\}$ would offer 45 to each player in $\{3,4,5\}$ (this is their continuation payoff after either coalition $\{1\}$ or coalition $\{2\}$ leaves the game). Assume that player 3 is the proposer of coalition $\{3,4,5\}$ in the first stage. Then, any acceptable proposal should satisfy $a_i(N,3)^\gamma = 180$ for all $i \in \{1,2\}$ and $a_j(N,3)^\gamma = 20$ for all $j \in \{4,5\}$ (so that $\frac{1}{5}a_j(N,1)^\gamma + \frac{1}{5}a_j(N,2)^\gamma + \frac{3}{5}a_j(N,3)^\gamma = 30$, that is, player j's continuation payoff after rejection). Hence $a_3(N,3)^\gamma \leq -40$. This leaves player 3 with a negative final expected payoff [5]. Hence, it is optimal for player 3 to make an unacceptable proposal and receive zero. The final equilibrium payoff would be $(150,150,20,20,20)$ in expected terms, whereas $\zeta^N = (153,153,18,18,18)$.

In equilibrium, making acceptable proposals is profitable if the conditions given in Proposition 3.4 hold. These conditions state that the aggregate payoff of the members of a coalition is higher than their aggregate payoff when one of its members (the proposer) leaves the game and receives r_i. This generates sufficient surplus to be profitable for the proposer to make an acceptable offer.

It is still possible to implement ζ for general TU games by imposing an additional feature to the mechanism: Assume that each excluded player i is charged with a penalty $p_i > 0$. Hence, the final payoff after exclusion is $r_i - p_i$. Under these circumstances, all the offers are accepted in equilibrium

[5] This payoff is at most -6, *not* -40, since with probabitity $\frac{2}{5}$ the offer in the second stage comes from coalition $\{1\}$ or coalition $\{2\}$.

as long as $\sum_{j\in C'_q}\zeta^S > \sum_{j\in C'_q\backslash i}\zeta^{S\backslash i}+r_i-p_i$ for all $S\subseteq N$ and $i\in C'_q\in\mathcal{C}_S$. Hence, for p high enough[6] the result in Theorem 4.1 holds for any TU game.

This penalty may have a justification in the model. As Hart and Mas-Colell (1996, Section 7) point out, $r_i = v(\{i\})$ may represent the total worth of player i assuming that he is the only member of the society and control a common resource, whereas $r_i - p_i$ (a lower amount) is what he would get if he leaves the society. However, I will not move in that direction, because the mechanism works well without penalty in both the example proposed by Krasa, Temimi and Yannelis (see next section) and unanimity games (see Proposition 4.5 below), where the Harsanyi paradox is defined.

The last result of this Section deals with pure bargaining problems:

Proposition 4.5 *There exists at least one stationary subgame perfect equilibrium in pure bargaining problems. Moreover, as ρ approaches 1, any stationary subgame perfect equilibrium payoffs $a(\rho)$ converge to the Nash solution.*

In particular, for unanimity games, the unique stationary subgame perfect equilibrium payoff is $x_i = 1/|N|$ for all $i\in N$ and any coalition structure.

Proof. Clearly, when the set of active players is $S\subsetneq N$, there exists a unique subgame perfect equilibrium payoff which equals r^S. Assume $S=N$. It is straightforward to check that the proposals corresponding to a stationary subgame perfect equilibrium are characterized by:

Q-1 $a_j(N,i) = \rho a_j(N) + (1-\rho) r_j$ for all $i,j\in N$, $i\neq j$; and

Q-2 $a(N,i)\in\partial V(N)$ for all $i\in N$.

Moreover, $a(N) = \frac{1}{|N|}\sum_{i\in N} a(N,i)$ (Proposition 4.1). These are the conditions in Proposition 1 in Hart and Mas-Colell (1996), and the result follows from Theorem 3 in Hart and Mas-Colell (1996). ■

[6] In the previous example, any $p_i > 6$ would suffice.

5 An eloquent example

Krasa, Temimi and Yannelis (2003) propose a three-person economy with differential information where two players bargain as one unit against the third one. When there is complete information, the economy can be expressed as a TU game (N, v) where $N = \{1, 2, 3\}$ and v is given by

$$\begin{aligned} v(\{1\}) &= v(\{2\}) = 1 \\ v(\{3\}) &= \frac{43}{16} \\ v(\{1,2\}) &= \frac{5}{2} \\ v(\{1,3\}) &= v(\{2,3\}) = \frac{31}{8} \\ v(N) &= \frac{83}{16}. \end{aligned}$$

When there is differential information, due to incentive incompatibility, 1 and 2 are only able to achieve $v(\{1,2\}) = 2$ by themselves. For any other $S \subseteq N$, $v(S)$ is the same as under complete information.

Krasa, Temimi and Yannelis take the Owen value ϕ^N as a measure of players' expectations when 1 and 2 join forces. Their result is that bargaining as one unit is advantageous if and only if information is complete, as the next table shows:

ϕ^N	complete information	differential information
$\mathcal{C} = \{\{1\}, \{2\}, \{3\}\}$	$\left(\frac{39}{32}, \frac{39}{32}, \frac{88}{32}\right)$	$\left(\frac{109}{96}, \frac{109}{96}, \frac{280}{96}\right)$
$\mathcal{C} = \{\{1,2\}, \{3\}\}$	$\left(\frac{40}{32}, \frac{40}{32}, \frac{86}{32}\right)$	$\left(\frac{108}{96}, \frac{108}{96}, \frac{282}{96}\right)$.

Consider now that we take ζ^N as a measure of players' expectations when 1 and 2 join forces. Then, bargaining as one unit is advantageous in any case, as the next table shows:

ζ^N	complete information	differential information
$\mathcal{C} = \{\{1\}, \{2\}, \{3\}\}$	$\left(\frac{39}{32}, \frac{39}{32}, \frac{88}{32}\right)$	$\left(\frac{109}{96}, \frac{109}{96}, \frac{280}{96}\right)$
$\mathcal{C} = \{\{1,2\}, \{3\}\}$	$\left(\frac{40}{32}, \frac{40}{32}, \frac{86}{32}\right)$	$\left(\frac{112}{96}, \frac{112}{96}, \frac{274}{96}\right)$.

This last situation corresponds to the assumption that players, by joining, do not loose their respective "rights to talk". Note also that the benefit from cooperation is $\frac{1}{32}$ for each player in both cases.

It may be argued that the noncooperative approach that supports ζ is not acceptable here. Certainly, this game is not strictly convex, and thus the condition of Theorem 4.1 does not hold. However, the condition of strictly convexity is only used in the proof of Proposition 3.4, whose result still holds for this game: For (say) player 1, $\zeta_1^{\{1,3\}} = \frac{35}{32}$. Hence

$$\zeta_1^{\{1,3\}} + r_2 = \frac{35}{32} + 1 < \frac{39}{16} = \sum_{i \in \{1,2\}} \zeta_i^N$$

under complete information, and

$$\zeta_1^{\{1,3\}} + r_2 = \frac{35}{32} + 1 < \frac{109}{48} = \sum_{i \in \{1,2\}} \zeta_i^N$$

under differential information.

Hence the result stated in Theorem 4.1 still holds for this game.

6 Discussion

The Owen value seems to be a good measure of players' expectations when the coalition structure is exogenously given. For example, wage bargaining between firms and labor unions, tariff bargaining between countries, bargaining between the member states of a federated country, etc. In these situations, players do not have to wonder whether they would do it better bargaining as a unit, because it is something out of their control.

On the other hand, Hart and Kurz (1983) followed the idea that players form coalition structures in order to improve their bargaining strength. They studied four reasonable properties, or axioms, that determine uniquely the Owen value. The only property that is not satisfied by ζ is Carrier (Hart and Kurz (1983, p. 1051)). The Carrier axiom in Hart and Kurz has two parts. The first part (i) can also be split into two properties: efficiency (the value is efficient for all coalition structures) and *dummy* (null players[7] get zero). ζ satisfies efficiency, but not *dummy.* The second part (ii) states that moving null players[8] does not affect the outcome of the rest of the agents. I will contest this property.

In bargaining problems, asymmetries in the final outcome may be due to the players' different bargaining powers. As Binmore (1998, p. 80) points out: "Bargaining powers are determined by the strategic advantages conferred on players by the circumstances under which they bargain." In this case, the coalition structure. Assume for example a game in which all the players are mutually substitutes[9]. Since no asymmetries are introduced in the model, the expectation *a priori* should be the same for substitute players, i.e. all players are supposed to have equal bargaining powers. In general games, however, *nothing is said about the bargaining power of the null players!* If we admit that null players do have bargaining power, then this fact can somehow affect the aggregate power of the coalition they join.

Take for example the unanimity game (N', v') where $N' = \{1, 2\}$ and $v'(N) = 1$, $v'(\{1\}) = v'(\{2\}) = 0$. By a symmetry argument, the value of each player should be $\frac{1}{2}$, i.e. the expectation of each player before any implementation of the game is the same.

Assume now we add a null player 3 (Example 4.1). We get the game (N, v) with $N = \{1, 2, 3\}$ and $v(S) = 1$ if $\{1, 2\} \subseteq S$ and $v(S) = 0$ otherwise. What would the players' expectation be in this new game?

It can be argued that the situation does not change with the presence of a player that does not contribute anything to any coalition. Hence, the value of (N, v) should be $\left(\frac{1}{2}, \frac{1}{2}, 0\right)$. However, the situation may significantly change if we assume that player 3 joins forces with player 2. In this case, the symmetry argument used to assign the value $\left(\frac{1}{2}, \frac{1}{2}\right)$ in the previous game

[7] A *null player* is a player i with $v(S \cup i) = v(S)$ for all S.

[8] The name *null players* is not very accurate in this context. Even though their marginal contributions are zero, actually they are not null players, because their sole presence changes the weight of their coalition. I thank Inés Macho-Stadler for pointing this out.

[9] Two players i, j are *substitutes* if $v(S \cup i) = v(S \cup j)$ for all S with $i, j \notin S$.

(N', v') vanishes. Player 1 and coalition $\{2,3\}$ are substitutes in the game between coalitions, but not completely symmetric. The fact that $\{2,3\}$ has two members introduces an endogenous asymmetry. Hart and Kurz (p. 1048) describe this situation as follows:

> As an everyday example of such a situation, "I will have to check this with my wife/husband" may (but not necessarily) lead to a better bargaining position, due to the fact that the other party has to convince *both* the player and the spouse.

If we accept that player 2 may benefit from the support of player 3, one may wonder how to quantify this benefit. The value ζ provides a possible answer, by assigning an allocation $\zeta^N = \left(\frac{4}{12}, \frac{7}{12}, \frac{1}{12}\right)$ when the coalition structure is $\mathcal{C} = \{\{1\}, \{2,3\}\}$. Notice that, in the game between coalitions, the allocation is $\frac{1}{3}$ for coalition $\{1\}$ and $\frac{2}{3}$ for coalition $\{2,3\}$. Hence, payoffs are proportional to coalition size (Corollary 3.2). Without player 3, player 2 can only expect to get $\frac{1}{2}$, whereas coalition $\{2,3\}$ would get $\frac{2}{3}$. Allocation ζ simply suggests to split the *benefit of cooperation* $(\frac{2}{3} - \frac{1}{2} = \frac{1}{6})$ equally between player 2 and player 3.

References

[1] Bergantiños G. and J.J. Vidal-Puga, 2005, The consistent coalitional value, Mathematics of Operations Research, forthcoming.

[2] Binmore K., 1998, Just playing, Game theory and the social contract, Volume 2, MIT Press.

[3] Calvo E., J. Lasaga and E. Winter, 1996, The principle of balanced contributions and hierarchies of cooperation, Mathematical Social Sciences 31, 171-182.

[4] Chae S. and P. Heidhues, 2004, A group bargaining solution, Mathematical Social Sciences 48(1), 37-53.

[5] Chae S. and H. Moulin, 2004, Bargaining among groups: an axiomatic viewpoint, Mimeo.

[6] Haeringer G., 2000, A new weight scheme for the Shapley value, BETA - Working paper #9910.

[7] Harsanyi J.C., 1963, A simplified bargaining model for the n-person cooperative game, International Economic Review 4, 194-220.

[8] Harsanyi J.C., 1977, Rational behavior and bargaining equilibrium in games and social situations, Cambridge University Press.

[9] Hart S. and M. Kurz, 1983, Endogenous formation of coalitions, Econometrica 51(4), 1047-1064.

[10] Hart S. and A. Mas-Colell, 1996, Bargaining and value, Econometrica 64, 357-380.

[11] Kalai E. and D. Samet, 1985, A monotonic solution to general cooperative games, Econometrica 53, 307-327.

[12] Kalai E. and D. Samet, 1987, On weighted Shapley values, International Journal of Game Theory 16, 205-222.

[13] Kalai E. and D. Samet, 1988, Weighted Shapley values, in: A. Roth, ed., The Shapley value: Essays in honor of L. Shapley (Cambridge), Cambridge University Press, 83-99.

[14] Krasa S., A. Temimi and N.C. Yannelis, 2003, Coalition structure values in differential information economies: is unity a strength? Journal of Mathematical Economics 39, 51-62.

[15] Levy A. and R.P. McLean, 1989, Weighted coalition structure values, Games and Economic Behavior 1, 234-249.

[16] Myerson R.B., 1980, Conference structures and fair allocation rules, International Journal of Game Theory 9, 169-182.

[17] Nash J.F., 1950, The bargaining problem. Econometrica 18, 155-162.

[18] Owen G., 1968, A note on the Shapley value, Management Science 14, 731-732.

[19] Owen G., 1977, Values of games with a priori unions, in: R. Henn and O. Moeschlin, eds., Essays in mathematical economics and game theory (Berlin), Springer-Verlag, 76-88.

[20] Pérez-Castrillo D. and D. Wettstein, 2001, Bidding for the surplus: A non-cooperative approach to the Shapley value, Journal of Economic Theory 100, 274-294.

[21] Shapley L.S., 1953a, Additive and non-additive set functions, Ph.D. Thesis, Princeton University.

[22] Shapley L.S., 1953b, A value for n-person games, in: H.W. Kuhn and A.W. Tucker, eds., Contributions to the Theory of Games II (Princeton NJ), Princeton University Press, 307-317.

[23] Vidal-Puga J.J., 2005, A bargaining approach to the Owen value and the Nash solution with coalition structure, Economic Theory 25(3), 679-701.

[24] Vidal-Puga J.J. and G. Bergantiños, 2003, An implementation of the Owen value, Games and Economic Behavior 44, 412-427.