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

Juan Vidal-Puga[†]

Research Group in Economic Analysis

Universidade de Vigo

Published on November 2012

Abstract

We describe a 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. Moreover, the Harsanyi paradox (forming a coalition may be disadvantageous) disappears for convex games.

JEL codes: C71, C78

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

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

1

players 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 "coalitional 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. There is an extensive literature on axiomatic characterization of coalitional values: Aumann and Dreze (1974); Owen (1977); Hart and Kurz (1983); Levy and Mc Lean (1989); Alonso-Meijide and Fiestras-Janeiro (2002); Amer et al. (2002); Młodak (2003); Carreras and Puente (2006); Kamijo (2009, 2013); Gómez-Rúa and Vidal-Puga (2010); Calvo and Gutiérrez (2010), among others.

The non-cooperative point of view leads to the study of the allocations which arise in a given non-cooperative environment. Some coalitional values have also been studied from the non-cooperative point of view: Vidal-Puga and Bergantiños (2003); Vidal-Puga (2005); Kamijo (2008). This paper also follows the non-cooperative approach.

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 in general. An individual can be worse off bargaining as a member of a coalition than bargaining alone. Formally stated, the Harsanyi paradox[1] 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 the symmetry 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 (in the sense of forming an *a priori* coalition) can be harmful in bargaining environments. Chae and

---

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

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 (2010)). However, it has a natural interpretation in a non-cooperative environment. Many non-cooperative mechanisms[2] (for example, Rubinstein (1982)) include a key stage in which one of the players should make a proposal. Hence, the "right to talk" can be interpreted as the "right to make a proposal". The Harsanyi paradox may arise when 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 should come from one of the members of the joined coalition with a probability $1/(n-1)$, whereas when no coalition is formed the proposal should come from one of them with probability $2/n$.

In this paper, we study the effects of maintaining the "rights to talk" of the players inside a coalition. Hence, the coalitions with more members have more chances to make proposals. In the previous example, this means that the proposal from a member of the joined coalition will come with a probability $1/n$, as if she were acting alone.

In particular, we generalize a non-cooperative mechanism by 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 our model, this procedure is played in two stages. First, agreements are negotiated within coalitions and then through delegates among coalitions. Each coalition acts 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. Moreover, the probability of a coalition being chosen as proposer in the second stage is proportional to its size[3].

As a result, the resulting equilibrium payoff coincides with the coalitional value described and axiomatized in Gómez-Rúa and Vidal-Puga (2010). In particular, we get the weighted Shapley value (Shapley, 1953a) in the game between coalitions, with weights

---

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

[3]As opposed, if we give equal probability to each coalition, we obtain the mechanism presented by Vidal-Puga (2005) which gives the Owen value as expected final outcome. However, our results are not implied by the results in Vidal-Puga (2005) and the proofs are also different.

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 4.1). However, this is not true in general games (see Example 3.1 and Example 4.1).

The new mechanism generalizes 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).

In Section 2 we present the notation used throughout the paper and some previous results. In Section 3 we describe the coalitional value. In Section 4 we present the formal mechanism and the main results. In Section 5 we discuss a generalization of the mechanism. The proofs are located in Section 6 (Appendix).

2   Preliminaries

A *non-transferable utility game*, or *NTU game*, is a pair (N, V) where N is a finite set of *players* and V is a correspondence which assigns to each $S \subset N$, $S \neq \emptyset$ 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 $S \subset N$, we maintain the notation V when referring to the application V restricted to S as player set. For simplicity, denote $V(i)$ instead of $V(\{i\})$, $S \cup i$ instead of $S \cup \{i\}$ and $N\backslash i$ instead of $N\backslash\{i\}$. The set of NTU games is denoted 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 $v(\emptyset) = 0$, we say (N, V) is a *transferable utility game* (or *TU game*) and we represent it as (N, v). As before, we 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 $S \cap T = \emptyset$. 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 $T \subset S \subset N\backslash i$. 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 $S \neq N$, where $r_i^S = r_i$ for all $i \in S$, and $r^N \in V(N)$, we say that (N, V) is a *pure bargaining problem*.

Unanimity games are both TU games and pure bargaining problems.

A *coalition structure* over N is 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 $q \neq r$. We also assume $C_q \neq \emptyset$ for all q. A coalition structure $\mathcal{C}$ over N is *trivial* if either $\mathcal{C}$ consists of singletons or $\mathcal{C} = \{N\}$. For any $S \subset N$, we 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 $Q \subset M$.

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

Given a subset G 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, we 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 Pareto frontier 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). We denote the Shapley value of the TU game (N, v) as $\varphi^N(v) \in \mathbb{R}^N$.

In NTU games that are both TU games and pure bargaining problems, the Shapley value and the Nash solution coincide. In unanimity games, $\varphi_i^N(v) = \frac{1}{|N|}$ 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). We denote the weighted Shapley value of the TU game (N, v) as $\varphi^{\omega N}(v) \in \mathbb{R}^N$, where $\omega \in \gneqq^N$ is a vector of weights. When $\omega_i = \frac{1}{|N|}$ for all i, the weighted Shapley value coincides with the Shapley value.

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 (2006) (Example 1).

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

We now focus on TU games with coalition structure. Fix $\mathcal{C} = \{C_1, \ldots, C_m\}$ and $M = \{1, ..., m\}$. Owen (1977) proposed an efficient value based on Shapley's which takes into account the coalition structure. We call this value the *Owen coalitional value*, or simply the *Owen value*, and we denote it as $\phi^N(v)$. When the coalition structure is trivial, i.e. $\mathcal{C} = \{\{i\}\}_{i \in N}$ or $\mathcal{C} = \{N\}$, the Owen value coincides with the Shapley value.

Levy and Mc Lean (1989) studied the *weighted coalitional value with intracoalitional*

symmetry, that we denote as $\phi^{\omega N}(v)$. 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, we write φ^N, ϕ^N, $\varphi^{\omega N}$, $\phi^{\omega N}$ instead of $\varphi^N(v)$, $\phi^N(v)$, $\varphi^{\omega N}(v)$, $\phi^{\omega N}(v)$, respectively. With some abuse of notation, given $S \subset N$, we denote as $\varphi^{\omega S}$ and $\phi^{\omega S}$ the weighted Shapley value and the Levy-McLean coalitional value, respectively, of the game (S, v) with weights $\omega_i' = \frac{|N|}{|S|}\omega_i$ for all $i \in S$.

We now define formally the Harsanyi paradox. Given $C_q, C_r \in \mathcal{C}$, we define the coalition structure $\mathcal{C}^{q+r}$ as $(\mathcal{C} \setminus \{C_q, C_r\}) \cup \{C_q \cup C_r\}$. 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 $G \subset CNTU$. Just in this case, we 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. We say that ψ is *joint-monotonic* in G if

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

for all $(N, V, \mathcal{C}) \in G$ and all $C_q, C_r \in \mathcal{C}$. 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[4].

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   The 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 Mc Lean (1989) (p.235) claim that this intercoalitional symmetry may not be a reasonable requirement for a value. A classical example (Kalai

---

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

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 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 $\omega_q = \frac{|C_q|}{|N|}$ for each $C_q \in \mathcal{C}$. However, we will use a different coalitional value ζ^N, that follows a similar idea as $\phi^{\omega N}$. This ζ^N is characterized in Gómez-Rúa and Vidal-Puga (2010, 2011), and it is defined following a two-step procedure. In the first step, we define a reduced TU game (C_q, v_q^{*N}) for each $C_q \in \mathcal{C}$ as follows: Given $T \subset C_q$, let $\lambda \in \mathbb{R}_{++}^M$ be the weight system given by $\lambda_q = \frac{|T|}{|N| - |C_q \setminus T|}$ and $\lambda_r = \frac{|C_r|}{|N| - |C_q \setminus T|}$ otherwise. We then define:

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

for all $T \subset C_q$. Notice that $\varphi^{\lambda M}$ is the weighted Shapley value of the game between coalitions.

We can interpreted $v_q^{*N}(T)$ as the worth of subcoalition T when players in $C_q \setminus T$ are out.

In the second step, we use the Shapley value to determine the final allocation. 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, we write ζ^N instead of $\zeta^N(v)$.

The following example will help to clarify the previous definition, comparing it to the Owen value and the Levy-McLean value:

Example 3.1 *Let (N, v) be the TU game defined as $N = \{1, 2, 3\}$, $v(\{1, 2\}) = 12$, $v(N) = 24$, and $v(S) = 0$ otherwise. Let $\mathcal{C} = \{C_1, C_2\}$ with $C_1 = \{1\}$ and $C_2 = \{2, 3\}$.*

In this case, the game between coalitions $(\{1, 2\}, v/\mathcal{C})$ is defined as $v/\mathcal{C}(\{1\}) = v/\mathcal{C}(\{2\}) = 0$ and $v/\mathcal{C}(\{1, 2\}) = 24$. The $\left(\left(\frac{1}{2}, \frac{1}{2}\right)\text{-weighted}\right)$ Shapley value is $(12, 12)$, i.e. 12 for coalition C_1 and 12 for coalition C_2. Similarly, the $\left(\frac{1}{3}, \frac{2}{3}\right)$-weighted Shapley value is $(8, 16)$. When player 2 leaves N, the game between coalitions becomes the null game $\left(\{1, 2\}, v/\mathcal{C}_{\{1, 3\}}\right)$ with $v/\mathcal{C}_{\{1, 3\}}(\{1\}) = v/\mathcal{C}_{\{1, 3\}}(\{2\}) = v/\mathcal{C}_{\{1, 3\}}(\{1, 2\}) = 0$. The $\left(\left(\frac{1}{2}, \frac{1}{2}\right)\text{-weighted}\right)$ Shapley value and the $\left(\frac{1}{3}, \frac{2}{3}\right)$-weighted Shapley value are both $(0, 0)$. When player 3 leaves N, the game between coalitions becomes $\left(\{1, 2\}, v/\mathcal{C}_{\{1, 2\}}\right)$ with $v/\mathcal{C}_{\{1, 2\}}(\{1\}) = v/\mathcal{C}_{\{1, 2\}}(\{2\}) = 0$ and $v/\mathcal{C}_{\{1, 2\}}(\{1, 2\}) = 12$. The $\left(\left(\frac{1}{2}, \frac{1}{2}\right)\text{-weighted}\right)$ Shapley value is $(6, 6)$ and the $\left(\frac{1}{3}, \frac{2}{3}\right)$-weighted Shapley value is $(4, 8)$.

The reduced game $\left(\{1\}, v_1^{*N}\right)$ is defined as

$$v_1^{*N}\left(\{1\}\right) = \varphi_2^{\left(\frac{1}{3}, \frac{2}{3}\right)\{1,2\}}\left(v/\mathcal{C}_{\{1,2\}}\right) = 8,$$

where $\left(\frac{1}{3}, \frac{2}{3}\right)$ is the vector of weights for coalitions $\{1\}$ and $\{2,3\}$.

Similarly, the reduced game $\left(\{2,3\}, v_2^{*N}\right)$ is defined as

$$v_2^{*N}\left(\{2\}\right) = \varphi_2^{\left(\frac{1}{2}, \frac{1}{2}\right)\{1,2\}}\left(v/\mathcal{C}_{\{1,2\}}\right) = 6$$

$$v_2^{*N}\left(\{3\}\right) = \varphi_2^{\left(\frac{1}{2}, \frac{1}{2}\right)\{1,2\}}\left(v/\mathcal{C}_{\{1,3\}}\right) = 0$$

$$v_2^{*N}\left(\{2,3\}\right) = \varphi_2^{\left(\frac{1}{3}, \frac{2}{3}\right)\{1,2\}}\left(v/\mathcal{C}\right) = 16.$$

Hence, $\zeta_1^N\left(v\right) = \varphi_1^{\{1\}}\left(v_1^{*N}\right) = 8$, $\zeta_2^N\left(v\right) = \varphi_2^{\{2,3\}}\left(v_2^{*N}\right) = 11$, and $\zeta_3^N\left(v\right) = \varphi_2^{\{2,3\}}\left(v_2^{*N}\right) = 5$, i.e. $\zeta^N\left(v\right) = (8, 11, 5)$.

As opposed, ϕ^N defines the reduced game using the Shapley value, so that $\phi^N\left(v\right) = (12, 9, 3)$ (see Owen (1977) for details) and $\phi^{\omega N}$ defines the reduced game using the $\left(\frac{1}{3}, \frac{2}{3}\right)$-weighted Shapley value, so that $\phi^{\omega N}\left(v\right) = (8, 12, 4)$ (see Levy and Mc Lean (1989) (Proposition C(2)) for details).

The critical difference between the definitions of ζ and the Levy-McLean value ϕ^ω is that the weights λ that appear in the definition of $v_q^{*N}\left(T\right)$ depend on T, whereas in the definition of ϕ^ω (see Levy and Mc Lean (1989) (Proposition C(2))) the weights are the same for each possible T. On the other hand, in the definition of the Owen value, the Shapley value is used in both steps.

Remark 3.1 *It follows from the definition of ζ that each coalition gets its weighted Shapley value of the game between coalitions, with weights given by their size. Namely, for any $C_q \in \mathcal{C}$, $\sum_{i \in C_q} \zeta_i^N = \varphi_q^{\lambda M}\left(v/\mathcal{C}\right)$.*

One practical problem with the above definition is that ζ^N is extremely laborious to compute, being necessary to calculate $\sum_{q=1}^m 2^{|C_q|}$ distinct weighted Shapley value vectors (to identify v_q^{*N}), and then calculate m distinct Shapley value vectors (to identify φ^{C_q}). In Section 6 (Proposition 6.1), we provide an easily implementable recursive formula to compute ζ that allows to overcome this difficulty[5].

A different issue would be to study the complexity in the computation of ζ. By Proposition 6.1, ζ^N is independent of the worth of coalitions $S \subset N$ that have proper

---

[5]There exists a similar formula for the (weighted) Shapley value (see Pérez-Castrillo and Wettstein (2001) (Lemma 1)).

intersection with more than one C_q. This property is called *Coordination* in Gómez-Rúa and Vidal-Puga (2010) and it is also satisfied by the Owen value and Levy-McLean value. Hence, computing complexity in the more general case (i.e. without restricting to any particular class of TU games) is not higher than for the Shapley value.

We now prove that with this coalitional value the Harsanyi paradox disappears.

Proposition 3.1 *The coalitional value ζ is joint-monotonic in convex games.*

Proof. We 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 $C_q, C_r \in \mathcal{C}$. Assume w.l.o.g. $q = m - 1$ 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 Remark 3.1, it is enough to prove that

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

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

Pérez-Castrillo and Wettstein (2001) (Lemma 1) proved that $\varphi_q^{\omega M}$ can be inductively computed as $\varphi_q^{\omega M} (v) = \omega_q v (M) - \omega_q v (M \backslash q) + \sum_{p \in M \backslash q} \omega_p \varphi_q^{\omega M \backslash p} (v)$. Hence,

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

and

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

Under the induction hypothesis, $\varphi_{m-1}^{\omega M \backslash p}(u) + \varphi_m^{\omega M \backslash p}(u) \leq \varphi_{m-1}^{\omega M^* \backslash p}(u^*)$ for all $p < m-1$. Hence, it is enough to prove,

$$\omega_{m-1} u(M) - \omega_{m-1} u(M \backslash (m-1)) + \omega_m u(M)$$
$$- \omega_m u(M \backslash m) + \omega_m \varphi_{m-1}^{\omega M \backslash m}(u) + \omega_{m-1} \varphi_m^{\omega M \backslash (m-1)}(u)$$
$$\leq (\omega_{m-1} + \omega_m) u(M) - (\omega_{m-1} + \omega_m) u(M \backslash \{m-1, m\}).$$

Simplifying and rearranging terms,

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

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

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

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. ∎

As opposed, Proposition 3.1 does not hold for the Owen value ϕ. Take the TU game (N, v) given in Example 3.1. This is a convex game with $\varphi^N(v) = (10, 10, 4)$. Since $\phi^N(v) = (12, 9, 3)$, forming coalition $\{2, 3\}$ is disadvantageous for both players 2 and 3.

Proposition 3.1 does not hold in general for nonconvex games, as the next example shows:

Example 3.2 *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 $T \subset \{3, 4, 5\}$, $T \neq \emptyset$. 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, $\zeta^N = (153, 153, 18, 18, 18)$.

4 The non-cooperative mechanism

In this section we describe the non-cooperative mechanism. Even though the model is defined for NTU games, we focus on TU games and bargaining problems.

Fix $(N, V, \mathcal{C}) \in CNTU$. For each $S \subset N$, we denote as 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, we denote $\gamma_q := \gamma\left(C_q'\right)$. Moreover, we denote as λ^S the weight vector in the subgame (S, V), i.e., $\lambda_q^S = \frac{|C_q'|}{|S|}$ for all $S \subset N$ and all $C_q' \in \mathcal{C}_S$.

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 $S \subset N$ 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 $a\left(S, \gamma_r\right)$ being chosen is $\lambda_r^S = \frac{|C_r'|}{|S|}$, 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.

A key feature is that, 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 generalizes Hart and Mas-Colell (1996)'s 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, we 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.

The main result of the paper, that provides a non-cooperative justification for ζ^N, is the following:

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

This result is an immediate consequence of Propositions 6.5, 6.6 and 6.7 proved in Section 6 (Appendix).

It is worthy to analyze a particular example.

Example 4.1 *Let $N = \{1, 2, 3\}$ and v be defined as $v(\{1, 2\}) = 12$, $v(N) = 24$ 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 $\zeta^{\{1,2\}} = (6, 6)$ 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 $a(N, 1) = (24, 0, 0)$, 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 $a(N, 3) = (0, 9, 15)$, 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 $\frac{1}{3}0 + \frac{2}{3}9 = 6$. Analogously, player 2 would propose $a(N, 2) = (0, 24, 0)$.*

Once these proposals are accepted in the first stage, in the second stage the proposal of coalition $\{2, 3\}$ is either $(0, 24, 0)$ (probability $\frac{1}{2}$), or $(0, 9, 15)$ (probability $\frac{1}{2}$). In the second stage, the final proposal will be $(24, 0, 0)$ with probability $\frac{1}{3}$, and either $(0, 24, 0)$ or

$(0, 9, 15)$ *with probability* $\frac{2}{3}$. *On average, the expected final payoff is*

$$\frac{1}{3} (24, 0, 0) + \frac{2}{3} \left(\frac{1}{2} (0, 24, 0) + \frac{1}{2} (0, 9, 15) \right) = (8, 11, 5) = \zeta^N.$$

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

Proposition 4.1 *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 $S \neq N$, 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, $a(N) = \frac{1}{|N|} \sum_{i \in N} a(N, i)$ (see Proposition 6.4 in Section 6). 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). ∎

5   Concluding remark

In general, the mechanism does not implement ζ for nonconvex games. Take $\rho = 0$. Take the TU game given in Example 3.2 with coalition structure $\{\{1\}, \{2\}, \{3, 4, 5\}\}$. Assume the only equilibrium payoff is ζ^S for all $S \neq N$. Some of these values are given in the following table:

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

We 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 $a_i\left(N,3\right)^\gamma = 180$ for all $i \in \{1,2\}$ and $a_j\left(N,3\right)^\gamma = 20$ for all $j \in \{4,5\}$ (so that $\frac{1}{5}a_j\left(N,1\right)^\gamma + \frac{1}{5}a_j\left(N,2\right)^\gamma + \frac{3}{5}a_j\left(N,3\right)^\gamma = 30$, that is, player j's continuation payoff after rejection). Hence $a_3\left(N,3\right)^\gamma \leq -40$. This leaves player 3 with a negative final expected payoff [6]. 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 6.3 in Section 6 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 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 $S \subset N$ and $i \in C'_q \in \mathcal{C}_S$. Hence, for p high enough[7] 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, $r_i = v\left(\{i\}\right)$ 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.

6 Appendix

Fix $(N,v,\mathcal{C})$. In the next proposition we describe an inductive formula to compute ζ:

Proposition 6.1 *The coalitional value 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 $T \subset S, T \neq S$. Then, $\zeta_i^S =$*

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

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

---

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

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

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

Claim 6.1 *Given $i, j \in C'_q \in \mathcal{C}_S$, $\varphi_i^{C'_q \backslash j}\left(v_q^{*S}\right) = \varphi_i^{C'_q \backslash j}\left(v_q^{*S \backslash j}\right)$.*

The proof is straightforward and we omit it.

Claim 6.2 *Given $q, r \in M'$, $\varphi_q^{\lambda^S M' \backslash r}\left(v/\mathcal{C}_S\right) = v_q^{*S \backslash C'_r}\left(C'_q\right)$.*

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 $Q \subset M' \backslash r$. 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 6.3 *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, $\sum_{j \in C'_q} \zeta_j^{S \backslash C'_r} = \sum_{j \in C'_q} \varphi_j^{C'_q}\left(v_q^{*S \backslash C'_r}\right) = v_q^{*S \backslash C'_r}\left(C'_q\right)$.

We now use the claims to prove the result. It follows from Pérez-Castrillo and Wettstein (2001) (Lemma 1) that the weighted Shapley value can be computed as

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

for all $i \in N$, $\omega \in \mathbb{R}_{++}^N$. Remark that the Shapley value φ^N coincides with $\varphi^{\omega N}$ for $\omega_i = \frac{1}{|N|}$ for all $i \in N$.

Given $i \in C'_q \in \mathcal{C}_S$,

$$\zeta_i^S = \varphi_i^{C'_q}\left(v_q^{*S}\right) \overset{(1)}{=} \frac{1}{|C'_q|}\left[v_q^{*S}\left(C'_q\right) + \sum_{j \in C'_q \backslash i}\left(\varphi_i^{C'_q \backslash j}\left(v_q^{*S}\right) - \varphi_j^{C'_q \backslash i}\left(v_q^{*S}\right)\right)\right]$$

$$\overset{(\text{Claim 6.1})}{=} \frac{1}{|C'_q|}\left[v_q^{*S}\left(C'_q\right) + \sum_{j \in C'_q \backslash i}\left(\zeta_i^{S \backslash j} - \zeta_j^{S \backslash i}\right)\right]. \tag{2}$$

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

$$\varphi_q^{\lambda^S M'}\left(v/\mathcal{C}_S\right) \overset{(1)}{=} \lambda_q^S v/\mathcal{C}_S\left(M'\right) + \sum_{r \in M' \backslash q}\left(\lambda_r^S \varphi_q^{\lambda^S M' \backslash r}\left(v/\mathcal{C}_S\right) - \lambda_q^S \varphi_r^{\lambda^S M' \backslash q}\left(v/\mathcal{C}_S\right)\right)$$

$$\overset{(\text{Claim 6.2})}{=} \lambda_q^S v\left(S\right) + \sum_{r \in M' \backslash q}\left(\lambda_r^S v_q^{*S \backslash C'_r}\left(C'_q\right) - \lambda_q^S v_r^{*S \backslash C'_q}\left(C'_r\right)\right)$$

$$\overset{(\text{Claim 6.3})}{=} \lambda_q^S v\left(S\right) + \sum_{C'_r \in \mathcal{C}_S \backslash C'_q}\left(\lambda_r^S \sum_{j \in C'_q} \zeta_j^{S \backslash C'_r} - \lambda_q^S \sum_{j \in C'_r} \zeta_j^{S \backslash C'_q}\right). \tag{3}$$

The result comes from combining (2) and (3). ∎

Corollary 6.1 *For any $S \subset N$ and $C'_q \in \mathcal{C}_S$,*

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

Proof. It follows from Proposition 6.1. ∎

The next property has the flavor of the balanced contributions property of Myerson's (Myerson, 1980), and it is also satisfied by the Owen value (Calvo et al., 1996; Bergantiños and Vidal-Puga, 2005):

Proposition 6.2 *For all $S \subset N$ and $i \in C'_q \in \mathcal{C}_S$, $\sum_{j \in C'_q \setminus i} \left(\zeta_i^S - \zeta_i^{S \setminus j} \right) = \sum_{j \in C'_q \setminus i} \left(\zeta_j^S - \zeta_j^{S \setminus i} \right).$*

Proof. It follows from Proposition 6.1 and Corollary 6.1. ∎

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 autarchy payoff.

Proposition 6.3 *For strictly convex TU games, $\sum_{j \in C'_q \setminus i} \zeta_j^{S \setminus i} + r_i < \sum_{j \in C'_q} \zeta_j^S$ 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 $\sum_{j \in C'_q \setminus i} \zeta_j^{S \setminus i} = 0$). Under Remark 3.1, 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 Remark 3.1, it is enough to prove

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

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

Hence, it is enough to prove

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

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

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 $\varphi_q^{\lambda^S M'}(v/\mathcal{C}_{S\setminus i}) + \varphi_q^{\lambda^S M'}(u_q) = \varphi_q^{\lambda^S M'}(v')$. Moreover, $\varphi_q^{\lambda^S M'}(u_q) = r_i$. Hence,

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

∎

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

$$a(S)^\gamma := \sum_{C'_q \in \mathcal{C}_S} \lambda_q^S a(S, \gamma_q)^\gamma \in V(S) \tag{4}$$

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

Denote

$$a(S) := \sum_{\gamma \in \Gamma_S} \frac{1}{|\Gamma_S|} a(S)^\gamma \in V(S) \tag{5}$$

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 $|\Gamma_S| = |\Gamma_{S,i}| |C'_q|$ for all $i \in C'_q \in \mathcal{C}_S$.

Let

$$a(S,i) := \sum_{\gamma \in \Gamma_{S,i}} \frac{1}{|\Gamma_{S,i}|} a(S,i)^\gamma \tag{6}$$

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 6.4 *for all $S \subset N$, $a(S) = \sum_{i \in S} \frac{1}{|S|} a(S,i)$.*

Proof. Given $S \subset N$,

$$a(S) \overset{(5)}{=} \sum_{\gamma \in \Gamma_S} \frac{1}{|\Gamma_S|} a(S)^\gamma \overset{(4)}{=} \sum_{\gamma \in \Gamma_S} \frac{1}{|\Gamma_S|} \sum_{C'_q \in \mathcal{C}_S} \lambda_q^S a(S, \gamma_q)^\gamma$$

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

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

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

Under (6),

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

∎

Proposition 6.5 *Assume a set of proposals* $\left(a\left(S,i\right)_{i \in S, \gamma \in \Gamma_{S,i}}^\gamma\right)_{S \subset N}$ *satisfies the following three conditions for all* $S \subset N$:

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* $S \subset N$.

Proof. By P-3,

$$\sum_{i \in S} a_i\left(S\right) = v\left(S\right). \tag{7}$$

Fix $i \in C_q' \in \mathcal{C}_S$. From (4) it is readily checked that, for any $j \in C_q'\backslash i$, $\gamma \in \Gamma_{S,i}$:

$$a_j\left(S,i\right)^\gamma = \frac{1}{\lambda_q^S} a_j\left(S\right)^\gamma - \sum_{C_r' \in \mathcal{C}_S\backslash C_q'} \frac{\lambda_r^S}{\lambda_q^S} a_j\left(S,\gamma_r\right)^\gamma.$$

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

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

$$= \rho a_j(S) + (1-\rho)\left[\frac{1}{\lambda_q^S} a_j(S\backslash i) - \sum_{C_r' \in \mathcal{C}_S\backslash C_q'} \frac{\lambda_r^S}{\lambda_q^S} a_j(S\backslash C_r')\right]. \tag{8}$$

Under Proposition 6.4 and (6),

$$\left|S\right| a_i\left(S\right) \overset{\text{(Proposition 6.4)}}{=} \sum_{j \in S} a_i\left(S,j\right) \overset{\text{(6)}}{=} \sum_{j \in S} \sum_{\gamma \in \Gamma_{S,j}} \frac{1}{\left|\Gamma_{S,j}\right|} a_i\left(S,j\right)^\gamma$$

$$= \sum_{\gamma \in \Gamma_{S,i}} \frac{1}{\left|\Gamma_{S,i}\right|} a_i\left(S,i\right)^\gamma + \sum_{j \in C_q'\backslash i} \sum_{\gamma \in \Gamma_{S,j}} \frac{1}{\left|\Gamma_{S,j}\right|} a_i\left(S,j\right)^\gamma + \sum_{j \in S\backslash C_q'} \sum_{\gamma \in \Gamma_{S,j}} \frac{1}{\left|\Gamma_{S,j}\right|} a_i\left(S,j\right)^\gamma.$$

18

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

$$\sum_{\gamma \in \Gamma_{S,i}} \frac{1}{|\Gamma_{S,i}|} a_i (S,i)^\gamma \overset{(P\text{-}3)}{=} v(S) - \sum_{\gamma \in \Gamma_{S,i}} \frac{1}{|\Gamma_{S,i}|} \sum_{j \in S \setminus i} a_j (S,i)^\gamma$$

$$= v(S) - \sum_{\gamma \in \Gamma_{S,i}} \frac{1}{|\Gamma_{S,i}|} \sum_{C'_r \in \mathcal{C}_S \setminus C'_q} \sum_{j \in C'_r} a_j (S,i)^\gamma - \sum_{\gamma \in \Gamma_{S,i}} \frac{1}{|\Gamma_{S,i}|} \sum_{j \in C'_q \setminus i} a_j (S,i)^\gamma$$

$$\overset{(P\text{-}1)\text{-}(8)}{=} v(S) - \sum_{C'_r \in \mathcal{C}_S \setminus C'_q} \sum_{j \in C'_r} \left[\rho a_j (S) + (1-\rho) a_j \left(S \setminus C'_q \right) \right]$$

$$- \sum_{j \in C'_q \setminus i} \left[\rho a_j (S) + (1-\rho) \left[\frac{1}{\lambda^S_q} a_j (S \setminus i) - \sum_{C'_r \in \mathcal{C}_S \setminus C'_q} \frac{\lambda^S_r}{\lambda^S_q} a_j (S \setminus C'_r) \right] \right]$$

under (7), $\sum_{j \in S \setminus i} \rho a_j (S) = \rho (v(S) - a_i (S))$ and thus

$$= v(S) - \rho (v(S) - a_i (S)) - (1-\rho) \sum_{C'_r \in \mathcal{C}_S \setminus C'_q} \sum_{j \in C'_r} a_j \left(S \setminus C'_q \right)$$

$$- (1-\rho) \sum_{j \in C'_q \setminus i} \left[\frac{1}{\lambda^S_q} a_j (S \setminus i) - \sum_{C'_r \in \mathcal{C}_S \setminus C'_q} \frac{\lambda^S_r}{\lambda^S_q} a_j (S \setminus C'_r) \right].$$

For the second term:

$$\sum_{j \in C'_q \setminus i} \sum_{\gamma \in \Gamma_{S,j}} \frac{1}{|\Gamma_{S,j}|} a_i (S,j)^\gamma \overset{(8)}{=}$$

$$\sum_{j \in C'_q \setminus i} \left[\rho a_i (S) + (1-\rho) \left[\frac{1}{\lambda^S_q} a_i (S \setminus j) - \sum_{C'_r \in \mathcal{C}_S \setminus C'_q} \frac{\lambda^S_r}{\lambda^S_q} a_i (S \setminus C'_r) \right] \right]$$

$$= \rho \left(|C'_q| - 1 \right) a_i (S)$$

$$+ (1-\rho) \left[\sum_{j \in C'_q \setminus i} \frac{1}{\lambda^S_q} a_i (S \setminus j) - \left(|C'_q| - 1 \right) \sum_{C'_r \in \mathcal{C}_S \setminus C'_q} \frac{\lambda^S_r}{\lambda^S_q} a_i (S \setminus C'_r) \right].$$

For the third term:

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

$$= \rho \left(|S| - |C'_q| \right) a_i (S) + (1-\rho) \sum_{C'_r \in \mathcal{C}_S \setminus C'_q} |C'_r| a_i (S \setminus C'_r).$$

19

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

$$v\left(S\right) - \rho v\left(S\right) - \left(1-\rho\right) \sum_{C'_r \in \mathcal{C}_S \backslash C'_q} \sum_{j \in C'_r} a_j\left(S\backslash C'_q\right)$$

$$- \sum_{j \in C'_q \backslash i} \left(1-\rho\right) \left[\frac{1}{\lambda^S_q} a_j\left(S\backslash i\right) - \sum_{C'_r \in \mathcal{C}_S \backslash C'_q} \frac{\lambda^S_r}{\lambda^S_q} a_j\left(S\backslash C'_r\right) \right]$$

$$+ \left(1-\rho\right) \left[\sum_{j \in C'_q \backslash i} \frac{1}{\lambda^S_q} a_i\left(S\backslash j\right) + \sum_{C'_r \in \mathcal{C}_S \backslash C'_q} \frac{\lambda^S_r}{\lambda^S_q} a_i\left(S\backslash C'_r\right) \right]$$

$$+ \rho\, |S|\, a_i\left(S\right).$$

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

$$= v\left(S\right) + \sum_{j \in C'_q \backslash i} \frac{1}{\lambda^S_q} \left(a_i\left(S\backslash j\right) - a_j\left(S\backslash i\right)\right)$$

$$+ \sum_{C'_r \in \mathcal{C}_S \backslash C'_q} \left(\frac{\lambda^S_r}{\lambda^S_q} \sum_{j \in C'_q} a_j\left(S\backslash C'_r\right) - \sum_{j \in C'_r} a_j\left(S\backslash C'_q\right) \right).$$

Hence,

$$a_i\left(S\right) = \frac{\lambda^S_q}{\left|C'_q\right|} v\left(S\right) + \sum_{j \in C'_q \backslash i} \frac{1}{\left|C'_q\right|} \left(a_i\left(S\backslash j\right) - a_j\left(S\backslash i\right)\right)$$

$$+ \sum_{C'_r \in \mathcal{C}_S \backslash C'_q} \left(\sum_{j \in C'_q} \frac{\lambda^S_r}{\left|C'_q\right|} a_j\left(S\backslash C'_r\right) - \sum_{j \in C'_r} \frac{\lambda^S_q}{\left|C'_q\right|} a_j\left(S\backslash C'_q\right) \right).$$

Under Proposition 6.1, $a\left(S\right) = \zeta^S$ is easily deduced following a standard induction argument. ∎

Proposition 6.6 *For any ρ, a set of proposals $\left(a\left(S, i\right)^\gamma_{i \in S, \gamma \in \Gamma_{S,i}} \right)_{S \subset N}$ 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. Besides the result, we will also prove that all the proposals are accepted and $a_i\left(S\right)^\gamma \geq \rho a_i\left(S\right) + \left(1-\rho\right) r_i$ for all $i \in S \subset N$ and $\gamma \in \Gamma_{S,i}$.

The argument is by induction. The result holds trivially when $|N| = 1$. Assume that it is true when there are at most $|N| - 1$ players.

Assume we are in a stationary subgame perfect equilibrium in a strictly convex game. Under the induction hypothesis, the expected payoff for the players in $S \neq N$ in any

stationary subgame perfect equilibrium with S as set of active players is $a(S)$. Let $b \in \mathbb{R}^N$ denote the expected final payoff allocation when N is the set of active players.

We proceed by a series of Claims:

Claim 6.4 *Given $C_q \in \mathcal{C}$ in the second stage, assume the proposers are determined by $\gamma \in \Gamma_N$ and the RP is γ_q. Then, all the players in $N \backslash C_q$ accept γ_q's proposal if $a_i(N, \gamma_q) > \rho b_i + (1-\rho) a_i(N \backslash C_q)$ for all $i \in N \backslash C_q$. If $a_i(N, \gamma_q) < \rho b_i + (1-\rho) a_i(N \backslash C_q)$ for some $i \in N \backslash C_q$, then the proposal is rejected.*

In the case of rejection in the second stage, the expected payoff of a player $i \in N \backslash C_q$ is, under the induction hypothesis, $\rho b_i + (1-\rho) a_i(N \backslash C_q)$. Thus, the result follows from a standard argument in this kind of bargaining.

Claim 6.5 *Let $\gamma \in \Gamma_N$ determine the set of proposers in the first stage. Given $C_q \in \mathcal{C}$, assume we are in the subgame that begins after player γ_q makes his proposal. Assume also that all the coalitions with choose representative after C_q are bound to choose their proposer as representative should γ_q's proposal be accepted. Let b^{γ_q} be the expected final payoff allocation if γ_q's proposal is accepted. Then, all the players in $C_q \backslash \gamma_q$ accept γ_q's proposal if $b_i^{\gamma_q} > \rho b_i + (1-\rho) a_i(N \backslash \gamma_q)$ for every $i \in C_q \backslash \gamma_q$. If $b_i^{\gamma_q} < \rho b_i + (1-\rho) a_i(N \backslash \gamma_q)$ for some $i \in C_q \backslash \gamma_q$, the proposal is rejected.*

The result follows the same arguments as in the proof of Claim 6.4. Under these hypothesis, in the case of rejection of γ_q's proposal in the first stage, the expected payoff to a player $i \in C_q \backslash \gamma$ is $\rho b_i + (1-\rho) a_i(N \backslash \gamma_q)$.

Claim 6.6 *All the offers in the first stage are accepted.*

Assume coalitions play the first stage in the order $(C_1, C_2, ..., C_m)$ and that the mechanism reaches coalition C_m, i.e. there has been no previous rejection. Assume γ_m's proposal is rejected. This means the final payoff for player γ_m is $\rho b_{\gamma_m} + (1-\rho) r_{\gamma_m}$.

Define a new proposal $a(N, \gamma_m)$ for player γ_m. First, given $\varepsilon > 0$, let $c^\varepsilon(N, \gamma_m)$ be defined as follows:

$$c_i^\varepsilon(N, \gamma_m) := \begin{cases} -\sum\limits_{C_q \in \mathcal{C} \backslash C_m} \frac{\lambda_q^N}{\lambda_m^N} a_i(N \backslash C_q) + \frac{1}{\lambda_m^N} r_i + \varepsilon & \text{if } i = \gamma_m \\ \frac{1}{\lambda_m^N} a_i(N \backslash \gamma_m) - \sum\limits_{C_q \in \mathcal{C} \backslash C_m} \frac{\lambda_q^N}{\lambda_m^N} a_i(N \backslash C_q) + \varepsilon & \text{if } i \in C_m \backslash \gamma_m \\ a_i(N \backslash C_m) + \varepsilon & \text{if } i \in N \backslash C_m. \end{cases} \tag{9}$$

For $\varepsilon > 0$ small enough, we prove that $\sum_{i \in N} c_i^\varepsilon (N, \gamma_m) \leq v(N)$: Under the induction hypothesis, $a(S, i)_{i \in S, \gamma \in \Gamma_{S,i}}^\gamma$ satisfies P-1, P-2 and P-3 for all $S \neq N$. Under Proposition 6.5, $a(S) = \zeta^S$ for all $S \neq N$. Hence, $c^\varepsilon (N, \gamma_m)$ can be re-written as

$$
c_i^\varepsilon (N, \gamma_m) = \begin{cases} -\sum\limits_{C_q \in \mathcal{C} \backslash C_m} \frac{\lambda_q^N}{\lambda_m^N} \zeta_i^{N \backslash C_q} + \frac{1}{\lambda_m^N} r_i + \varepsilon & \text{if } i = \gamma_m \\[2ex] \frac{1}{\lambda_m^N} \zeta_i^{N \backslash \gamma_m} - \sum\limits_{C_q \in \mathcal{C} \backslash C_m} \frac{\lambda_q^N}{\lambda_m^N} \zeta_i^{N \backslash C_q} + \varepsilon & \text{if } i \in C_m \backslash \gamma_m \\[2ex] \zeta_i^{N \backslash C_m} + \varepsilon & \text{if } i \in N \backslash C_m. \end{cases}
$$

Adding terms,

$$
\lambda_m^N \sum_{i \in N} c_i^\varepsilon (N, \gamma_m) = |N| \lambda_m^N \varepsilon + r_{\gamma_m} + \sum_{i \in C_m \backslash \gamma_m} \zeta_i^{N \backslash \gamma_m}
$$
$$
+ \sum_{C_q \in \mathcal{C} \backslash C_m} \sum_{i \in C_q} \lambda_m^N \zeta_i^{N \backslash C_m} - \sum_{C_q \in \mathcal{C} \backslash C_m} \sum_{i \in C_m} \lambda_q^N \zeta_i^{N \backslash C_q}
$$

$$
\overset{(\text{Proposition } 6.1)}{=} |N| \lambda_m^N \varepsilon + r_{\gamma_m} + \lambda_m^N v(N) + \sum_{i \in C_m \backslash \gamma_m} \zeta_{\gamma_m}^{N \backslash i} - |C_m| \zeta_{\gamma_m}^N
$$

$$
= |N| \lambda_m^N \varepsilon + r_{\gamma_m} + \lambda_m^N v(N) + \sum_{i \in C_m \backslash \gamma_m} \left(\zeta_{\gamma_m}^{N \backslash i} - \zeta_{\gamma_m}^N \right) - \zeta_{\gamma_m}^N
$$

$$
\overset{(\text{Proposition } 6.2)}{=} |N| \lambda_m^N \varepsilon + r_{\gamma_m} + \lambda_m^N v(N) + \sum_{i \in C_m \backslash \gamma_m} \left(\zeta_i^{N \backslash \gamma_m} - \zeta_i^N \right) - \zeta_{\gamma_m}^N
$$

$$
= |N| \lambda_m^N \varepsilon + r_{\gamma_m} + \lambda_m^N v(N) + \sum_{i \in C_m \backslash \gamma_m} \zeta_i^{N \backslash \gamma_m} - \sum_{i \in C_m} \zeta_i^N
$$

$$
\overset{(\text{Proposition } 6.3)}{<} |N| \lambda_m^N \varepsilon + \lambda_m^N v(N).
$$

Hence,

$$
\sum_{i \in N} c_i^\varepsilon (N, \gamma_m) < |N| \varepsilon + v(N)
$$

and hence $\sum_{i \in N} c_i^\varepsilon (N, \gamma_m) \leq v(N)$ for ε small enough.

Define $a(N, \gamma_m) = \rho b + (1 - \rho) c^\varepsilon (N, \gamma_m)$ as the new proposal for player γ_m.

In case of rejection, the expected final payoff for any player $i \in C_m \backslash \gamma_m$ is $\rho b_i + (1 - \rho) a_i (N \backslash \gamma_m)$.

If all the players in $C_m \backslash \gamma_m$ accept $a(N, \gamma_m)$ and the proposal chosen in the second stage is from $C_q \neq C_m$ (probability λ_q^N), then any player $i \in C_m \backslash \gamma_m$ can obtain $\rho b_i + (1 - \rho) a_i (N \backslash C_q)$ by rejecting it. If the proposal chosen in the second stage is from C_m (probability λ_m^N), then it is accepted (by Claim *6.4*).

Thus, if all the players in $C_m \backslash \gamma_m$ accept $a(N, \gamma_m)$, their expected final payoff is at least

$$\sum_{C_q \in \mathcal{C} \backslash C_m} \lambda_q^N \left[\rho b_i + (1-\rho) a_i (N \backslash C_q) \right] + \lambda_m^N a_i (N, \gamma_m)$$

$$= \sum_{C_q \in \mathcal{C} \backslash C_m} \lambda_q^N \left[\rho b_i + (1-\rho) a_i (N \backslash C_q) \right] + \lambda_m^N \rho b_i$$

$$+ (1-\rho) \left[a_i (N \backslash \gamma_m) - \sum_{C_q \in \mathcal{C} \backslash C_m} \lambda_q^N a_i (N \backslash C_q) + \lambda_m^N \varepsilon \right]$$

$$= \rho b_i + (1-\rho) a_i (N \backslash \gamma_m) + (1-\rho) \lambda_m^N \varepsilon$$

for each $i \in C_m \backslash \gamma_m$. Thus, it is optimal for players in $C_m \backslash \gamma_m$ to accept $a(N, \gamma_m)$. Analogously, the expected final payoff for player γ_m after acceptance is at least

$$\sum_{C_q \in \mathcal{C} \backslash C_m} \lambda_q^N \left[\rho b_{\gamma_m} + (1-\rho) a_{\gamma_m} (N \backslash C_q) \right] + \lambda_m^N a_{\gamma_m} (N, \gamma_m)$$

$$= \sum_{C_q \in \mathcal{C} \backslash C_m} \lambda_q^N \left[\rho b_{\gamma_m} + (1-\rho) a_{\gamma_m} (N \backslash C_q) \right] + \lambda_m^N \rho b_{\gamma_m}$$

$$+ (1-\rho) \left(- \sum_{C_q \in \mathcal{C} \backslash C_m} \lambda_q^N a_{\gamma_m} (N \backslash C_q) + r_{\gamma_m} + \lambda_m^N \varepsilon \right)$$

$$= \rho b_{\gamma_m} + (1-\rho) r_{\gamma_m} + (1-\rho) \lambda_m^N \varepsilon.$$

So, it is optimal for γ_m to change his proposal. This contradiction proves that no proposals are rejected in the first stage in C_m. By going backwards, the same reasoning shows that no proposal is rejected in the first stage in $C_{m-1}, ..., C_1$.

Claim 6.7 *All the offers in the second stage are accepted.*

Suppose the proposal of γ_q is bound to be rejected in the second stage. Then, the final payoff for the members of C_q (including γ_q) is r^{C_q} with probability λ_m^N. Under Claim 6.5 and Claim 6.6, we know that $b_i^{\gamma_q} \geq \rho b_i + (1-\rho) a_i (N \backslash \gamma_q)$ for all $i \in C_q \backslash \gamma_q$. Assume that γ_q changes his strategy and proposes

$$a_i (N, \gamma_q) = \begin{cases} r_i + \varepsilon & \text{if } i \in C_q \\ \rho b_i + (1-\rho) a_i (N \backslash C_q) + \varepsilon & \text{otherwise.} \end{cases} \tag{10}$$

We proof that this proposal is feasible for $\varepsilon > 0$ small enough: When S is the set of active players, each $i \in S$ can assure himself at least r_i. Since b is a subgame equilibrium payoff allocation when N is the set of active players, $b_i \geq r_i$ for all $i \in N$. Moreover, the

induction hypothesis implies that $a(S)$ is the only subgame equilibrium payoff allocation when $S \neq N$ is the set of active players. Hence, $a_i(S) \geq r_i$ for all $i \in S \neq N$.

On the other hand, under P-3, $\sum_{i \in N \setminus C_q} a_i(N \setminus C_q) = v(N \setminus C_q)$.

Hence,

$$
\begin{aligned}
\sum_{i \in N} a_i(N, \gamma_q) = \quad & |N|\varepsilon + \sum_{i \in C_q} r_i + \sum_{i \in N \setminus C_q} \rho b_i + \sum_{i \in N \setminus C_q} (1-\rho) a_i(N \setminus C_q) \\
= \quad & |N|\varepsilon + \rho \left(\sum_{i \in C_q} r_i + \sum_{i \in N \setminus C_q} b_i \right) + (1-\rho) \left(\sum_{i \in C_q} r_i + v(N \setminus C_q) \right) \\
\leq \quad & |N|\varepsilon + \rho \sum_{i \in N} b_i + (1-\rho) \left(\sum_{i \in C_q} r_i + v(N \setminus C_q) \right) \\
\leq \quad & |N|\varepsilon + \rho v(N) + (1-\rho) \left(\sum_{i \in C_q} v(\{i\}) + v(N \setminus C_q) \right)
\end{aligned}
$$

since (N, v) is strictly convex,

$$
< |N|\varepsilon + \rho v(N) + (1-\rho) v(N) = |N|\varepsilon + v(N).
$$

Hence, $a(N, \gamma_q)$ is feasible for ε small enough.

Under Claim 6.4, the new proposal is bound to be accepted should γ_q be the RP in the second stage. However, $\left(b_i^{\gamma_q} \right)_{i \in C_q \setminus \gamma_q}$ increases in all coordinates. So, under Claim 6.5, $a(N, \gamma_q)$ is also accepted in the first stage. Moreover, the expected final payoff for γ_q also increases. Hence, we are not in a subgame perfect equilibrium. This contradiction proves that the proposals in the second stage are always accepted.

Since all the proposals are accepted, we can assure that $b = a(N)$ and $b^{\gamma_q} = a(N)^\gamma$ for all $\gamma \in \Gamma_N$.

We show now that P-1, P-2 and P-3 hold.

Suppose P-3 does not hold, i.e. there exists a player $i \in C_q$ such that $\sum_{j \in N} a_j(N, i) < v(N)$. Thus, there exists $\varepsilon > 0$ such that $d \in \mathbb{R}^N$ defined as $d_j = a_j(N, i) + \varepsilon$ for all $j \in N$ satisfies $\sum_{j \in N} d_j < v(N)$.

Notice that, since the proposal $a(N, i)$ of player i is accepted, under Claim 6.5, together with Claim 6.6 and Claim 6.7, we know that, given $\gamma \in \Gamma_{N,i}$, $a_j(N)^\gamma \geq \rho b_j + (1-\rho) a_j(N \setminus i)$ for every $j \in C_q \setminus i$ and, under Claim 6.4, $a_j(N, i) \geq \rho b_j + (1-\rho) a_j(N \setminus C_q)$ for every $j \in N \setminus C_q$. So, if player i changes his proposal to d, it is bound to be accepted and his expected final payoff improves by $\lambda_q^N \varepsilon > 0$. This contradiction proves P-3.

Suppose P-2 does not hold. Let $\gamma \in \Gamma_{N,i}$ and let $j_0 \in C_q \setminus i$ such that $a_{j_0}(N)^\gamma = \rho a_{j_0}(N) + (1-\rho) a_{j_0}(N \setminus i) + \alpha$ with $\alpha \neq 0$. Under Claim 6.5, $\alpha > 0$.

Assume player i changes his proposal so that $a_{j_0}(N,i)$ decreases β and $a_j(N,i)$ increases $\frac{\beta}{|N|-1}$ for all $j \in N\backslash i$, with β strictly between 0 and α. The new proposal $a(N,i)$ satisfies the conditions of Claim 6.4 and Claim 6.5, and thus it is due to be accepted. Also, player i improves his expected payoff by $\frac{\lambda_q^N \beta}{|N|-1} > 0$. This contradiction proves P-2.

The reasoning for P-1 is similar to that for P-2 so it is omitted.

It remains to show that $a_i(N)^\gamma \geq \rho a_i(N) + (1-\rho)r_i$ for all $i \in N$ and $\gamma \in \Gamma_{N,i}$. Notice that player $i \in N$ can guarantee himself a payoff of at least $\rho a_i(N) + (1-\rho)r_i$ by proposing r and accepting only proposals which give him at least $\rho a_i(N) + (1-\rho)r_i$ as expected final payoff. Thus, $a_i(N)^\gamma \geq \rho a_i(N) + (1-\rho)r_i$.

The next step is to show that proposals $(a(S,i)_{i \in S})_{S \subset N}$ satisfying P-1, P-2 and P-3 can be supported as a subgame perfect equilibrium, and $a_i(S)^\gamma \geq \rho a_i(S) + (1-\rho)r_i$ for all $i \in S \subset N$ and $\gamma \in \Gamma_{S,i}$.

Under the induction hypothesis, these results are true for any $S \neq N$. We prove first that $a_i(N)^\gamma \geq \rho a_i(N) + (1-\rho)r_i$ for all $i \in N$ and $\gamma \in \Gamma_{N,i}$.

Given $i \in C_q \in \mathcal{C}$, the vector $c(N,i)$, defined as

$$
c_j(N,i) = \begin{cases}
-\sum\limits_{C_r \in \mathcal{C}\backslash C_q} \frac{\lambda_r^N}{\lambda_q^N} a_i(N\backslash C_r) + \frac{1}{\lambda_q^N} r_i & \text{if } j = i \\
\frac{1}{\lambda_q^N} a_j(N\backslash i) - \sum\limits_{C_r \in \mathcal{C}\backslash C_q} \frac{\lambda_q^N}{\lambda_m^N} a_j(N\backslash C_r) & \text{if } j \in C_q\backslash i \\
a_j(N\backslash C_q) & \text{if } j \in N\backslash C_q.
\end{cases}
$$

is a feasible payoff allocation (analogous to (9)). Hence $\widetilde{c}(N,i) := \rho a(N) + (1-\rho)c(N,i)$ is feasible, too.

Let $i \in C_q \in \mathcal{C}$ and $\gamma \in \Gamma_{N,i}$. Let $j \in C_q\backslash i$. Since $a(N,i)$ satisfies P-1 and P-2,

$$
\lambda_q^N a_j(N,i)^\gamma = a_j(N)^\gamma - \sum_{C_r \in \mathcal{C}\backslash C_q} \lambda_r^N a_j(N,\gamma_r)^\gamma
$$

$$
= \rho a_j(N) + (1-\rho)a_j(N\backslash i) - \sum_{C_r \in \mathcal{C}\backslash C_q} \lambda_r^N [\rho a_j(N) + (1-\rho)a_j(N\backslash C_r)]
$$

$$
= \lambda_q^N \rho a_j(N) + (1-\rho)\left[a_j(N\backslash i) - \sum_{C_r \in \mathcal{C}\backslash C_q} \lambda_r^N a_j(N\backslash C_r) \right]
$$

and hence

$$
a_j(N,i)^\gamma = \rho a_j(N) + (1-\rho)\left[\frac{1}{\lambda_q^N} a_j(N\backslash i) - \sum_{C_r \in \mathcal{C}\backslash C_q} \frac{\lambda_r^N}{\lambda_q^N} a_j(N\backslash C_r) \right]
$$

for all $j \in C_q \backslash i$. But this expression does not depend on γ. Hence, $a_j(N, i) = a_j(N, i)^\gamma$ for all $\gamma \in \Gamma_{N,i}$. Under P-1 and P-3, $a(N, i) = a(N, i)^\gamma$ for all $\gamma \in \Gamma_{N,i}$.

Furthermore, $a_j(N, i) = \widetilde{c}_j(N, i)$ for all $j \in N \backslash i$. Hence, $a(N, i) \geq \widetilde{c}(N, i)$ because $\sum_{j \in N} a_j(N, i) = v(N)$ and $\sum_{j \in N} \widetilde{c}_j(N, i) \leq v(N)$. Thus,

$$a_i(N, i) \geq \widetilde{c}_i(N, i) = \rho a_i(N) + (1 - \rho) \left[- \sum_{C_r \in \mathcal{C} \backslash C_q} \frac{\lambda_r^N}{\lambda_q^N} a_i(N \backslash C_r) + \frac{1}{\lambda_q^N} r_i \right]. \qquad (11)$$

Under P-1 and (11),

$$
\begin{aligned}
a_i(N)^\gamma = & \sum_{C_r \in \mathcal{C}} \lambda_r^N a_i(N, \gamma_r)^\gamma = \sum_{C_r \in \mathcal{C} \backslash C_q} \lambda_r^N a_i(N, \gamma_r)^\gamma + \lambda_q^N a_i(N, i) \\
\geq & \sum_{C_r \in \mathcal{C} \backslash C_q} \lambda_r^N \left[\rho a_i(N) + (1 - \rho) a_i(N \backslash C_r) \right] \\
& + \lambda_q^N \rho a_i(N) + (1 - \rho) \left[- \sum_{C_r \in \mathcal{C} \backslash C_q} \lambda_r^N a_i(N \backslash C_r) + r_i \right] \\
= & \rho a_i(N) + (1 - \rho) r_i.
\end{aligned}
$$

The last step is to prove that the strategies corresponding to these proposals form a subgame perfect equilibrium. The reasoning is analogous to those used by Hart and Mas-Colell (1996) (Proposition 1). Under the induction hypothesis, the result hold in any subgame after a player (or coalition) has dropped out. Fix a player $i \in C_q \in \mathcal{C}$. If he rejects the offer from a proposer $j \in C_q \backslash i$, his expected final payoff is $\rho a_i(N) + (1 - \rho) a_i(N \backslash j)$. If he rejects the offer from a RP $j \in N \backslash C_q$, his expected final payoff is $\rho a_i(N) + (1 - \rho) a_i(N \backslash C_q)$. In any case, his expected final payoff is the same as that the other player is offering, and he does not improve by rejecting it. If the proposer is player i himself (i.e. $\gamma \in \Gamma_{N,i}$), the strategies of the other players do not allow him to decrease his proposal to any of them (since it would be rejected under Claim 6.4 and Claim 6.5). Moreover, increasing one or more of his offers to the other players keeping the rest unaltered implies his own payoff decreases (under P-3). Finally, if he proposes an unacceptable offer, his expected final payoff will be at most $\rho a_i(N) + (1 - \rho) r_i$, whereas the proposed strategy gives him $a_i(N)^\gamma$. Since $a_i(N)^\gamma \geq \rho a_i(N) + (1 - \rho) r_i$, he does not improve. ∎

Proposition 6.7 *There always exists a stationary subgame perfect equilibrium for convex games.*

Proof. Under Proposition 6.6, it is enough to prove that there exits a set of proposals satisfying P-1, P-2 and P-3. We define $a_i(\{i\}, i) = r_i$ for all $i \in S$. Assume we have

26

defined $a(T, i)$ for all $i \in T \subset, T \neq S$. We define:

$$a_j (S, i)^\gamma = \rho a_j (S) + (1 - \rho) a_j (S \backslash C_q')$$

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. ∎

References

Alonso-Meijide, J. and Fiestras-Janeiro, M. (2002). Modification of the banzhaf value for games with a coalition structure. *Annals of Operations Research*, 109:213–227.

Amer, R., Carreras, F., and Giménez, J. (2002). The modified Banzhaf value for games with coalition structure: an axiomatic characterization. *Mathematical Social Sciences*, 43:45–54.

Aumann, R. and Dreze, J. (1974). Cooperative games with coalition structure. *International Journal of Game Theory*, 3(4):217–237.

Bergantiños, G. and Vidal-Puga, J. (2005). The consistent coalitional value. *Mathematics of Operations Research*, 30(4):832–851.

Calvo, E. and Gutiérrez, E. (2010). Solidarity in games with coalition structure. *Mathematical Social Sciences*, 60(3):196–206.

Calvo, E., Lasaga, J. J., and Winter, E. (1996). The principle of balanced contributions and hierarchies of cooperation. *Mathematical Social Sciences*, 31(3):171–182.

Carreras, F. and Puente, M. (2006). A parametric family of mixed coalitional values. In Seeger, A., editor, *Recent Advances in Optimization*, volume 563 of *Lecture Notes in Economics and Mathematical Systems*, pages 323–339. Springer, Berlin Heidelberg.

Chae, S. and Heidhues, P. (2004). A group bargaining solution. *Mathematical Social Sciences*, 48(1):37–53.

Chae, S. and Moulin, M. (2010). Bargaining among groups: an axiomatic viewpoint. *International Journal of Game Theory*, 39(1-2):71–88.

Gómez-Rúa, M. and Vidal-Puga, J. (2010). The axiomatic approach to three values in games with coalition structure. *European Journal of Operational Research*, 207(2):795–806.

Gómez-Rúa, M. and Vidal-Puga, J. (2011). Balanced per capita contributions and levels structure of cooperation. *Top*, 19:167–176.

Haeringer, G. (2000). A new weight scheme for the Shapley value. Beta - working paper 9910, Bureau d'Economie Théorique et Appliquée.

Haeringer, G. (2006). A new weight scheme for the Shapley value. *Mathematical Social Sciences*, 52:88–98.

Harsanyi, J. (1977). *Rational behavior and bargaining equilibrium in games and social situations.* Cambridge University Press.

Hart, S. and Kurz, M. (1983). Endogenous formation of coalitions. *Econometrica*, 51(4):1047–1064.

Hart, S. and Mas-Colell, A. (1996). Bargaining and value. *Econometrica*, 64(2):357–380.

Kalai, E. and Samet, D. (1985). Monotonic solutions to general cooperative games. *Econometrica*, 53:307–327.

Kalai, E. and Samet, D. (1987). On weighted Shapley values. *International Journal of Game Theory*, 16(3):205–222.

Kalai, E. and Samet, D. (1988). Weighted Shapley values. In Roth, A. E., editor, *The Shapley value: Essays in honour of Lloyds S. Shapley*, pages 83–100. Cambridge University Press, Cambridge.

Kamijo, Y. (2008). Implementation of weighted values in hierarchical and horizontal cooperation structures. *Mathematical Social Sciences*, 56(3):336–349.

Kamijo, Y. (2009). A two-step Shapley value for cooperative games with coalition structures. *International Game Theory Review*, 11(02):207–214.

Kamijo, Y. (2013). The collective value: a new solution for games with coalition structures. *Top*, 21(3):572–589.

Levy, A. and Mc Lean, R. (1989). Weighted coalitional structure values. *Games and Economic Behavior*, 1:234–249.

Młodak, A. (2003). Three additive solutions of cooperative games with a priori unions. *Applicationes Mathematicae*, 30(1):69–87.

Myerson, R. B. (1980). Conference structures and fair allocation rules. *International Journal of Game Theory*, 9(3):169–182.

Nash, J. (1950). The bargaining problem. *Econometrica*, 18(2):155–162.

Owen, G. (1968). A note on the Shapley value. *Management Science*, 14:731–732.

Owen, G. (1977). Values of games with a priori unions. In Henn, R. and Moeschlin, O., editors, *Mathematical Economics and Game Theory*, volume 141 of *Lecture Notes in Economics and Mathematical Systems*, pages 76–88. Springer-Verlag, Berlin.

Pérez-Castrillo, D. and Wettstein, D. (2001). Bidding for the surplus: a non-cooperative approach to the Shapley value. *Journal of Economic Theory*, 100(2):274–294.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica*, 50(1):97–109.

Shapley, L. S. (1953a). *Additive and non-additive set functions*. PhD thesis, Princeton University.

Shapley, L. S. (1953b). A value for n-person games. In Kuhn, H. and Tucker, A., editors, *Contributions to the theory of games*, volume II of *Annals of Mathematics Studies*, pages 307–317. Princeton University Press, Princeton NJ.

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

Vidal-Puga, J. and Bergantiños, G. (2003). An implementation of the Owen value. *Games and Economic Behavior*, 44(2):412–427.