On the Ordinal Invariance of Power Indices on Coalitional Games

Jean-Paul Doignon[1*], **Stefano Moretti**[2], **Meltem Öztürk**[2]
[1]Département de Mathématique, c.p. 216, Université Libre de Bruxelles, Bruxelles, Belgium
[2]Université Paris Dauphine, PSL Research University, CNRS, Lamsade, 75016 Paris, France
Jean-Paul.Doignon@ulb.be, {stefano.moretti, meltem.ozturk}@dauphine.fr

Abstract

In a coalitional game, the coalitions are weakly ordered according to their worths in the game. When moreover a power index is given, the players are ranked according to the real numbers they are assigned by the power index. If any game inducing the same ordering of the coalitions generates the same ranking of the players then, by definition, the game is (ordinally) stable for the power index, which in turn is ordinally invariant for the game. If one is interested in ranking players of a game which is stable, re-computing the power index numbers when the coalitional worths slightly fluctuate or are uncertain becomes useless. Bivalued games are easy examples of games stable for any power index which is linear. Among general games, we characterize those that are stable for a given linear index. Note that the Shapley and Banzhaf indices, frequently used in AI, are particular semivalues, and all semivalues are linear indices. To check whether a game is stable for a specific semivalue, it suffices to inspect the ordering of the coalitions and to perform some direct computation based on the semivalue parameters.

1 Introduction

Classical power indices and, more in general, solutions or values for coalitional games [Shapley, 1953; Banzhaf III, 1964; Dubey *et al.*, 1981] aim at converting the information about the performance of coalitions into a personal attribution of "power" representing each player's role over all possible coalitions. Game theory is arguably the field which has dealt with the notion of power index the most [Holler and Nurmi, 2013]. The notion has also been extensively applied in voting theory (for instance, in order to define the "power" of different countries in the European Council [Algaba *et al.*, 2003]), network theory (for measuring the centrality of network elements [Michalak *et al.*, 2013]), belief merging/revision (for the responsibility of a formula in the inconsistency of a belief base [Hunter and Konieczny, 2006]), multicriteria decision aiding (the impact and synergy of some criteria [Grabisch and Roubens, 2000], machine learning (selecting best features, for instance using SHAP [Lundberg and Lee, 2017]), argumentation theory (influence of an argument in a debate [Amgoud *et al.*, 2017]), etc.

Due to the high number of power indices in the literature and their heterogeneous nature, some studies looked for conditions under which distinct power indices rank players in the same manner (see, for instance, [Carreras and Freixas, 2008; Lambo and Moulen, 2002]).

Unfortunately, power indices computed on the characteristic function of a game often are very sensitive to small fluctuations of coalitions' performance. Moreover, due to multiple and barely measurable attributes of performance to coalitions, it is more reasonable to expect that only a ranking over coalitions is given as input for the computation of the relative power of players [Haret *et al.*, 2018; Bernardi *et al.*, 2019]. There results our main question: For a given power index, characterize the coalitional games v for which the ranking over the players set remains the same when the game v is replaced with any game w such that v and w induce the same ranking over the coalitions. By definition, such a game v is stable for the given power index, and the power index is then ordinally invariant for the game. Consider for example a coalitional game with three players. If we are interested in ranking two players i and j using the Banzhaf index [Banzhaf III, 1964], it is easy to verify that a little variation in the worths of coalitions containing either i or j (but not both of them) may completely change the relative ranking of the two players (see for instance Example 2 in Section 3). To the contrary, any game on two players is always stable, not only for the Banzhaf index but also for any semivalue (in the sense of [Dubey *et al.*, 1981; Lucchetti *et al.*, 2015]); see Proposition 1.

As an answer to our main question, Theorem 1 offers a characterization of stable games under the mundane assumption that the power index is linear (linearity allows us to take advantage of a certain decomposition of the game function). Note that the Shapley and Banzhaf indices, as well as the more general semivalues, all are linear indices. For the Banzhaff index, Theorem 2 provides a better characterization of game stability (in that it involves a lesser number of coalitions worths). It is extended to all semivalues in Theorem 3.

Before stating the three theorems in Sections 5 and 6, we introduce in Section 2 some necessary notation and defini-

*Contact Author

tions, and show some results on particular cases in Sections 3 and 4. We conclude by some perspectives in Section 7.

2 Notations and Definitions

2.1 Set Functions

A **set function** (N, v) consists of a finite set N, of cardinality n, together with a mapping $v : 2^N \to \mathbb{R}$; it is a (*Transferable Utility* or *coalitional*) **game** when $v(\varnothing) = 0$. The elements of the set N are the **players**, the subsets of N are the **coalitions**, the map v is the **characteristic function**, and $v(S)$ is the **worth** of coalition S. When N is clear from the context, we denote the set function (N, v) simply by v. The collection of all set functions on N is denoted as $\mathbb{R}^{2^N}$. A set function v is **bivalued** when there exist two real numbers λ and μ such that $v(S) = \lambda$ or $v(S) = \mu$ for all S in 2^N, and **binary** if moreover $\lambda = 0$ and $\mu = 1$.

2.2 Scorings

A **social scoring rule** or **scoring** π on N attributes to any set function[1] v on N a **scoring vector** in $\mathbb{R}^N$. The **score** of player i for the set function v under the scoring π is the component $\pi_i(v)$ of $\pi(v)$.

The **Shapley scoring** of a set function v on N, with $|N| = n$, attributes to player i the score

$$\pi_i^{\text{Sh}}(v) \;=\; \sum_{S \in 2^N :\, i \notin S} \frac{1}{n \binom{n-1}{|S|}} \big(v(S \cup \{i\}) - v(S)\big). \quad (1)$$

The **Banzhaf scoring** attributes to player i the score

$$\pi_i^{\text{Ban}}(v) \;=\; \sum_{S \in 2^N :\, i \notin S} \frac{1}{2^{n-1}} \big(v(S \cup \{i\}) - v(S)\big). \quad (2)$$

Here is a first family of scorings which generalizes both the Shapley and the Banzhaf scorings. Let q be a vector whose components $q(k)$, for $k = 0, 1, \ldots, n-1$, are positive[2] real numbers such that $\sum_{k=0}^{n-1} q(k) \binom{n-1}{k} = 1$. A **(regular) semivalue** $\pi^{(q)}$ **with parameter vector** q attributes, for any given set function v, to player i the score

$$\pi_i^{(q)}(v) \;=\; \sum_{S \in 2^N :\, i \notin S} q(|S|) \, \big(v(S \cup \{i\}) - v(S)\big). \quad (3)$$

To build as in [Lucchetti *et al.*, 2015] a more general family of scorings, let Λ be a family of positive real numbers $p_i(S)$, with $i \in N$ and $i \notin S \subset N$, called **parameters**, such that $\sum_{S \in 2^N :\, i \notin S} p_i(S) = 1$ for each player i. A **(regular) probabilistic scoring rule with parameter family** Λ attributes to player i the score

$$\pi_i^{\Lambda}(v) \;=\; \sum_{S \in 2^N :\, i \notin S} p_i(S) \, \big(v(S \cup \{i\}) - v(S)\big). \quad (4)$$

Any scoring rule π as in Equations (1), (2), (3), (4) is **steady**, that is, for the particular set function $\mathbf{1}$ mapping any coalition to 1, it assigns equal scores to all players. It is also **linear** in the sense that for all real numbers α_1, α_2 and set functions v_1, v_2 on N,

$$\pi(\alpha_1 \, v_1 + \alpha_2 \, v_2) \;=\; \alpha_1 \pi(v_1) + \alpha_2 \, \pi(v_2). \quad (5)$$

Figure 1 summarizes relations between different types of scorings.

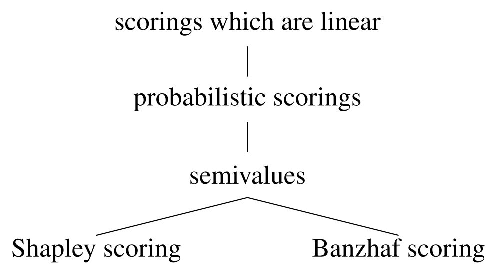

Figure 1: Hierarchy of various families of scorings.

2.3 Set Ordering and Player Ranking

Any set function v on N **induces** the **set ordering** on 2^N which is the weak order $\sqsupseteq_v$ satisfying for S, T in 2^N

$$S \sqsupseteq_v T \iff v(S) \geqslant v(T).$$

Two set functions u and v on the same set N are **ordinally equivalent** if they induce the same set ordering on 2^N. The **equivalence class** $\mathcal{C}(v)$ of v consists of all set functions u on N such that

$$u(S) \geqslant u(T) \iff v(S) \geqslant v(T).$$

There is thus a canonical bijection from equivalence classes of set functions to weak orders on 2^N.

Any scoring π on N induces for any set function v a **player ranking**, which is the weak order $\succsim_v^\pi$ **induced** on N by the scores under π, that is, for i, j in N,

$$i \succsim_v^\pi j \iff \pi_i(v) \geqslant \pi_j(v).$$

3 Stable Set Functions

When does the set ordering induced by a set function determine the player ranking for a given scoring? Let us first turn the question into the next formal problem, for a fixed set N of players.

Definition 1. *Let π be a scoring. A set function v is **π-stable** when any set function ordinally equivalent to v produces under π the same player ranking as v does. We then also say that π is ordinally invariant for the set function v.*

Problem 1. *Given a scoring π on N, characterize the set functions on N that are π-stable.*

Here is a kind of 'local' form of stability.

Definition 2. *Let π be a scoring on N, and i, j two players in N. A set function v on N is **two-player π-stable** for i, j when for any set function u ordinally equivalent to v the comparison of the scores of i and j under π is the same for u as for v, that is*

$$\pi_i(u) \geqslant \pi_j(u) \quad \iff \quad \pi_i(v) \geqslant \pi_j(v). \quad (6)$$

[1] Authors often restrict here the set functions to games, and use terms as power indices, solutions or values instead of scorings.

[2] By 'positive', we mean 'strictly positive'.

Many of our results on stability arise first for two-player stability; results on general stability follow then easily. Note also that two-player π-stability for players h, i with $\pi_h(v) \geqslant \pi_i(v)$ together with two-player π-stability for players i, j with $\pi_i(v) \geqslant \pi_j(v)$ implies two-player π-stability for h, j.

Remark 1. *Let $\mathcal{G}^{(N)}$ be the collection of all games on N. A **solution** on N is any mapping $\sigma : \mathcal{G}^{(N)} \to \mathbb{R}^N$. Any solution σ on N extends to the scoring $\pi^\sigma : \mathbb{R}^{2^N} \to \mathbb{R}^N$ such that $\pi^\sigma(v) = \sigma(v - v(\varnothing))$, for all $v \in \mathbb{R}^{2^N}$. A game w is **two-player σ-solid** for players i, j when for any game w' ordinally equivalent to w there holds $\sigma_i(w) \geqslant \sigma_j(w)$ exactly when $\sigma_i(w') \geqslant \sigma_j(w')$. Note that the latter happens if and only if the set function w is two-player π^σ-stable for i, j. Thus solidity reduces to stability, so we focus on the latter notion.*

We start with the simple case in which there are only two players.

Proposition 1. *When $|N| = 2$, any set function v on N is $\pi^{(q)}$-stable for any semivalue $\pi^{(q)}$ with parameter vector q.*

Proof. Let $N = \{i, j\}$. By Equation (3) we have

$$\pi_i^{(q)}(v) = q(0)\big(v(\{i\}) - v(\varnothing)\big) + q(1)\big(v(\{i,j\}) - v(\{j\})\big),$$
$$\pi_j^{(q)}(v) = q(0)\big(v(\{j\}) - v(\varnothing)\big) + q(1)\big(v(\{i,j\}) - v(\{i\})\big),$$

and thus

$$\pi_i^{(q)}(v) - \pi_j^{(q)}(v) \;=\; \big((q(0) + q(1)\big)\big(v(\{i\}) - v(\{j\})\big).$$

Now note

$$\pi_i^{(q)}(v) \geqslant \pi_j^{(q)}(v) \iff v(\{i\}) \geqslant v(\{j\})$$

because $q(0)$ and $q(1)$ are positive. Thus the player ranking is determined by the coalition ordering. ☐

The conclusion in Proposition 1 does not remain valid for all probabilistic scorings, as the next example shows.

Example 1. *Let $N = \{a, b\}$ and consider the probabilistic scoring π^Λ with parameter family Λ given by*

$$p_a(\varnothing) = 0.9, \qquad p_a(\{b\}) = 0.1,$$
$$p_b(\varnothing) = 0.2, \qquad p_b(\{a\}) = 0.8.$$

For real numbers α and β, the set function v specified by

S	$\{a,b\}$	$\{a\}$	$\{b\}$	$\varnothing$
$v(S)$	β	α	0	0

leads to

$$\pi_a^\Lambda - \pi_b^\Lambda = 1.7\,\alpha - 0.7\,\beta.$$

The last quantity is negative for some choice of the numbers α and β, and positive form some other choice, with both choices satisfying $\beta > \alpha > 0$. Thus v is not π^Λ-stable.

Proposition 1 is no more satisfied if there are more than two players. For the Banzhaf scoring, the two set functions in Example 2 lead to opposite player rankings although the two set functions are ordinally equivalent.

Example 2. *Let $N = \{a, b, c\}$ and v be the set function specified by*

S	N	$\{a,b\}$	$\{a,c\}$	$\{b,c\}$	$\{c\}$	$\{b\}$	$\{a\}$	$\varnothing$
$v(S)$	9	8	6	4	3	2	1	0
$w(S)$	9	8	7	6	5	3	1	0

For the Banzhaf scoring, the player rankings are $a \succ_v^{Ban} b \succ_v^{Ban} c$ and $c \succ_w^{Ban} b \succ_w^{Ban} a$. Thus neither v nor w is Banzhaf-stable.

4 First Results

The next two propositions present two families of set functions which are stable for fairly general scorings.

Proposition 2. *Any bivalued set function v on N is π-stable for any steady and linear scoring π (thus in particular for any probabilistic scoring, the Shapley scoring and the Banzhaf scoring).*

Proof. Let $v(2^N) = \{\alpha, \beta\}$ with $\alpha > \beta$. The set ordering $\sqsupseteq_v$ is a weak order with two equivalence classes (namely $v^{-1}(\alpha)$ and $v^{-1}(\beta)$ with $v^{-1}(\alpha) \sqsupset_v v^{-1}(\beta)$). Thus the same must hold for any set function w ordinally equivalent to v. As a consequence, there exist some real numbers λ and μ, with $\lambda > 0$, such that $w = \lambda\, v + \mu\, \mathbf{1}$. Linearity of π then gives

$$\pi_i(w) = \lambda\, \pi_i(v) + \mu\, \pi_i(\mathbf{1}). \tag{7}$$

Because π is steady, $\pi_i(\mathbf{1})$ is independent of i. Thus v and w induce the same player ranking under π. ☐

The following example (with only three distinct coalition worths) shows that when there are more than two worths assigned by a set function we can no longer guarantee the stability.

Example 3. *Let $N = \{a, b, c\}$. Take any set function v on N with set ordering $\sqsupseteq$ satisfying $\{a, b, c\} \simeq \{a, b\} \simeq \{a, c\} \sqsupset \{b, c\} \simeq \{c\} \simeq \{b\} \sqsupset \{a\} \simeq \varnothing$. Letting α, β, γ respectively be the numbers assigned by v to the coalitions in the three equivalence classes, there must only hold $\alpha > \beta > \gamma$. For any semivalue $\pi^{(q)}$ with parameter q, we then have*

$$\pi_a^{(q)}(v) - \pi_b^{(q)}(v) \;=\; \big(q(0) + q(1)\big)(\gamma - \beta) + \big(q(1) + q(2)\big)(\alpha - \beta).$$

By setting $\beta = 1$ and $\gamma = 0$, and taking $\alpha > 1$, with either α sufficiently close to 1 or sufficiently high, we obtain set functions ordinally equivalent to v, whose player rankings have either b or a before the other (details are left to the reader). Thus the monotone set function v is unstable for any semivalue.

The next Lemma and Proposition assume properties of set functions close to responsiveness[3].

[3]According to [Roth, 1985], a set function v satisfies **responsiveness** if and only if $v(\{j\}) \geqslant v(\{i\}) \iff v(S \cup \{j\} \geqslant v(S \cup \{i\})$ for all $i, j \in N$ and $S \subseteq N \setminus \{i, j\}$.

Lemma 1. *Let i, j be players in N, let v be a set function on N and let $\pi^{(q)}$ be any semivalue. If for all S in $2^{S\setminus\{i,j\}}$ there holds $v(S \cup \{j\}) \geqslant v(S \cup \{i\})$, then $\pi_j^{(q)}(v) \geqslant \pi_i^{(q)}(v)$. Moreover, if for at least one S the inequality is strict, then $\pi_j^{(q)}(v) > \pi_i^{(q)}(v)$.*

Proof. For any semivalue $\pi^{(q)}$ with parameter vector q we have

$$\pi_j^{(q)}(v) - \pi_i^{(q)}(v) = \sum_{S \in 2^N : i,j \notin S} \big(q(|S|) + q(|S|+1)\big)\big(v(S \cup \{j\}) - v(S \cup \{i\})\big) \tag{8}$$

Both assertions immediately follow, because all numbers $q(|S|) + q(|S|+1)$ are positive. □

Proposition 3. *Let v be any set function satisfying the following double condition: for all i, j in N and $S \in 2^{N\setminus\{i,j\}}$,*

$v(\{i\}) > v(\{j\})$ *implies* $v(S \cup \{i\}) \geqslant v(S \cup \{j\})$;

$v(\{i\}) = v(\{j\})$ *implies* $v(S \cup \{i\}) = v(S \cup \{j\})$.

Then the set function v is $\pi^{(q)}$-stable for any semivalue $\pi^{(q)}$ (thus in particular for the Shapley scoring and the Banzhaf scoring). Moreover the player ranking $\succsim_v^{\pi^{(q)}}$ satisfies

$$i \succsim_v^{\pi^{(q)}} j \iff v(\{i\}) \geqslant v(\{j\}). \tag{9}$$

However there are Banzhaf-stable games satisfying (9) *which do not meet the double condition.*

Proof. Assume that the set function v satisfies the double condition. Then for all i, j in N, by Lemma 1, $v(\{i\}) > v(\{j\})$ implies $\pi_i^{(q)}(v) > \pi_j^{(q)}(v)$. Similarly, $v(\{i\}) = v(\{j\})$ implies $\pi_i^{(q)}(v) = \pi_j^{(q)}(v)$ and $v(\{i\}) < v(\{j\})$ implies $\pi_i^{(q)}(v) < \pi_j^{(q)}(v)$. Thus the player ranking $\succsim_v^{\pi^{(q)}}$ satisfies Equation (9).

All set functions w ordinally equivalent to v also satisfy the double condition (where we substitute w for v). Therefore w satisfies Equation (9) when v is replaced with w. The first two assertions follow, because v and w induces the same ordering of one-element sets.

The game v in the following example is Banzhaf-stable and satisfies (9) for $i = a$ and $j = b$, but not the double condition (indeed, $v(\{a\}) > v(\{b\})$ and $v(\{a, c\}) < v(\{b, c\})$).

S	N	$\{a,b\}$	$\{a\}$	$\{b,c\}$	$\{a,c\}$	$\{b\}$	$\{c\}$	$\varnothing$
$v(S)$	7	6	5	4	3	2	1	0

□

5 Necessary and Sufficient Conditions for Two-player Stability

Until now we met some special cases guarantying stability of a set function for a linear scoring. Theorem 1 in this section delivers a necessary and sufficient condition for stability (again for a fixed set N of players).

Given any set function v on N and any coalition A, we define a binary set function v_A by letting for $S \in 2^N$

$$v_A(S) = \begin{cases} 1 & \text{if } S \sqsupseteq_v A, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

Fix a nondecreasing labelling of all coalitions

$$S_1, \quad S_2, \quad \ldots, \quad S_{2^n} \tag{11}$$

in the sense that

$$v(S_1) \leqslant v(S_2) \leqslant \cdots \leqslant v(S_{2^n}). \tag{12}$$

For $1 \leqslant k \leqslant 2^n$, we then write v_k for the binary set function v_{S_k}. Thus $v_k(S_j) = 1$ for all $j \in \{k, k+1, \ldots, 2^n\}$, and $v_k(S_j) = 0$ if and only if $S_j \sqsubset_v S_k$.

Proposition 4. *With the above notations, there holds the equality of set functions*

$$v = v(S_1)v_1 + \sum_{k=2}^{2^n} \big(v(S_k) - v(S_{k-1})\big)\, v_k. \tag{13}$$

Definition 3. *We say that v_1, v_2, …, v_{2^n} are the* binary set functions *attached to the given set function v for the chosen, nondecreasing labelling of coalitions, and call the right-hand side of Equation* (13) *the corresponding* binary set function decomposition *of v or in short the* binary decomposition *of v.*

Proposition 5. *Given a set function v, select a nondecreasing labelling of all coalitions and form the corresponding decomposition of v in binary set functions v_1, v_2, …, v_{2^n} as in Equation* (13)*. For any scoring π which is linear (in particular for all probabilistic scorings π^Λ), there holds the equality of vectors in $\mathbb{R}^N$*

$$\pi(v) = v(S_1)\pi(v_1) + \sum_{k=2}^{2^n} \big(v(S_k) - v(S_{k-1})\big)\, \pi(v_k). \tag{14}$$

Proof. Equation (14) follows from applying the linear mapping π to both sides of Equation (13). □

Here is a necessary and sufficient condition for two-player stability, which ultimately relies only on the set ordering induced by v.

Theorem 1. *Let v be a set function on N, π be a scoring on N which is steady and linear, and i, j be players in N. Then the two following conditions are equivalent, where v_A is as in Equation* (10)*:*

(i) *for all set functions u on N ordinally equivalent to v:*

$$\pi_i(u) \geqslant \pi_j(u);$$

(ii) *for all coalitions A in 2^N:*

$$\pi_i(v_A) \geqslant \pi_j(v_A). \tag{15}$$

Proof. Let $S_1, S_2, \ldots, S_{2^n}$ be the sequence of coalitions labeled in such a way that $v(S_1) \leqslant v(S_2) \leqslant \ldots \leqslant v(S_{2^n})$. For any set function u ordinally equivalent to v, notice $u(S_1) \leqslant u(S_2) \leqslant \ldots \leqslant u(S_{2^n})$. So for $k = 1, 2, \ldots, 2^n$ we can make $v_k = u_k$, where v_k and u_k are the binary set functions attached to v and u, respectively. Hence by Equation (14) together with $\pi_i(v_1) = \pi_j(v_1)$ (because v_1 is the constant mapping $\mathbf{1}$ and the scoring π is steady), we have

$$\pi_i(v) - \pi_j(v) = \sum_{k=2}^{2^n} (v(S_k) - v(S_{k-1}))\big(\pi_i(v_k) - \pi_j(v_k)\big) \tag{16}$$

$$\pi_i(u) - \pi_j(u) = \sum_{k=2}^{2^n} (u(S_k) - u(S_{k-1}))\big(\pi_i(v_k) - \pi_j(v_k)\big). \tag{17}$$

($\Leftarrow$) Condition (ii) implies Condition (i) in view of Equation (17), because for $k = 2, 3, \ldots, 2^n$, both $u(S_k) - u(S_{k-1}) \geqslant 0$ and $v_k = v_A$ for $S_k = A$.

($\Rightarrow$) Proceeding by contradiction, assume Condition (i) and also $\pi_i(v_A) < \pi_j(v_A)$ for some $A \in 2^N$. Select the smallest k such that $v(A) = v(S_k)$, then $k > 1$ and $v(S_{k-1}) < v(S_k)$. Moreover, it is not restrictive to assume $A = S_k$ (indeed, if $v(S_k) = v(S_\ell)$, then $v_k = v_\ell$). So $\pi_i(v_k) < \pi_j(v_k)$.

Let

$$\alpha = \pi_i(v) - \pi_j(v) \qquad \text{and} \qquad \beta = \frac{2\alpha + 1}{\pi_j(v_k) - \pi_i(v_k)},$$

then $\alpha \geqslant 0$ (by the assumed Condition (i)) and $\beta > 0$.

Define a new set function u by letting for $S \in 2^N$

$$u(S) = \begin{cases} v(S) & \text{if } v(S) \leqslant v(S_{k-1}), \\ v(S) + \beta & \text{otherwise.} \end{cases}$$

Then u is ordinally equivalent to v, and moreover for $\ell \neq k$

$$u(S_\ell) - u(S_{\ell-1}) = v(S_\ell) - v(S_{\ell-1}), \tag{18}$$

$$u(S_k) - u(S_{k-1}) = v(S_k) - v(S_{k-1}) + \beta > 0. \tag{19}$$

Then by Equations (17), (18), (19) and (16) successively

$$\begin{aligned} \pi_i(u) - \pi_j(u) &= \sum_{k=2}^{2^n} (u(S_k) - u(S_{k-1}))\big(\pi_i(v_k) - \pi_j(v_k)\big) \\ &= \sum_{k=2}^{2^n} (v(S_k) - v(S_{k-1}))\big(\pi_i(v_k) - \pi_j(v_k)\big) \\ &\qquad + \beta\,(\pi_i(v_k) - \pi_j(v_k)) \\ &= \pi_i(v) - \pi_j(v) - (2\alpha + 1) \\ &= \alpha - 2\alpha - 1 < 0, \end{aligned}$$

in contradiction with Condition (i). □

Remark 2. *Here is a straightforward variant of Theorem 1 for games (using terms and notations introduced in Remark 1). For a solution σ on N, the game w in $\mathcal{G}^{(N)}$ is two-player σ-solid for players i, j if and only for all coalitions A in 2^N:*

$$\pi_i^\sigma(w_A) \geqslant \pi_j^\sigma(w_A)$$

(the w_A's are the binary set functions attached to the set function w; note that the steadyness assumption is not needed anymore because $\pi^\sigma(\mathbf{1}) = \sigma(\mathbf{0}) = \mathbf{0}$ by linearity of σ—here the first $\mathbf{0}$ is the null game, the second $\mathbf{0}$ is the origin in $\mathbb{R}^N$).

We can easily turn Theorem 1 into a characterization of π-stability; in doing this, we do not need to require Condition (ii) for all players i and j (see after Definition 2).

It seems difficult to reformulate Condition (ii) in terms of v alone. To give more weight to the assertion, notice that any collection of distinct set functions v_A is linearly independent in the real vector space $\mathbb{R}^{2^N}$, therefore we can let the linear mapping π assign any score vectors to distinct v_A's.

For any players i and j, Theorem 1 particularized to any semivalue π delivers technically simpler conditions equivalent to two-player π-stability for i and j. However, the resulting conditions involve all images of v. To the contrary, Equation (8) shows that only the images by v of sets containing exactly one of i and j suffice to test two-player π-stability for i and j. It is exploided in the next section to provide better conditions for two-player stability.

6 Two-player Stability for a Semivalue

For two given players i and j, the next theorem characterizes the two-player stability for i and j for the Banzhaf scoring: after the numbers $v(S \cup \{i\})$, for $S \in 2^{N \setminus \{i,j\}}$, are sorted in nondecreasing order, and similarly the numbers $v(S \cup \{j\})$ are sorted in nondecreasing order, we have that the k-th $v(S \cup \{i\})$ is not larger than the k-th $v(S \cup \{j\})$. Theorem 3 provides the extension of the result to semivalues, with a more technical condition involving the parameters of the semivalue.

Theorem 2. *Let $v : 2^N \to \mathbb{R}$ be a set function, and i, j be players in N. Then the three following conditions are equivalent for the Banzhaf scoring π^{Ban}:*

(i) *$\pi_i^{Ban}(u) \leqslant \pi_j^{Ban}(u)$, for all set functions u ordinally equivalent to v;*

(ii) *let $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_{2^{n-2}}$ be the sequence of 2^{n-2} numbers $v(S \cup \{i\})$ written in nondecreasing order, where $S \in 2^{N \setminus \{i,j\}}$, and similarly let $y_1 \leqslant y_2 \leqslant \ldots \leqslant y_{2^{n-2}}$ be the sequence of 2^{n-2} numbers $v(S \cup \{j\})$ written in nondecreasing order; then*

$$x_k \leqslant y_k, \quad \text{for } k = 1, 2, \ldots, 2^{n-2};$$

(iii) *for each S in $2^{N \setminus \{i,j\}}$ there holds*

$$\begin{gathered} |\{U \in 2^{N \setminus \{i,j\}} : v(U \cup \{j\}) \leqslant v(S \cup \{j\})\}| \\ \leqslant \\ |\{T \in 2^{N \setminus \{i,j\}} : v(T \cup \{i\}) \leqslant v(S \cup \{j\})\}|. \end{gathered}$$

Proof. By Equation (8), $\pi_j^{\text{Ban}}(v) - \pi_i^{\text{Ban}}(v)$ is equal to

$$\frac{1}{2^{n-2}} \sum_{S \in 2^N : i,j \notin S} \big(v(S \cup \{j\}) - v(S \cup \{i\})\big) \tag{20}$$

$$= \frac{1}{2^{n-2}} \sum_{\ell=1}^{2^{n-2}} \big(y_\ell - x_\ell\big); \tag{21}$$

the last equality holds because its sides are each a rewriting of $\frac{1}{2^{n-2}} \left(\sum_{\ell=1}^{2^{n-2}} y_\ell - \sum_{\ell=1}^{2^{n-2}} x_\ell\right)$.

(ii) $\Rightarrow$ (i) Let $f : \mathbb{R} \to \mathbb{R}$ be a strictly increasing function such that $u(S) = f\big(v(S)\big)$ for all S in 2^N. The (nondecreasingly) sorted sequence of numbers $u(S \cup \{i\})$, for $S \in 2^{N\setminus\{i,j\}}$, coincides with the nondecreasing sequence of the $f(x_\ell)$'s. Similarly the sorted sequence of numbers $u(S \cup \{j\})$, for $S \in 2^{N\setminus\{i,j\}}$, coincides with the nondecreasing sequence of the $f(y_\ell)$'s. Then by replacing v with u in (21) we obtain

$$\pi_j^{\text{Ban}}(u) - \pi_i^{\text{Ban}}(u) \;=\; \frac{1}{2^{n-2}} \sum_{\ell=1}^{n} \big(f(y_\ell) - f(x_\ell)\big).$$

From (ii) and the nondecreasingness of f there follows $f(y_\ell) - f(x_\ell) \geqslant 0$, and thus (i).

(i) $\Rightarrow$ (ii). Proceeding by contraposition, assume (ii) does not hold. Take the smallest k such that $y_k < x_k$. Then

$$x_1 \leqslant x_2 \leqslant \ldots \leqslant x_{k-1} \leqslant y_{k-1} \leqslant y_k < x_k, \qquad (22)$$

$$y_1 \leqslant y_2 \leqslant \ldots \leqslant y_k. \qquad (23)$$

Now for any positive real number M define the strictly increasing function $f : \mathbb{R} \to \mathbb{R}$ by letting

$$f(t) \;=\; \begin{cases} t & \text{if } t \leqslant y_k, \\ t + M & \text{if } y_k < t. \end{cases}$$

Then

$$f(y_\ell) - f(x_\ell) \;=\; \begin{cases} y_\ell - x_\ell + M, & \text{if } x_\ell \leqslant y_k < y_\ell; \\ y_\ell - x_\ell - M, & \text{if } y_\ell \leqslant y_k < x_\ell; \\ y_\ell - x_\ell, & \text{in all other cases.} \end{cases}$$

In view of (22), the number of (index values) ℓ's such that $x_\ell \leqslant y_k$ equals $k-1$. Let m be the number of ℓ's such that $y_\ell \leqslant y_k$; then by (23) we have $m \geqslant k$ (note that $y_k = y_{k+1}$ could happen). By subtracting from each of $k-1$ and m the number of ℓ's such that both $x_\ell \leqslant y_k$ and $y_\ell \leqslant y_k$, we derive that the number of ℓ's such that $x_\ell \leqslant y_k < y_\ell$ minus the number of ℓ's such that $y_\ell \leqslant y_k < x_\ell$ equals $(k-1) - m$. Hence

$$\sum_{\ell=1}^{n} \big(f(y_\ell) - f(x_\ell)\big) = \left(\sum_{\ell=1}^{n} (y_\ell - x_\ell)\right) + ((k-1) - m)\, M,$$

where $(k-1) - m < 0$. With M large enough, both sides become strictly negative. This establishes the negation of (i), because for the set function $u = f \circ v$ we now derive $\pi_j^{\text{Ban}}(u) - \pi_i^{\text{Ban}}(u) < 0$ from Equation (21) rewritten for u in place of v.

(ii) $\iff$ (iii) Because the x_ℓ's are the sorted $v(S \cup \{i\})$'s, and the y_ℓ's are the sorted $v(S \cup \{j\})$'s, the equivalence of (ii) and (iii) is easily established. □

We now aim at generalizing Theorem 2 to semivalues. Let $\pi^{(q)}$ be a semivalue with parameter vector q, and set $q'(s) = q(s) + q(s+1)$. For any S_0 in $2^{N\setminus\{i,j\}}$, define successively (with $\sqsubseteq_v$ denoting as before the set ordering induced by v)

$$\mathcal{T}_{i,j}^{+} = \{T \in 2^{N\setminus\{i,j\}} : T \cup \{i\} \sqsubseteq_v S_0 \cup \{j\} \sqsubset_v T \cup \{j\}\},$$

$$\mathcal{T}_{i,j}^{-} = \{U \in 2^{N\setminus\{i,j\}} : U \cup \{j\} \sqsubseteq_v S_0 \cup \{j\} \sqsubset_v U \cup \{i\}\},$$

$$D_{i,j}^{\pi^{(q)}}(\sqsubseteq_v, S_0) = \sum_{T \in \mathcal{T}_{i,j}^{+}} q'(|T|) \;-\; \sum_{U \in \mathcal{T}_{i,j}^{-}} q'(|U|).$$

Note that the quantity $D_{i,j}^{\pi^{(q)}}(\sqsubseteq_u, S_0)$ takes the same value for all set functions u ordinally equivalent to v.

Theorem 3. *Let $\pi^{(q)}$ be a semivalue with parameter vector q, let i, j be two players, and let v be a set function for which*

$$\pi_i^{(q)}(v) \;\leqslant\; \pi_j^{(q)}(v).$$

Then the two following conditions are equivalent:

(i) *for all $u \in \mathbb{R}^{2^N}$ ordinally equivalent to v:*

$$\pi_i^{(q)}(u) \;\leqslant\; \pi_j^{(q)}(u);$$

(ii) $\forall S_0 \in 2^{N\setminus\{i,j\}} :\; 0 \leqslant D_{i,j}^{\pi^{(q)}}(\sqsubseteq_v, S_0)$.

Proof. The proof relies on two lemmas (requiring one and half more page) which we present in a supplementary document for sake of clarity. □

7 Conclusion

We define ordinal stability of a set function with respect to any given scoring rule (thus covering also games and so-called power index, solution or value). Stability captures coalitional situations that are more robust to possible fluctuations in the data, and for which the accuracy in determining the precise worths of coalitions is less crucial for the correct computation of the players' ranking. For any semivalue, we show that (i) a set function on a two-player set is stable (Proposition 1), and (ii) all bivalued set functions are stable (Proposition 2). Our main contributions are twofold. First we characterize stability for any given steady and linear scoring, by relying on a special decomposition of any set function. Theorem 1 covers in particular semivalues, with the Banzhaf and Shapley scorings as particular cases. Second, for semivalues, Theorem 3 provides an improved characterization of stability. It shows the operational advantage to build, for any two players i and j, a fixed set of coalitions from ordinal comparisons of coalitions (namely, $\mathcal{T}_{i,j}^{+}$ and $\mathcal{T}_{i,j}^{-}$ in Section 6). Then, to verify the stability of the relative rankings of the players i and j for any proposed semivalue π, it suffices to check nonnegativity of algebraic sums of semivalue parameters (Condition (ii) in Theorem 3).

In future research, it would be interesting to explore the ordinal stability of coalitional games for solutions which are not semivalue-like, for instance, the nucleolus [Schmeidler, 1969] or other power indices for simple games which are based on minimal winning coalitions [Deegan and Packel, 1978; Holler, 1982]. Another potential generalization of our ordinal framework integrates restricted possibilities of cooperation [Myerson, 1977] and calls for an investigation of their impact on the stability of the player ranking with respect to a power index.

Acknowledgements

We acknowledge a financial support from the project THEMIS ANR-20-CE23-0018 of the French National Research Agency (ANR).

References

[Algaba *et al.*, 2003] Encarnación Algaba, Jesús Mario Bilbao, Fernández Garcıa, and J.J. López. Computing power indices in weighted multiple majority games. *Mathematical Social Sciences*, 46(1):63–80, 2003.

[Amgoud *et al.*, 2017] Leila Amgoud, Jonathan Ben-Naim, and Srdjan Vesic. Measuring the intensity of attacks in argumentation graphs with Shapley value. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017*, pages 63–69, 2017.

[Banzhaf III, 1964] John F. Banzhaf III. Weighted voting doesn't work: A mathematical analysis. *Rutgers University Law Review*, 19:317–343, 1964.

[Bernardi *et al.*, 2019] Giulia Bernardi, Roberto Lucchetti, and Stefano Moretti. Ranking objects from a preference relation over their subsets. *Social Choice and Welfare*, 52(4):589–606, 2019.

[Carreras and Freixas, 2008] Francesc Carreras and Josep Freixas. On ordinal equivalence of power measures given by regular semivalues. *Mathematical Social Sciences*, 55(2):221–234, 2008.

[Deegan and Packel, 1978] John Deegan and Edward W. Packel. A new index of power for simple n-person games. *International Journal of Game Theory*, 7(2):113–123, 1978.

[Dubey *et al.*, 1981] Pradeep Dubey, Abraham Neyman, and Robert James Weber. Value theory without efficiency. *Mathematics of Operations Research*, 6(1):122–128, 1981.

[Grabisch and Roubens, 2000] Michel Grabisch and Marc Roubens. Application of the Choquet integral in multicriteria decision making. *Fuzzy Measures and Integrals-Theory and Applications*, pages 348–374, 2000.

[Haret *et al.*, 2018] Adrian Haret, Hossein Khani, Stefano Moretti, and Meltem Öztürk. Ceteris paribus majority for social ranking. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018*, pages 303–309, 2018.

[Holler and Nurmi, 2013] Manfred J. Holler and Hannu Nurmi. *Power, Voting, and Voting Power: 30 years after*. Springer, 2013.

[Holler, 1982] Manfred J. Holler. Forming coalitions and measuring voting power. *Political Studies*, 30(2):262–271, 1982.

[Hunter and Konieczny, 2006] Anthony Hunter and Sébastien Konieczny. Shapley inconsistency values. In *Proceedings of Knowledge Representation and Reasoning, KR 2006*, volume 6, pages 249–259. 2006.

[Lambo and Moulen, 2002] Lawrence Diffo Lambo and Joël Moulen. Ordinal equivalence of power notions in voting games. *Theory and Decision*, 53(4):313–325, 2002.

[Lucchetti *et al.*, 2015] Roberto Lucchetti, Stefano Moretti, and Fioravante Patrone. Ranking sets of interacting objects via semivalues. *TOP*, 23(2):567–590, 2015.

[Lundberg and Lee, 2017] Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *31st International Conference on Neural Information Processing Systems*, pages 4768–4777, 2017.

[Michalak *et al.*, 2013] Tomasz P. Michalak, Karthik V. Aadithya, Piotr L. Szczepanski, Balaraman Ravindran, and Nicholas R. Jennings. Efficient computation of the Shapley value for game-theoretic network centrality. *Journal of Artificial Intelligence Research*, 46:607–650, 2013.

[Myerson, 1977] Roger Myerson. Graphs and cooperation in games. *Mathematics of Operations Research*, 2(3):225–229, 1977.

[Roth, 1985] Alvin E. Roth. The college admissions problem is not equivalent to the marriage problem. *Journal of Economic Theory*, 36(2):277 – 288, 1985.

[Schmeidler, 1969] David Schmeidler. The nucleolus of a characteristic function game. *SIAM Journal on Applied Mathematics*, 17(6):1163–1170, 1969.

[Shapley, 1953] Lloyd S. Shapley. A value for n-person games. *Contributions to the Theory of Games*, 2(28):307–317, 1953.