An axiomatisation of the Banzhaf value and interaction index for multichoice games

Mustapha Ridaoui, Michel Grabisch, Christophe Labreuche

An axiomatisation of the Banzhaf value and interaction index for multichoice games

Mustapha Ridaoui[1], Michel Grabisch[1], Christophe Labreuche[2]

[1] Paris School of Economics, Université Paris I - Panthéon-Sorbonne, Paris, France
{mustapha.ridaoui,michel.grabisch}@univ-paris1.fr
[2] Thales Research & Technology, Palaiseau, France
christophe.labreuche@thalesgroup.com

Abstract. We provide an axiomatisation of the Banzhaf value (or power index) and the Banzhaf interaction index for multichoice games, which are a generalisation of cooperative games with several levels of participation. Multichoice games can model any aggregation model in multicriteria decision making, provided the attributes take a finite number of values. Our axiomatisation uses standard axioms of the Banzhaf value for classical games (linearity, null axiom, symmetry), an invariance axiom specific to the multichoice context, and a generalisation of the 2-efficiency axiom, characteristic of the Banzhaf value.

Keywords: Banzhaf value, multicriteria decision aid, multichoice games, interaction

1 Introduction

In cooperative game theory, a central problem is to define a *value*, that is, a payoff to be given to each player, taking into account his contribution into the game. Among the many values proposed in the literature, two of them have deserved a lot of attention, namely the Shapley value [24] and the Banzhaf value [1]. Both of them satisfy basic properties as linearity, symmetry, which means that the payoff given does not depend on the way the players are numbered, and the null player property, saying that a player who does not bring any contribution in coalitions he joins should receive a zero payoff. A value satisying these three properties has necessarily the form of a weighted average of the marginal contribution of a given player into coalitions. The Shapley and Banzhaf values differ on the weights used when computing the average. In the Shapley value, the marginal contributions are weighted according to the size of the coalition, in order to satisfy efficiency, that is, the total payoff given to the players is equal to the total worth of the game. In other words, the "cake" is divided among the players with no waste. For the Banzhaf value, the weights are simply equal, and so do not depend on the size of the coalition. As a consequence, the Banzhaf value is not efficient in general.

Lack of efficiency could be perceived, in the context of cooperative game theory, as an undesirable feature. This explains why in this domain, the Shapley value is much more popular. However, there are contexts where efficiency is not a relevant issue or even does not make sense. This is the case for voting games and in multicriteria decision aid (MCDA). A voting game is a cooperative game which is 0-1-valued, the value 1 indicating that the coalition wins the election. In this context, the relevant notion is the *power index*, and the Banzhaf value is used as such. A power index indicates how central a player is for making a coalition winning (this is called a swing). Banzhaf [1, 5] has shown that for counting swings, no weight should be applied, and this directly leads to the Banzhaf value (called in this context Banzhaf power index or Banzhaf index). In MCDA, criteria can be interpreted as voters in a voting game, and here a power index becomes an *importance index*, quantifying how important in the final decision a criterion is. In both domains, efficiency simply does not make sense, so that the Banzhaf value/index should be considered perhaps more relevant than the Shapley value.

There are other reasons to consider the Banzhaf value as a natural concept. In order to establish this, we need to generalize the notion of value or power index to the notion of interaction index, especially meaningful in a MCDA context [7, 12, 13, 20]. The interaction index for a set S of criteria quantifies the way the criteria in S interact, that is, how the scores on criteria in S contribute to the overall score. It can be considered that the interaction index when S is a singleton amounts to the importance index, which leads to two types of interaction indices, one based on the Shapley value and the other based on the Banzhaf index. This being said, aggregation models in MCDA which are based on capacities (monotone cooperative games) can be of the Choquet integral type, multilinear type, or other integrals like Pan-integral, concave integral, decomposition integral, etc. (beside other types such as the Sugeno integral, suitable in an ordinal context). It has been proved by Grabisch et al. [11] that if the Choquet integral is used, the relevant interaction index is the Shapley interaction index, while in the case of the multilinear model, the Banzhaf index should be used. In addition, in computer sciences, the notion of Fourier Transform is defined and widely used, e.g., in cryptography (see, e.g., [4]). It turns out that the Banzhaf interaction index and the Fourier transform differ only by some coefficient (see details in [8, Ch. 2.16.2]. Other connections exist, e.g., with the Sobol indices in statistics (see [10]).

The aim of the paper is to establish the Banzhaf index and Banzhaf interaction index for multichoice games, which are a generalisation of cooperative games. Multichoice games allow each player to choose a certain level of participation, among k possible levels. Their counterpart in MCDA are very interesting since they encode any aggregation model with discrete attributes [22, 21]. To our knowledge, there is no definition of an interaction index for multichoice games. Nevertheless, Lange and Grabisch [17] have provided a general form of interaction index for games on lattices. This does not fit our analycis, that focuses on interaction index defined for groups of criteria. Our approach is to build these indices in an axiomatic way, using an approach similar to Weber [26].

2 Preliminary definitions

We consider throughout a finite set of elements $N = \{1, \ldots, n\}$, which could be players, agents in cooperative game theory, criteria, attributes in multi-criteria decision analysis, voters or political parties in voting theory. We often denote cardinality of sets $S, T, \ldots$ by corresponding small letters $s, t, \ldots$, otherwise by the standard notation

$|S|, |T|, \ldots$. Moreover, we will often omit braces for singletons, e.g., writing $N \setminus i$ instead of $N \setminus \{i\}$.

Let $L_i := \{0, 1, \ldots, k_i\}, (k_i \in \mathbb{N}, k_i \geq 1)$ and define $L = \times_{i \in N} L_i$. The set L is endowed with the usual partial order $\leq$: for any $x, y \in L$, $x \leq y$ if and only if $x_i \leq y_i$ for every $i \in N$. For each $x \in L$, we define the support of x by $\Sigma(x) = \{i \in N | x_i > 0\}$ and the kernel of x by $\mathrm{K}(x) = \{i \in N | x_i = k_i\}$. Their cardinalities are respectively denoted by $\sigma(x)$ and $\kappa(x)$. For any $x \in L$ and $S \subseteq N$, x_S denotes the restriction of x to the set S, while x_{-S} denotes the restriction of x to the set $N \setminus S$. For all alternatives $x, y \in L$ and $S \subseteq N$, the notation (x_S, y_{-S}) denotes the compound alternative z such that $z_i = x_i$ if $i \in S$ and y_i otherwise. The same meaning is intended for L_S and L_{-S}.

In cooperative game theory, the set L_i is interpreted as the set of activity levels of player $i \in N$, and any $x \in L$ is called an *activity profile*. In an MCDA context, L_i is the set of all possible values taken by (discrete) attribute $i \in N$, while $x \in L$ is called an *alternative*. Throughout the paper, we adopt without limitation the terminology of game theory.

For convenience, we assume that all players have the same number of levels, i.e., $k_i = k$ for every $i \in N$, $(k \in \mathbb{N})$.

A *(cooperative) game* on N is a set function $v : 2^N \to \mathbb{R}$ vanishing on the empty set. A game v is said to be a *capacity* [2] or *fuzzy measure* [25] if it satisfies the monotonicity condition: $v(A) \leq v(B)$ for every $A \subseteq B \subseteq N$.

Cooperative games can be seen as pseudo-Boolean functions vanishing at 0_N. A *pseudo-Boolean function* [3, 14] is any function $f : \{0,1\}^N \to \mathbb{R}$. Noting that any subset S of N can be encoded by its characteristic function 1_S, where $1_S = (x_1, \ldots, x_n)$, with $x_i = 1$ if $i \in S$ and $x_i = 0$ otherwise, there is a one-to-one correspondence between set functions and pseudo-Boolean functions: $f(1_S) = v(S)$ for every $S \subseteq N$. Therefore, a natural generalisation of games is multichoice games. A *multichoice game* [15] on N is a function $v : L \to \mathbb{R}$ such that $v(0, \ldots, 0) = 0$. A multichoice game v is monotone if $v(x) \leq v(y)$ whenever $x \leq y, \forall x, y \in L$. A monotonic multichoice game is called a *k-ary capacity* [9]. In a MCDA context, $v(x)$ is the overall score of alternative x. For any $x \in L$, $x \neq 0_N$, the *Dirac game* δ_x is defined by $\delta_x(y) = 1$ iff $y = x$, and 0 otherwise. We denote by $\mathcal{G}(L)$ the set of all multichoice games defined on L.

The *derivative* of $v \in \mathcal{G}(L)$ w.r.t. $T \subseteq N$ at $x \in L$ such that for any $i \in T, x_i < k_i$ is given by: $\Delta_T v(x) = \sum_{S \subseteq T} (-1)^{t-s} v(x + 1_S)$.

3 Banzhaf value and interaction indices

In this section we recall the concepts of value and interaction indices introduced in cooperative game theory. The notion of power index or value is one of the most important concepts in cooperative game theory. A *value* [24] on N is a function $\phi : \mathcal{G}(2^N) \to \mathbb{R}^N$ which assigns to each player $i \in N$ in a game $v \in \mathcal{G}(2^N)$ a payoff $\phi_i(v)$, which is most often a share of $v(N)$, the total worth of the game. In the context where N is the set of voters, $\phi_i(v)$ can be interpreted as the voting power of player $i \in N$ in game $v \in \mathcal{G}(2^N)$, i.e., to what extent the fact that i votes 'yes' makes the final decision to be ‘yes‘. In such a case, ϕ is called a power index. Obviously, power indices in voting theory are close to importance indices in MCDA. In cooperative game theory, diverse kinds of values/power indices have been proposed, among which a large part have the following form: $\phi_i(v) = \sum_{S\subseteq N\setminus i} p^i_S\big(v(S\cup i) - v(S)\big), p^i_S \in \mathbb{R}$. If the family of real constants $\{p^i_S, S \subseteq N \setminus i\}$ forms a probability distribution, the value ϕ_i is said to be a *probabilistic value* [26].

The exact form of a value/power index depends on the axioms that are imposed on it. The two best known are due to Shapley [24] and Banzhaf [1]. The Banzhaf value [5] of a player $i \in N$ in a game $v \in \mathcal{G}(2^N)$ is defined by

$$\phi^B_i(v) = \sum_{S\subseteq N\setminus i} \frac{1}{2^{n-1}}\big(v(S\cup i) - v(S)\big).$$

It is uniquely axiomatized by a set of four axioms [5, 18]: linearity axiom, dummy axiom, symmetry axiom and 2-efficiency axiom. They will be recalled below.

Another interesting concept is that of interaction among criteria. An *interaction index* on N of the game $v \in \mathcal{G}(2^N)$ is a function $I^v : 2^N \to \mathbb{R}$ that represents the amount of interaction (it can be positive or negative) among any subset of players. Grabisch and Roubens [12] proposed an axiomatic characterisation of the Shapley and the Banzhaf interaction indices. For this, they introduce the following definitions:

Let v be a game on N, and T a nonempty subset of N. The restriction of v to T is a game of $\mathcal{G}(2^T)$ defined by $v^T(S) = v(S), \forall S \subseteq T$. The restriction of v to T in the presence of a set $A \subseteq N \setminus T$ is a game $\mathcal{G}(2^T)$ defined by $v^T_{\cup A}(S) = v(S \cup A) - v(A)$ for every $S \subseteq T$. The reduced game with respect to T is a game denoted $v_{[T]}$ defined on the set $(N \setminus T) \cup [T]$ where $[T]$ indicates a single hypothetical player, which is the union (or representative) of the players in T. It is defined as follows for any $S \subseteq N\setminus T$:

$$v_{[T]}(S) = v(S),$$
$$v_{[T]}(S \cup [T]) = v(S \cup T).$$

The following axioms have been considered by Grabisch and Roubens [12] :

– Linearity axiom (L): $I^v(S)$ is linear on $\mathcal{G}(2^N)$ for every $S \subseteq N$.

  $i \in N$ is said to be *dummy* for $v \in \mathcal{G}(2^N)$ if $\forall S \subseteq N \setminus i$, $v(S \cup i) = v(S) + v(i)$.
– Dummy player axiom (D): If $i \in N$ is a dummy player for $v \in \mathcal{G}(2^N)$, then
  1. $I^v(i) = v(i)$,
  2. for every $S \subseteq N \setminus i$, $S \neq \varnothing$, $I^v(S \cup i) = 0$.
– Symmetry axiom (S): for all $v \in \mathcal{G}(2^N)$, for all permutation π on on N,

$$I^v(S) = I^{\pi v}(\pi S).$$

– 2-efficiency axiom (2-E): For any $v \in \mathcal{G}(2^N)$,

$$I^v(i) + I^v(j) = I^{v_{[ij]}}([ij]), \forall i, j \in N.$$

– Recursive axiom (R): For any $v \in \mathcal{G}(2^N)$,

$$I^v(S) = I_{\cup j}^{v^{N\setminus j}}(S \setminus j) - I^{v^{N\setminus j}}(S \setminus j), \forall S \subseteq N, s \geq 2, \forall j \in S.$$

Theorem 1. *(Grabisch and Roubens [12]) Under (L), (D), (S), (2-E) and (R),*

$$\forall \in \mathcal{G}(2^N), I^v(S) = \sum_{T \subseteq N \setminus S} \frac{1}{2^{n-t}} \sum_{L \subseteq S} (-1)^{s-l} v(T \cup L), \forall S \subseteq N, S \neq \varnothing.$$

In particular, for a pair $S = \{i, j\}$, we obtain $I^v(\{i,j\}) = \sum_{T \subseteq N \setminus \{i,j\}} \frac{1}{2^{n-t}} \delta_{i,j} v(S)$, where $\delta_{i,j} v(S) := v(S \cup \{i,j\}) - v(S \cup \{i\}) - v(S \cup \{j\}) + v(S)$. Moulin interprets the quantity $v(\{i,j\}) - v(\{i\}) - v(\{j\})$ as the cost/surplus of mutual externalities of players i and j [19]. More generally, $\delta_{i,j} v(S)$ can be seen as the cost/surplus of mutual externalities of players i and j, in the presence of coalition S. The interaction index $I_v(\{i,j\})$ is thus the *expected cost/surplus of mutual externalities* of players i and j.

In MCDA, recall that $v(S)$ is the overall score of an option that is perfectly satisfactory (with score 1) on criteria S and completely unacceptable (with score 0) on the remaining criteria. The interaction index $I_v(\{i,j\})$ can also be interpreted as the variation of the mean weight of criterion i when criterion j switches from the least satisfied criterion to the most satisfied criterion [16]. Positive interaction depicts situations where there is *complementarity* among criteria i and j: criteria i and j deserve to be well-satisfied together (the more criterion i is satisfied, the more it is important to satisfy as well criterion j). On the opposite side, negative interaction occurs when there is *substitutability* among criteria i and j: it is not rewarding to improve both criteria i and j together.

Now we present an axiomatization of Fujimoto et al. [6] based on the concept of partnership coalition. For this, they introduce the following axiom:

Reduced-partnership-consistency axiom (RPC): If P is a partnership in a game v then $I^v(P) = I^{v_{[P]}}([P])$.

A coalition $P \subseteq N, P \neq \varnothing$, is said to be a partnership in a game $v \in \mathcal{G}(2^N)$ if, for all $S \subsetneq P, v(S \cup T) = v(T)$, for all $T \subseteq N \setminus P$.

Theorem 2. *(Fujimoto et al. [6]) Under the linear axiom, the dummy axiom, the symmetry axiom, the 2-efficiency axiom and the reduced-partnership-consistency axiom,*

$$\forall v \in \mathcal{G}(2^N), I^v(S) = \sum_{T \subseteq N \setminus S} \frac{1}{2^{n-t}} \sum_{L \subseteq S} (-1)^{s-l} v(T \cup L), \forall S \subseteq N, S \neq \varnothing.$$

4 Axiomatisation of the Banzhaf value for multichoice games

In this section, we give a characterisation of Banzhaf value for multichoice games, in the spirit of what was done by Weber [26] for cooperative games. Ridaoui et al. [22] have already generalized and axiomatized the Shapley value for multichoice games. The axiomatisation given in [22] is based on five axioms, linearity, nullity, symmetry, invariance and efficiency. We present the first four axioms used in [22], as some of them will be used in our characterisation. It is worth mentioning that the use of such axioms is common in axiomatisation of values. Let ϕ be a value defined for any $v \in \mathcal{G}(L)$.

Linearity axiom (L) : ϕ is linear on $\mathcal{G}(L)$, i.e., $\forall v, w \in \mathcal{G}(L), \forall \alpha \in \mathbb{R}$,

$$\phi_i(v + \alpha w) = \phi_i(v) + \alpha \phi_i(w), \forall i \in N.$$

A player $i \in N$ is said to be *null* for $v \in \mathcal{G}(L)$ if $v(x + 1_i) = v(x), \forall x \in L, x_i < k$.

Null axiom (N): If a player i is null for $v \in \mathcal{G}(L)$, then $\phi_i(v) = 0$.

Let π be a permutation on N. For all $x \in L$, we denote $\pi(x)_{\pi(i)} = x_i$. For all $v \in \mathcal{G}(L)$, the game $\pi \circ v$ is defined by $\pi \circ v(\pi(x)) = v(x)$.

Symmetry axiom (S): For any permutation π of N,

$$\phi_{\pi(i)}(\pi \circ v) = \phi_i(v), \forall i \in N.$$

Invariance axiom (I): Let us consider two games $v, w \in \mathcal{G}(L)$ such that, for some $i \in N$,

$$v(x + 1_i) - v(x) = w(x) - w(x - 1_i), \forall x \in L, x_i \notin \{0, k\}$$
$$v(x_{-i}, 1_i) - v(x_{-i}, 0_i) = w(x_{-i}, k_i) - w(x_{-i}, k_i - 1), \forall x_{-i} \in L_{-i},$$

then $\phi_i(v) = \phi_i(w)$.

The linearity axiom means that if several multichoice games are combined linearly, the value of the resulting multichoice game is a linear combination of the values of each individual multichoice game. Axiom (N) states that a player having no influence on a multichoice game is not important. Axiom (S) says that the numbering of the players plays no role in the computation of value. Axiom (I) indicates that the computation of the value does not depend on the position on the grid. More precisely, if the game w is simply a shift of v of one unit on the grid, then v and w shall have the same value (importance).

Ridaoui et al. [22] have shown the following result.

Theorem 3. *Let ϕ be a value defined for any $v \in \mathcal{G}(L)$. If ϕ fulfils (L), (N), (I) and (S) then there exists a family of real constants $\{b_{n(x_{-i})}, x_{-i} \in L_{-i}\}$ such that*

$$\phi_i(v) = \sum_{x_{-i} \in L_{-i}} b_{n(x_{-i})}\big(v(x_{-i}, k_i) - v(x_{-i}, 0_i)\big), \forall i \in N, \tag{1}$$

where $n(x_{-i}) = (n_0, \ldots, n_k)$ with n_j the number of components of x_{-i} being equal to $j \in \{0, 1, \ldots, k\}$.

We introduce two additional axioms, and first some notation. For $i, j \in N$, and $v \in \mathcal{G}(L)$, denote by $v^{[ij]}$ the multichoice game defined on the set $(N^{[ij]} = N \setminus \{i, j\} \cup [ij])$, where $[ij]$ indicates a single player, which is the merge of the distinct players i and j. The multichoice game $v^{[ij]}$ is defined as follows,

$$\forall y \in \{0, 1, \ldots, k\}^{N^{[ij]}}, v^{[ij]}(y) = v(y_{-ij}, \ell_{ij}) \quad \text{if } y_{[ij]} = \ell, \ell \in \{0, 1, \ldots, k\}.$$

2-Restricted efficiency (2-RE): For all $x \in L \setminus 0_N$,

$$\phi_i(\delta_x) + \phi_j(\delta_x) = \phi_{[ij]}(\delta_x^{[ij]}),$$

where, $\forall y \in \{0, 1, \dots, k\}^{N^{[ij]}}$, with $y_{[ij]} = \ell, \ell \in \{0, 1, \dots, k\}$,

$$\delta_x^{[ij]}(y) = \begin{cases} \delta_x(y_{-ij}, l_{ij}) & \text{if } x_i, x_j \in \{1, 2, \dots, k-1\}, \text{ or } \{x_i, x_j\} = \{0, k\}, \\ \delta_{(x_{-ij}, k_i, k_j)}(y_{-ij}, l_{ij}) & \text{else if } x_i \vee x_j = k, \\ \delta_{(x_{-ij}, 0_i, 0_j)}(y_{-ij}, l_{ij}) & \text{otherwise (i.e., if } x_i \wedge x_j = 0). \end{cases}$$

The original 2-Efficiency [18] says that the worth alloted to a coalition of two players when they form a partnership shall be divided into the worth alloted to its members. Here this axiom is considered only for the Dirac multichoice games. In the definition of $\delta_x^{[ij]}$, we need to change x by adding some symmetry between i and j in the last two cases. The 2-Restricted efficiency axiom means that for the Dirac multichoice game, the sum of the values of two players equals to the value of the merge of these players in the corresponding reduced game. The first situation of $\delta_x^{[ij]}(y)$ is standard and generalizes the classical case. The last two cases are limit cases. If only one of the elements x_i, x_j belong to $\{0, k\}$ but not the other one, then one shall take, for symmetry reasons, the same value for i and j. We need to take, for consistency reasons, the extreme value 0 or k that is reached by x_i or x_j.

For the classical Banzhaf value, the *dummy player axiom* (stronger than the null axiom) is used as a calibration property. When there is only one player left, the player shall get its worth $v(\{i\})$. We generalize this idea by the following calibration axiom restricted to Dirac games.

Calibration axiom (C): Let $i \in N$, with $n = 1$. $\phi_i(\delta_{k_N}) = 1$.

Theorem 4. *Under axioms* **(L)**, **(N)**, **(I)**, **(S)**, **(2-RE)** *and* **(C)**, *for all* $v \in \mathcal{G}(L)$

$$\phi_i(v) = \frac{1}{2^{n-1}} \sum_{x_{-i} \in L_{-i}} 2^{\sigma(x_{-i}) - \kappa(x_{-i})} \big(v(x_{-i}, k_i) - v(x_{-i}, 0_i)\big), \forall i \in N \qquad (2)$$

Proof : It is easy to check that the formula (2) satisfies the axioms.

Conversely, we consider ϕ satisfying the axioms (L), (N), (I), (S), (2-RE) and (C). Let $x \in L$, we write $x = (0_{N \setminus S \cup T}, x_S, k_T)$, with $x_S \in L_S \setminus \{0, k\}^S$, $S = \Sigma(x) \setminus \mathrm{K}(x)$, and $T = \mathrm{K}(x)$. From axioms (L), (N), (I) and (S) and Theorem 3, we have

$$\phi_i(\delta_x) = b_{n(x_{-i})}\big(\delta_x(x_{-i}, k_i) - \delta_x(x_{-i}, 0_i)\big),$$

then we obtain,

$$\phi_i(\delta_{(x_{-i}, k_i)}) = b_{n(x_{-i})} = -\phi_i(\delta_{(x_{-i}, 0_i)}), \qquad (3)$$

and

$$\phi_i(\delta_{(x_{-i}, x_i)}) = 0, \text{for } x_i \in L_i \setminus \{0, k\}. \qquad (4)$$

From (3) and (4), we have, for any $i \in T$

$$\phi_i(\delta_x) + \phi_j(\delta_x) = b_{(n-s-t, n(x_S), t-1)}, \forall j \in S, \qquad (5)$$

$$\phi_i(\delta_x) + \phi_j(\delta_x) = 2b_{(n-s-t,n(x_S),t-1)}, \forall j \in T, \tag{6}$$

and,

$$\phi_i(\delta_x) + \phi_j(\delta_x) = b_{(n-s-t,n(x_S),t-1)} - b_{(n-s-t-1,n(x_S),t)}, \forall j \in N \setminus S \cup T. \tag{7}$$

By axiom (2-RE) we have,

- from (7), $\forall s \in \{0, \ldots, n-1\}$, $\forall t \in \{1, \ldots, n\}$, with $s + t \leq n - 1$,

$$b_{(n-s-t-1,n(x_S),t)} = b_{(n-s-t,n(x_S),t-1)}, \tag{8}$$

- from (6), $\forall s \in \{0, \ldots, n-1\}$, $\forall t \in \{2, \ldots, n\}$, with $s + t \leq n$,

$$b_{(n-s-t,n(x_S),t-2)} = 2b_{(n-s-t,n(x_S),t-1)}, \tag{9}$$

- from (5), $\forall s \in \{1, \ldots, n-1\}$, $\forall t \in \{1, \ldots, n-1\}$, with $s + t \leq n$,

$$b_{(n-s-t,n(x_S),t-1)} = b_{(n-s-t,n(x_{S\setminus j}),t-1)}, j \in S, \tag{10}$$

and from (C) and (9), we have

$$b_{0,\ldots,n-1} = \frac{1}{2^{n-1}}, \forall i \in N. \tag{11}$$

We distinguish the two following cases:

1. If $S = \varnothing$,
   - from (11) and (8), we have

   $$b_{n-1,0,\ldots,0} = b_{n-2,0,\ldots,0,1} = \ldots = b_{1,0,\ldots,0,n-2} = b_{0,\ldots,0,n-1} = \frac{1}{2^{n-1}}, \tag{12}$$

   then, for every $\ell \in \{1, \ldots, n\}$,

   $$b_{n-\ell,0,\ldots,0,\ell-1} = \frac{1}{2^{n-1}}, \tag{13}$$

   - by (9) and (12), we have: $b_{n-2,0,\ldots,0} = \ldots = b_{0,\ldots,0,n-2} = \dfrac{1}{2^{n-2}}$,

   then, for every $\ell \in \{2, \ldots, n\}$, $b_{n-\ell,0,\ldots,0,\ell-2} = \dfrac{1}{2^{n-1}}$,
2. If $S \neq \varnothing$, by (10) and (9), we have

$$\begin{aligned}
b_{(n-s-t,n(x_S),t-1)} &= 2b_{(n-1-s_1-t,n(x_{S_1}),t)}, S_1 = S \setminus j, j \in S \\
b_{(n-1-s_1-t,n(x_{S_1}),t)} &= 2b_{(n-2-s_2-t,n(x_{S_2}),t+1)}, S_2 = S_1 \setminus j, j \in S_1 \\
&\vdots \\
b_{(n-s-t,n(x_j),t-s)} &= 2b_{(n-s-t,0,\ldots,0,t+s-1)},
\end{aligned}$$

hence, by (13) we have, $\forall s \in \{1, \ldots, n-1\}$, $\forall t \in \{1, \ldots, n-1\}$, with $s + t \leq n$,

$$b_{(n-s-t,n(x_S),t-1)} = \frac{2^s}{2^{n-1}}.$$

The result is proved. ■

We finally show that our value $\phi_i(v)$ can be written as the sum of Banzhaf values over games derived from the multichoice game. This is related to some additivity property. More precisely, the power index $\phi_i(v)$ takes the form of the sum over $x \in \{0, \ldots, k-1\}^N$ of a classical Banzhaf value over the restriction of function v on $\times_{i \in N}\{x_i, x_i + 1\}$.

Proposition 1. *For every* $v \in \mathcal{G}(L), \phi_i(v) = \sum_{x \in \{0,\ldots,k-1\}^N} \phi_i^B(\mu_x^v), \forall i \in N,$ *with,* $\mu_x^v(S) = v(x + 1_S) - v(x), \forall S \subseteq N, \forall x \in L,$ *such that* $x_i < k, \forall i \in N$.

Proof : Let $v \in \mathcal{G}(L)$ and for any $x \in L$, such that $x_i < k, \forall i \in N$, we define the game μ_x^v for every $S \subseteq N$ by $\mu_x^v(S) = v(x + 1_S) - v(x)$. We have

$$\begin{aligned}
\phi_i(v) &= \frac{1}{2^{n-1}} \sum_{x_{-i} \in L_{-i}} 2^{\sigma(x_{-i}) - \kappa(x_{-i})} \big(v(x_{-i}, k_i) - v(x_{-i}, 0_i)\big) \\
&= \sum_{\substack{y_{-i} \in L_{-i} \\ \forall j \in N \setminus i, y_j < k}} \frac{1}{2^{n-1}} \sum_{y_{-i} \le x_{-i} \le (y+1)_{-i}} \big(v(x_{-i}, k_i) - v(x_{-i}, 0_i)\big) \\
&= \sum_{\substack{y \in L \\ \forall j \in N, y_j < k}} \frac{1}{2^{n-1}} \sum_{x_{-i} \in \{0,1\}^{N \setminus i}} \big(v(x_{-i} + y_{-i}, y_i + 1) - v(x_{-i} + y_{-i}, y_i)\big) \\
&= \sum_{\substack{y \in L \\ \forall j \in N, y_j < k}} \frac{1}{2^{n-1}} \sum_{A \subseteq N \setminus i} \big(\mu_y^v(A \cup i) - \mu_y^v(A)\big).
\end{aligned}$$

■

5 Axiomatisation of the Banzhaf interaction index

An interaction index of a multichoice game v is a function $I^v : 2^N \to \mathbb{R}$. The interaction of a single player i is the value related to player i. In this section, we present an axiomatisation of the interaction index based on the Banzhaf value. To this aim, we use the following generalised axioms introduced in [23] :

Linearity axiom (L) : I^v is linear on $\mathcal{G}(L)$, i.e., $\forall v, w \in \mathcal{G}(L), \forall \alpha \in \mathbb{R}$,

$$I^{v+\alpha w} = I^v + \alpha I^w.$$

Null axiom (N): If a player i is null for $v \in \mathcal{G}(L)$, then for all $T \subseteq N$ such that $T \ni i$, $I^v(T) = 0$.

Invariance axiom (I): Let us consider two functions $v, w \in \mathcal{G}(L)$ such that, for all $i \in N$,

$$v(x + 1_i) - v(x) = w(x) - w(x - 1_i), \forall x \in L, x_i \notin \{0, k\}$$

$$v(x_{-i},1_i)-v(x_{-i},0_i)=w(x_{-i},k_i)-w(x_{-i},k_i-1),\forall x_{-i}\in L_{-i},$$

then $I^v(T\cup i)=I^w(T\cup i),\forall T\subseteq N\setminus i$.

Symmetry axiom (S): For all $v\in\mathcal{G}(L)$, for all permutation π on N,

$$I^{\pi\circ v}(\pi(T))=I^v(T),\forall T\subseteq N,T\neq\varnothing.$$

Let v be a multichoice game in $\mathcal{G}(L)$ and $S\subseteq N$. The restriction of v to $N\setminus S$, denoted by v^{-S}, is defined by $v^{-S}(x_{-S})=v(x_{-S},0_S),\forall x_{-S}\in L_{-S}$. The restriction of v on $N\setminus i$ in the presence of i denoted by v_i^{-i} is the multichoice game on L_{-i} defined by $v_i^{-i}(x_{-i})=v(x_{-i},k_i)-v(0_{-i},k_i),\forall x_{-i}\in L_{-i}$.

Recursivity axiom (R): For any $v\in\mathcal{G}(L)$,

$$I^v(T)=I^{v_i^{-i}}(T\setminus i)-I^{v^{-i}}(T\setminus i),\forall T\subseteq N,T\neq\varnothing,\forall i\in T.$$

The Recursivity axiom is the exact counterpart of the one for classical games in [12].

Ridaoui et al. [23] proved the following Lemma.

Lemma 1. *Under axioms* **(L)**, **(N)**, **(I)** **(S)** *and* **(R)**, *for any* $v\in\mathcal{G}(L)$, $\forall T\subseteq N,T\neq\varnothing$,

$$I^v(T)=\sum_{\substack{A\subseteq T\\A\neq\emptyset}}(-1)^{t-a}I^{v_{[A]}^{(-T)\cup[A]}}([A]),\tag{14}$$

where $v_{[A]}^{(-T)\cup[A]}$ *is the restriction of* v *to* T *with respect to* $A\subseteq T$ *defined on the set* $\{0,\dots,k\}^{(N\setminus T)\cup[A]}$ *as follows:* $v_{[A]}^{(-T)\cup[A]}(x_{-T},\ell_{[A]})=v(x_{-T},\ell_A,0_{T\setminus A})$.

Our main result shows that there is a unique index fulfilling the previous axioms.

Theorem 5. *Under axioms* **(L)**, **(N)**, **(I)**, **(S)**, **(C)**, **(2-E)** *and* **(R)**, *for all* $v\in\mathcal{G}(L)$

$$I^v(T)=\sum_{x_{-T}\in L_{-T}}\frac{2^{\sigma(x_{-T})-\kappa(x_{-T})}}{2^{n-t}}\sum_{A\subseteq T}(-1)^{t-a}v(0_{T\setminus A},k_A,x_{-T}),\forall T\subseteq N,T\neq\varnothing.$$

Proof : Let $v\in\mathcal{G}(L)$, and $T\subseteq N,T\neq\varnothing$. By axioms **(L)**, **(N)**, **(I)**, **(S)**, **(C)** and **(2-E)**, we have $I^{v_{[A]}^{(-T)\cup[A]}}([A])=\sum_{x_{-T}\in L_{-T}}b_{n(x_{-T})}\big(v_{[A]}^{(-T)\cup[A]}(x_{-T},k_{[A]})-v_{[A]}^{(-T)\cup[A]}(x_{-T},0_{[A]})\big)$, with $b_{n(x_{-T})}=\dfrac{2^{\sigma(x_{-T})-\kappa(x_{-T})}}{2^{n-t}}$.

By Lemma (1), we have

$$
\begin{aligned}
I^v(T) &= \sum_{\substack{A\subseteq T\\ A\neq\emptyset}} (-1)^{t-a} I^{v_{[A]}^{(-T)\cup[A]}}([A]) \\
&= \sum_{\substack{A\subseteq T\\ A\neq\emptyset}} (-1)^{t-a} \sum_{x_{-T}\in L_{-T}} b_{n(x_{-T})}\big(v(x_{-T},k_A,0_{T\setminus A}) - v(x_{-T},0_T)\big) \\
&= \sum_{x_{-T}\in L_{-T}} b_{n(x_{-T})} \sum_{\substack{A\subseteq T\\ A\neq\emptyset}} (-1)^{t-a}\big(v(x_{-T},k_A,0_{T\setminus A}) - v(x_{-T},0_T)\big) \\
&= \sum_{x_T\in L_{-T}} b_{n(x_{-T})} \sum_{A\subseteq T} (-1)^{t-a} v(k_A,0_{T\setminus A},x_{-T}).
\end{aligned}
$$

■

As for the power index ϕ_i, the interaction index $I^v(T)$ can be written as the sum of Banzhaf interaction indices over games derived from the multichoice game.

Proposition 2. *Let* $v \in \mathcal{G}(L)$. $I^v(T) = \sum_{x\in\{0,\ldots,k-1\}^N} I_B^{\mu_x^v}(T), \forall T \subseteq N, T \neq \varnothing$, *with,* $\mu_x^v(S) = v(x+1_S) - v(x), \forall S \subseteq N, \forall x \in L$, *such that* $x_i < k_i, \forall i \in N$.

Proof :

$$
\begin{aligned}
I_B^{\mu_x^v}(T) &= \frac{1}{2^{n-1}} \sum_{x\in\{0,\ldots,k-1\}^N} \sum_{S\subseteq N\setminus T} \Delta_T \mu_x^v(S) \\
&= \frac{1}{2^{n-1}} \sum_{x\in\{0,\ldots,k-1\}^N} \sum_{S\subseteq N\setminus T} \Delta_T v(x+1_S) \\
&= \frac{1}{2^{n-1}} \sum_{S\subseteq T} (-1)^{t-s} \sum_{x_{-T}<k_{-T}} \big(v(0_{T\setminus A},k_A,x_{-T}) + v(0_{T\setminus A},k_A,x_{-T}+1_{-T})\big) \\
&= \frac{1}{2^{n-1}} \sum_{S\subseteq T} (-1)^{t-s} \sum_{x_{-T}\leq k_{-T}} 2^{\sigma(x_{-T})-\kappa(x_{-T})} v(0_{T\setminus A},k_A,x_{-T}).
\end{aligned}
$$

■

References

1. J. Banzhaf. Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review*, 19:317–343, 1965.
2. G. Choquet. Theory of capacities. *Annales de l'institut Fourier*, 5:131–295, 1953.
3. Y. Crama and P. Hammer. *Boolean functions. Number 142 in Encyclopedia of Mathematics and Its Applications*. Cambridge University Press, 2011.
4. R. de Wolf. A brief introduction to Fourier analysis on the Boolean cube. *Theory of Computing Library Graduate Surveys*, 1:1–20, 2008.

5. P. Dubey and L. S. Shapley. Mathematical properties of the Banzhaf power index. *Mathematics of Operations Research*, 4(2):99–131, 1979.
6. K. Fujimoto, I. Kojadinovic, and J.-L. Marichal. Axiomatic characterizations of probabilistic and cardinal-probabilistic interaction indices. *Games and Economic Behavior*, 55(1):72 – 99, 2006.
7. M. Grabisch. k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, 92(2):167–189, 1997.
8. M. Grabisch. *Set functions, games and capacities in decision making*. Springer Publishing Company, Incorporated, 2016.
9. M. Grabisch and C. Labreuche. Capacities on lattices and k-ary capacities. In *Int. Conf. Of the Euro Society for Fuzzy Logic and Technology (EUSFLAT)*, Zittau, Germany, September 10-12 2003.
10. M. Grabisch and C. Labreuche. A note on the Sobol' indices and interactive criteria. *Fuzzy Sets and Systems*, 315:99 – 108, 2017.
11. M. Grabisch, J.-L. Marichal, and M. Roubens. Equivalent representations of set functions. *Mathematics of Operations Research*, 25(2):157–178, 2000.
12. M. Grabisch and M. Roubens. An axiomatic approach to the concept of interaction among players in cooperative games. *International Journal of Game Theory*, 28(4):547–565, 1999.
13. M. Grabisch and M. Roubens. Application of the Choquet integral in multicriteria decision making. *Fuzzy Measures and Integrals-Theory and Applications*, pages 348–374, 2000.
14. P. Hammer and S. Rudeanu. *Boolean methods in operations research and related areas*, volume 7 of *Econometrics and Operations Research*. Springer, 1986.
15. C. R. Hsiao and T. E. S. Raghavan. Shapley value for multi-choice cooperative games, I. *Discussion paper of the University of Illinois at Chicago, Chicago*, 1990.
16. I. Kojadinovic. A weight-based approach to the measurement of the interaction among criteria in the framework of aggregation by the bipolar Choquet integral. *European journal of operational research*, 179:498–517, 2007.
17. F. Lange and M. Grabisch. The interaction transform for functions on lattices. *Discrete Mathematics*, 309(12):4037 – 4048, 2009.
18. E. Lehrer. An axiomatization of the Banzhaf value. *International Journal of Game Theory*, 17(2):89–99, 1988.
19. H. Moulin. *Fair division and collective welfare*. MIT Press, 2003.
20. T. Murofushi and S. Soneda. Techniques for reading fuzzy measures (iii): interaction index. In *9th fuzzy system symposium*, pages 693–696. Sapporo, Japan, 1993.
21. M. Ridaoui, M. Grabisch, and C. Labreuche. An alternative view of importance indices for multichoice games. In *Algorithmic Decision Theory*, pages 81–92, 2017.
22. M. Ridaoui, M. Grabisch, and C. Labreuche. Axiomatization of an importance index for generalized additive independence models. In *14th European Conference, ECSQARU 2017, Lugano, Switzerland, July 2017*, pages 340–350, 2017.
23. M. Ridaoui, M. Grabisch, and C. Labreuche. An interaction index for multichoice games. arXiv:1803.07541, 2018.
24. L. S. Shapley. A value for n-person games. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games, Vol. II*, number 28 in Annals of Mathematics Studies, pages 307–317. Princeton University Press, 1953.
25. M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.
26. R. J. Weber. Probabilistic values for games. In A. E. Roth, editor, *The Shapley Value: Essays in Honor of Lloyd S. Shapley*, pages 101–120. Cambridge University Press, 1988.