DATE DUE _____

New Mathematical Properties
of the Banzhaf Value

by

Irinel Dragan

TR #300

Department of Mathematics
University of Texas at Arlington
January, 1995

ABSTRACT

In a paper by P. Dubey and L.S. Shapley an axiomatic definition of the Banzhaf value has been extracted from an axiomatic definition of the Banzhaf power index (see [6]). Briefly speaking, the Banzhaf value axioms can be obtained from the Shapley value axioms by replacing efficiency by a similar axiom. This fact suggest that other recently discovered properties of the Shapley value can lead to similar properties for the Banzhaf value and this is the motivation for the present work.

In our paper, we introduce a potential function for the Banzhaf value and based upon some properties of this function similar to those obtained by S. Hart and A. Mas-Colell (see [8]), we find a potential basis for the space of T.U. games with the set of players N . This basis allows us to determine the null space of the Banzhaf value and more general, to solve the problem of giving an explicit expression for the set of games in the above mentioned space which have an apriori given Banzhaf value. A similar approach has been used in an earlier paper of the author for the weighted Shapley value (see [2]). Further, we introduce a concept of reduced game for the Banzhaf value and give an axiomatic characterization of this value based upon the reduced game property and the standard property, similar to the characterization obtained for the Shapley value by S. Hart and A. Mas-Colell (see [8]).

In the first section, we give the basic concepts as well as the major results obtained for the Shapley value in [8]. In the second section, we introduce the potential of the Banzhaf value and find the potential in terms of the coalitional form and in terms of the dividend form of the game

(Ths. 1 and 4). The power of a game, which is almost the average of all marginals for all players in the game, is the main tool in the proofs, because an explicit formula for the power has been obtained (Lemma 2). Then, we prove that the marginals of the potential function are the components of the Banzhaf value (Th.3). Further we give the major results obtained in [2] for the Shapley value, for which similar results on the Banzhaf value will be obtained in the next section. In the third section, we determine a basis of the space G^N (Th. 6), which is a potential basis for the Banzhaf value, i.e. the coordinates of any $v \in G^N$ relative to this basis are the potentials of v and its subgames (Th. 7). To be able to do this, we determine first the potentials of the basic vectors (Lemma 5). Then, we compute the Banzhaf values of the basic vectors (Th. 8) and determine the null space of the Banzhaf value and more general, the set of all games having an a priori given Banzhaf value (Th. 9). In the fourth section, we define a reduced game for the Banzhaf value (Definition 11), by using a functional equation which contains the power and uniquely determines the reduced game (Lemma 10). Further, we show that the Banzhaf value has the reduced game property relative to the reduced game just defined (Th. 12) and it is standard for two person games (Lemma 13). Finally, the axiomatic characterization by means of the reduced game property and standard property is proved (Th. 14).

Some examples help the reader in writing the potential basis for G^N with $|N|=3$ (Example 1), expressing explicitly the set of all three person games having an a priori given Banzhaf value (Example 2), and finding a reduced game by solving a functional equation for a three person game (Example 3).

1. Notations and Basic Concepts

A cooperative transferable utility game in coalitional form is a pair $G=(N,v)$, where N is a finite set with $|N|=n$, and v is a function defined on the power set of N with real values and $v(\emptyset)=0$. For each $S \subseteq N, S \neq \emptyset$, called a coalition, the number $v(S)$ represents the worth of coalition S . Denote by G^N the set of all games with the set of players N ; as in the following N will be fixed, we can use a notation like $v \in G^N$, which means that the game (N,v) is in G^N . Recall that G^N is a vector space with $\dim(G^N)=2^n-1$, if we define the addition of games by $(v_1+v_2)(S)=v_1(S)+v_2(S)$ and the scalar multiplication by $(av)(S)=av(S), \forall S \subseteq N$.

Any functional $\chi:G^N \rightarrow R^n$ is called a value; here $\chi_i(v), \forall i \in N$, is interpreted as the win of player i in game $v \in G^N$. In the classical game theory, a fundamental problem is that of splitting fairly the worth of the grand coalition $v(N)$ among the players, in case that N forms. Therefore, many values are supposed to satisfy among others the axiom $\sum_{i \in N} \chi_i(v) = v(N), \forall v \in G^N$, and in this case χ is called efficient. Among these values is the Shapley value, introduced by L.S. Shapley in [12], and the multiweighted Shapley values introduced in [3]. However, there are also non-efficient values. For any game $v \in G^N$ denote the sum of marginals for all players by $m(N,v)$; consider

$$(1.1) \quad \pi(N,v) = \frac{1}{2^n-1} \quad m(N,v) = \frac{1}{2^n-1} \sum_{i \in N} \sum_{S: i \in S \subseteq N} [v(S) - v(S - \{i\})].$$

We shall call $\pi(N, v)$ the power of the game (N, v) . The Banzhaf value of $v \in G^N$ is the value defined by

$$(1.2) \quad B_i(N, v) = \frac{1}{2^{n-1}} \sum_{S: i \in S \subseteq N} [v(S) - v(S - \{i\})], \forall i \in N,$$

(see [11]). Obviously, we have

$$(1.3) \quad \sum_{i \in N} B_i(N, v) = \pi(N, v);$$

for example, for a game $\bar{v}$ with $\bar{v}(N) = 1$ and $\bar{v}(S) = 0, \forall S \subset N$, we have $\pi(N, \bar{v}) = n/2^{n-1} \neq \bar{v}(N)$ if $n > 2$, hence the Banzhaf value is not efficient.

Note that (1.3) is a property similar to efficiency, here $v(N)$ has been replaced by $\pi(N, v)$ which does not depend on any solution. Note also that we can construct an efficient value by multiplying each component of the Banzhaf value by $[\pi(N, v)]^{-1} v(N)$, however the new value will be nonlinear and other properties will be lost.

A game $v \in G^N$ is monotonic, if $S \subseteq T$ implies $v(S) \leq v(T), \forall S \subseteq N, \forall T \subseteq N$; a game $v \in G^N$ is simple, if $v(S) \in \{0, 1\}, \forall S \subseteq N$. Note that in some publications (like [11]) the term simple is used for games having both properties. For monotonic simple games the marginal $v(S) - v(S - \{i\})$ is always equal to 0 or 1; precisely, the value 1 is obtained if coalition S is winning while $S - \{i\}$ is losing, and the value 0 occurs in all other cases. Therefore, such a difference is called a swing. The Banzhaf power index, from which the Banzhaf value emerged for general games, has been defined on the set of monotonic simple games by $m_i(N, v) = \sum_{S: i \in S \subseteq N} [v(S) - v(S - \{i\})], \forall i \in N$, i.e. the

sum of swings for each player (see [1]). In [6] an axiomatic definition of the Banzhaf power index has been proved; in a note at the end of the paper, the Banzhaf value is discussed, and it is stated that on G^N this value, like the Shapley value, enjoys the symmetry, dummy and linearity properties, but the efficiency axiom fails. In [7], which is a natural continuation of [5] and [6], the Shapley value and the Banzhaf value are discussed in parallel and it has essentially been shown that the efficiency axiom is the one making the difference between the two values. Precisely, for the Banzhaf value the efficiency axiom is replaced by $\sum_{i \in N} B_i(N, v) = \pi(N, v)$. This similarity and recent results in the theory of the Shapley value which have no similar results in the theory of the Banzhaf value is the motivation for our present work.

In [8], S. Hart and A. Mas-Colell introduced the potential of the Shapley value and gave a new axiomatic characterization for the Shapley value by means of the potential. For each game $v \in G^N$ the potential of the Shapley value is recursively defined on the power set of N by $P(\emptyset, v) = 0$ and

$$(1.4) \quad P(T, v) = \frac{1}{t} [v(T) + \sum_{j \in T} P(T - \{j\}, v)], \forall T \subseteq N,$$

where $t = |T|$. We mention the following major results to be of interest in the next section: (a) the potential of the Shapley value satisfies

$$(1.5) \quad P(T, v) = \sum_{S \subseteq T} \frac{(s-1)!(t-s)!}{t!} v(S), \forall T \subseteq N;$$

(b) If ϕ is the Shapley value of $v \in G^N$, then we have

$$(1.6) \quad P(N, v) - P(N - \{j\}, v) = \phi_j(N, v), \forall j \in N.$$

Note that in the sum (1.5) occur the numbers $\gamma_S(T) = \frac{(s-1)!(t-s)!}{t!}$, $\forall T \subseteq N$, $S \subseteq T$, and for the restriction of v to T we have

$$(1.7) \quad \phi_j(T, v) = \sum_{S: j \in S \subseteq T} \gamma_S(T) [v(S) - v(S - \{j\})], \forall j \in T.$$

In the next section, we shall introduce a potential of the Banzhaf value having properties similar to (a) and (b) and before going to an axiomatic characterization in the fourth section, we shall return at the end of the next section with other recent results for which the similar ones will be discussed in the third section.

2. The Potential of the Banzhaf Value.

For each game $v \in G^N$ the potential of the Banzhaf value is a function Q recursively defined on the power set of N by $Q(\emptyset, v) = 0$ and

$$(2.1) \quad Q(T, v) = \frac{1}{t} [\pi(T, v) + \sum_{j \in T} Q(T - \{j\}, v)], \forall T \subseteq N,$$

where $\pi(T, v)$ is the power of the subgame (T, v) . Here and in other similar situations we understand that in the subgame we consider the restriction of v to T .

Theorem 1: If Q is the potential of the Banzhaf value, then we have

$$(2.2) \quad Q(T, v) = \frac{1}{2^{t-1}} \sum_{S \subseteq T} v(S), \forall T \subseteq N,$$

where $t = |T|$.

Note the similarity of formulas (2.1) and (1.4), $v(T)$ has been replaced by $\pi(T, v)$; also, Theorem 1 is comparable to the above stated result (a). To prove Theorem 1 we need the following:

Lemma 2: For all $T \subseteq N$ we have

$$(2.3) \quad \pi(T, v) = \frac{1}{2^{t-1}} [tv(T) + \sum_{U \subset T} (2u - t)v(U)],$$

where $\pi(T, v)$ is the power of the subgame (T, v) .

Proof: We get

$$(2.4) \quad \begin{aligned} \pi(T, v) &= \sum_{i \in T} B_i(T, v) = \frac{1}{2^{t-1}} \sum_{U \subseteq T} [uv(U) - (t-u)v(U)] = \\ &= \frac{1}{2^{t-1}} [tv(T) + \sum_{U \subset T} (2u - t)v(U)], \end{aligned}$$

where the first equality sign is given by the definition of $\pi(T, v)$, the second equality sign follows from the fact that for each coalition $U \subseteq T$ the term $v(U)$ occurs in u components $B_i(T, v)$ with a positive sign and in $t-u$ components with a negative sign (if $u < t$), and the third equality sign is obtained by separating the term with $v(T)$.

Note that we shall use Lemma 2 also in the fourth section because $\pi(T, v)$ enters in the definition of the reduced game.

Proof of Theorem 1: We use an induction over the size of coalition T . If $|T| = 1$, i.e. $T = \{i\}$, then from (2.1) we get $Q(\{i\}, v) = \pi(\{i\}, v) = v(\{i\})$ which proves (2.2) in this case. Assume that (2.2) holds for all coalitions of size at most $t-1$; then, we get

$$(2.5) \quad \sum_{j \in T} Q(T - \{j\}, v) = \frac{1}{2^{t-2}} \sum_{U \subset T} (t-u)v(U),$$

because a coalition $U \subset T$ is a subset of $t-u$ coalitions $T - \{j\}$. The last formula and Lemma 2 give:

$$(2.6) \quad \pi(T, v) + \sum_{j \in T} Q(T - \{j\}, v) = \frac{t}{2^{t-1}} v(T) + \frac{1}{2^{t-1}} \sum_{U \subset T} (2u - t) v(U) + \\ + \frac{1}{2^{t-2}} \sum_{U \subset T} (t - u) v(U) = \frac{t}{2^{t-1}} v(T) + \frac{t}{2^{t-1}} \sum_{U \subset T} v(U)$$

and from here and (2.1) we obtain (2.2) for T . Hence, (2.2) holds in general.

Theorem 3: If Q is the potential of the Banzhaf value, then we have:

$$(2.7) \quad Q(N, v) - Q(N - \{j\}, v) = B_j(N, v), \forall j \in N,$$

where B_j are the components of the Banzhaf value.

Proof: We compute the difference by using (2.2); for each $j \in N$ we have

$$(2.8) \quad Q(N, v) - Q(N - \{j\}, v) = \frac{1}{2^{n-1}} \sum_{S \subseteq N} v(S) - \frac{1}{2^{n-2}} \sum_{S \subseteq N - \{j\}} v(S) = \\ = \frac{1}{2^{n-1}} \left[\sum_{S \subseteq N - \{j\}} v(S) + \sum_{S: j \in S \subseteq N} v(S) \right] - \frac{1}{2^{n-2}} \sum_{S \subseteq N - \{j\}} v(S) = \\ = \frac{1}{2^{n-1}} \sum_{S: j \in S \subseteq N} [v(S) - v(S - \{j\})] = B_j(N, v).$$

Note that in the sum of Theorem 1 occur the numbers $\gamma_S(T) = \frac{1}{2^{t-1}}, \forall T \subseteq N, S \subseteq T$,

and for the restriction of v to T we have

$$(2.9) \quad B_j(T, v) = \sum_{S: j \in S \subseteq T} \gamma_S(T) [v(S) - v(S - \{j\})], \forall j \in T,$$

which is a formula similar to formula (1.7) for the Shapley value.

Obviously, Theorem 3 is similar to the major result (b). Note that in [8]

the results for the Shapley value have been proved in the converse order, using in the proof of (b) the well known axioms for the Shapley value (see [11],[12]). Instead, we have used simple algebraic manipulations with the Banzhaf value formula which led to Lemma 2. Clearly, the results for the Shapley value can also be presented in the same way.

Note that the representation of the potential of the Banzhaf value in terms of the dividend form is quite different:

Theorem 4: If Q is the potential of the Banzhaf value and α_T are the coordinates of $v \in G^N$ in the unanimity basis, then

$$(2.10) \quad Q(N, v) = \sum_{T \subseteq N} \frac{1}{2^{t-1}} \alpha_T,$$

where $t = |T|$.

Proof: Q is linear on G^N , so that to prove the result we should compute the potentials of the basic vectors. Theorem 1 for $T=N$ and $v = u_R$, $R \subseteq N$, gives

$$(2.11) \quad Q(N, u_R) = \frac{1}{2^{n-1}} \sum_{S \subseteq N} u_R(S),$$

where u_R is the unanimity game associated with coalition R , i.e. $u_R(S) = 1$ if $S \supseteq R$ and $u_R(S) = 0$ otherwise. Taking into account the definition of u_R the sum makes 2^{n-r} , so that from (2.11) we get $Q(N, u_R) = 2^{1-r}$. Then, from $v = \sum_{R \subseteq N} \alpha_R u_R$ the linearity of Q and the result just proved, we get (2.10). Recall that the expression for the Shapley value was $P(N, v) = \sum_{T \subseteq N} t^{-1} \alpha_T$, (see [8]).

Let us return to other recent results on the Shapley value, for which we shall be looking for similar results on the Banzhaf value in the next section. In [2], the author has introduced the concept of potential basis for the Shapley value and the following major results have been proved:

(c) There is a set of games $w_S \in G^N, \forall S \subseteq N, S \neq \emptyset$, given by

$$(2.12) \quad w_S(S) = s, \quad w_S(S \cup \{j\}) = -1, \quad \forall j \notin S, \quad w_S = 0 \text{ otherwise}$$

which form a basis for G^N , such that the coordinates of each $v \in G^N$ in that basis are exactly the potentials of the Shapley value for v and its subgames; (d) If ϕ is the Shapley value, then

$$(2.13) \quad \begin{aligned} \phi(N, w_S) &= 0, \forall S \subset N, \quad |S| \leq n-2, \quad \phi(N, w_N) = e \\ \phi(N, w_{N-\{j\}}) &= -e_j, \forall j \in N, \end{aligned}$$

where e_j are the vectors of the standard basis in R^n and $e \in R^n$ has all components 1; (e) the null space of the Shapley value is generated by $w_N + \sum_{j \in N} w_{N-\{j\}}$ and all w_S with $|S| \leq n-2$, and each $v \in G^N$ can be represented as

$$(2.14) \quad v = \sum_{|S| \leq n-2} P(S, w_S) w_S + P(N, w_N) (w_N + \sum_{j \in N} w_{N-\{j\}}) - \sum_{j \in N} \phi_j(N, v) w_{N-\{j\}}.$$

Note that the last formula gives all games in G^N for which the Shapley value equals ϕ , if we consider ϕ as an apriori given vector in R^n . In the next section, we shall see that the concept of potential for the Banzhaf value introduced above is helpful in getting results similar to (c), (d), and (e), in this case.

3. Potential Basis, Null Space, The Inverse Problem.

A basis for G^N is a potential basis of some value χ , if the coordinates of each $v \in G^N$ relative to that basis are exactly the potentials of χ for its subgames.

Lemma 5: If χ is a value which has a potential H , and there is a potential basis $C = \{c_S : S \subseteq N, S \neq \emptyset\}$, of χ , then we should have for each coalition S :

$$(3.1) \quad H(S, c_S) = 1, \quad H(T, c_S) = 0, \quad \forall T \neq S.$$

Proof: If C is a basis for G^N , then each $v \in G^N$ can be expressed as $v = \sum_{T \subseteq N} \alpha_T c_T$; so, for $v = c_S$ there are coefficients α_T such that $c_S = \sum_{T \subseteq N} \alpha_T c_T$. Then, we have $\alpha_S = 1$ and $\alpha_T = 0, \forall T \neq S$, because C is a basis. As C is a potential basis, we have $\alpha_T = H(T, c_S), \forall T \subseteq N$, and (3.1) follows.

Note that Theorem 1 gives the potentials of the Banzhaf value for v and its subgames when the game is known, but the same Theorem 1 determines recursively the game whenever the potentials are known, by the formula

$$(3.2) \quad v(T) = 2^{t-1} Q(T, v) - \sum_{U \subset T} v(U), \quad \forall T \subseteq N,$$

derived from (2.2).

Theorem 6: For each $S \subseteq N$, $S \neq \emptyset$, there is a unique game c_S such that $Q(S, c_S) = 1$ and $Q(T, c_S) = 0 \quad \forall T \neq S$. Moreover, we have for each S :

$$(3.3) \quad c_S(T) = \begin{cases} (-1)^p \cdot 2^{s-1} & \text{if } T \supseteq S, p = t - s, \\ 0 & \text{otherwise.} \end{cases}$$

Proof: If $|T| < |S|$, then formula (3.2) written for $v = c_S$ shows that $c_S(T) = 0$; we have $c_S(S) = 2^{s-1}$ and $c_S(T) = 0, \forall T \neq S$, if $|T| = |S|$. If $t > s$, then in (3.2) the first term is zero and we should use

$$(3.4) \quad c_S(T) = - \sum_{U \subset T} c_S(U).$$

Proceed by induction over $p = t - s$; check the case $p = 1$. If $T \not\supseteq S$, then all $c_S(U), \forall U \subset T$, in (3.4) are zero, hence $c_S(T) = 0$; if $T \supset S$, then in the sum only $c_S(S)$ is different of zero and we get $c_S(T) = -2^{s-1}$, i.e. (3.3) holds. Assume that for $t = s + p$, with $1 \leq p < k$, formula (3.3) holds and consider the case $p = k$. If $T \not\supseteq S$, then each $U \subset T$ does not contain S and again $c_S(T) = 0$; if $T \supset S$, then the nonzero terms in the right hand side of (3.4) can be grouped according to the size of U : there is one term for $U = S$, then $C(k, 1)$ terms for $U = S \cup \{j\}, \forall j \notin S, j \in T$, and so on, $C(k, k-1)$ terms for $S = T - \{j\}, j \notin S, j \in T$. In each group all terms are equal and by using the induction assumption we get

$$(3.5) \quad \begin{aligned} c_S(T) &= -[2^{s-1} + C(k, 1)(-1) \cdot 2^{s-1} + \dots + C(k, k-1)(-1)^{k-1} \cdot 2^{s-1}] = \\ &= -2^{s-1}[1 - C(k, 1) + \dots + (-1)^{k-1} C(k, k-1)] = (-1)^k \cdot 2^{s-1} \end{aligned}$$

where we have used a well known combinatorial identity. Hence, formula (3.3) holds for k and the result follows.

Example 1: If $|N| = 3$, then formula (3.3) gives the following games

$$\begin{aligned}
c_{\{1\}} &= (1, 0, 0, -1, -1, 0, 1), & c_{\{1,2\}} &= (0, 0, 0, 2, 0, 0, -2) \\
c_{\{2\}} &= (0, 1, 0, -1, 0, -1, 1), & e_{\{1,3\}} &= (0, 0, 0, 0, 2, 0, -2), \\
c_{\{3\}} &= (0, 0, 1, 0, -1, -1, 1), & c_{\{2,3\}} &= (0, 0, 0, 0, 0, 2, -2), \\
c_{\{1,2,3\}} &= (0, 0, 0, 0, 0, 0, 4),
\end{aligned}$$

where the components of each c_S have been taken in the order $\{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}$, and $\{1,2,3\}$. Note that these vectors are linearly independent, hence they form a basis of R^7 .

Theorem 7: The set of games $c_S \in G^N, \forall S \subseteq N, S \neq \emptyset$, form a basis for G^N and this is a potential basis for the Banzhaf value.

Proof: It is easy to see, like in Example 1, that the set C is linearly independent, hence it is a basis for G^N because $|C| = 2^n - 1 = \dim(G^N)$. Thus we should show that C is a potential basis for the Banzhaf value, i.e. for any $v \in G^N$, in $v = \sum_{T \subseteq N} \alpha_T c_T$ we have $\alpha_T = Q(T, v), \forall T \subseteq N$, where Q is the potential of the Banzhaf value. From the expansion $v(S) = \sum_{U \subseteq N} \alpha_U c_U(S)$ by adding up as shown in formula (2.2) we get $Q(T, v) = \sum_{U \subseteq N} \alpha_U Q(T, c_U)$. The games c_U satisfy $Q(T, c_U) = 0, \forall U \neq T$, and $Q(T, c_T) = 1$, hence $Q(T, v) = \alpha_T$.

Note that for the space G^N with $|N| = 3$, we can check the fact that the games of Example 1 form a potential basis. Indeed, from the expansion

$v = \sum_{U \subseteq N} \alpha_U c_U$ we get componentwise:

$$\begin{aligned}
v(1) &= \alpha_1, \quad v(2) = \alpha_2, \quad v(3) = \alpha_3, \quad v(1,2) = 2\alpha_{1,2} - \alpha_1 - \alpha_2, \quad v(1,3) = 2\alpha_{1,3} - \alpha_1 - \alpha_3 \\
v(2,3) &= 2\alpha_{2,3} - \alpha_2 - \alpha_3, \quad v(1,2,3) = 4\alpha_{1,2,3} - 2\alpha_{1,2} - 2\alpha_{1,3} - 2\alpha_{2,3} + \alpha_1 + \alpha_2 + \alpha_3, \quad \text{and for} \\
\text{example we can compute by using (2.2):}
\end{aligned}$$

$$Q(\{1,2\}, v) = \frac{1}{2} [v(1) + v(2) + v(1,2)] = \alpha_{1,2}$$

$$Q(\{1,2,3\}, v) = \frac{1}{4} [v(1) + v(2) + v(3) + v(1,2) + v(1,3) + v(2,3) + v(1,2,3)] = \alpha_{1,2,3}$$

Now we intend to compute the Banzhaf value for the basic vectors in the potential basis. This gives

Theorem 8: If $c_S, \forall S \subseteq N, S \neq \emptyset$, are the games in the potential basis of the Banzhaf value, then their Banzhaf values are:

$$(3.6) \quad \begin{aligned} B(N, c_S) &= 0, \forall S \subset N, |S| \leq n-2, \\ B(N, c_{N-\{j\}}) &= -e_j, \forall j \in N, \\ B(N, c_N) &= e, \end{aligned}$$

where $e_j, \forall j \in N$, are the vectors of the standard basis for R^n , and e is a vector with all components equal to 1.

Proof: To find the Banzhaf values we should compute $Q(N, c_S)$ and $Q(N - \{j\}, c_S), \forall j \in N$, and use Theorem 3. If $S = N$, then $Q(N, c_N) = 1, Q(N - \{j\}, c_N) = 0, \forall j \in N$, hence $B(N, c_N) = e$. If $S = N - \{j\}$, then $Q(N, c_{N-\{j\}}) = 0$ and $Q(N - \{i\}, c_{N-\{j\}}) = \delta_i^j$, where $\delta_i^j = 0$ when $i \neq j$, and $\delta_i^i = 1$, hence $B(N, c_{N-\{j\}}) = -e_j, \forall j \in N$. If $|S| \leq n-2$, then both $Q(N, c_S) = 0$ and $Q(N - \{j\}, c_S) = 0, \forall j \in N$, hence $B(N, c_S) = 0, \forall S \subset N$ with $|S| \leq n-2$.

Note that the Shapley values of the basic vectors in the potential basis of the Shapley value (2.13), are the same as the Banzhaf values (3.6) of the basic vectors in the potential basis of the Banzhaf value.

Now, we intend to solve the following inverse problem: a vector $\Psi \in R^n$ is given; find all games $v \in G^N$ such that $B(N, v) = \Psi$. In particular, we shall determine the null space of the Banzhaf value.

Theorem 9: The set of games $\{c_N + \sum_{j \in N} c_{N-\{j\}}\} \cup \{c_S : |S| \leq n-2\}$, generates the null space of the Banzhaf value. Moreover, the set of games $v \in G^N$ such that $B(N, v) = \Psi$, where Ψ is apriori given, can be represented by

$$(3.7)v = \sum_{|S| \leq n-2} q_S c_S + q_N (c_N + \sum_{j \in N} c_{N-\{j\}}) - \sum_{j \in N} \Psi_j c_{N-\{j\}}.$$

Proof: By Theorem 8, we have $B(N, c_S) = 0, \forall S \subset N$, with $|S| \leq n-2$, and we can get easily $B(N, c_N + \sum_{j \in N} c_{N-\{j\}}) = 0$; therefore, any linear combination of these vectors is in the null space. It remains to show that any $v \in G^N$ with $B(N, v) = 0$ is such a linear combination. As C is a potential basis, for any $v \in G^N$ we can write the expansion

$$(3.8) \quad v = \sum_{|S| \leq n-2} Q(S, v) c_S + Q(N, v) (c_N + \sum_{j \in N} c_{N-\{j\}}) + \sum_{j \in N} [Q(N, v) - Q(N - \{j\}, v)] c_{N-\{j\}};$$

if $B(N, v) = \Psi$, and we denote $Q(S, v) = q_S, \forall S \subset N$ with $|S| \leq n-2$, and $Q(N, v) = q_N$, then by theorem 3 we get the representation (3.7) for v . Conversely, if a game $v \in G^N$ is represented by (3.7), then for any numbers $q_S, |S| \leq n-2$, and q_N , we have $B(N, v) = \Psi$, by Theorem 8. Hence (3.7) gives the set of all games having the apriori known Banzhaf value Ψ . If $\Psi = 0$, then v is a linear combination of $c_N + \sum_{j \in N} c_{N-\{j\}}$ and $c_S, \forall S \subset N$ with $|S| \leq n-2$, so that these vectors generate the null space.

Example 2: Let $\Psi \in R^3$ be given; taking into account (3.7) and the expressions of the basic vectors shown in Example 1, we get an explicit representation of all games in G^N with $|N| = 3$, with $B(N, v) = \Psi$, as follows:

$$\begin{aligned} v(1) &= q_1, & v(1, 2) &= 2q_N - q_1 - q_2 - 2\Psi_3 \\ v(2) &= q_2, & v(1, 3) &= 2q_N - q_1 - q_3 - 2\Psi_2 \\ v(3) &= q_3, & v(2, 3) &= 2q_N - q_2 - q_3 - 2\Psi_1, \\ v(1, 2, 3) &= q_1 + q_2 + q_3 - 2q_N + 2(\Psi_1 + \Psi_2 + \Psi_3). \end{aligned}$$

One may check that the parameters q_1, q_2, q_3 and q_N are potentials and the Banzhaf value is Ψ .

We return to an axiomatic characterization based upon the reduced game property, similar to that given by S. Hart and A. Mas-Colell in [8] for the Shapley value; however, it is easy to show an example in which the Banzhaf value does not possess a reduced game property relative to the reduced game used in [8]. Therefore, we shall introduce a new reduced game in the next section and we shall prove that an axiomatic characterization of the Banzhaf value relative to this game can be given.

4. The reduced game and an axiomatic characterization.

For any $v \in G^N$ and any $S \subseteq N, S \neq \emptyset$, let $\pi(S, v)$ denotes the power of the subgame (S, v) , i.e. $\pi(S, v) = \sum_{i \in S} B_i(S, v)$, where $B_i(S, v)$ are the components of the Banzhaf value for this subgame. By Lemma 2, if v is fixed then $\pi(S, v)$ are uniquely defined for all $S \subseteq N, S \neq \emptyset$, by formula

$$(4.1) \quad \pi(S, v) = \frac{1}{2^s - 1} [sv(S) + \sum_{U \subset S} (2u - s)v(U)] ;$$

we can say that π is a functional which assigns to each $v \in G^N$ a power game $(\pi(S, v)) \in G^N$. But from (4.1) we get

$$(4.2) \quad v(S) = \frac{1}{s} [2^{s-1} \pi(S, v) - \sum_{U \subset S} (2u - s)v(U)] , \forall S \subseteq N,$$

which shows that conversely, if we know the values of the functional π for v and all its subgames, then we can obtain from (4.2) the unique game v for

which the power game is the given one. In other words, we have:

Lemma 10: The functional $\pi: G^N \rightarrow G^N$ is a one-to-one correspondence on G^N . This result makes valid the following:

Definition 11: For any $T \subset N$ and any value χ , the reduced game $v_T^\chi \in G^T$ is the game defined by the functional equation

$$(4.3) \quad \pi(S, v_T^\chi) = \pi(S \cup T^c, v) - \sum_{j \in T} c_j \chi_j(S \cup T^c, v), \forall S \subseteq T,$$

where $T^c = N - T$.

Note the interpretation of (4.3): the power of the subgame (S, v_T^χ) equals the power of the game that S would have made with the players who left, minus the total payoff taken by those players before leaving. Note also that if the solution is the Banzhaf value,

$$(4.3)' \quad \pi(S, v_T^B) = \sum_{j \in S} B_j(S \cup T^c, v), \forall S \subseteq T,$$

taking into account the property (1.3) of the value.

Recall that a solution χ has the reduced game property relative to the reduced game v_T^χ , if for all $T \subset N$ we have $\chi_j(T, v_T^\chi) = \chi_j(N, v), \forall j \in T$. In other words, in the reduced game v_T^χ the players in T will get the same payoff as in v .

Theorem 12: The Banzhaf value has the reduced game property relative to the game v_T^B defined by (4.3)' for all $T \subset N$.

Proof: In the first stage, we find the relationship between the potentials of the game $v \in G^N$ and those of the game $v_T^B \in G^T$ for all $S \subseteq T$. The potentials of v_T^B by (2.1) occur in

$$(4.4) \quad \pi(S, v_T^B) = \sum_{j \in S} [Q(S, v_T^B) - Q(S - \{j\}, v_T^B)], \forall S \subseteq T.$$

In the definition (4.3)' we use Theorem 3 to get

$$(4.5) \quad \pi(S, v_T^B) = \sum_{j \in S} [Q(S \cup T^c, v) - Q(S \cup T^c - \{j\}, v)], \forall S \subseteq T.$$

From (4.4) and (4.5) we obtain the equality

$$(4.6) \quad \sum_{j \in S} [Q(S, v_T^B) - Q(S - \{j\}, v_T^B)] = \sum_{j \in S} [Q(S \cup T^c, v) - Q(S \cup T^c - \{j\}, v)], \forall S \subseteq T.$$

Now, we intend to prove by induction the formula

$$(4.7) \quad Q(S, v_T^B) = Q(S \cup T^c, v) - Q(T^c, v), \forall S \subseteq T$$

by using (4.6). Obviously, (4.7) holds for $S = \emptyset$; formula (4.7) is valid also for $S = \{i\}$, as shown by (4.6) for $S = \{i\}, \forall i \in N$. Assume that (4.7) is valid for all coalitions L with $|L| < s$ and consider a coalition $\tilde{S}$ with $|\tilde{S}| = s$; clearly, $\tilde{S} = L \cup \{i\}$ for some L with $|L| = s-1$ and $i \notin L$. We actually write (4.6) for $S = \tilde{S} = L \cup \{i\}$:

$$sQ(L \cup \{i\}, v_T^B) = sQ(L \cup T^c \cup \{i\}, v) + \sum_{j \in L \cup \{i\}} [Q(L \cup \{i\} - \{j\}, v_T^B) - Q(L \cup T^c \cup \{i\} - \{j\}, v)].$$

In the right hand side, the potential which occurs as the first in the bracket is applied to a coalition of size $|L \cup \{i\} - \{j\}| = s-1$, hence (4.7) holds for $S = L \cup \{i\} - \{j\}$, and each term of the sum equals $-Q(T^c, v)$; there are s terms, hence the right hand side gives

$$Q(L \cup \{i\}, v_T^B) = Q(L \cup T^c \cup \{i\}, v) - Q(T^c, v),$$

i.e. (4.7) holds for $\tilde{S} = L \cup \{i\}$.

Formula (4.7) will be used for proving the reduced game property of the Banzhaf value as follows:

$$B_j(T, v_T^B) = Q(T, v_T^B) - Q(T - \{j\}, v_T^B) = Q(N, v) - Q(N - \{j\}, v) = B_j(N, v), \forall j \in N,$$

where the first and the last equality follow from Theorem 3 and the second one follows from (4.7) written for $S = T$ and $S = T - \{j\}$.

Note that the proof has the same type of steps as the proof of the similar theorem in [8]; the new part is an appropriate definition for the potential and for the reduced game corresponding to the Banzhaf value. The next example shows the computation of a reduced game and illustrates the reduced game property.

Example 3 Consider the monotonic simple game: $v(1) = v(2) = v(3) = v(1, 2) = v(1, 3) = 0$ and $v(2, 3) = v(1, 2, 3) = 1$. We intend to compute the reduced game for $T = \{1, 2\}$ and the Banzhaf value. We get the reduced game by using (4.2) for $v = v_T^B$, where $\pi(S, v_T^B)$ is given by (4.3)', i.e. we use

$$(4.8) \quad v_T^B(S) = \frac{1}{s} [2^{s-1} \sum_{j \in S} B_j(S \cup T^c, v) - \sum_{U \subset S} (2u - s)v(U)], \forall S \subseteq T.$$

To compute $v_T^B(1)$ we need $B_1(\{1, 3\}, v)$, to compute $v_T^B(2)$ we need $B_2(\{2, 3\}, v)$ and to compute $v_T^B(1, 2)$ we need $B_1(N, v)$ and $B_2(N, v)$. By using (1.2) we get $B_1(\{1, 3\}, v) = 0, B_2(\{2, 3\}, v) = \frac{1}{2}, B_1(N, v) = 0$ and $B_2(N, v) = \frac{1}{2}$. Now, from (4.8) we obtain the reduced game

$$(4.9) \quad v_T^B(1) = 0, \quad v_T^B(2) = \frac{1}{2}, \quad v_T^B(1, 2) = \frac{1}{2}.$$

We can illustrate the reduced game property by computing $B_1(T, v_T^B)$, $B_2(T, v_T^B)$.

We get $B_1(T, v_T^B) = 0 = B_1(N, v)$, and $B_2(T, v_T^B) = \frac{1}{2} = B_2(N, v)$.

Recall from [8]: a solution χ is standard for two person games if for all $v \in G^N$ with $|N| = 2$, we have

$$(4.10) \quad \chi_i(N, v) = v(i) + \frac{1}{2}[v(N) - v(i) - v(j)], \forall i \in N,$$

where $N = \{i, j\}$, so that $j \neq i$.

Lemma 13: The Banzhaf value is standard for two person games.

Proof: Use (1.2) to write $B_i(N, v) = \frac{1}{2}[v(i) + v(N) - v(j)]$, which is exactly (4.10).

Now, we intend to show that the reduced game property and the standard property uniquely determine the Banzhaf value.

Theorem 14: Let $\chi: G^N \rightarrow R^n$. Then

- (i) χ has the reduced game property relative to the reduced game (4.3), and
- (ii) χ is standard for two person games,

if and only if χ is the Banzhaf value.

Proof: The “if” part has been proved in Theorem 12 and Lemma 13, so that the “only if” part remains to be proved. By the first part of the proof, we know that the Banzhaf value has the properties (i) and (ii); therefore, we have to prove that if χ has also properties (i) and (ii), then $\chi = B$. In the first stage, we show that these properties imply for χ :

$$(4.11) \quad \sum_{k \in S} \chi_k(S, v) = \pi(S, v), \forall S \subseteq N.$$

If $|S|=1$, i.e. $S=\{i\}$, let j be a dummy to be added to form the game $(\{i,j\},v)$, where $v(i,j)=v(i)+v(j)$ because j is a dummy. By (ii), we have $\chi_i(\{i,j\},v)=v(i)$; by (i), we have $\chi_i(\{i\},v_{\{i\}}^x)=\chi_i(\{i,j\},v)$, hence $\chi_i(\{i\},v_{\{i\}}^x)=v(i)$, where $v_{\{i\}}^x$ is the reduced game. From the definition of the reduced game we have $\pi(\{i\},v_{\{i\}}^x)=v(i)$ and from the definition of π we have $\pi(\{i\},v_{\{i\}}^x)=v_{\{i\}}^x(i)$, hence we have $v_{\{i\}}^x(i)=v(i)$. The result used in the equality shown above gives $\chi_i(\{i\},v)=\pi(\{i\},v)$, where we used again the definition of π in the right hand side. This shows that (4.11) holds for $|S|=1$. If $|S|=2$, then by (ii) we have $\chi_k(S,v)=B_k(S,v), \forall i \in S$, hence by adding up these two equalities we get (4.11) satisfied for $|S|=2$, because B satisfies (4.11). If $|S|>2$, consider a coalition U with $|U|=2$ and let $U^c=S-U$; we can write

$$\begin{aligned} \sum_{j \in S} \chi_j(S,v) &= \sum_{j \in U} \chi_j(S,v) + \sum_{j \in U^c} \chi_j(S,v) = \sum_{j \in U} \chi_j(U, v_U^x) + \\ &\quad + \sum_{j \in U^c} \chi_j(S,v) = \pi(U, v_U^x) + \sum_{j \in U^c} \chi_j(S,v) = \pi(S,v), \end{aligned}$$

where the second equality follows from (i), the third follows from (4.11) for $|S|=2$, which has already been proved, and the last equality follows from the definition (4.3) of the reduced game. This shows that (4.11) has been proved for all $S \subseteq N$.

Note that in cases $|S|=1$ and $|S|=2$, we have proved more, namely we have

$$(4.12) \quad \chi_k(S,v) = B_k(S,v), \forall k \in S.$$

Now, we intend to prove the equality (4.12) for all $S \subseteq N$, by induction over the size of S ; this will be the second stage of the proof, the uniqueness of B . We assume that (4.12) holds for all coalitions L with $|L|<s$, and we

shall prove it for S ; the assumption holds for $s=3$ as we noticed above, hence $|S| \geq 3$.

Consider any $i \in S$ and $j \in S$, $i \neq j$, and denote $U = \{i, j\}$, and $U^c = S - U$; let v_U^X and v_U^B be the reduced games defined by (4.3) for χ and B . From definition (4.3) and (4.11) we have:

$$\begin{aligned}\pi(\{i\}, v_U^X) &= \pi(S - \{j\}, v) - \sum_{k \in U^c} c_k \chi_k(S - \{j\}, v) = \chi_i(S - \{j\}, v), \\ \pi(\{i\}, v_U^B) &= \pi(S - \{j\}, v) - \sum_{k \in U^c} c_k B_k(S - \{j\}, v) = B_i(S - \{j\}, v);\end{aligned}$$

by the induction assumption the last side on each row is the same, hence we proved $v_U^X(i) = v_U^B(i)$, taking into account the definition of π . Obviously, we have also $v_U^X(j) = v_U^B(j)$. This result shows that $\chi_i(U, v_U^X) \geq \chi_i(U, v_U^B)$ if and only if $v_U^X(U) \geq v_U^B(U)$, and also if and only if $\chi_j(U, v_U^X) \geq \chi_j(U, v_U^B)$. We proved

$$(4.13) \quad \chi_i(U, v_U^X) \geq \chi_i(U, v_U^B) \Leftrightarrow \chi_j(U, v_U^X) \geq \chi_j(U, v_U^B),$$

for any coalition $U \subseteq S$ with $|U| = 2$. Consider now again $U = \{i, j\}$; we have

$$\chi_i(S, v) = \chi_i(U, v_U^X) \geq \chi_i(U, v_U^B) = B_i(U, v_U^B) = B_i(S, v)$$

if and only if

$$\chi_j(S, v) = \chi_j(U, v_U^X) \geq \chi_j(U, v_U^B) = B_j(U, v_U^B) = B_j(S, v),$$

where the first equality on each row follows from (i), the second equality follows from the fact that $\chi = B$ for two person games, the last equality

follows again from (i) and the double implication is given by (4.13). For each pair $i, j \in S$ we have

$$(4.14) \quad \chi_i(S, v) \geq B_i(S, v) \Leftrightarrow \chi_j(S, v) \geq B_j(S, v);$$

from (4.11) we have also

$$(4.15) \quad \sum_{k \in S} \chi_k(S, v) = \sum_{k \in S} B_k(S, v)$$

hence, (4.14) can hold only if (4.12) holds, so that (4.12) is proved.

Note that the proof has the same type of steps as the proof of the similar result in [8], but the new definition of the reduced game has introduced more difficulties.

Remarks: a) T. Driessen (1991), gives a comprehensive survey of axiomatizations based upon reduced game properties, for some values; b) S. Hart and A. Mas-Colell (1989), have also discussed extensions to the weighted Shapley value of the properties connected to their potential of the Shapley value; c) An alternative axiomatization of the Banzhaf value, containing another reduced game and the reduced game property, is introduced by E. Lehrer (1988).

References

1. J.F. Banzhaf III, (1965), Weighted voting doesn't work: a mathematical analysis, *Rutgers Law Review*, 19, 317-343.
2. I. Dragan, (1991), The potential basis and the weighted Shapley value, *Libertas Mathematica*, 11, 139-150.
3. I. Dragan, (1992), Multiweighted Shapley values and random order values, Technical Report, #286, Univ. of Texas at Arlington.
4. T. Driessen, (1991), A survey of consistency properties in cooperative game theory, *SIAM Review*, 33, 43-59.
5. P. Dubey, (1975), On the uniqueness of the Shapley value, *Int. Journal of Game Theory*, 4, 131-139.
6. P. Dubey and L.S. Shapley, (1979), Mathematical properties of the Banzhaf power index, *Math. Oper. Res.*, 4, 99-131.
7. V. Feltkamp, (1993), Alternative axiomatic characterizations of the Shapley and Banzhaf values, Discussion Paper #9353, Center Tilburg University.
8. S. Hart and A. Mas-Colell, (1988), The potential of the Shapley value, the The Shapley value, Essays in Honor of L.S. Shapley, A.E. Roth, (ed.), Cambridge Univ. Press, 127-137.
9. S. Hart and A. Mas-Colell, (1989), Potential, value and consistency, *Econometrica*, 57, 3, 589-614.
10. E. Lehrer, (1988), An axiomatization of the Banzhaf value, *Int. Journal of Game Theory*, 17, 89-99.
11. G. Owen, (1982), *Game Theory*, 2nd edition, Academic Press.
12. L.S. Shapley, (1953), A value for n-person games, *Ann. Math. Studies*, 28, 307-317.