The Shapley value, the Owen value, and the veil of ignorance

André Casajus†

(January 2007, this version: October 3, 2008, 16:54)

Abstract

We show that the Owen value for TU games with a cooperation structure extends the Shapley value in a consistent way. In particular, the Shapley value is the expected Owen value for all symmetric distributions on the partitions of the player set. Similar extensions of the Banzhaf value do not show this property.

Journal of Economic Literature Classification Number: C71.

Key Words: Shapley value, Owen value, Banzhaf value, extension, consistency.

†Chair of Economics and Information Systems, HHL Leipzig Graduate School of Management, Jahnallee 59, D-04109 Leipzig, Germany.
‡Professur für Mikroökonomik, Wirtschaftswissenschaftliche Fakultät, Universität Leipzig, PF 100920, D-04009 Leipzig, Germany. e-mail: casajus@wifa.uni-leipzig.de

I thank an anonymous referee and an associate editor for comments on an earlier version of this note. Of course, the usual disclaimer applies.
1. Introduction

Suppose all players of a TU game cooperate with respect to the production of worth. Then, they face the problem of distributing the grand coalition’s worth among them. In the bargaining process on this distribution, the players may find themselves arranged in groups which may influence the resulting distribution of worth. Ex post, the players know which group they belong to and how the other players are organized, but ex ante, they are behind a Rawlsian (1971) “veil of ignorance”, only being able to form expectations on how the players are split into groups. These expectations, however, cannot be arbitrary. At least, they should treat “similar” arrangements in a “similar” way.

The most influential concept dealing with the ex-ante problem is the Shapley (1953) value. In order to respond to the ex post problem, the Shapley value was extended by Owen (1977) to TU games with a coalition structure (partition of the player set) (henceforth CS-games). Further, he indicated an extension to probability distributions on the set of coalition structures (p. 85).

One could argue that a “good” extension should exhibit the following consistency property: The expected payoffs of the extended (ex post) value for any “fair” probability distribution on the set of coalition structures should equal the payoff under the original (ex ante) value. Hart and Kurz (1983, Corollary 2.5) show that the Owen value meets this property in two cases: The probability of either the atomic partition or the trivial partition is 1. Both of these probability distributions are “fair”—they treat all players symmetrically. In this paper, we make explicit the notion of “fairness” and then show that the Owen value is a consistent extension of the Shapley value. However, the Owen value is not unique in this respect. Further, we show that the extensions of the Banzhaf (1965) value to CS-games by Owen (1981) and by Alonso-Meijide and Fiestras-Janeiro (2002) do not exhibit this property.

The plan of this note is as follows: The second section provides basic definitions and notation. In the third section, we present the positive result for the Shapley value and the negative one for the Banzhaf value.

2. Basic definitions and notation

A TU game is a pair \((N, v)\) where \(N \neq \emptyset \) is finite and \(v \in V(N) := \{f : 2^N \rightarrow \mathbb{R} | v(\emptyset) = 0\} \). Members of \(N\) are called player; subsets of \(N\) are called coalitions; \(v(K)\) is called the worth of \(K \subseteq N\). A player \(i\) is called a Null player in \((N, v)\) iff
$v(K \cup \{i\}) = v(K)$ for all $K \subseteq N \setminus \{i\}$. For $v, v' \in V(N)$, $v + v' \in V(N)$ is given by $(v + v')(K) = v(K) + v'(K)$ for all $K \subseteq N$. A value is an operator $\varphi$ that assigns payoff vectors to all games, $\varphi(N, v) \in \mathbb{R}^N$. For $K \subseteq N$, we denote $\sum_{i \in K} \varphi_i(N, v)$ by $\varphi_K(N, v)$. The Shapley (1953) value, $\text{Sh}$, is characterized by the well-known axioms $\text{E, A, N,}$ and $\text{S}$ below ($\text{Sh}$ satisfies $\text{S}$, but a weaker version already suffices). Since we only make use of these axioms, we do not provide the well-known formulas for the Shapley value and for the Owen value below.

**Efficiency, E.** $\varphi_N(N, v) = v(N)$.

**Symmetry, S.** For all bijections $\pi : N \rightarrow N$ and $i \in N$, $\varphi_\pi(i)(N, v) = \varphi_i(N, v \circ \pi)$.

**Null player, N.** If $i \in N$ is a Null player then $\varphi_i(N, v) = 0$.

**Additivity, A.** $\varphi(N, v + v') = \varphi(N, v) + \varphi(N, v')$ for all $v, v'$.

A coalition structure for $(N, v)$ is a partition $\mathcal{P} \subseteq 2^N$ of $N$, elements of $\mathcal{P}$ are referred to as components, and $\mathcal{P}(i)$ denotes the component containing player $i$. Set $\langle N \rangle := \{\{i\} | i \in N\}$; for $K \subseteq N$, set $\mathcal{P}(K) := \bigcup_{i \in K} \mathcal{P}(i)$. Components $P, P' \in \mathcal{P}$ are called symmetric in $(N, v, \mathcal{P})$ iff $v(\mathcal{P}(K) \cup P) = v(\mathcal{P}(K) \cup P')$ for all $K \subseteq N \setminus (P \cup P')$. A CS-game is a game together with a coalition structure, $(N, v, \mathcal{P})$. A CS-value is an operator $\varphi$ that assigns payoff vectors to all CS-games, $\varphi(N, v, \mathcal{P}) \in \mathbb{R}^N$. The Owen (1977) value is characterized by the axioms $\text{E, A, N,}$ as well as $\text{CS}$ and $\text{SC}$, below (Winter, 1992) (Ow satisfies $\text{CS}$, but a weaker version already suffices).

**CS-Symmetry, CS.** For all bijections $\pi : N \rightarrow N$ and $i \in N$, $\varphi_\pi(i)(N, v, \mathcal{P}) = \varphi_i(N, v \circ \pi, \pi(\mathcal{P}))$.

**Symmetry between components, SC.** If $P, P' \in \mathcal{P}$ are symmetric in $(N, v, \mathcal{P})$ then

$$\varphi_P(N, v, \mathcal{P}) = \varphi_{P'}(N, v, \mathcal{P}).$$

### 3. The Shapley value as an average Owen value

Owen (1977, Section 6) already indicated the obvious extension of his CS-value to probability distributions of coalition structures. Let $\mathbb{P}(N)$ denote the set of all partitions on $N$, and let further $W(\mathbb{P}(N))$ denote the set of all probability distributions $p$ on $\mathbb{P}(N)$. The Owen value is extended to $W(\mathbb{P}(N))$ by

$$\text{Ow}(N, v, p) := \sum_{\mathcal{P} \in \mathbb{P}(N)} p(\mathcal{P}) \text{Ow}(N, v, \mathcal{P}), \quad p \in W(\mathbb{P}(N)).$$
Remains to make precise which \( p \in W (\mathcal{P} (N)) \) treat similar coalitions similarly. The following definition requires such \( p \) not to depend on the players’ arbitrary names given by the game theorist. This implies that these probabilities only respond to the cardinalities of the components of the coalition structures. In a sense, this reflects the situation behind the “veil of ignorance”. A probability distribution \( p \in W (\mathcal{P} (N)) \) is called symmetric if we have \( p (\mathcal{P}) = p (\pi (\mathcal{P})) \) for all \( \mathcal{P} \in \mathcal{P} (N) \) and all bijections \( \pi : N \to N \) where \( \pi (\mathcal{P}) := \{ \pi (P) \, | \, P \in \mathcal{P} \} \).

**Theorem 1.** If \( p \in W (\mathcal{P} (N)) \) is symmetric then \( \text{Ow} (N, v, p) = \text{Sh} (N, v) \).

**Proof.** Let \( p \in W (\mathcal{P} (N)) \) be symmetric. Since the Shapley value is characterized by \( E, S, N, \) and \( A \) on a fixed player set, it suffices to show that the value \( \text{Ow} (N, \cdot, p) \) given by (1) satisfies all these axioms. By (1), it is clear that \( \text{Ow} (N, \cdot, p) \) inherits \( E, A, \) and \( N \) from the Owen value. Let \( \pi : N \to N \) be an arbitrary bijection.

\[
\text{Ow}_{\pi(i)} (N, v, p) \overset{(1)}{=} \sum_{\mathcal{P} \in \mathcal{P} (N)} p (\mathcal{P}) \text{Ow}_{\pi(i)} (N, v, \mathcal{P})
\]

\[
= \sum_{\mathcal{P} \in \mathcal{P} (N)} p (\pi (\mathcal{P})) \text{Ow}_{\pi (\mathcal{P})} (N, v \circ \pi, \pi (\mathcal{P})) \overset{(1)}{=} \text{Ow}_{\pi (\mathcal{P})} (N, v \circ \pi, p).
\]

Thus, \( \text{Ow} (N, \cdot, p) \) satisfies \( S \). \( \square \)

Some remarks seem to be in order.

**Remark 1.** Since Theorem 1 considers all symmetric \( p \in W (\mathcal{P} (N)) \), it expresses a rather strong consistency property. Compare this with Remarks 5 and 6.

**Remark 2.** For \( p (\{ N \}) = 1 \) or \( p (\langle N \rangle) = 1 \), we have \( \text{Ow} (N, v, p) = \text{Ow} (N, v, \{ N \}) \) and \( \text{Ow} (N, v, p) = \text{Ow} (N, v, [N]) \), respectively. Hence, Theorem 1 is an extension of Hart and Kurz (1983, Corollary 2.5).

**Remark 3.** Wiese (2005) constructs a one-parameter family \( p_\alpha \in W (\mathcal{P} (N)) \), \( \alpha \in [0, 1] \). The idea behind this construction is the following: The players enter a room in random order where all these orders are equally probable; and upon entering, they either join an existing coalition or open a new one. With probability \( \alpha \in [0, 1] \), a player closes the coalition he joined or formed for subsequent players. For \( \mathcal{P} \in \mathcal{P} (N) \), this gives the probability

\[
p_\alpha (\mathcal{P}) = |N|^{-1} \cdot |\mathcal{P}|! \cdot \prod_{P \in \mathcal{P}} |P|! \cdot (1 - \alpha)^{|\mathcal{P}|-1} \cdot \prod_{P \in \mathcal{P}} \alpha^{|P|-1}.
\]
Since \( p_{\alpha}(P) \) depends on the cardinalities \(|P|\), \( P \in P \) only, is clear that \( p_{\alpha} \) is symmetric. For \( \alpha = 0 \), of course, we have \( p(\{N\}) = 1 \), and for \( \alpha = 1 \), we have \( p([N]) = 1 \) as in the previous remark.

**Remark 4.** In the proof of Theorem 1, we only employ the fact that the Owen value satisfies \( E, CS, N, \) and \( A \). Since axiom \( SC \) from the characterization of the Owen value is missing in this list, there are other CS-values that may replace \( Ow \) in Theorem 1. In the definition of the Owen value, for example, one may restrict attention to those orders of components where smaller components come precede larger ones.

**Remark 5.** Banzhaf (1965) introduces another value, \( Ba \), initially for voting games, which later was extended to general TU games by Owen (1975). Two extensions of the Banzhaf value to CS-games in the spirit of the Owen value have been proposed. The Owen (1981) extension, \( Ba^O \), controls the distribution within the components as well as between the components in the same fashion as the Banzhaf value. In contrast, the Alonso-Meijide and Fiestras-Janeiro (2002) extension, \( Ba^{AF} \), controls the distribution within the components as the Shapley value. All these concepts do not satisfy \( E \), which already indicates that consistency in the sense of Theorem 1 may not hold in general.

Indeed, this is what the following example reveals. Consider the TU game \((N, v)\), \( N = \{1, 2, 3\} \), \( v(K) = 1 \) if \(|K| > 1 \) and else \( v(K) = 0 \). We then have \( Ba_1(N, v) = \frac{1}{2} \) but

\[
\begin{array}{|c|c|c|c|c|}
\hline
\mathcal{P} & \langle N \rangle & \{\{1\}, \{2, 3\}\} & \{\{2\}, \{1, 3\}\} & \{\{3\}, \{1, 2\}\} & \{N\} \\
\hline
Ba^O_1(N, v, \mathcal{P}) & \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\
Ba^{AF}_1(N, v, \mathcal{P}) & \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\
\hline
\end{array}
\]

where the other players’ payoffs follow from symmetry arguments. Since a symmetric \( p \in W(\mathcal{P}(N)) \) puts equal weight on \( \{\{1\}, \{2, 3\}\} \), \( \{\{2\}, \{1, 3\}\} \), and \( \{\{3\}, \{1, 2\}\} \), the only symmetric \( p \) satisfying a consistency property similar to Theorem 1 are characterized by \( p(\langle N \rangle) + p(\{N\}) = 1 \) for \( Ba^O \) and by \( p(\langle N \rangle) = 1 \) for \( Ba^{AF} \).

So it seems to be an open question whether there are a non-trivial extensions of the Banzhaf value which are consistent with the Banzhaf value itself in the above sense.

**Remark 6.** Kamijo (2007) suggests the *two-step Shapley value*, \( Sh^{2\text{step}} \), an extension of the Shapley value and a relative of the Owen value, which satisfies \( E, A, CS, \)
and SC, but not N. The latter seems to be the reason why this concept is not consistent with Shapley value in the sense of (1) and Theorem 1. Consider the following example: \( N = \{1, 2, 3\} \), \( v(K) = 1 \) if \( \{1, 2\} \subseteq K \) and else \( v(K) = 0 \). We have \( \text{Sh}_3(N, v) = 0 \), but

\[
\begin{array}{c|cccc}
\mathcal{P} & \{1\}, \{2, 3\} & \{2\}, \{1, 3\} & \{3\}, \{1, 2\} & \{N\} \\
\text{Sh}_{3}^{2\text{step}}(N, v, \mathcal{P}) & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0
\end{array}
\]

By arguments as in Remark 5, the only symmetric \( p \) for which the above consistency property could be satisfied are characterized by \( p(\{N\}) + p(\{N\}) = 1 \).

References


