Membership Separability: A New Axiomatization of the Shapley Value*

Geoffroy de Clippel

August 2016

Abstract

The paper shows that Shapley's axiomatic characterization of his value can be strengthened considerably. Indeed, his additivity axiom can be replaced by a simple accounting property whereby a player's payoff is the difference of a reward based on the worth of coalitions to which she belongs, and a tax based on the worth of coalition to which she does not belong, without placing any restriction whatsoever on the functional relationship between the reward or the tax and the worths that determine them.

1 Introduction

The paper sheds new light on the Shapley value, one of the most succesfull solution concepts in cooperative game theory, and in particular how Shapley's original axiomatic result can be strengthened considerably.

His characterization result rests on three axioms. Both his anonymity and carrier axioms admit strong normative and positive interpretations. If two players play a same role in creating the surplus, then they will and should receive the same payoff (anonymity). If a group of players do not contribute to creating any surplus, then they will not or should not receive any payoff (carrier). Some have proposed justifications of the additivity axiom,[1] but it

*The author thanks two anonymous referees for their useful comments on the paper.

[1]Shapley thinks for instance of his value as determining an expected payoff in a game, and suggests addivity as a property of expected utility.

often comes across as a more technical requirement, in which case Shapley's result essentially amounts to investigating the carrier and anonimity axioms within the simpler class of linear[2] solutions.

We know that additivity does play a key role in Shapley's result. Schmeidler's (1969) nucleolus provides an example of a non-linear value that satisfies both the anonimity and carrier axioms. This note shows, however, that carrier and anonymity do characterize the Shapley value within a remarkably large class of non-linear values. In other words, Shapley's axiomatic result survives if one replaces his additivity axiom by a weaker axiom that does not rule out as many values. Shapley's core ideas are thus mathematically more robust than originally thought, as the anonimity and carrier axioms alone appear harder to meet than originally thought.

The new axiom that replaces additivity can be thought of as a simple accounting method. When considering a player, I propose to consider separately a reward she will receive based on the worth of coalitions to which she belongs, and a tax she will pay based on the worth of coalitions to which she does not belong. The axiom requires that her payoff be the reward minus the tax, without placing any restriction at all on the functional relationship between the reward or the tax and the worths that determine them.

There have been other attemps in the past to dispense with the additivity axiom. Young's (1985) marginality axiom postulates that a player's payoff depends only on her marginal contributions to the different coalitions to which she belongs. Thus it allows the value to be any function applied to inputs that are derived as differences of coalitional worths. By contrast, the axiom proposed here encompasses a different class of values where coalitional worths are not manipulated, but the impact of coalitional worths on the final payments can be separated based on membership. I will also show that the new axiom is weaker than Feltkamp's (1995) transfer axiom (first introduced by Dubey (1975) on the class of simple games). In addition to a few more axiomatic results in that vein, there are also other approaches to characterize the Shapley value without relying on a property of additivity, using for instance ideas of balanced contributions, reduced game properties, or following the Nash program. The interested reader is referred to Myerson (1980), Hart and Mas-Colell (1989), Chun (1989), van den Brink (2001), Hamiache (2001), Kamijo and Kongo (2010), and Casajus (2011, 2014). Fi-

[2]As is well-known, additivity is not exactly the same as linearity, but the distinction between the two properties is not relevant in the context of our discussion.

nally, Eisenman (1967) and Béal et al. (2012) show how the Shapley value can be computed as some average of 'compensations' where members of a coalition receive an equal share of its worth, and pay an equal share of the complement's worth. They thus provide instances where the Shapley value is presented explicitly via a tax-reward structure.

2 Reminder

The set of players is denoted by N. *Coalitions* are nonempty subsets of N. The set of all coalitions is denoted by $P(N)$. A characteristic function associates to every coalition a real number that represents the amount to be shared by its member should they cooperate. A *value* associates a payoff vector in $\mathbb{R}^N$ to every characteristic function. The Shapley value for instance is a weighted sum of the players's marginal contributions:

$$Sh_i(v) = \sum_{S \in P(N) | i \in S} \frac{(s-1)!(n-s)!}{n!}(v(S) - v(S \setminus \{i\})),$$

for each player i and each characteristic function v, where $v(\emptyset) = 0$, $s = \#S$, $n = \#N$.

Let v be a characteristic function and let x a payoff vector in $\mathbb{R}^N$. If π is a permutation of N, then $\pi(v)$ is the characteristic function defined as follows:

$$\pi(v)(S) = v(\pi(S)),$$

for each coalition S. Similarly, the vector $\pi(x)$ is defined as follows:

$$\pi(x)_i = x_{\pi(i)},$$

for each player i. A coalition S is said to be a *carrier* for a characteristic function v if $v(T) = v(S \cap T)$, for all coalitions T.

Shapley introduced the following axioms for a value σ.[3]

Anonymity (AN) *If π is a permutation of N, then $\sigma(\pi(v)) = \pi(\sigma(v))$, for each characteristic function v.*

Carrier (C) *If S is a carrier, then $\sum_{i \in S} \sigma_i(v) = v(S)$.*

[3]The carrier axiom is equivalent to the combination of the axioms of "efficiency" and "null player" found sometimes in presentations of Shapley's result.

3 Main Result

I start by stating a new axiom.

Difference Formula (DF) *For each $i \in N$, there exist a function*

$$r_i : \mathbb{R}^{\{S \in P(N) | i \in S\}} \to \mathbb{R}$$

and a function

$$t_i : \mathbb{R}^{\{S \in P(N) | i \notin S\}} \to \mathbb{R}$$

such that

$$\sigma_i(v) = r_i([v(S)]_{S|i \in S}) - t_i([v(S)]_{S|i \notin S}),$$

for each characteristic function v.[4]

We will see in the next section how permissive this axiom is in perspective of Shapley's additivity axiom and subsequent relaxations of it. As explained in the introduction, DF can be thought of as a rather intuitive accounting method. Each player i's payoff is seen as the difference of a reward and a tax, where the reward depends only on the worths of coalitions to which she belongs and the tax depends only on the worths of coalitions that do not contain i. It is thus a separability property based on membership. Beyond restricting the set of variables that can impact the functions r_i and t_i, their functional forms is left entirely unspecified.

Our main result goes as follows. Its proof, which is available in the appendix, differs quite a bit from Shapley's proof. Indeed, Shapley's argument essentially boils down to finding a natural basis of the (linear) space of characteristic functions over which the value is pinned down by AN and C, and using additivity to derive the value over the entire space of characteristic functions. Additivity thus plays a central role in his proof that has no equivalent when using DF instead.

Theorem *The Shapley value is the only value that satisfies AN, C, and DF.*

[4] One could require the functions r_i and t_i to take only nonnegative values, so as to interpretation of a reward and a tax. The characterization result holds even without requiring these functions to be positive.

4 Discussion

1. A characteristic function v is *superadditive* if $v(S) + v(T) \leq v(S \cup T)$, for all coalitions S and T. It is a *simple game* if it is superadditive, and the coalitions' worths are 0 or 1. I will argue at the end of the proof of the Theorem that the axiomatic characterization remains valid both on the class of simple games and on the class of non-negative superadditive characteristic functions (with coalitions' worths representing profits, for instance). At this time it remains an open question whether the result also holds on the class of all superadditive characteristic functions.

2. As explained before, Shapley's original axiomatization is based on the additivity axiom.

Additivity (ADD) $\sigma(v+w) = \sigma(v)+\sigma(w)$*, for each characteristic functions* v *and* w.

To see that ADD implies DF, simply consider for any individual i the decomposition of any characteristic function v as the sum of two functions v^i and v^{-i}, where v^i coincides with v for all coalitions that contain i, and attributes a zero worth to all other coalitions, while v^{-i} coincides with v for all coalitions that do not contain i and attributes a zero worth to all other coalitions. Additivity implies that $\sigma_i(v) = \sigma_i(v^i)+\sigma_i(v^{-i})$, which proves that DF holds.

This reasoning does not hold, though, when considering only simple games, because the function v^{-i} in the decomposition does not define a simple game. In fact, it is well-known that Shapley's axiomatic characterization does not hold on the class of simple games. For instance, the Banzhaf index satisfies Shapley's axioms on that smaller class of characteristic functions. By contrast, our characterization result with DF instead of ADD does apply on the class of simple games.

3. Dubey (1975) proposes the 'transfer axiom' to characterize the Shapley value on the class of simple games. Feltkamp (1995) showed that the result is also valid on the class of all characteristic functions. For any two characteristic functions v and w, let $v \vee w$ and $v \wedge w$ be the characteristic functions defined as follows:

$$(v \vee w)(S) = \max\{v(S), w(S)\},$$

$$(v \wedge w)(S) = \min\{v(S), w(S)\},$$

for each coalition S. The transfer axiom then reads as follows.

Transfer (T) $\sigma(v \vee w) + \sigma(v \wedge w) = \sigma(v) + \sigma(w)$, *for each characteristic functions v and w.*

To show that T implies DF both on the class of all characteristic functions and the class of simple games, consider for any given i the characteristic function v^i and v^{-i} defined as before. We have $v^i \vee v^{-i} = v$ and $v^i \wedge v^{-i} = 0$. Thus T implies that $\sigma_i(v) = \sigma_i(v^i) + \sigma_i(v^{-i}) - \sigma_i(0)$, which is equivalent to DF. On the other hand, there are of course many values that satisfy DF, but not T.

4. Following van den Brink and Gilles (1996), say that a player i is *necessary* in a characteristic function v if it belongs to all coalitions with a strictly positive worth: $v(S) = 0$ if $i \notin S$. As should be clear from the proof of the Theorem, it remains valid when AN is replaced by the weaker requirement that all necessary players get a same payoff.

5. The axioms appearing in the Theorem are independent. First, the nucleolus satisfies AN and C, but violates DF. Second, the equal split solution that gives $v(N)/n$ to each player satisfies DF and AN, but violates C. There are also many non-additive solutions with this feature. For each coalition S different from N, fix a function $f_S : \mathbb{R} \to \mathbb{R}$. For each characteristic function v, let then $\hat{v}$ be the characteristic function defined by $\hat{v}(N) = v(N)$ and $\hat{v}(S) = f_S(v(S))$ for each strict subset S of N. Then the solution that associates to each characteristic function v the Shapley value of the characteristic function $\hat{v}$ also satisfies DF and AN, but not C. Third, the weighted Shapley values satisfies DF and C, but not AN. There are also many non-additive solutions with this feature. Fix a function $g : \mathbb{R} \to \mathbb{R}$ such that $g(0) = 0$. Consider then the solution that coincides with the Shapley value for all $i \neq 1, 2$, that pays 1 his Shapley value plus the amount

$$M = [g(v(\{1,2\})) - g(v(\{1\})) - g(v(\{2\}))] - [v(\{1,2\}) - v(\{1\}) - v(\{2\})],$$

and that pays 2 his Shapley value minus M. It satisfies DF and C, but not AN.

Appendix: Proof of the Theorem

It is straightforward to check that the Shapley value satisfies the three axioms. It remains to show that it is the only value satisfying them.

For each charateristic function v, let $\alpha(v)$ be the number of non-necessary players in v, $\beta(v)$ be the number of coalitions that have a non-zero worth under v, and $s(v)$ be the sum of these two numbers.

Consider now a value σ that satisfies the three axioms, and yet differs from the Shapley value. Among all characteristic functions for which σ and Sh disagree, let v be one for which $s(v)$ is minimal. Let i be a player such that $\sigma_i(v) \neq Sh_i(v)$.

To derive a contradiction, suppose first that i is not necessary in v. Let v^i be the characteristic function defined as follows:

$$v^i(S) = \begin{cases} v(S) & \text{if } i \in S \\ 0 & \text{if } i \notin S, \end{cases}$$

for each coalition S. Observe that $s(v^i) < s(v)$ (since i is not necessary). We have:

$$\begin{aligned} r_i([v(S)]_{S|i \in S}) &= r_i([v^i(S)]_{S|i \in S}) \\ &= \sigma_i(v^i) + t_i(0) \\ &= Sh_i(v^i) + t_i(0) \end{aligned} \tag{1}$$

The first equality follows from the fact that v^i coincides with v for all coalitions containing i. The second equality follows from DF. The last equality follows from the fact that $s(v^i) < s(v)$, in which case σ and Sh coincide.

Let w^i and $\hat{w}^i$ be the two characteristic functions defined as follows:[5]

$$w^i(S) = \begin{cases} v(S \setminus \{i\}) & \text{if } i \in S \\ v(S) & \text{if } i \notin S, \end{cases}$$

and

$$\hat{w}^i(S) = \begin{cases} w^i(S) & \text{if } i \in S \\ 0 & \text{if } i \notin S, \end{cases}$$

for each coalition S. Observe that i is necessary in $\hat{w}^i$, and non-necessary players in $\hat{w}^i$ except i are non-necessary in v as well, implying that $s(\hat{w}^i) <$

[5]The characteristic function w^i is obtained from v after i's nullification to use the terminology of Béal et al. (2016).

$s(v)$. We have:

$$\begin{aligned}
t_i([v(S)]_{S|i\notin S}) &= t_i([w^i(S)]_{S|i\notin S}) \\
&= r_i([w^i(S)]_{S|i\in S}) \\
&= r_i([\hat{w}^i(S)]_{S|i\in S}) \\
&= \sigma_i(\hat{w}^i) + t_i(0) \\
&= Sh_i(\hat{w}^i) + t_i(0). \qquad (2)
\end{aligned}$$

The first equality follows from the fact that w^i coincides with v for all coalitions that do not contain i. The second equality follows from DF and C (player i is null in w^i). The third equality follows from the fact that $\hat{w}^i$ coincides with w^i for all coalitions containing i. The fourth equality follows from DF. The last equality follows from the fact that $s(\hat{w}^i) < s(v)$, in which case σ and Sh coincide.

Combining equations (1) and (2) with DF, we get $\sigma_i(v) = Sh_i(v^i) - Sh_i(\hat{w}^i)$, which in turn is equal to $Sh_i(v)$. Thus σ and the Shapley value coincide for all non-necessary players. Both being anonymous, $\sigma_i(v) = \sigma_j(v)$ and $Sh_i(v) = Sh_j(v)$, for every couple (i, j) of necessary players. This implies that $\sigma(v) = Sh(v)$, since σ and the Shapley value are efficient (by C).

Notice how v^i, w^i and $\hat{w}^i$ are superadditive whenever v is, define non-negative worths whenever v does, and define a worth in $\{0, 1\}$ whenever v does. Thus, the theorem remains valid, as claimed in first point of the discussion, when restricting attention to simple games, or non-negative superadditive characteristic function. ■

References

Béal, S., E. Rémila, and P. Solal, (2012): "Compensations in the Shapley Value and the Compensation Solutions for Graph Games," *International Journal of Game Theory* **41**, 157-178.

Béal, S., S. Ferrires, E. Rmila, and P. Solal, (2016): "Axiomatic Characterizations under Players Nullification," *Mathematical Social Sciences* **80**, 47-57.

Casajus, A., (2011): "Differential Marginality, van den Brink Fairness, and the Shapley Value," *Theory and Decision* **71**, 163-174.

Casajus, A., (2014): "The Shapley Value Without Efficiency and Additivity," *Mathematical Social Sciences* **68**, 1-4.

Chun, Y., (1989): "A New Axiomatization of the Shapley Value," *Games and Economic Behavior* **1**, 119-130.

Dubey, P., (1975): "On the Uniqueness of the Shapley Value," *International Journal of Game Theory* **4**, 131-139.

Eisenman, R. L., (1967): "A Profit-Sharing Interpretation of Shapley Value for n-Person Games," *Behavioral Sciences* **12**, 396-398.

Feltkamp, V., (1995): "Alternative Axiomatic Characterizations of the Shapley and Banzhaf Values," *International Journal of Game Theory* **24**, 179-186.

Hamiache, G., (2001): "Associated Consistency and Shapley Value," *International Journal of Game Theory* **30**, 279-289.

Hart, S., and A. Mas-Colell, (1989): "Potential, value, and consistency," *Econometrica* **57**, 589-614.

Kamijo, Y., and T. Kongo, (2010): "Axiomatization of the Shapley value using the balanced cycle contributions property," *International Journal of Game Theory* **39**, 563-571.

Myerson, R. B., (1980): "Conference Structures and Fair Allocation Rules," *International Journal of Game Theory* **9**, 169-182.

Schmeidler, D., (1969): "The Nucleolus of a Characteristic Function Game," *SIAM J. Appl. Math.* **17**, 1163-1170.

Shapley, L. S., (1953): "A Value for n-Person Games," in *Contributions to the Theory of Games II*, ed. by A. W. Tucker and R. D. Luce. Princeton, NJ: Princeton University Press, 307-317.

van den Brink, R., (2001): "An Axiomatization of the Shapley Value Using a Fairness Property," *International Journal of Game Theory* **30**, 309-319.

Young, H. P., (1985): "Monotonic Solutions of Cooperative Games," *International Journal of Game Theory* **14**, 65-72.