
The potential of the Shapley value

Sergiu Hart and Andreu Mas-Colell

1 Introduction

We study multiperson games in characteristic function form with transferable utility. The problem is to *solve* such a game (i.e., to associate to it payoffs to all the players).

Three main solution concepts are as follows. The first was introduced by von Neumann and Morgenstern: A “stable set” of a given game is a *set* of payoff *vectors*; such a set, if it exists, need *not* be *unique*. Next came the “core,” due to Shapley and Gillies, which is a *unique set* of payoff *vectors*. Finally, the Shapley “value” consists of just *one* payoff *vector*. There is thus an apparent historical trend from “complexity” to “simplicity” in the structure of the solution.^{1,2}

We propose now an even simpler construction: Associate to each game just *one number*! How would the payoffs to *all* players then be determined? by using the “marginal contribution” principle, an approach with a long tradition (especially in economics). Thus, we assign to each player his or her marginal contribution according to the numbers defined earlier. The surprising fact is that only one requirement, that the resulting payoff vector be “efficient” (i.e., that the payoffs add up to the worth of the grand coalition), determines this procedure uniquely.

Clearly it is not possible in general to assign to each player his or her direct marginal contribution to the grand coalition (according to the given characteristic function of the game). This is true simply because these marginal contributions need not add up to the worth of the grand coalition; namely, they will either be not feasible or, if feasible, not Pareto

Dedicated with great admiration to Lloyd S. Shapley on his 65th birthday. This chapter is based on the paper “Potential, Value and Consistency” (1987) and its previous versions (1985,1986). Financial support by the National Science Foundation and the U.S.-Israel Binational Science Foundation is gratefully acknowledged.

optimal (as usual, we refer to this “adding up” requirement as “efficiency”). The way these two principles—marginal contribution and efficiency—are reconciled is by introducing the function that associates to each game a real number, called the *potential* of the game, and computing marginal contributions according to it. The main result can now be stated.

Theorem A. There exists a unique³ real function on games, called the *potential function*, with respect to which the marginal contributions of all players are always efficient. Moreover, these marginal contributions are precisely the Shapley (1953) value.

Thus the Shapley value, viewed as a vector-valued function on games, is just the (discrete) gradient of the potential function (this explains our choice of name for it). The Shapley value has therefore been singled out as the unique efficient solution concept that admits a potential. More detailed discussions, together with additional results and interpretations, are the subject of Section 2.

The potential, although by its definition just an analytical tool, has nonetheless turned out to be most suggestive and productive. See Hart and Mas-Colell (1987) for details. In Section 3, we present one result obtained by the potential approach: a new way to characterize the Shapley value by an internal “consistency” property.

2 The potential

In this section we formalize the previous discussion and study various properties of the potential.

A cooperative game with sidepayments (or with transferable utility)—in short, a *game*—consists of a pair (N, v) , where N is a finite set of *players* and⁴ $v: 2^N \rightarrow \mathbf{R}$ is the *characteristic function*, satisfying $v(\emptyset) = 0$. A subset⁵ $S \subset N$ is called a *coalition*, and $v(S)$ is the *worth* of the coalition S . Given a game (N, v) and a coalition $S \subset N$, we write (S, v) for the *subgame* obtained by restricting v to (the subsets of) S ; that is, the domain of the function v is restricted to 2^S .

Let Γ denote the set of all games. Given a function $P: \Gamma \rightarrow \mathbf{R}$ that associates a real number⁶ $P(N, v)$ to every game (N, v) , the *marginal contribution* of a player i in a game (N, v) is defined to be

$$D^i P(N, v) = P(N, v) - P(N \setminus \{i\}, v),$$

where $i \in N$; recall that the game $(N \setminus \{i\}, v)$ is just the restriction of (N, v) to $N \setminus \{i\}$.

A function $P: \Gamma \rightarrow \mathbf{R}$ with⁷ $P(\emptyset, v) = 0$ is called a *potential function* if it satisfies the following condition:

$$\sum_{i \in N} D^i P(N, v) = v(N) \tag{1}$$

for all games (N, v) . Thus, a potential function is such that its marginals are always efficient; that is, they add up to the worth of the grand coalition. The main result is the next theorem.

Theorem A. There exists a unique potential function P . For every game (N, v) the resulting payoff vector $(D^i P(N, v))_{i \in N}$ of marginal contributions coincides with the Shapley value of the game. Moreover, the potential of a game (N, v) is uniquely determined by (1) applied only to the game and its subgames (i.e., to (S, v) for all $S \subset N$).

Proof: Rewrite (1) as⁸

$$P(N, v) = \frac{1}{|N|} \left[v(N) + \sum_{i \in N} P(N \setminus \{i\}, v) \right]. \tag{2}$$

Starting with $P(\emptyset, v) = 0$, (2) determines $P(N, v)$ recursively. This proves the existence and uniqueness of the potential function P and that $P(N, v)$ is uniquely determined by (1) (or (2)) applied just to (S, v) for all $S \subset N$.

It remains to show that $D^i P(N, v) = \text{Sh}^i(N, v)$ for all games (N, v) and all players $i \in N$, where P is the (unique) potential function and $\text{Sh}^i(N, v)$ denotes the Shapley value of player i in the game (N, v) . We prove that all the axioms that uniquely determine the Shapley value are satisfied by⁹ $D^i P$. Efficiency is just (1); the other three axioms—dummy (null) player, symmetry, and additivity—are proved inductively using (2). Indeed, let i be a null player in the game (N, v) (i.e., $v(S) = v(S \setminus \{i\})$ for all S). We claim that this implies $P(N, v) = P(N \setminus \{i\}, v)$; hence $D^i P(N, v) = 0$. Assume the assertion holds for all games with less than $|N|$ players; in particular, $P(N \setminus \{j\}, v) = P(N \setminus \{j, i\}, v)$ for all $j \neq i$. Now subtract (2) for $N \setminus \{i\}$ from (2) for N , to obtain

$$\begin{aligned} |N|[P(N, v) - P(N \setminus \{i\}, v)] &= [v(N) - v(N \setminus \{i\})] \\ &\quad + \sum_{j \neq i} [P(N \setminus \{j\}, v) - P(N \setminus \{j, i\}, v)] \\ &= 0. \end{aligned}$$

Next, assume players i and j are substitutes in the game (N, v) . This implies that $P(N \setminus \{i\}, v) = P(N \setminus \{j\}, v)$ (use (2), noting that i and j are substitutes in $(N \setminus \{k\}, v)$ for all $k \neq i, j$); thus $D^i P(N, v) = D^j P(N, v)$. Finally, another in-

ductive argument on (2) shows that $P(N, v + w) = P(N, v) + P(N, w)$, implying additivity. ■

Remark 1: The potential approach may be viewed as a *new axiomatic characterization* for the Shapley value. Its significance is twofold: First, *only one* axiom, (1), is needed (though one may view it as the combination of two postulates: “efficiency” and “marginal contributions”; note, however, that no additivity, symmetry, and so on, are assumed). Second, one needs to consider *the given game only*; the potential and a fortiori the Shapley values are uniquely determined by (1) applied just to the game and its subgames (thus, only one characteristic function is taken into account). This is particularly important in applications, where typically just one specific problem is considered. In contrast, all the standard axiomatizations of the Shapley value require, in order to uniquely determine it for any single game, the application of the various axioms (additivity, symmetry, etc.) to a large domain (e.g., all games or all simple games).¹⁰

Remark 2: Formula (2) yields a simple and straightforward recursive procedure for the computation of the potential and of the Shapley values of the game as well as all its subgames. This seems to be a most efficient algorithm for computing Shapley values (note that (2) has to be applied just once for each one of the $2^{|M|} - 1$ nonempty coalitions).

We now present another way of viewing the potential. Given a game (N, v) , the allocation of marginal contributions (i.e., $v(N) - v(N \setminus \{i\})$ to player i) is, in general, not efficient. One way to resolve this difficulty is to add a new player, say player 0, and extend the game to $N_0 = N \cup \{0\}$ in such a way that the allocation of marginal contributions in the extended game becomes efficient. Formally, let (N_0, v_0) be an extension of (N, v) (i.e., $v_0(S) = v(S)$ for all $S \subset N$). Then the requirement is

$$\begin{aligned} v_0(N_0) &= \sum_{i \in N_0} [v_0(N_0) - v_0(N_0 \setminus \{i\})] \\ &= [v_0(N_0) - v(N)] + \sum_{i \in N} [v_0(N_0) - v_0(N_0 \setminus \{i\})]. \end{aligned} \quad (3)$$

This reduces to

$$v(N) = \sum_{i \in N} [v_0(N_0) - v_0(N_0 \setminus \{i\})], \quad (4)$$

which, when compared to (1), yields the following restatement of the result of Theorem A.

Corollary 1. There exists a unique extension v_0 of v whose marginal contributions to the grand coalition are always efficient (more precisely, (3) is satisfied for the game and all its subgames); it is given by $v_0(S \cup \{0\}) = P(S, v)$ for all $S \subset N$, where P is the potential function.

Note that the payoffs to the original players (in N) add up correctly to $v(N)$ (see (4); these are the Shapley values). Player 0, whose payoff is the residual $P(N, v) - v(N)$, may be regarded as a “hidden player,” similarly to the “hidden factor” introduced by McKenzie¹¹ in the study of production functions in order to explain the residual profit (or loss).¹²

In (1) and (2) the potential is only given implicitly. We now present two explicit formulas. The T -unanimity game u_T (where T is a nonempty finite set) is defined by $u_T(S) = 1$ if $S \supset T$, and $u_T(S) = 0$ otherwise. It is well known that these games form a linear basis for Γ : Each game (N, v) has a unique representation (e.g., see Shapley 1953)

$$v = \sum_{T \subset N} \alpha_T u_T,$$

where, for all $T \subset N$,

$$\alpha_T \equiv \alpha_T(N, v) = \sum_{S \subset T} (-1)^{|T|-|S|} v(S). \tag{5}$$

Proposition 1. The potential function P satisfies

$$P(N, v) = \sum_{T \subset N} \frac{1}{|T|} \alpha_T$$

for all games (N, v) , where α_T is given by (5).

Proof: Let $Q(N, v)$ denote the right-hand side in the preceding formula. Then $Q(\emptyset, v) = 0$ and $Q(N, v) - Q(N \setminus \{i\}, v) = \sum_{T \ni i} \alpha_T / |T|$, which when summed up over i shows that Q satisfies (1). Therefore, by Theorem A, Q coincides with the unique potential function P . ■

The number $\delta_T = \alpha_T / |T|$ is called the *dividend* of each member of the coalition T and $\text{Sh}^i(N, v) = \sum_{T \ni i} \delta_T$ (cf. Harsanyi 1963).

Proposition 2. The potential function P satisfies

$$P(N, v) = \sum_{S \subset N} \frac{(s-1)!(n-s)!}{n!} v(S),$$

where $n = |N|$ and $s = |S|$.

Proof: The marginal contributions of the function on the right side are easily seen to yield the Shapley value. ■

To interpret this last formula, consider the following probabilistic model of choosing a random nonempty coalition $S \subset N$: First, choose a size $s = 1, 2, \dots, n = |N|$ uniformly (i.e., with probability $1/n$ each). Second, choose a subset S of size s , again uniformly (i.e., each of the $\binom{n}{s}$ subsets has the same probability). Equivalently, choose a random order of the n elements of N (with probability $1/n!$ each), choose a cutting point s ($1 \leq s \leq n$), and let S be the first s elements in that order. The probability of choosing of a set S with $|S| = s$ is

$$\pi_s = \frac{s!(n-s)!}{n \cdot n!} = \frac{s}{n} \frac{(s-1)!(n-s)!}{n!}.$$

Therefore the formula of Proposition 2 may be rewritten as

$$P(N,v) = \sum_{S \subset N} \pi_s \frac{n}{s} v(S) = E \left[\frac{|N|}{|S|} v(S) \right], \tag{6}$$

where E denotes expectation over S with respect to the foregoing probability model. The interpretation of (6) is that the potential is the *expected normalized worth*, or, equivalently, the *per capita potential* $P(N,v)/|N|$ equals the *average per capita worth* $v(S)/|S|$. This shows that the potential may be viewed as an appropriate “summary” of the characteristic function into one number (from which marginal contributions are then computed).¹³

To study some further properties of the potential function, we regard it as an operator on games. Fix N and let Γ_N be the set of all games with player set N . Let \mathbf{P} be the operator from Γ_N into itself that associates to each game v another game $\mathbf{P}v$ given by $(\mathbf{P}v)(S) = P(S,v)$ for all $S \subset N$.

Proposition 3. The operator $\mathbf{P}: \Gamma_N \rightarrow \Gamma_N$ has the following properties (for all $v, w \in \Gamma_N$ and all scalars α, β):

- (i) \mathbf{P} is *linear*: $\mathbf{P}(\alpha v + \beta w) = \alpha \mathbf{P}v + \beta \mathbf{P}w$.
- (ii) \mathbf{P} is *symmetric*: $\mathbf{P}(\theta v) = \theta(\mathbf{P}v)$ for every one-to-one mapping θ of N into itself (i.e., a permutation of the players; for a game w the “permuted” game θw is defined by $(\theta w)(S) = w(\theta S)$ for all $S \subset N$).
- (iii) \mathbf{P} is *positive*: $v \geq 0$ implies $\mathbf{P}v \geq 0$ (where $w \geq 0$ means $w(S) \geq 0$ for all $S \subset N$).
- (iv) \mathbf{P} is *one-to-one* and *onto*.

- (v) The *fixed points* of \mathbf{P} are the inessential games (or additive games – games v such that $v(S) = \sum_{i \in S} v(\{i\})$ for all $S \subset N$).

Proof: Proposition 1 implies that if we decompose v as $v = \sum \alpha_T u_T$, then $\mathbf{P}v = \sum (\alpha_T / |T|) u_T$ (both sums are over $T \subset N$ and α_T is given by (5)). From this (i), (ii), (iv), and (v) follow easily, and (iii) is implied by Proposition 2. ■

From these basic properties additional ones may be derived. For example:

Corollary 2. If the core of (N, v) is not empty (i.e., (N, v) is balanced), then $P(N, v) \leq v(N)$. If (N, v) is a market game (i.e., totally balanced), then $\mathbf{P}v \leq v$.

Proof: Let $x = (x^i)_{i \in N}$ be a payoff vector in the core of (N, v) and consider the inessential game (N, w) given by $w(S) = \sum_{i \in S} x^i$ for all $S \subset N$. Then $v \leq w$, implying $\mathbf{P}v \leq \mathbf{P}w = w$ (apply (i), (iii), and (iv) of Proposition 3); hence $P(N, v) = (\mathbf{P}v)(N) \leq w(N) = v(N)$. In a market game this argument applies to all subgames as well. ■

3 Consistency

This section presents one of the results obtained by the potential approach. It shows that the Shapley value enjoys an internal consistency property, similarly to most solution concepts (see Hart and Mas-Colell 1987, sec. 4 for references).

The “consistency” requirement may be described informally as follows: Let ϕ be a “solution function” that associates a payoff to every player in every game. For any group of players in a game, one defines a “reduced game” among them by considering the amounts remaining after the rest of the players are given the payoffs prescribed by ϕ . Then ϕ is said to be *consistent* if, when it is applied to any reduced game, it always yields the same payoffs as in the original game.

Formally, a *solution function* ϕ is a function defined on Γ , the set of all games, that associates to every $(N, v) \in \Gamma$ a payoff vector $\phi(N, v) = (\phi^i(N, v))_{i \in N} \in \mathbf{R}^N$. Given a solution function ϕ , a game (N, v) , and a coalition $T \subset N$, the *reduced game* (T, v^ϕ) is defined by

$$v^\phi(S) = v(S \cup T^c) - \sum_{i \in T^c} \phi^i(S \cup T^c, v) \tag{7}$$

for all $S \subset T$, where $T^c = N \setminus T$. The solution function ϕ is *consistent* if

$$\phi^j(T, v \upharpoonright_T) = \phi^j(N, v) \quad (8)$$

for every game (N, v) , every coalition $T \subset N$, and all $j \in T$.

These definitions may be interpreted as follows. Fix ϕ , (N, v) , and $T \subset N$. The members of T —more precisely, every (sub)coalition $S \subset T$ —need to consider the total payoff remaining after the players in T^c are paid according to ϕ . To compute the worth of S (in this reduced game), one assumes that the complementary coalition $T \setminus S$ is not present. Therefore, the game to be considered is $(S \cup T^c, v)$, in which the payoffs are distributed according to ϕ . The amount that remains for S is then precisely that given by (7). Finally, note that, if ϕ is efficient, then

$$v_T^\phi(S) = \sum_{i \in S} \phi^i(S \cup T^c, v). \quad (9)$$

There are alternative definitions of consistency (used for various solution concepts). They differ only in the definition of the reduced game. The appropriateness of definition (7) here depends on the specific situation being modeled, particularly on the concrete assumptions underlying the determination of the characteristic function.

One example where (7) appears to be the natural definition is the problem of allocating joint costs among various projects (or departments, tasks, etc.); these projects are now the “players.” The cost imputations are *not* to be interpreted as some kind of “efficiency prices” that are used to make optimal decisions on which projects to undertake, but as an equitable way to distribute exactly the total costs once the set of projects is fixed.

Such an instance arises in multistate corporations that for tax purposes have to allocate the joint costs (and benefits) among the projects in the various states in which they operate. As an example of such a corporation and, say, its Tennessee division, let T be the set of projects in Tennessee. For every subset $S \subset T$ of Tennessee’s projects, the “local accountant” has to determine its cost, assuming that it was the only subset of projects to be undertaken in Tennessee. In addition, there is the set T^c of all the projects outside Tennessee (which are not in the domain of this local “gedanken experiment”). Therefore, the cost of S is the amount imputed to it by the “accounting procedure” (= solution function) under consideration when the set of projects to be implemented is $S \cup T^c$. This is exactly given by formula (9). Consistency requires that for T the imputation obtained by the local accountant be no different than that of the general (national) accountant.

It turns out that this consistency requirement is the one satisfied by the Shapley value. Moreover, together with the appropriate initial conditions for two-person games, consistency uniquely characterizes the Shapley value.

A solution function ϕ is *standard for two-person games* if

$$\phi^i(\{i,j\},v) = v(\{i\}) + \frac{1}{2}[v(\{i,j\}) - v(\{i\}) - v(\{j\})]$$

for all v and all $i \neq j$; thus, in a two-person game each player first gets his own guaranteed payoff and then the remaining “surplus” is divided equally. Most solutions satisfy this condition. The main result can now be stated.

Theorem B. Let ϕ be a solution function. Then ϕ is (i) consistent and (ii) standard for two-person games, if and only if ϕ is the Shapley value.

Proof: First, we show that the Shapley value is a consistent solution function. Let (N,v) be a game and $T \subset N$ a nonempty coalition. The reduced game $v_T = v_T^\phi$ (where ϕ is the Shapley value) is given by (see (9) and Theorem A)

$$v_T(S) = \sum_{i \in S} [P(S \cup T^c, v) - P(S \cup T^c \setminus \{i\}, v)] \tag{10}$$

for every $S \subset T$. By Theorem A, the potential of the game (T, v_T) is uniquely determined by formula (1) applied to the game and all its subgames. Comparing this with (10) implies that

$$P(S, v_T) = P(S \cup T^c, v) + c$$

for all $S \subset T$, where c is an appropriate constant (so as to make $P(\emptyset, v_T) = 0$). Therefore

$$\begin{aligned} \text{Sh}^i(T, v_T) &= P(T, v_T) - P(T \setminus \{i\}, v_T) \\ &= P(N, v) - P(N \setminus \{i\}, v) = \text{Sh}^i(N, v) \end{aligned}$$

for every $i \in T$, proving that the Shapley values of the players in the reduced game and in the original game coincide.

For the converse, one may show that if ϕ satisfies (i) and (ii), then ϕ must admit a potential function (see the proof of Theorem B in Hart and Mas-Colell 1987).¹⁴ We provide here an alternative direct proof (see Lemma 6.8 in Hart and Mas-Colell 1987).

First, we show that (i) and (ii) imply that ϕ is efficient; that is,

$$\sum_{i \in N} \phi^i(N, v) = v(N)$$

for all games (N, v) . This holds for $|N| = 2$, by (ii), and for $|N| = 1$ (add a dummy player and apply (ii) and (i) to the resulting two-person game). Now consider a game (N, v) with $|N| \geq 3$. Let $T \subset N$ be a two-person coalition; by consistency

$$\sum_{j \in N} \phi^j(N, v) = \sum_{j \in T} \phi^j(T, v_T^\phi) + \sum_{i \in T^c} \phi^i(N, v).$$

Because $|T| = 2$, the first sum equals $v_T^\phi(T)$; the definition (7) of the reduced game now implies that the right side is $v(N)$.

Second, assume ϕ and ψ are two solution functions satisfying (i) and (ii), and assume by induction that they coincide for all games with less than n players (this is true for $n = 3$). Let (N, v) be an n -person game, and let $i, j \in N, i \neq j$. Consider the two reduced games $(\{i, j\}, v_{\{i, j\}}^\phi)$ and $(\{i, j\}, v_{\{i, j\}}^\psi)$, which we denote by v^ϕ and v^ψ , respectively. They coincide for singletons (by induction, because only $n - 1$ players matter); therefore, by (ii), $\phi^i(v^\phi) \cong \phi^i(v^\psi)$ if and only if $\phi^j(v^\phi) \cong \phi^j(v^\psi)$ (if and only if $v^\phi(\{i, j\}) \cong v^\psi(\{i, j\})$). Now $\phi = \psi$ for two-person games, and both ϕ and ψ are consistent; therefore

$$\phi^i(N, v) = \phi^i(v^\phi) \cong \phi^i(v^\psi) = \psi^i(v^\psi) = \psi^i(N, v)$$

if and only if, similarly,

$$\phi^j(N, v) \cong \psi^j(N, v).$$

This applies, however, to any two players i and j ; because both ϕ and ψ are efficient, we must therefore have $\phi^i(N, v) = \psi^i(N, v)$ for all i . ■

See Hart and Mas-Colell (1987) for additional results regarding consistency and the Shapley value (and its extensions, the weighted Shapley values and the nontransferable utility case).

NOTES

- 1 Other solution concepts (e.g., bargaining set, kernel, nucleolus, etc.) may also fit this description.
- 2 A simpler solution may be easier to study and apply, which increases its usefulness. However, it is important to look at different solution concepts, based on different postulates, because they illuminate the problem from different angles.
- 3 Up to an additive constant (which does not change the marginal contributions).
- 4 \mathbf{R} denotes the set of real numbers, \emptyset is the empty set, and the symbol \setminus is used for set subtraction.

- 5 All subset inclusions should be understood in the weak sense (i.e., equality is possible).
- 6 We write $P(N,v)$ rather than $P((N,v))$.
- 7 There is only one "empty game" (\emptyset,v) , because $v(\emptyset) = 0$ always.
- 8 $|A|$ denotes the number of elements of the finite set A .
- 9 For another proof see Hart and Mas-Colell (1987).
- 10 Note that the explicit formula for the Shapley value (= average marginal contribution) also applies to just one game. However, this seems too complex to be viewed as a "basic postulate" (in particular, one has to justify the specific probabilistic model).
- 11 McKenzie, L. (1959), "On the Existence of General Equilibrium for a Competitive Market," *Econometrica* 27, 54–71.
- 12 A similar construction is embodied in this story: A sheik died, leaving a will directing his three sons to divide the property as follows: one-half goes to the oldest son, one-third to the middle son, and one-ninth to the youngest. The property consisted of seventeen camels. The problem was solved by the "wise man" who rode into town on his camel. He added his own camel to the seventeen, and then the three brothers took their shares out of the eighteen camels: nine, six, and two, respectively. One camel was left, on which the wise man rode out of town.
- 13 Formula (6) should hardly be surprising. The Shapley value is the expected marginal contribution; we obtain it here as the marginal contribution to the appropriate expectation—the potential.
- 14 Another proof has been independently obtained by Michael Maschler.

REFERENCES

- Harsanyi, J. C. [1963], "A Simplified Bargaining Model for the n -Person Cooperative Game," *International Economic Review* 4, 194–220.
- Hart, S. and A. Mas-Colell [1987], "Potential, Value and Consistency," forthcoming in *Econometrica*.
- Shapley, L. S. [1953], "A Value for n -Person Games," *Contributions to the Theory of Games*, vol. II (*Annals of Mathematics Studies* 28), H. W. Kuhn and A. W. Tucker (eds.), Princeton University Press, Princeton, pp. 307–17.