# Chapter VIII

# The Harsanyi Value

SERGIU HART[1]

## 1 Introduction

We study multiperson games in coalitional (or, characteristic) function form. The problem we address here is that of developing general principles for solving such a game.

Consider first transferable utility (TU)-games. An approach with a long tradition in economics would proceed by assigning to every player his direct marginal contribution to the grand coalition (*i.e.,* the set of all players). This is not possible in general since 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 optimal. Nonetheless, this is the spirit of our approach: we associate to each game a single number — called the potential of the game — and then assign to each player his marginal contribution computed according to these numbers. The surprising fact is: the requirement that a feasible and efficient allocation should always be obtained determines the procedure uniquely. Moreover, the resulting solution is well-known: it is the Shapley [1953b] value.

The potential, although presented here just as a technical tool[2], has turned out to be most productive. In particular, the potential approach has suggested further ways to characterize the Shapley value. As an example, the Shapley value is characterized by an internal consistency property: eliminating some of the players, after paying them according to the solution, does not change the outcome for the remaining ones. Another approach, also suggested by the potential, is via a "preservation of differences" postulate.

In section 3, we extend the potential approach to the general case, where utility need not be additively transferable. Clearly, the computation of the marginal contributions according to the potential leads to interpersonal comparison of utilities, since all players use the same real-valued potential.

This suggests the following construction: fix first a vector $w$ of positive weights, and use the potential function approach to get a solution $x_w$. Second, require that $w$ represent the appropriate marginal rates of efficient substitution between the players' payoffs *at* $x_w$. This is a standard procedure for obtaining solutions in the nontransferable utility case. One first

[1]Student notes, taken during the author's lecture, and partially revised by the author.
[2]For an interpretation, see Hart and Mas-Colell [1992].



*J.-F. Mertens and S. Sorin (eds.), Game-Theoretic Methods in General Equilibrium Analysis,* 105–111.
© 1994 *Kluwer Academic Publishers. Printed in the Netherlands.*

assumes that the utility scales of the players are comparable (according to the weights $w$) and then requires that these are indeed the "right" weights at the resulting solution. This makes the final solution correspond to a fixed-point (of the mapping $w \mapsto x_w \mapsto w'$) and, most important, independent of rescaling utilities (for each player separately).

It turns out that this procedure leads to the Harsanyi value.

# 2 The TU case

## 2.1 The potential

A (cooperative) *game* (with transferable utility) is a pair $(N, v)$, where $N$ is a finite set of *players* and $v : 2^N \to \mathbb{R}$ is a *coalitional function* satisfying $v(\emptyset) = 0$. We refer to a subset $S$ of $N$ as a *coalition*, and to $v(S)$ as the *worth* of $S$. Given a game $(N, v)$ and a coalition $S$, we write $(S, v)$ for the subgame obtained by restricting $v$ to subsets of $S$ only (*i.e.*, to $2^S$).

Let $G$ be the set of all games. Given a function $P : G \to \mathbb{R}$ which associates the real number $P(N, v)$ to every game $(N, v)$, the marginal contribution of a player $i$ in a game $(N, v)$ is defined to be

$$D^i P(N, v) := P(N, v) - P(N \setminus \{i\}, v). \tag{1}$$

A function $P : G \to \mathbb{R}$ with $P(\phi, v) = 0$ is called a *potential function* if it satisfies the following condition:

$$\sum_{i \in N} D^i P(N, v) = v(N) \tag{2}$$

for all games $(N, v)$. Thus, a potential function is such that the allocation of marginal contributions (according to the potential function) always adds up exactly to the worth of the grand coalition.

**Theorem 1** *(Hart–Mas-Colell [1989], [1988]). There exists a unique potential function $P$. For every game $(N, v)$, the resulting payoff vector $(D^i P(N, v))_{i \in N}$ coincides with the Shapley value of the game. Moreover, the potential of any game $(N, v)$ is uniquely determined by (2) applied only to the game and its subgames (i.e. to $(S, v)$ for all $S \subset N$).*

**Proof:** Rewrite (2) as

$$P(N, v) = \frac{1}{|N|}\left[v(N) + \sum_{i \in N} P(N \setminus \{i\}, v)\right]. \tag{3}$$

Starting with $P(\emptyset, v) = 0$, (3) determines $P(N, v)$ recursively. This proves the existence and uniqueness of the potential function $P$, and also that $P(N, v)$ is uniquely determined by (2), applied to all subgames of $(N, v)$.

It remains to show that $D^i P(N, v) = Sh^i(N, v)$, where $(Sh^i(N, v))_{i \in N}$ stands for the Shapley value of the game $(N, v)$. This may be proved by using an axiomatic approach: one may show, inductively, that the payoff vector $(D^i P(N, v))_{i \in N}$ satisfies all the axioms that uniquely characterize the Shapley value: efficiency, dummy (null) player, symmetry and additivity.

Another possibility is to prove that the Shapley value derives from a potential function. The result then follows from the uniqueness of the potential function.

## 2.2 Consistency

This section is devoted to another characterization of the Shapley value by means of an (internal) consistency property. This is a standard approach that has been successfully applied to many solution concepts (for a survey, see Thomson [1990]).

The consistency requirement may be described as follows: let $\phi$ be a 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 giving to the other players their payoffs according to $\phi$. Then $\phi$ is said to be consistent if, when it is applied to any reduced game, it yields the same payoffs as in the original game. Note that one gets different requirements by modifying the definition of the reduced game.

Formally, let $\phi$ be a function defined on the set of games, *i.e.*, $\phi(N,v) \in \mathbb{R}^N$ for all $(N,v)$. Such a function is called a solution function. Let $(N,v)$ be a game, and $T$ a subset of $N$. The reduced game is defined as follows:

$$v_T^{\phi}(S) := v(S \cup T^c) - \sum_{i \in T^c} \phi^i(S \cup T^c, v), \text{ for all } S \subset T,$$

where $T^c := N \setminus T$. The function $\phi$ is *consistent* if, for every game $(N,v)$ and every coalition $T \subset N$, one has

$$\phi^j(T, v_T^{\phi}) = \phi^j(N,v), \text{ for all } j \in T.$$

The interpretation is as follows. Given $\phi$, a game $(N,v)$ and a coalition $T$, every subcoalition of $T$ needs to consider the total payoff remaining after paying the members of $T^c$ according to $\phi$. To compute the worth of a coalition $S$, we assume that the members of $T \setminus S$ are not present; in other words, one considers the game $(S \cup T^c, v)$.

The property of consistency is essentially equivalent to the existence of a potential function.

**Theorem 2** *(Hart–Mas-Colell [1989], [1988]) Let $\phi$ be a solution function. Then $\phi$ is consistent and $\phi$ is standard*[3] *for two-person games if and only if $\phi$ is the Shapley value for all games.*

# 3 The NTU case

A *nontransferable-utility game* —an NTU game, for short— is a pair $(N,V)$, with $V(S)$ a subset of $\mathbb{R}^S$ for all coalitions $S$ of $N$. The interpretation is that $x = (x^i)_{i \in S} \in V(S)$ if and only if there is an outcome attainable by the coalition $S$, whose utility to each member $i$ of $S$ is $x^i$. A TU game $(N,v)$ in $G$ corresponds to the NTU game $(N,V)$, where

$$V(S) = \{x \in \mathbb{R}^S : \sum_{i \in S} x^i \leq v(S)\}.$$

We make the following (standard) assumptions: all sets $V(S)$ are nonempty, not the whole space $\mathbb{R}^S$, convex, closed and comprehensive. We assume furthermore that *bd* $V(S)$, the Pareto-efficient boundary of $V(S)$, is smooth and non-level. We denote by $\Gamma$ the set of games $(N,V)$.

[3]That is, it "divides the surplus equally": $\phi^i(\{1,2\},v) = v(\{i\}) + \frac{1}{2}(v(\{1,2\}) - v(\{1\}) - v(\{2\}))$ for $i = 1,2$.

## 3.1 Axiomatizations

We briefly review the axiomatizations of the Shapley [1969] and the Harsanyi [1963] NTU solution concepts.

We will now distinguish between between "solution" and "value"; the former yields a *payoff configuration*, that is a collection $x = (x_S)_{S \subset N}$ with $x_S \in V(S)$ for all $S$, and the latter a payoff vector for $N$ (i.e., $x_N \in V(N)$).

### 3.1.1 The Shapley NTU-value

The axiomatization of the Shapley NTU-value is due to Aumann [1985].

First, we say that a game $(N, V)$ is monotone if, for each coalition $S$, there is a payoff vector $x$ such that

$$V(S) \times \{0^{N \setminus S}\} \subset V(N) + x$$

(this is a very weak kind of monotonicity).

**Definition 1** *A payoff configuration $x$ is a* Shapley NTU-solution *of a game $(N, V)$ if there exists a vector $\lambda \in \mathbb{R}^N$, such that:*

$$x_S \in \partial V(S) \text{ for all } S \subset N; \tag{4}$$

$$\lambda^S \cdot x_S \geq \lambda^S \cdot y \text{ for all } y \in V(S) \text{ and all } S \subset N; \tag{5}$$

$$\text{for all } i \in N, \lambda^i x^i_N = Sh^i(N, v), \text{ where } v(S) := \lambda^S \cdot x_S \text{ for all } S \subset N. \tag{6}$$

If $x$ is a Shapley NTU-solution of $(N, V)$, then its $N$-coordinate $x_N$ is a Shapley NTU-*value* of $(N, V)$.

### 3.1.2 Axioms

Let $\phi$ denote a value function.

(A0)-Nonemptiness: $\phi(V) \neq \emptyset$.

(A1)-Efficiency: $\phi(V) \subset \partial V(N)$.

(A2)-Scale Covariance: $\phi(\lambda V) = \lambda \phi(V)$, for all $\lambda \in \mathbb{R}^N_{++}$.

(A3)-Conditional additivity: if $U = V + W$, then

$$\phi(U) \supset [\phi(V) + \phi(W)] \cap \partial U(N).$$

(A4)-Independance of Irrelevant Alternatives: if $V(N) \subset W(N)$ and $V(S) = W(S)$ for all $S \neq N$, then

$$\phi(V) \supset \phi(W) \cap V(N).$$

(A5)-Unanimity Games: for every non-empty coalition $T$,

$$\phi(U_T) = \left\{ \frac{1_T}{|T|} \right\},$$

where $U_T$ is the $T$-unanimity game.

Let $\Gamma_L$ be the subset of games in $\Gamma$ that are monotone and have a Shapley NTU-value.

**Theorem 3** *(Aumann [1985]) There exists a unique value function on $\Gamma_L$ that satisfies Axioms A0–A5: it is the Shapley NTU-value function.*

### 3.1.3 The Harsanyi NTU-solution

**Definition 2** *A payoff configuration $x$ is a* Harsanyi NTU-solution *of a game $(N, V)$ if there exists a vector $\lambda \in \mathbb{R}^N$ and real numbers $\xi_T$ for all $T \subset N$, such that:*

$$\text{for each } S \subset N,\ x_S \in \partial V(S); \tag{7}$$

$$\lambda \cdot x_N \geq \lambda \cdot y \text{ for all } y \in V(N); \tag{8}$$

$$\text{for each } S \subset N \text{ and each } i \in S,\ \lambda^i x_S^i = \sum_{T \subset S, i \in T} \xi_T. \tag{9}$$

### 3.1.4 Axioms

Let $\Gamma_H$ be the set of games that have at least one Harsanyi solution, and let $\psi$ be a solution function, that is a set-valued function that assigns to each game $(N, V)$ a set of *payoff configurations.*

(B0)-Nonemptiness: $\psi(V) \neq \emptyset$.

(B1)-Efficiency: $\psi(V) \subset \partial V$: every solution $x \in \psi(V)$ satisfies Pareto efficiency for all coalitions $S$.

(B2)-Scale Covariance: $\psi(\lambda V) = \lambda \psi(V)$, for all $\lambda \in \mathbb{R}^N_{++}$.

(B3)-Conditional additivity: if $U = V + W$, then

$$\psi(U) \supset [\psi(V) + \psi(W)] \cap \partial U.$$

(B4)-Independance of Irrelevant Alternatives: if $V \subset W$, then

$$\psi(V) \supset \psi(W) \cap V.$$

(B5)-Unanimity Games: for every non-empty coalition $T$,

$$\psi(U_T) = \left\{ \frac{1_T}{|T|} \right\}.$$

**Theorem 4** *(Hart [1985]) There exists a unique solution function on $\Gamma_H$ that satisfies axioms B0-B5: it is the Harsanyi NTU-solution function.*

Note that the two axiom systems are essentially the same. The difference is that 3.1.2 applies to payoff vectors (for $N$), whereas 3.1.4 applies to payoff configurations (for all $S$). Further discussion of this can be found in Hart [1985].

### 3.2 The potential

As suggested in the introduction, let $w = (w^i)_{i\in N} \in \mathbb{R}^N_{++}$ be a collection of positive weights. The $w$−potential function $P_w$ associates with every NTU game $(N, V)$ a real number $P_w(N, V)$ such that

$$(w^i D^i P_w(N, V))_{i\in N} \in bd\ V(N). \tag{10}$$

Without loss of generality, let again $P_w(\emptyset, V) = 0$. Thus, (10) is the exact counterpart, in the NTU case, of (2) in the TU case: the vector of (rescaled) marginal contributions is efficient.

**Theorem 5** *(Hart–Mas-Colell [1989]) For every collection $w = (w^i)_{i\in N}$ of positive weights there exists a unique $w$-potential function on the class of NTU games.*

**Proof:** The assumptions we made above imply that, for each $S$, the set $V(S)$ is bounded from above in any strictly positive direction, hence $bd\ V(S)$ intersects any such line in a unique point. The proof proceeds by induction. Consider a game $(N, V)$ and assume the potential has been defined on all the subgames of $(N, V)$. Define, for $i \in N$, $y^i := -w^i P_w(N \setminus \{i\}, V)$ and let $y = (y^i)_{i\in N}$. $P_w(N, V)$ is then the unique $t$ (not necessarily positive) such that

$$y + tw \in bd\ V(N).$$

**Theorem 6** *(Hart–Mas-Colell [1989]) The solution function, resulting from the potential approach, that associates the payoff vector*

$$(w^i D^i P_w(N, V))_{i\in N}$$

*to the NTU game $(N, V)$, coincides with the $w$-egalitarian solution.*

The *$w$-egalitarian* solution has been introduced by Shapley [1953a]; see Kalai–Samet [1985] for an extensive study.

A (Pareto) efficient payoff vector $x \in bd\ V(N)$ is called *$w$-utilitarian* if it maximizes the sum of the utilities over the feasible set $V(N)$, rescaled according to $w$:

$$\sum_{i\in N} \frac{1}{w^i} x^i \geq \sum_{i\in N} \frac{1}{w^i} y^i, \text{ for all } y \in V(N).$$

Finally, $x$ is a *Harsanyi NTU-value* if there exist weights $w$ such that $x$ is simultaneously $w$-egalitarian and $w$-utilitarian. This is essentially the original definition of Harsanyi [1963].

From this we get the following characterization.

**Theorem 7** *(Hart–Mas-Colell [1989]) For every NTU game $(N, V)$, the payoff vector $x \in \mathbb{R}^N$ is a Harsanyi NTU-value of $(N, V)$ if and only if there exist positive weights $w = (w^i)_i$ such that*

$$\frac{1}{w^i} x^i = D^i P_w(N, V), \text{ for all } i \in N,$$

*and*

$$\sum_{i\in N} \frac{1}{w^i} x^i \geq \sum_{i\in N} \frac{1}{w^i} y^i, \text{ for all } y \in V(N),$$

*where $P_w$ is the $w$-potential function.*

## References

[1] R.J. Aumann [1985], An axiomatization of the non-transferable utility value. *Econometrica*, 53:599–612.

[2] J.C. Harsanyi [1963], A simplified bargaining model for the $n$-person cooperative game. *International Economic Review*, 4:194–220.

[3] S. Hart [1985], An axiomatization of Harsanyi's nontransferable utility solution. *Econometrica*, 53:1295–1313.

[4] S. Hart and A.Mas-Colell [1988], The potential of the Shapley value. In *The Shapley Value, Essays in Honor of L.S.Shapley*, A.E. Roth (ed.), 127–137, Cambridge, Cambridge University Press.

[5] S. Hart and A. Mas-Colell [1989], Potential, value and consistency. *Econometrica*, 57:589–614.

[6] S. Hart and A. Mas-Colell [1992], A Non-cooperative interpretation of value and potential. In *Rational Interactions, Essays in Honor of J.C. Harsanyi*, R. Selten (ed.), 83–93, Springer–Verlag.

[7] E. Kalai and D. Samet [1985], Monotonic solutions to general cooperative games. *Econometrica*, 53:307–327.

[8] L.S. Shapley [1953a], Additive and non-additive set functions, Ph.D. thesis, Princeton University.

[9] L.S. Shapley [1953b], A value for $n$-person games. In *Contributions to the Theory of Games, II*, 307–317. Princeton University Press.

[10] L.S. Shapley [1969], Utility comparisons and the theory of games. In *La Décision*, Ed. du CNRS, Paris, 251–263.

[11] W. Thomson [1990], The consistency principle. In *Game Theory and Applications*, T. Ichiishi, A. Neyman and Y. Tauman (eds.), 187–215, Academic Press.