Econometrica, Vol. 57, No. 3 (May, 1989), 589–614

POTENTIAL, VALUE, AND CONSISTENCY[1]

BY SERGIU HART AND ANDREU MAS-COLELL

Let P be a real-valued function defined on the space of cooperative games with transferable utility, satisfying the following condition: In every game, the marginal contributions of all players (according to P) are efficient (i.e., add up to the worth of the grand coalition). It is proved that there exists just one such function P—called *the potential*—and moreover that the resulting payoff vector coincides with the Shapley *value*. The potential approach is also shown to yield other characterizations for the Shapley value, in particular, in terms of a new internal *consistency* property. Further results deal with weighted Shapley values (which emerge from the above consistency) and with the nontransferable utility case (where the egalitarian solutions and the Harsanyi value are obtained).

KEYWORDS: Shapley value, potential, consistency, n-person games, weighted values.

1. INTRODUCTION

CONSIDER THE PROBLEM of allocating some resource (or costs, profits, etc.) among n participants (economic agents, projects, departments, ...). Assume that the situation is well described as an n-person game in characteristic function form. The problem we address here is that of developing general principles for performing this allocation.

An approach with a long tradition in economics[2] would proceed by assigning to every player his direct marginal contribution to the grand coalition (i.e., the set of all players). It is obvious, however, that it is not possible in general to solve the problem in this way. This is simply because these marginal contributions may not add up to the worth of the grand coalition (namely, they will either be not feasible, or, if feasible, not efficient); from now on, we will refer to this "adding up" requirement simply as "efficiency."

In this paper we introduce a new analytical concept with clear affinities to the marginal contribution approach. We propose that to every allocation problem described by an n-person game be associated a single number (called the potential of the game) and that each player receive his marginal contribution (computed according to these numbers). The surprising fact is: the requirement that a feasible and efficient allocation (one that exactly shares everything available) should always be obtained determines the procedure uniquely. Namely, there exists just one such allocation procedure. Moreover, the resulting solution is well known: it is the Shapley (1953b) value.

[1] This paper supersedes our earlier papers, "Value and Potential: Marginal Pricing and Cost Sharing Reconciled" and "The Potential: A New Approach to the Value in Multi-Person Allocation Problems" (HIER, Harvard University, DP-1127, January, 1985 and DP-1157, June, 1985). We want to acknowledge useful discussions with Michael Maschler and Lloyd S. Shapley. Financial support by the National Science Foundation and by the U.S.-Israel Binational Science Foundation is gratefully acknowledged.

[2] For a modern point of view, see, for example, Ostroy (1984).

In summary, our central result is the following:

THEOREM A: *There exists a unique real function on games—called the potential—such that the marginal contributions of all players (according to this function) are always efficient. Moreover, the resulting payoff vector is precisely the Shapley value.*

This theorem is discussed at length in Section 2.

Although the potential is in its essence just a technical tool, it is nonetheless a powerful and suggestive one. In particular, the potential approach has suggested to us two further, substantive ways to characterize the Shapley value. The first (in Section 3) uses a "preservation of differences" principle, which is a straightforward generalization of the "divide the surplus equally" idea for two-person situations. The second (in Section 4) considers an internal "consistency" property: eliminating some of the players, after paying them according to the solution, does not change the outcome for any of the remaining ones. The main result here is as follows:

THEOREM B: *Consider the class of solutions that, for two-person games, divide the surplus equally. Then the Shapley value is the unique consistent solution in this class.*

In Section 4 we also discuss alternative consistency requirements (including one for various bargaining solutions).

Next, we broaden the class of consistent solutions by dropping symmetry. We obtain the following theorem (in Section 5):

THEOREM C: *Consider the class of solutions that, for two-person games, are efficient, invariant (in a very weak sense), and monotone. Then the weighted Shapley values are the only consistent solutions in this class.*

The weighted Shapley values were introduced by Shapley (1953a) as nonsymmetric generalizations of the Shapley value. They are based on relative weights among the players. By our result, only one requirement—consistency—suffices to endogenously generate these weights (with the appropriate very weak initial conditions for two-person games being assumed).

Finally, we address also the nontransferable utility case (in Section 6): the potential approach leads naturally to the egalitarian solutions studied by Myerson (1980) and Kalai-Samet (1985). These can be seen as the first step in the construction of the Harsanyi (1963) NTU-value (Theorem D provides the exact characterization). Theorem E generalizes Theorem C. It shows that consistency (together with the correct initial conditions for two-person transferable utility games) characterizes the egalitarian solutions for nontransferable utility games too.

2. THE POTENTIAL

A (cooperative) *game* (with side payments) is a pair (N, v), where N is a finite set of *players* and $v\colon 2^N \to R$ is[3] a *characteristic function* satisfying $v(\phi) = 0$. We will 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 $\boldsymbol{G}$ be the set of all games. Given a function $P\colon \boldsymbol{G} \to R$ which associates a real number[4] $P(N, v)$ to every game (N, v), the *marginal contribution* of a player in a game is defined to be

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

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

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

$$\sum_{i \in N} D^iP(N, v) = v(N) \tag{2.1}$$

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. From now on, we refer to this property as "efficiency."

THEOREM A: *There exists a unique potential function P. For every game (N, v), the resulting payoff vector $(D^iP(N, v))_{i \in N}$ coincides with the Shapley value of the game. Moreover, the potential of any game (N, v) is uniquely determined by* (2.1) *applied only to the game and its subgames (i.e., to (S, v) for all $S \subset N$).*

Let $Sh^i(N, v)$ denote the *Shapley value* of player i in the game (N, v). We thus have

$$D^iP(N, v) = Sh^i(N, v)$$

for every game (N, v) and each player i in N.

PROOF OF THEOREM A: Formula (2.1) can be rewritten as

$$P(N, v) = \frac{1}{|N|}\left[v(N) + \sum_{i \in N} P(N \setminus \{i\}, v)\right]. \tag{2.2}$$

Starting with $P(\phi, v) = 0$, it determines $P(N, v)$ recursively. This proves the existence of the potential function P, and moreover that $P(N, v)$ is uniquely determined by (2.1) (or (2.2)) applied to (S, v) for all $S \subset N$.

[3] R denotes the real line.

[4] We write $P(N, v)$ rather than the more cumbersome $P((N, v))$.

Next, express (N, v) as a linear combination of unanimity games: $v = \sum_{T \subset N} a_T u_T$, where u_T is the *T-unanimity* game, defined by $u_T(S) = 1$ if S contains T and $= 0$ otherwise (it is well known that this decomposition exists and is unique). Define $d_T = a_T / |T|$ (this is Harsanyi's "dividend" to the members of coalition T), and put

$$(2.3) \qquad P(N, v) = \sum_T d_T.$$

It is easily checked that (2.1) is satisfied by this P; hence (2.3) defines the unique potential function. The result now follows since $Sh^i(N, v) = \sum_{\{T:\, i \in T\}} d_T$ (i.e., T ranges over all subsets T of N that contain i).[5] *Q.E.D.*

For further interpretation of the potential we derive an explicit formula ((2.1) defines it only implicitly). Consider the following (standard) model of choosing a random subset S of a given set N with $n = |N|$ elements: First, a size $s = 1, 2, \ldots, n$ is chosen randomly (with probability $1/n$ each). Second, a subset S of N of size s is chosen randomly (with probability $1/\binom{n}{s}$, where $s = |S|$). Equivalently, one can order the n elements (there are $n!$ orders), choose a cutting point s (there are n choices), and take the first s elements in the order. Let E denote expectation with respect to this probability distribution.

PROPOSITION 2.4: *Let P be the potential function. Then*

$$P(N, v) = E\left[\frac{|N|}{|S|} v(S)\right]$$

for every game (N, v).

PROOF: Using the explicit probabilities described above, we have to show that

$$(2.5) \qquad P(N, v) = \sum_{S \subset N} \frac{(s-1)!(n-s)!}{n!} v(S),$$

where $n = |N|$ and $s = |S|$. The marginal contributions $D^i P$ of (2.5) are easily seen to coincide with the Shapley value; therefore (2.5) is the potential function. *Q.E.D.*

Thus, the potential is the *expected normalized worth*; equivalently, the per-capita potential $P(N, v)/|N|$ equals the average per-capita worth $v(S)/|S|$. Hence, the potential provides a most natural one-number summary of the game. The Shapley value is known to be an expected marginal contribution. We obtain it here as a marginal contribution to an expectation (the potential).[6]

[5] An alternative proof is to show that $(D^i P)_i$ satisfy the standard axioms for the Shapley value (this is done inductively, using (2.1) or (2.2); see Hart and Mas-Colell (1988), which includes alternative proofs for other results here, as well as further discussions).

[6] The idea of reversing the order of integration and differentiation has been fruitfully used by Mertens (1980) in the context of values of nonatomic games.

REMARK 2.5: Our approach can be regarded as a *new characterization* for the Shapley value. Only one axiom, (2.1), suffices. Moreover, only the game itself and its subgames have to be considered; this is important particularly in applications, where typically only *one* specific problem is considered. In contrast, the standard axiomatizations require the application of the axioms to a large class of games (e.g., all games; or, all simple games; etc.) in order to uniquely determine it for any single game. Finally, note that although marginal contributions appear implicitly in the formulae for the Shapley value, our approach uses this principle explicitly (i.e., in the axioms), and in a very simple form.

REMARK 2.6: Formula (2.2) yields a simple and straightforward recursive procedure for the computation of the potential, and thus, a fortiori, for the Shapley value of the game as well as of all its subgames. It seems to be a most efficient algorithm for computing Shapley values.

REMARK 2.7: It is clear that P is, formally, just a mathematical potential function for the Shapley value (taking discrete rather than infinitesimal differences); this explains our choice of name for it. Thus, the Shapley value vector function is the (discrete) gradient of the potential function.[7] Moreover, if we do not require $P(\phi, v)=0$, then P is only determined up to an additive constant (which of course does not change the payoff vectors).[8]

REMARK 2.8: One may regard P as an operator that associates to each game (N, v) another game (N, Pv), given by $(Pv)(S)=P(S, v)$ for all $S \subset N$. It is easily checked that this is a linear, positive, and symmetric operator; it is one-to-one and onto, and its fixed points are exactly the inessential (additive) games. Additional properties are implied by these; for example, if (N, v) is a market game (i.e., totally balanced), then $Pv \leqslant v$.

The potential approach can be extended to games with a continuum of players. We consider here only the finite type case: the game is given by a pair (f, z), where f: $R^n_+ \to R$ (with $f(0)=0$) is the characteristic function, and $z=(z^1, \ldots, z^n) \in R^n_+$ represents the grand coalition (z^i being the "number" of

[7] This "summability" condition that is satisfied by the Shapley value can be stated as follows: For every game (N, v) and every $S \subset N$ with $|S|=s$

$$\sum_{t=1}^{s} Sh^{i_t}(\{i_1, i_2, \ldots, i_t\}, v)$$

is the same for all orderings $i_1, i_2, \ldots, i_s$ of the elements of S (this number is just $P(S, v)$). This condition may be viewed as "path-independence."

[8] We note that D'Aspremont, Jacquemin, and Mertens (1984) have recently defined a class of real functions on games: "aggregate (power) indices." It can be easily checked that the potential function does not belong to this class. However, it would be included if their "Normalization axiom" is dropped (the potential function is obtained when, in their notations, $d\mu(\alpha)=d\alpha/\alpha$—i.e., ν is the Lebesgue measure).

players of type i). This is, for instance, the standard setup for cost allocation problems, f being either the "cost" or the "production" function (see Billera and Heath (1982) and Mirman and Tauman (1982)). A (type-symmetric) feasible and efficient solution to (f, z) is then a vector $\phi(f, z) \in R^n$ with the property $z \cdot \phi(f, z) = f(z)$. A *potential* for f is a differentiable function F: $R^n_+ \to R$ such that for any z we have $z \cdot \partial F(z) = f(z)$, where $\partial F(z)$ is the gradient of F at z. All the results in this section generalize: Given f, there is one and only one potential[9] F. Moreover, the solution $\phi(f, z) = \partial F(z)$ corresponds to the Aumann–Shapley (1974) value for games with a continuum of players (it is thus called "the Aumann–Shapley price vector"). Hence, this is the only integrable vector field ψ_f on R^n_+ satisfying $z \cdot \psi_f(z) = f(z)$ for all $z \in R^n_+$. Finally, the potential F can be derived explicitly and takes a familiar diagonal form (compare with Proposition 2.4):

$$F(z) = \int_0^1 \frac{1}{t} f(tz)\, dt.$$

3. PRESERVATION OF DIFFERENCES

We now take a different tack to the payoff allocation problem. It will, however, lead us to the same solution.

Consider the grand coalition N in the game (N, v). Suppose we are given constants d^{ij} for all $i, j \in N$ which are *compatible*, in the sense that $d^{ii} = 0$, $d^{ij} = -d^{ji}$ and $d^{ij} + d^{jk} = d^{ik}$ for all $i, j, k \in N$. We say that a payoff vector $(x^i)_{i \in N}$ *preserves differences according to* $\{d^{ij}\}$ if

$$x^i - x^j = d^{ij} \quad \text{for all } i, j.$$

It is trivial to verify that, given compatible constants $\{d^{ij}\}$, there exists a single *efficient* payoff vector x that preserves differences:

$$x^i = \frac{1}{|N|}\left[v(N) + \sum_j d^{ij}\right], \quad x^j = x^i - d^{ij}.$$

To have a well defined solution we therefore only need to specify, for every game, differences $\{d^{ij}\}$. We do this recursively. Suppose that payoffs have been determined for all strict subgames of (N, v); let $x^i(S)$ be the payoff of player i in (S, v), for $i \in S \subset N$. Then we convene that the difference d^{ij} to be preserved is:

$$(3.1) \qquad d^{ij} = x^i(N \setminus \{j\}) - x^j(N \setminus \{i\});$$

that is, the difference between what i would get if j was not around and what j would get if i was not around. (It will be seen below that these differences are indeed compatible.) This seems to be a most natural way to compare the "relative position" (or, "relative strengths") of the players. It is notable that, as it

[9] Minor regularity conditions are required; for example, f continuous and $f(z) = 0(\|z\|)$ as $z \to 0$ suffice. (See the formula for F below.)

will be shown below, one obtains *a unique efficient outcome which simultaneously preserves all these differences.*[10]

The preservation of differences principle can be regarded as a straightforward generalization of the "equal division of the surplus" idea for two-person problems. Indeed, in that case, the payoffs are

$$x^i \equiv x^i(\{i,j\}) = v(\{i\}) + \tfrac{1}{2}[v(\{i,j\}) - v(\{i\}) - v(\{j\})],$$

thus

$$x^i - x^j = v(\{i\}) - v(\{j\}) = x^i(\{i\}) - x^j(\{j\}),$$

or: differences are preserved.

Observe that the proposed solution satisfies

$$(3.2) \qquad \sum_{i \in N} x^i(N) = v(N), \qquad \text{and}$$

$$(3.3) \qquad x^i(N) - x^i(N \setminus \{j\}) = x^j(N) - x^j(N \setminus \{i\}).$$

Condition (3.3) has already been used by Myerson (1980), under the name of "balanced contributions." The mathematically inclined reader will recognize it as a finite difference analog of the Frobenius integrability condition (i.e., symmetry of the cross partial derivatives), which suggests that the solution admits a potential function. Conversely, if a solution is generated by a potential, then (3.3) is clearly satisfied.

THEOREM 3.4: *The construction of the above solution, according to the principle of preservation of differences, is well-defined. Moreover, the solution is generated by a potential function—thus it coincides with the Shapley value.*

PROOF: Consider an n-person game (N, v). Assume, by induction, that the solution has been already determined for all strict subgames of (N, v), and that moreover

$$(3.5) \qquad x^i(S) = P(S) - P(S \setminus \{i\})$$

for all $i \in S \subset N$, $S \neq N$ (where P is the unique potential function, and $P(S)$ is short for $P(S, v)$). Note that (3.5) is of course satisfied for singletons ($|S| = 1$) by efficiency (3.2).

Applying (3.5) to coalitions of size $n - 1$ implies that the differences d^{ij} are indeed compatible, since

$$\begin{aligned} d^{ij} &= x^i(N \setminus \{j\}) - x^j(N \setminus \{i\}) \\ &= [P(N \setminus \{j\}) - P(N \setminus \{i,j\})] - [P(N \setminus \{i\}) - P(N \setminus \{i,j\})] \\ &= P(N \setminus \{j\}) - P(N \setminus \{i\}). \end{aligned}$$

[10] One wants to preserve differences rather than, say, ratios, since the resulting outcome should not depend on the choice of origin of a player's utility scale.

Therefore $(x^i(N))_i$ is well defined, and we have

$$x^i(N) - x^j(N) = d^{ij} = P(N\setminus\{j\}) - P(N\setminus\{i\}).$$

Fix i, and average over j:

$$x^i(N) - \frac{1}{n}v(N) = \frac{1}{n}\sum_{j\in N} P(N\setminus\{j\}) - P(N\setminus\{i\})$$

(by (3.2)). Thus

$$x^i(N) + P(N\setminus\{i\}) = \frac{1}{n}\left[v(N) + \sum_{j\in N} P(N\setminus\{j\})\right].$$

By (2.2), the right-hand side is exactly $P(N)$; thus (3.5) is satisfied for $S = N$ also. *Q.E.D.*

4. CONSISTENCY

This section is devoted to a characterization of the Shapley value by means of an (internal) consistency property. Such an approach may be traced back to Harsanyi (1959). It has been successfully applied to a wide variety of solution concepts: Davis and Maschler (1965), Sobolev (1975), Lensberg (1982), Balinsky and Young (1982), Thomson (1984), Peleg (1985, 1986), Aumann and Maschler (1985), Moulin (1985), etc. We will discuss the connections between some of these approaches later in this section.

The consistency requirement may be described informally as follows: Let ϕ 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 the rest of the players payoffs according to ϕ. Then ϕ is said to be consistent if, when it is applied to any reduced game, it yields the same payoffs as in the original game. The various consistency requirements differ in the precise definition of the reduced game (i.e., exactly how are the players outside being paid off).

Formally, let ϕ be a function defined on $\boldsymbol{G}$, the set of all games (see Section 2), with $\phi(N,v)$ a vector in R^N for all (N,v) in $\boldsymbol{G}$; such a function is called a *solution function*. We will write $\phi^i(N,v)$ for its ith coordinate; thus, $\phi(N,v) = (\phi^i(N,v))_{i\in N}$.

Let ϕ be a solution function, (N,v) a game, and $T \subset N$. We define the *reduced game* (T, v_T^ϕ) as follows:

$$v_T^\phi(S) = v(S\cup T^c) - \sum_{i\in T^c}\phi^i(S\cup T^c, v), \quad \text{for all} \quad S\subset T, \tag{4.1}$$

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

$$\phi^j\left(T, v_T^\phi\right) = \phi^j(N,v), \quad \text{for all} \quad j\in T. \tag{4.2}$$

The interpretation is as follows. Given the solution function ϕ, a game (N,v) and a coalition $T\subset N$, the members of T (or, more precisely, every subcoalition of T) need to consider the total payoff remaining after paying the members of T^c according to ϕ. To compute the worth of a coalition $S \subset T$ (in the reduced game), we assume that the members of $T \setminus S$ are not present; in other words, one considers the game $(S \cup T^c, v)$, in which payoffs are distributed according to ϕ. The appropriateness of this definition of reduced games depends, of course, on the particular situation being modelled; more specifically, on the concrete assumptions underlying the determination of the characteristic function. An example will be discussed later on in this section.

Note that one usually deals with efficient solution functions. In this case, (4.1) can be rewritten as

$$(4.3) \qquad v_T^{\phi}(S) = \sum_{i \in S} \phi^i(S \cup T^c, v).$$

Furthermore, if ϕ is an efficient and consistent solution function, then necessarily

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

which is exactly (4.1) for $S = T$. If one wants the definition of $v_T^{\phi}(S)$ to be always according to the same rule, then (4.1) results for all $S \subset T$.

The following lemma shows that whether ϕ is consistent or not may be determined by considering only singleton coalitions T^c.

LEMMA 4.4: *ϕ is a consistent solution function if and only if* (4.2) *is satisfied for all games* (N, v) *and all* $T \subset N$ *with* $|T^c| = 1$.

PROOF: Use induction on the size of T^c. *Q.E.D.*

The first result is as follows:

PROPOSITION 4.5: *The Shapley value is a consistent solution function.*

PROOF: We will use the potential P, as defined by (2.1). Let (N, v) be a game and let $i \in N$; we will write v_{-i} for $v_{N\setminus\{i\}}^{\phi}$, where $\phi = Sh$. Since Sh is efficient, we have for all $S \subset N \setminus \{i\}$:

$$\begin{aligned} v_{-i}(S) &= v(S \cup \{i\}) - Sh^i(S \cup \{i\}, v) = \sum_{j \in S} Sh^j(S \cup \{i\}, v) \\ &= \sum_{j \in S} \left[P(S \cup \{i\}, v) - P(S \cup \{i\} \setminus \{j\}, v) \right]. \end{aligned}$$

By Theorem A, formula (2.1) applied to $(N \setminus \{i\}, v_{-i})$ and all its subgames *uniquely* determines their potential. Comparing this with the above equalities, we obtain that $P(S, v_{-i})$ and $P(S \cup \{i\}, v)$ may differ only by a constant[11]

$$P(S, v_{-i}) = P(S \cup \{i\}, v) + c.$$

[11] From $P(\phi, v_{-i}) = 0$ it follows that $c = -P(\{i\}, v)$; this is however not needed in the sequel.

Thus

$$\begin{aligned} Sh^j(N\setminus\{i\}, v_{-i}) &= P(N\setminus\{i\}, v_{-i}) - P(N\setminus\{i,j\}, v_{-i}) \\ &= P(N,v) - P(N\setminus\{j\}, v) \\ &= Sh^j(N,v). \end{aligned} \qquad Q.E.D.$$

Next, we will show that the property of consistency is essentially equivalent to the existence of a potential function; thus, consistency almost characterizes the Shapley value. "Almost" refers to the "initial conditions," namely, the behavior of the solution for two-person games.[12]

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

$$(4.6) \qquad \phi^i(\{i,j\}, v) = v(\{i\}) + \tfrac{1}{2}[v(\{i,j\}) - v(\{i\}) - v(\{j\})]$$

for all $i \neq j$ and all v. Thus, the "surplus" $[v(\{i,j\}) - v(\{i\}) - v(\{j\})]$ is equally divided among the two players. Most solutions satisfy this requirement, in particular, the Shapley value and the nucleolus.

THEOREM B: *Let ϕ be a solution function. Then*:

(i) *ϕ is consistent*; *and*

(ii) *ϕ is standard for two-person games*;

if and only if ϕ is the Shapley value.

PROOF:[13] One direction is immediate (recall Proposition 4.5).

For the other direction, assume ϕ satisfies (i) and (ii). We claim first that ϕ is efficient, i.e.,

$$(4.7) \qquad \sum_{i\in N} \phi^i(N,v) = v(N)$$

for all (N, v). This indeed holds for $|N| = 2$ by (4.6). Let $n \geqslant 3$, and assume (4.7) holds for all games with less than n players. For a game (N, v) with $|N| = n$, let $i \in N$; by consistency

$$\sum_{j\in N} \phi^j(N,v) = \sum_{j\in N\setminus\{i\}} \phi^j(N\setminus\{i\}, v_{-i}) + \phi^i(N,v),$$

where $v_{-i} \equiv v^{\phi}_{N\setminus\{i\}}$. By assumption, ϕ is efficient for games with $n-1$ players; thus

$$= v_{-i}(N\setminus\{i\}) + \phi^i(N,v) = v(N)$$

(by definition of v_{-i}). Therefore ϕ is efficient for all $n \geqslant 2$.

[12] In contrast to the potential approach (in Section 2), consistency requires one to consider a large domain of games rather than a single one. Moreover, the reduced games may not share some of the properties of the original game; e.g., super-additivity (consider the three-person majority game).

[13] An independent proof (not based on the potential) has been communicated to us by Michael Maschler. For yet another proof of uniqueness, see Lemma 6.8.

Finally, for $n=1$, we have to show that $\phi^i(\{i\},v)=v(\{i\})$. Indeed, let $v(\{i\})=c$, and consider the game $(\{i,j\},\bar{v})$ (for some $j\neq i$), with $\bar{v}(\{i\})=\bar{v}(\{i,j\})=c$, $\bar{v}(\{j\})=0$. By (ii), $\phi^i(\{i,j\},\bar{v})=c$ and $\phi^j(\{i,j\},\bar{v})=0$; hence $\bar{v}_{-j}(\{i\})=c-0=c=v(\{i\})$, and $c=\phi^i(\{i,j\},\bar{v})=\phi^i(\{i\},\bar{v}_{-j})=\phi^i(\{i\},v)$ by consistency. This concludes the proof of the efficiency of ϕ.

Next, we show that ϕ admits a potential. To that end, define a real function Q on the set of all games of at most two players by

$$Q(\phi,v)=0,$$
$$Q(\{i\},v)=v(\{i\}),$$
$$Q(\{i,j\},v)=\tfrac{1}{2}[v(\{i\})+v(\{j\})+v(\{i,j\})],$$

for all v and all $i\neq j$. It is straightforward to check that, for all (N,v) with $|N|=1,2$:

$$\phi^i(N,v)=Q(N,v)-Q(N\setminus\{i\},v) \tag{4.8}$$

for all $i\in N$.

We will now show that Q can be extended to all games (N,v), in such a way that (4.8) always holds. Together with efficiency (4.7), this implies (2.1); therefore Q is actually the potential P, and ϕ is the Shapley value.

We again use induction: Let $n\geqslant 3$, and assume Q has been defined, and moreover satisfies (4.8), for all games of at most $n-1$ players. Fix a game (N,v) with $|N|=n$. We have to show that

$$\phi^i(N,v)+Q(N\setminus\{i\},v)$$

is the same for all $i\in N$ (and this will then be $Q(N,v)$). Let $i,j\in N$, $i\neq j$, and let $k\in N$, $k\neq i,j$ (such a k exists since $|N|\geqslant 3$).

We have (by consistency and (4.8) for $n-1$)

$$\begin{aligned}
&\phi^i(N,v)-\phi^j(N,v)\\
&\quad=\phi^i(N\setminus\{k\},v_{-k})-\phi^j(N\setminus\{k\},v_{-k})\\
&\quad=[Q(N\setminus\{k\},v_{-k})-Q(N\setminus\{i,k\},v_{-k})]\\
&\qquad-[Q(N\setminus\{k\},v_{-k})-Q(N\setminus\{j,k\},v_{-k})]\\
&\quad=[Q(N\setminus\{j,k\},v_{-k})-Q(N\setminus\{i,j,k\},v_{-k})]\\
&\qquad-[Q(N\setminus\{i,k\},v_{-k})-Q(N\setminus\{i,j,k\},v_{-k})].
\end{aligned}$$

Apply again (4.8) (for $n-2$) and consistency:

$$\begin{aligned}
&\quad=\phi^i(N\setminus\{j,k\},v_{-k})-\phi^j(N\setminus\{i,k\},v_{-k})\\
&\quad=\phi^i(N\setminus\{j\},v)-\phi^j(N\setminus\{i\},v)\\
&\quad=[Q(N\setminus\{j\},v)-Q(N\setminus\{i,j\},v)]\\
&\qquad-[Q(N\setminus\{i\},v)-Q(N\setminus\{i,j\},v)]\\
&\quad=Q(N\setminus\{j\},v)-Q(N\setminus\{i\},v),
\end{aligned}$$

where we used (4.8) once more (for $n-1$). This completes the proof. *Q.E.D.*

REMARK 4.9: Theorem B (and all our other results on consistency in this paper) applies to any fixed finite number of players; i.e., if one considers only games with at most n players, then consistency together with the appropriate initial condition suffice to characterize the solution. This should be contrasted with other consistency results, e.g., Sobolev (1975), Lensberg (1982), Peleg (1986), where an unbounded number of players is needed.

The standard solution for two-person games is very natural; it may, however, be derived from more basic postulates.

A solution function ϕ is *transferable-utility invariant* (TU-invariant)[14] if, for any two games (N, v) and (N, u) (with the same set of players) and real constants $a > 0$ and $\{b^i\}_{i\in N}$,

$$u(S) = av(S) + \sum_{i\in S} b^i \quad \text{for all} \quad S \subset N$$

implies

$$\phi^i(N, u) = a\phi^i(N, v) + b^i \quad \text{for all} \quad i \in N.$$

(Note: (N, v) and (N, u) are called *TU-equivalent* games.) TU-invariance requires, first, that a change in scale *common* to all players should affect the solution accordingly; and second, that adding a fixed amount, whenever a player i appears, should lead to just adding this amount to his final payoff.

A solution function ϕ satisfies the *equal treatment property* if, for any game (N,v) and any two players $i, j \in N$,

$$v(S \cup \{i\}) = v(S \cup \{j\}) \quad \text{for all} \quad S \subset N \setminus \{i, j\}$$

implies

$$\phi^i(N, v) = \phi^j(N, v).$$

Note that, for two-person games, this amounts to just

$$\phi^i(\{i, j\}, v) = \phi^j(\{i, j\}, v) = \tfrac{1}{2}v(\{i, j\})$$

whenever $v(\{i\}) = v(\{j\})$.

THEOREM B′: *Let ϕ be a solution function. Then*:

(i) *ϕ is consistent*;

(ii) *for two-person games*: (a) *ϕ is efficient*, (b) *ϕ is TU-invariant*, (c) *ϕ satisfies the equal treatment property*;

if and only if ϕ is the Shapley value.

PROOF: Efficiency (a) and equal treatment (c) imply that

$$\phi^i(\{i, j\}, cu_{\{i,j\}}) = \tfrac{1}{2}c$$

[14] This is sometimes referred to as "strategic invariance;" we prefer to call it "TU-invariance" instead, in order to avoid confusion with "NTU-invariance," where scales of different players may change independently.

for all $i \neq j$ and all real c, where $u_{\{i,j\}}$ denotes the $\{i, j\}$-unanimity game. Using this for $c = 1$, 0 and -1 together with TU-invariance (b) yields (4.6), and Theorem B applies. *Q.E.D.*

It is instructive to compare our approach with Sobolev's (1975) consistency treatment of the (pre)nucleolus. The two concepts of consistency—namely, (4.2)—are the same. The difference lies in the definition of the reduced game.[15] Following Davis and Maschler (1965), Sobolev uses the following definition for the reduced game (i.e., instead of (4.1)):

$$(4.10) \qquad v_T^{\phi}(S) = \begin{cases} \sum_{j \in T} \phi^j(N, v), & S = T; \\ \operatorname*{Max}_{Q \subset T^c} \left\{ v(S \cup Q) - \sum_{i \in Q} \phi^i(N, v) \right\}, & S \neq T, \phi; \\ 0, & S = \phi. \end{cases}$$

There are two important distinctions between the two definitions of the reduced game: first, the maximum over all $Q \subset T^c$ in (4.10) vs. just $Q = T^c$ in (4.1); and second, the payoffs to the players in T^c are taken, for every S, according to the solution of the game (N, v) in (4.10), and according to the solution of the subgame $(S \cup T^c, v)$ in (4.1).

These two differences may be understood as follows. The first indicates that, in (4.10), some sort of strategic freedom is available to each coalition S: they may choose which of the members of T^c (if any) to take along and pay them according to the solution. In comparison, in (4.1) it is assumed that *all* members of T^c have to be paid off. The second difference lies in the way these payoffs are computed. In (4.10), the members of T^c are paid according to the solution *allocation* (of the grand coalition), whereas in (4.1) according to the solution *function*, each time applied to the appropriate situation (namely, the solution allocation of the coalition $S \cup T^c$).

As we have already stated above, which definition is more appropriate will depend on the context being modeled (and the way the characteristic function is defined). An example where our definition seems natural is the problem of allocating joint costs among several "projects" (or, departments, tasks, etc.); these are now interpreted as the players. It is to be emphasized that these cost imputations are *not* meant to be "efficiency prices" (i.e., usable to make investment decisions on which of the projects to undertake); in fact, except in trivial cases, cost allocations satisfying the adding-up condition would generate inefficient decisions if used as investment guides. In summary the problem is: we are given a fixed set of projects, and required (by the legal and/or administrative environment—e.g., the tax authorities) to obtain an exact distribution of total costs. To this end, one should, of course, take into account all available information, in particular, the cost of any subset S of projects (assuming that these are

[15] Also, Sobolev uses the other assumptions (e.g., symmetry) in an essential way (for all n).

the *only* ones to be undertaken). We assume that this information (that may not be always easy to get) is available to the "accountants" in charge of computing the cost imputations.

What does consistency mean in this framework? Consider a multi-state company and a restricted set T of projects, say those in Tennessee. For every subset $S \subset T$ of Tennessee's projects, the local accountant has to determine their cost, assuming that these are the only projects in Tennessee to be undertaken. In addition, there is the set T^c of all the projects outside Tennessee (which are not in the domain of the local "gedanken experiment"). The cost of S is therefore the amount imputed to S by the general accounting procedure under consideration (solution function) when the projects to be implemented are $S \cup T^c$. This is exactly formula (4.3), and consistency requires that the local accountant's imputation be no different than the one obtained (for the Tennessee projects) by the general (national) accountant.

As suggested by the comparison of (4.1) and (4.10) there are two additional possible definitions of the reduced game. One is

$$\text{(4.11)} \qquad v_T^{\phi}(S) = v(S \cup T^c) - \sum_{i \in T^c} \phi^i(N, v)$$

for all $S \subset T$. This definition is similar to (4.1) according to the first criterion mentioned above (no maximum), and similar to (4.10) according to the second criterion ($\phi^i(N)$ rather than $\phi^i(S \cup T^c)$). The resulting consistency has been studied by Moulin (1985). He shows (Lemma 6 there) that it characterizes the so-called "equal allocation of nonseparable costs."

Finally, the fourth possible definition of a reduced game is:

$$\text{(4.12)} \qquad v_T^{\phi}(S) = \begin{cases} v(N) - \sum_{i \in T^c} \phi^i(N, v), & S = T, \\ \operatorname*{Max}_{Q \subset T^c} \left\{ v(S \cup Q) - \sum_{i \in Q} \phi^i(S \cup Q, v) \right\}, & S \subset T. \end{cases}$$

However, there is no solution function that is standard for two-person games and is consistent according to (4.12).[16]

5. WEIGHTS

The result in the previous section, Theorem B′, characterizes the Shapley value by means of consistency[17] together with natural initial conditions for two-person games. In this section we drop the equal treatment property, and characterize the resulting consistent solutions. They turn out to be the *weighted Shapley values*,

[16] A counterexample is provided by the following four-person game: $N = \{1,2,3,4\}$; player 4 is a null player; $v(\{1,2\}) = v(\{2,3\}) = 12$, $v(\{1,3\}) = 24$, $v(\{1,2,3\}) = 27$, and $v(S) = 0$ for all other $S \subset \{1,2,3\}$. One may, however, restrict the class of games under consideration. For example, a class of games can be found (containing, in particular, all convex games), at which the Shapley value is the unique solution that is standard for two-person games and consistent according to (4.12).

[17] From now on we will only deal with the original consistency (i.e., according to (4.1)).

introduced by Shapley (1953a) (see also Owen (1968), Shapley (1981), Kalai and Samet (1987)).

To present these solutions, we will first assume that a collection of weights is *exogenously* given. All the results of the previous sections (with the appropriate modifications) will be seen to remain valid in this more general setup: they are all related to a notion of weighted potential. We will then obtain our main result (Theorem C): consistency and the initial conditions for two-person games of Theorem B′, with equal treatment replaced by monotonicity, imply the existence of weights such that the solution is precisely the corresponding weighted Shapley value. Thus, weights are obtained *endogenously*, from the solution itself.

Formally, suppose first that for each player i we are given a weight $w^i > 0$; let[18] $w = (w^i)_i$. These weights can be interpreted as "a-priori measures of importance;" they are taken to reflect considerations *not* captured by the characteristic function. For example, we may be dealing with a problem of cost allocation among investment projects. Then the weights w^i could be associated to the profitability of the different projects. In a problem of allocating travel costs among various institutions visited (cf. Shapley (1981)), the weights may be the number of days spent at each one.

In line with the above interpretation, we would now desire that in any unanimity game the worth be distributed among the players in proportion to their weights. In a unanimity game, every player has exactly the same (marginal) contribution. Therefore, to obtain the w-proportional allocation we need to weight the marginal contribution of each player i by his w^i.[19]

We are thus led to the definition of a *w-potential* P_w, as a function P_w: $G \to R$ with $P_w(\phi, v) = 0$, satisfying the following condition:

$$(5.1) \qquad \sum_{i \in N} w^i D^i P_w(N, v) = v(N)$$

for all games (N, v). The potential of Section 2 of course corresponds to $w^i = 1$ for all i.

THEOREM 5.2: *For every collection $w = (w^i)_i$ of positive weights there exists a unique w-potential function P_w. Moreover, the resulting solution function, associating the payoff vector $(w^i D^i P_w(N, v))_{i \in N}$ to the game (N, v), coincides with the w-weighted Shapley value Sh_w. Finally, P_w can be computed recursively by the formula*

$$P_w(N, v) = \left[v(N) + \sum_{i \in N} w^i P_w(N \setminus \{i\}, v) \right] \Big/ \sum_{i \in N} w^i.$$

[18] One may think of a universe U of players, such that N is always a finite subset of U (cf. Shapley (1953b)).

[19] A player with weight 2 may be thought of (in some contexts) as two players with weight 1 (see Shapley (1981)).

PROOF: The only thing to check is that the w-potential indeed yields the weighted Shapley value Sh_w. It is easily seen (by the recursive formula above, for example) that $P_w(N, v) = 0$ if (N, v) is the null game (with $v(S) = 0$ for all S), and $P_w(N, cu_N) = c/\Sigma_{i \in N} w^i$ for an N-unanimity game (c is a constant). Therefore $w^i D^i P_w(N, cu_N) = cw^i/\Sigma_{i \in N} w^i = Sh_w^i(N, cu_N)$. Together with additivity (shown inductively using the recursive formula), the proof is completed. *Q.E.D.*

REMARK 5.3: We are assuming throughout that all the weights are positive numbers. One may extend this to allow zero or infinite weights (see Kalai and Samet (1987)).

Note that the w-potential is (positively) homogeneous of degree minus one on the weights, and the corresponding payoff vector (the weighted Shapley value) is homogeneous of degree zero. One could thus think of the w-potential as being measured in per-unit-weight terms. It is to be again emphasized that the weights are unrelated to the characteristic function and that the players (or projects, etc.) are to be regarded as indivisible.

An explicit formula for the w-potential function (of which Proposition 2.4 is a special case) is the following: For each i, let X^i be a random variable with distribution function $\mathrm{Prob}(X^i \leqslant t) = t^{w^i}$ for all $t \in [0,1]$, and assume all the X^i's are independent. For every $t \in [0,1]$, let $N(t) = \{i \in N | X^i \leqslant t\}$ (if one interprets X^i as the arrival time of i, then $N(t)$ is the set of players that arrived up to time t; see Owen (1968)). Then it may be checked that

$$P_w(N, v) = E\left[\int_0^1 \frac{1}{t} v(N(t))\, dt\right].$$

The preservation of differences principle (see Section 3) also applies here; it just becomes

$$\frac{1}{w^i} x^i(N) - \frac{1}{w^j} x^j(N) = d^{ij} = \frac{1}{w^i} x^i(N \setminus \{j\}) - \frac{1}{w^j} x^j(N \setminus \{i\}) \tag{5.4}$$

(thus, the differences between the normalized—i.e., per-unit-weight—payoffs are preserved).

Up to this point the introduction of weights may appear as a rather ad-hoc construction. For practical purposes we may want to add some flexibility to the Shapley value, obtaining nonsymmetric generalizations. But, why do it in this particular way? We will now see that the consistency postulate leads us directly to this class of weighted Shapley values.

Recall the definition of a reduced game (4.1) and the consistency postulate (4.2). The first observation is the following:

PROPOSITION 5.5: *For every collection of positive weights $w = (w^i)_i$, the corresponding weighted Shapley value Sh_w is a consistent solution function.*

PROOF: Mutatis mutandis as for Proposition 4.5, replacing the potential by the w-potential. *Q.E.D.*

Next, we generalize the standard solution to two-person games ("divide the surplus equally") in a nonsymmetric way as follows:

Let $w = (w^i)_i$ be positive weights. Then a solution function ϕ is *w-proportional for two-person games* if

$$(5.6) \qquad \phi^i(\{i,j\},v) = v(\{i\}) + \frac{w^i}{w^i + w^j}[v(\{i,j\}) - v(\{i\}) - v(\{j\})]$$

for all $i \neq j$.

We then have the generalization of Theorem B.

THEOREM 5.7: *Let ϕ be a solution function and w a collection of positive weights. Then*:

(i) *ϕ is consistent*; *and*

(ii) *ϕ is w-proportional for two-person games*;

if and only if ϕ is the w-weighted Shapley value Sh_w.

PROOF: Mutatis mutandis as for Theorem B. *Q.E.D.*

In Theorem 5.7 the weights are introduced exogeneously through the initial conditions for two-person games. We shall now see that, with some very weak conditions on the behavior of the solution for two-person games, consistency actually enables us to rule out all but the proportional solutions. Thus, we obtain a completely endogenous generation of the weights, and, a fortiori, of the weighted Shapley values.

We need the following definition (which will actually be applied only to two-person games). A solution function ϕ is *monotonic* if, for any two games (N, v) and (N, u),

$$u(N) > v(N) \quad \text{and} \quad u(S) = v(S) \quad \text{for all} \quad S \neq N$$

imply

$$\phi^i(N,u) > \phi^i(N,v) \quad \text{for all} \quad i \in N.$$

Thus, if the (grand coalition's) total payoff increases, but nothing else changes, then every player should get an increase in his own payoff. Note that, for two-person games, monotonicity is just $\phi^i(\{i, j\}, u) > \phi^i(\{i, j\}, v)$ and $\phi^j(\{i, j\}, u) > \phi^j(\{i, j\}, v)$ when $u(\{i\}) = v(\{i\})$, $u(\{j\}) = v(\{j\})$ and $u(\{i, j\}) > v(\{i, j\})$.

We now obtain the main result of this section.

THEOREM C: *Let ϕ be a solution function. Then*:

(i) *ϕ is consistent*; *and*

(ii) *for two-person games*: (a) *ϕ is efficient*, (b) *ϕ is TU-invariant*, (c) *ϕ is monotonic*;

if and only if there exist positive weights $w = (w^i)_i$ such that ϕ is the w-weighted Shapley value.

The proof of Theorem C can be found in the Appendix. It is important to point out that, in contrast to Theorem B′, the conditions (ii) (a)–(c) for two-person games do *not* imply that ϕ is a w-proportional solution for $n = 2$. One needs to use consistency repeatedly (for $n = 3$) in order to obtain the initial condition for $n = 2$.

6. NONTRANSFERABLE UTILITY

We now extend the potential function approach to the general case, where utility need not be additively transferable.

A *nontransferable-utility game*—an *NTU-game*, for short—is a pair (N, V), with $V(S)$ a subset of 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 member i of S is x^i. From now on, we will refer to games in $\boldsymbol{G}$ as *TU* or *transferable-utility games*. A TU-game (N, v) in $\boldsymbol{G}$ corresponds to the NTU-game (N, V), where

$$V(S) = \left\{ x \in R^S : \sum_{i \in S} x^i \leqslant v(S) \right\}.$$

We make the following (standard) assumptions: All sets $V(S)$ are (i) nonempty, (ii) not the whole space (R^S), and (iii) comprehensive.

How are marginal contributions computed in an NTU-game? Clearly, the formula $P(N, V) - P(N \setminus \{i\}, V)$ leads to interpersonal comparison of utilities, since all players use the same number $P(N, V)$. It is thus appropriate to use it only when the weights of the players are equal.

This suggests the following construction: First, use the potential function approach to obtain, for each collection $w = (w^i)_i$ of positive weights, a solution x_w. Second, require that w represent the appropriate marginal rates of efficient substitution between the players, *at* x_w. This is a standard procedure for obtaining solutions in the nontransferable utility case. One first 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 $w \to x_w \to w'$), and, most important, independent of rescaling utilities (for each player separately).

Formally, let $w = (w^i)_i$ 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

$$(6.1) \qquad \left(w^i D^i P_w(N, V) \right)_{i \in N} \in \operatorname{bd} V(N)$$

where $\operatorname{bd} V(N)$ denotes the (Pareto efficient) boundary of the set $V(N)$ (recall assumption (iii)); without loss of generality, let again $P_w(\phi, V) = 0$. Thus, (6.1) is the exact counterpart in the NTU-case of (2.1) (or, (5.1)) in the TU-case: the vector of (rescaled) marginal contributions is efficient.[20]

THEOREM 6.2: *For every collection $w=(w^i)_i$ of positive weights there exists a unique w-potential function on the class of NTU-games. Moreover, the resulting solution function, associating 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 *egalitarian solutions* have been studied by Myerson (1980) and Kalai and Samet (1985). They may be viewed as the first step in the construction of the *Harsanyi (NTU-) value* (cf. Harsanyi (1963)). Note that the w-egalitarian solution coincides with the w-Shapley value for TU-games.

PROOF: Assumptions (i)–(iii) above imply that, for each S, the set $V(S)$ is bounded from above in any strictly positive direction, hence $\operatorname{bd} V(S)$ intersects any such line in a unique point. The proof proceeds by induction as in Theorem A, yielding $P_w(N,V)$ as the unique t such that $y+tw\in \operatorname{bd} V(N)$, where $y^i=-w^iP_w(N\setminus\{i\},V)$ for all $i\in N$. This construction is easily seen to give the w-egalitarian solution (see Kalai and Samet (1985)). *Q.E.D.*

A (Pareto) efficient payoff vector $x\in \operatorname{bd} V(N)$ is called *w-utilitarian* (for given positive weights $w=(w^i)_i$) if it maximizes the sum of the utilities over the feasible set $V(N)$, rescaled according to w:

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

Finally, x is a *Harsanyi* (1963) *value* if there exist weights w such that it is simultaneously w-egalitarian and w-utilitarian. We thus finally obtain the following theorem.

THEOREM D: *For every NTU-game (N,V), the payoff vector $x\in R^N$ is a Harsanyi value[21] of (N,V) if and only if there exist positive weights $w=(w^i)_i$ such that*

$$\frac{1}{w^i}x^i=D^iP_w(N,V), \quad \textit{for all } i\in N, \textit{ and}$$

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

[20] It is easy to check that, for TU-games (represented in NTU form as above), we obtain the same w-potential of Section 5.

[21] We actually obtain only the nondegenerate Harsanyi values (i.e., those corresponding to positive weights). Again, this can be easily fixed (see Remark 5.3).

Since the egalitarian solutions are obtained by the potential approach, one is naturally led to investigate whether they can in fact be also characterized by the preservation of differences principle, and by consistency. It should come as no surprise that these alternative approaches apply just as in the TU-case.

In what concerns the preservation of differences principle, for every given set of weights w, it is easily seen that it uniquely determines the w-egalitarian solution. Of course, the efficiency condition (3.2) becomes $x(N) \in \mathrm{bd}\, V(N)$; as for (3.3) and its weighted counterpart (5.4), they remain unchanged. The application of this principle when the utilities are not transferable and are only assumed to be comparable (by the weights w) seems to yield a good foundation for the egalitarian solutions (and thus, a fortiori, for the Harsanyi value). In particular, there is no need for "dividends" (cf. Harsanyi (1963)) which accumulate (and which may sometimes be negative); instead, one considers *only the relative positions* of the players as more and more coalitions are taken into account.

Next, consider *consistency*. The definition of the reduced game is the natural extension of (4.1) to the NTU-case (see Maschler and Owen (1986)). For every solution function ϕ, a game (N, V), and a subset $T \subset N$, the reduced game (T, V_T^ϕ) is defined by:

$$V_T^\phi(S) = \left\{ x \in R^S : \left(x, \left(\phi^i(S \cup T^c, V) \right)_{i \in T^c} \right) \in V(S \cup T^c) \right\}$$

for all $S \subset T$. Thus, $V_T^\phi(S)$ is the S-section of $V(S \cup T^c)$ when the coordinates of all players outside T are fixed at their solution imputation (in the $S \cup T^c$ subgame). The solution function ϕ is consistent if

$$\phi^j(T, V_T^\phi) = \phi^j(N, V)$$

for all games (N, V) and all $j \in T \subset N$. Note that, in the case of an NTU-game that corresponds to a TU-game, this definition coincides with the one used in Sections 4 and 5.

From now on we restrict our attention to *nonlevel* NTU-games, defined as those games (N, V) such that, for all $S \subset N$, if $x, y \in \mathrm{bd}\, V(S)$ and $x \geqslant y$, then $x = y$. This condition has been widely used in the study of NTU-games; it means that strong and weak efficiency coincide. We use it here in order to guarantee that, if $x = (x^i)_{i \in N}$ is efficient, then $x_S = (x^i)_{i \in S}$ is efficient in the corresponding reduced game.

The main result we obtain is the following:

THEOREM E: *Let ϕ be a solution function on nonlevel NTU-games and let w be a collection of positive weights. Then*:

(i) *ϕ is consistent*; *and*

(ii) *ϕ is the w-proportional solution for two-person TU-games*;

if and only if ϕ is the w-egalitarian solution.

REMARK 6.3: The initial conditions (ii) apply to two-person *TU*-games only. It is therefore noteworthy that the consistency requirement enables one to obtain the w-egalitarian solution without having to assume it for two-person *NTU*-games, where it is a much stronger primitive postulate than in the two-person TU-case. Moreover, it will be seen in Lemma 6.9 below that, if one assumes that ϕ is the w-egalitarian solution for two-person *NTU*-games (i.e., two-person bargaining problems) then consistency easily implies that it coincides with the w-egalitarian solution for all games (this is exactly as in the TU-case: see Theorems B and 5.7).

REMARK 6.4: Theorems B′ and C show that the initial conditions (ii) follow from more primitive postulates using consistency. In particular, the weights are obtained *endogenously*. For example, Theorems C and E imply the following theorem:

THEOREM 6.5: *Let ϕ be a solution function on nonlevel NTU-games. Then*:
(i) *ϕ is consistent*;
(ii) *for two-person TU-games*: (a) *ϕ is efficient*, (b) *ϕ is TU-invariant*, (c) *ϕ is monotonic*;
if and only if there exist positive weights w such that ϕ is the w-egalitarian solution.

REMARK 6.6: Since Theorem E and Theorem 6.5 characterize only the w-egalitarian solutions, it follows that one cannot require in addition invariance under *independent* utility rescalings (i.e., where players may rescale differently). This has been shown by Maschler and Owen (1986), who also discuss a weaker form of consistency.

PROOF OF THEOREM E: It will follow from Lemmata 6.7–6.9 below.

LEMMA 6.7: *The w-egalitarian solution is consistent.*

PROOF: The proof follows the same lines as Proposition 4.5, since it is generated by a potential function. *Q.E.D.*

LEMMA 6.8: *There exists a unique consistent solution that is w-egalitarian on two-person NTU-games.*

PROOF: One may use an argument similar to that in Theorem B. We will instead provide an alternative direct proof (which also works for Theorem B, and Theorem 5.7). First, as in the TU-case, one shows by induction that if a solution function is efficient for $n=2$ and is consistent, then it is efficient for all n.

Let ϕ and ψ be two solution functions satisfying the hypothesis, and assume by induction that they coincide for all games of at most $n-1$ players. Let (N, V) be an n-player game, and let $i, j \in N$, $i \neq j$. Consider the two reduced games $(\{i, j\}, V^{\phi}_{\{i,j\}})$ and $(\{i, j\}, V^{\psi}_{\{i,j\}})$, which we will denote for short by V^{ϕ} and V^{ψ}. They coincide for singletons (by induction, since only $n-1$ players matter); therefore, since ϕ is w-egalitarian for two-person games, we have $\phi^i(V^{\phi}) \gtreqless \phi^i(V^{\psi})$ if and only if $\phi^j(V^{\phi}) \gtreqless \phi^j(V^{\psi})$ (both lie on the same strictly positive ray). Now

$\phi = \psi$ for two-person games, and both ϕ and ψ are consistent; therefore

$$\phi^i(V) = \phi^i(V^\phi) \gtreqless \phi^i(V^\psi) = \psi^i(V^\psi) = \psi^i(V)$$

if and only if, similarly,

$$\phi^j(V) \gtreqless \psi^j(V).$$

This applies to any two players, and both $\phi(V)$ and $\psi(V)$ are efficient; thus $\phi(V) = \psi(V)$. *Q.E.D.*

LEMMA 6.9: *If ϕ is consistent and w-proportional for two-person TU-games, then it is w-egalitarian for two-person NTU-games.*

PROOF: Let $(\{i, j\}, V)$ be a two-person NTU-game. Denote $z = (z^i, z^j) = \phi(V)$. We will show that z is the w-egalitarian solution of V, i.e., $w^i(z^i - \alpha^i) = w^j(z^j - \alpha^j)$, where $\alpha^i = \sup\{x^i\colon x^i \in V(\{i\})\}$ and $\alpha^j = \sup\{x^j\colon x^j \in V(\{j\})\}$.

Define a three-person NTU-game $(\{i, j, k\}, U)$ as follows:

$$U(S) = \begin{cases} V(S), & \text{if } S \subset \{i, j\}, \\ \left\{x \in R^S\colon \sum_{l \in S} x^l \leqslant \sum_{l \in S} \alpha^l\right\}, & \text{otherwise} \end{cases}$$

where we put $\alpha^k = 0$.

Let $(y^i, y^j, y^k) = \phi(U)$.

Consider the reduced games U^ϕ_{-i}, U^ϕ_{-j}, U^ϕ_{-k}. It can be easily checked that they all correspond to TU-games, which we will denote by u_{-i}, u_{-j}, and u_{-k}, respectively:[22]

$$\begin{aligned} &u_{-i}(j) = z^j, \quad u_{-i}(k) = 0, \quad u_{-i}(jk) = y^j + y^k, \\ &u_{-j}(i) = z^i, \quad u_{-j}(k) = 0, \quad u_{-j}(ik) = y^i + y^k, \\ &u_{-k}(i) = \alpha^i, \quad u_{-k}(j) = \alpha^j, \quad u_{-k}(ij) = y^i + y^j. \end{aligned}$$

Therefore, since ϕ is consistent, and is the w-proportional solution for two-person TU-games, we obtain:

$$\begin{aligned} &w^j(y^j - z^j) = w^k(y^k - 0), \\ &w^i(y^i - z^i) = w^k(y^k - 0), \\ &w^i(y^i - \alpha^i) = w^j(y^j - \alpha^j), \end{aligned}$$

from which the required equality follows. *Q.E.D.*

[22] We write $u(i)$, $u(ij)$, etc., instead of $u(\{i\})$, $u(\{i, j\})$, ...

School of Mathematical Sciences, Beverly and Raymond Sackler Faculty of Exact Sciences, Tel Aviv University, 69978 Tel-Aviv, Israel
and
Department of Economics, Harvard University, Cambridge, MA 02138, U.S.A.

Manuscript received July, 1985; revision received July, 1988.

APPENDIX

Proof of Theorem C: In view of Theorem 5.7, we have to prove that (i) and (ii) imply that ϕ is a proportional solution for two-person games: namely, there exist positive weights $\{w^i\}_i$ such that (5.6) holds.

Let $i \neq j$, and consider the $\{i, j\}$-unanimity game $(\{i, j\}, u_{\{i,j\}})$: $u_{\{i,j\}}(i) = u_{\{i,j\}}(j) = 0$ and $u_{\{i,j\}}(i, j) = 1$. Put

$$\text{(A.1)} \qquad \begin{aligned} \alpha^i_{\{i,j\}} &= \phi^i(\{i,j\}, u_{\{i,j\}}), \quad \text{and} \\ \beta^i_{\{i,j\}} &= -\phi^i(\{i,j\}, -u_{\{i,j\}}). \end{aligned}$$

By efficiency (ii) (a), we have

$$\text{(A.2)} \qquad \alpha^i_{\{i,j\}} + \alpha^j_{\{i,j\}} = \beta^i_{\{i,j\}} + \beta^j_{\{i,j\}} = 1.$$

A two-person game $(\{i, j\}, v)$ is *essential* if the "surplus" $\sigma = v(i, j) - v(i) - v(j)$ is not zero; it is then strategically equivalent to either $u_{\{i,j\}}$ or $-u_{\{i,j\}}$, depending on the sign of σ. By TU-invariance (ii)(b) and (A.1) we therefore obtain, for any essential game $(\{i, j\}, v)$,

$$\text{(A.3)} \qquad \phi^i(\{i,j\}, v) = v(i) + \delta^i[v(i,j) - v(i) - v(j)],$$

where $\delta^i = \alpha^i_{\{i,j\}}$ if the surplus $\sigma > 0$ and $\delta^i = \beta^i_{\{i,j\}}$ if $\sigma < 0$.

By monotonicity (ii)(c), the coefficient δ^i of $v(i, j)$ must be positive; thus

$$\text{(A.4)} \qquad \alpha^i_{\{i,j\}} > 0, \qquad \beta^i_{\{i,j\}} > 0.$$

Let $(\{i, j\}, v)$ be a two-person inessential game: $v(i, j) = v(i) + v(j)$. Let v^+_ε, v^-_ε be defined as follows ($\varepsilon > 0$):

$$\begin{aligned} v^+_\varepsilon(i,j) &= v(i,j) + \varepsilon, \\ v^-_\varepsilon(i,j) &= v(i,j) - \varepsilon, \\ v^+_\varepsilon(i) &= v^-_\varepsilon(i) = v(i), \\ v^+_\varepsilon(j) &= v^+_\varepsilon(j) = v(j). \end{aligned}$$

Monotonicity yields:

$$\begin{aligned} \phi^i(\{i,j\}, v) &< \phi^i(\{i,j\}, v^+_\varepsilon) = v(i) + \varepsilon\alpha^i_{\{i,j\}}. \\ &> \phi^i(\{i,j\}, v^-_\varepsilon) = v(i) + \varepsilon\beta^i_{\{i,j\}}. \end{aligned}$$

As ε decreases to 0 we obtain $\phi^i(\{i, j\}, v) = v(i)$. Thus, formula (A.3) applies to inessential games as well.

In particular, we note that (A.3) and (A.4) imply

(A.5) $\quad \phi^i(\{i,j\}, v) - v(i)$ and $\phi^j(\{i,j\}, v) - v(j)$ have always the same sign $(+, -,$ *or* $0)$.

LEMMA A.6: *There exist positive numbers* $\{w^i\}_i$ *such that*

$$\alpha^i_{\{i,j\}} = \frac{w^i}{(w^i + w^j)} \quad \textit{for all} \quad i \neq j.$$

PROOF: Fix a player k, and define:

$$w^k = 1,$$

$$w^i = \frac{\alpha^i_{\{i,k\}}}{\alpha^k_{\{i,k\}}} \qquad \text{for} \quad i \neq k.$$

We have to show that

$$\text{(A.7)} \qquad \frac{\alpha^i_{\{i,j\}}}{\alpha^j_{\{i,j\}}} = \frac{w^i}{w^j} = \frac{\alpha^i_{\{i,k\}}}{\alpha^k_{\{i,k\}}} \cdot \frac{\alpha^k_{\{j,k\}}}{\alpha^j_{\{j,k\}}};$$

together with (A.2), it will imply our result.

Let (N, v) be the $\{i, j, k\}$-unanimity game: $N = \{i, j, k\}, v(N) = 1$, $v(S) = 0$ for all $S \neq N$. Let $x = \phi(N, v)$ be its solution.

Consider now the reduced game $(N\setminus\{i\}, v_{-i})$, where $v_{-i} \equiv v^{\phi}_{N\setminus\{i\}}$:

$$v_{-i}(j,k) = v(i,j,k) - \phi^i(\{i,j,k\}, v) = v(N) - x^i = x^j + x^k,$$

$$v_{-i}(j) = v(i,j) - \phi^i(\{i,j\}, v) = 0 - 0 = 0$$

(use (A.3), or note directly that $(\{i,j\}, v)$ is the null game, thus inessential). Similarly,

$$v_{-i}(k) = 0.$$

Consistency and (A.3) now imply that

$$\text{(A.8)} \qquad \begin{aligned} x^j &= \phi^j(N, v) = \phi^j(N\setminus\{i\}, v_{-i}) = \delta^j(x^j + x^k), \quad \text{and} \\ x^k &= \phi^k(N, v) = \phi^k(N\setminus\{i\}, v_{-i}) = \delta^k(x^j + x^k), \end{aligned}$$

where $\delta^j = \alpha^j_{\{j,k\}}$ or $\beta^j_{\{j,k\}}$, and $\delta^k = \alpha^k_{\{j,k\}}$ or $\beta^k_{\{j,k\}}$. Therefore x^j and x^k have the same sign $(+, -, \text{or } 0)$: recall (A.5) or (A.4).

This holds for any two players in N; therefore x^i, x^j, and x^k all have the same sign. Together with $x^i + x^j + x^k = v(N) = 1$, this yields $x^i, x^j, x^k > 0$. From (A.8) we obtain

$$\alpha^j_{\{j,k\}} = \delta^j = \frac{x^j}{x^j + x^k} \quad \text{and}$$

$$\alpha^k_{\{j,k\}} = \delta^k = \frac{x^k}{x^j + x^k}.$$

This is true for any two players in N, from which (A.7) follows immediately. *Q.E.D.*

LEMMA A.9: *There exist positive numbers* $\{u^i\}_i$ *such that*

$$\beta^i_{\{i,j\}} = \frac{u^i}{u^i + u^j} \quad \textit{for all } i \neq j.$$

PROOF: The same as for Lemma A.6, using $-v$ instead of v. *Q.E.D.*

LEMMA A.10: *There exists a positive number* ρ *such that* $u^i = \rho w^i$ *for all* i.

PROOF: Let i, j, k be all different, and consider the following game: $(\{i, j, k\}, v)$, with $v(i, k) = w^i + w^k$, $v(j, k) = w^j + w^k$, $v(i, j, k) = c$ for some c satisfying $w^k < c < w^i + w^j + w^k$, and $v(S) = 0$

otherwise. Let (x^i, x^j, x^k) be its ϕ-solution. Consider the reduced games v_{-i}, v_{-j}, v_{-k}:

$$v_{-i}(j,k) = v(i,j,k) - \phi^i(\{i,j,k\}, v) = v(i,j,k) - x^i = x^j + x^k,$$

$$v_{-i}(j) = v(i,j) - \phi^i(\{i,j\}, v) = 0$$

(since $(\{i,j\}, v)$ is the null game), and

$$v_{-i}(k) = v(i,k) - \phi^i(\{i,k\}, v) = \phi^k(\{i,k\}, v)$$

$$= 0 + \frac{w^k}{w^k + w^i}\left[(w^i + w^k) - 0 - 0\right] = w^k$$

(recall (A.3) and Lemma A.6). Similarly:

$$v_{-j}(i,k) = x^i + x^k,$$

$$v_{-j}(i) = 0,$$

$$v_{-j}(k) = w^i.$$

Finally:

$$v_{-k}(i,j) = x^i + x^j,$$

$$v_{-k}(i) = v(i,k) - \phi^k(\{i,k\}, v) = w^i,$$

$$v_{-k}(j) = v(j,k) - \phi^k(\{j,k\}, v) = w^j.$$

Assume $x^k \leqslant w^k$. By (A.5) applied to v_{-i} we obtain $x^j \leqslant 0$; and by considering v_{-j}, we obtain $x^i \leqslant 0$. But this contradicts $x^i + x^j + x^k = c > w^k$. Therefore $x^k > w^k$, $x^j > 0$, $x^i > 0$, and moreover

$$\frac{x^j - 0}{x^k - w^k} = \frac{\alpha^j_{\{j,k\}}}{\alpha^k_{\{j,k\}}} = \frac{w^j}{w^k} \qquad (\text{from } v_{-i}),$$

$$\frac{x^i - 0}{x^k - w^k} = \frac{w^i}{w^k} \qquad (\text{from } v_{-j}).$$

This implies

$$\text{(A.11)} \qquad \frac{x^i}{x^j} = \frac{w^i}{w^j}.$$

Next, consider v_k. Since $x^k > w^k$ and $x^i + x^j + x^k = c < w^i + w^j + w^k$, we have $x^i + x^j < w^i + w^j$, and

$$\frac{x^i - w^i}{x^j - w^j} = \frac{\beta^i_{\{i,j\}}}{\beta^j_{\{i,j\}}} = \frac{u^i}{u^j}.$$

But (A.11) implies $(x^i - w^i)/(x^j - w^j) = w^i/w^j$; hence $w^i/w^j = u^i/u^j$, or $w^i/u^i = w^j/u^j$. This holds for all $i \neq j$, proving our claim. *Q.E.D.*

The three lemmata and (A.3) thus imply (5.6): ϕ is the w-proportional solution for two-person games, completing the proof of the theorem. *Q.E.D.*

REFERENCES

AUMANN, R. J., AND M. MASCHLER (1985): "Game Theoretic Analysis of a Bankruptcy Problem from the Talmud," *Journal of Economic Theory*, 36, 195–213.

BALINSKY, M. L., AND H. P. YOUNG (1982): *Fair Representation*. New Haven, CT: Yale University Press.

AUMANN, R. J., AND L. S. SHAPLEY (1974): *Values of Non-Atomic Games*. Princeton, NJ: Princeton University Press.

BILLERA, L. J., AND D. C. HEATH (1982): "Allocation of Shared Cost: A Set of Axioms Yielding a Unique Procedure," *Mathematics of Operations Research*, 7, 32–39.

D'ASPREMONT, C., A. JACQUEMIN, AND J.-F. MERTENS (1984): "A Measure of Aggregate Power in Organizations," CORE DP-8416.

DAVIS, M., AND M. MASCHLER (1965): "The Kernel of a Cooperative Game," *Naval Research Logistics Quarterly*, 12, 223–259.

HARSANYI, J. C. (1959): "A Bargaining Model for the Cooperative n-Person Game," in *Contributions to the Theory of Games IV* (*Annals of Mathematics Studies* 40), ed. by A. W. Tucker and R. D. Luce. Princeton: Princeton University Press, pp. 325–355.

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 (1988): "The Potential of the Shapley Value," in *The Shapley Value, Essays in Honor of Lloyd S. Shapley*, ed. by A. E. Roth. Cambridge: Cambridge University Press, pp. 127–137.

KALAI, E., AND D. SAMET (1985): "Monotonic Solutions to General Cooperative Games," *Econometrica*, 53, 307–327.

——— (1987): "On Weighted Shapley Values," *International Journal of Game Theory*, 16, 205–222.

LENSBERG, T. (1982): "Stability and the Nash Solution," mimeo, Norwegian School of Economics and Business Administration, Bergen. *Journal of Economic Theory*, forthcoming.

MASCHLER, M., AND G. OWEN (1986): "The Consistent Shapley Value for Hyperplane Games," mimeo, RM-74, The Hebrew University, Jerusalem.

MERTENS, J.-F. (1980): "Values and Derivatives," *Mathematics of Operations Research*, 4, 521–552.

MIRMAN, L. J., AND Y. TAUMAN (1982): "Demand Compatible Equitable Cost-Sharing Prices," *Mathematics of Operations Research*, 7, 40–56.

MOULIN, H. (1985): "The Separability Axiom and Equal-Sharing Methods," *Journal of Economic Theory*, 36, 120–148.

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

OSTROY, J. M. (1984): "A Reformulation of the Marginal Productivity Theory of Distribution," *Econometrica*, 52, 599–630.

OWEN, G. (1968): "A Note on the Shapley Value," *Management Science*, 14, 731–732.

PELEG, B. (1985): "An Axiomatization of the Core of Cooperative Games Without Side Payments," *Journal of Mathematical Economics*, 14, 203–214.

——— (1986): "On the Reduced Game Property and its Converse," *International Journal of Game Theory*, 15, 187–200.

SHAPLEY, L. S. (1953a): "Additive and Non-Additive Set Functions," Ph.D. Thesis, Princeton University.

——— (1953b): "A Value for n-Person Games," in: *Contributions to the Theory of Games II* (*Annals of Mathematics Studies* 28), ed. by H. W. Kuhn and A. W. Tucker. Princeton: Princeton University Press, 307–317.

——— (1981): "Comments on R. D. Banker's 'Equity Considerations in Traditional Full Cost Allocation Practices: An Axiomatic Perspective'," in *Joint Cost Allocations*, ed. by S. Moriarity. Norman, OK: University of Oklahoma, pp. 131–136.

SOBOLEV, A. I. (1975): "The Characterization of Optimality Principles in Cooperative Games by Functional Equations" (in Russian), in *Mathematical Methods in Social Sciences*, ed. by N. N. Vorobev., Vilnius, 94–151.

THOMSON, W. (1984): "Monotonicity, Stability and Egalitarianism," *Mathematical Social Sciences*, 8, 15–28.