Bureau d'économie théorique et appliquée (BETA)
UMR 7522

Documents de travail

«Data Games : Sharing public goods with exclusion»

Auteurs

Pierre Dehez, Daniela Tellone

Document de Travail n° 2009 - **31**

Octobre 2009

Faculté des sciences économiques et de gestion
Pôle européen de gestion et d'économie (PEGE)
61 avenue de la Forêt Noire
F-67085 Strasbourg Cedex

Secétariat du BETA
Géraldine Manderscheidt
Tél. : (33) 03 68 85 20 69
Fax : (33) 03 68 85 20 70
g.manderscheidt@unistra.fr
http://cournot2.u–strasbg.fr/beta

cnrs

Nancy-Université
Université Nancy 2

UNIVERSITÉ DE STRASBOURG

DATA GAMES

SHARING PUBLIC GOODS WITH EXCLUSION

Pierre Dehez* and Daniela Tellone**

October 2009

ABSTRACT

A group of firms consider collaborating on a project which requires a combination of elements which are owned by some of them. These elements are nonrival but excludable goods i.e. public goods with exclusion like for instance knowledge, data or informations, patents or copyrights. We address the question of how firms should be compensated for the goods they own. We shown that this problem can be framed as a cost sharing game to which standard allocation rules like the Shapley value, the nucleolus or accountings formulas can be applied and compared. Our analysis is inspired by the need for a cooperation between European chemical firms within the regulation program REACH which requires them to submit by 2018 a detailed analysis of the substances they produce or import.

JEL: C71, D46, M41

Keywords: cost sharing, Shapley value, core, nucleolus

This is a revised version of CORE Discussion Paper 2008-10. The authors are grateful to Alexandre Bailly from TRASYS for having drawn their attention to the cost sharing problem faced by the European chemical industry within EU regulation "REACH". Thanks are due to Pier Mario Pacini, Eve Ramaekers and Gisèle Umbhauer for useful comments on earlier versions. Financial support from the Belgian *Interuniversity Pole of Attraction* (PAI) program is gratefully acknowledged.

* CORE, University of Louvain, Voie du roman pays 34, 1348 Louvain-la-Neuve, Belgium.
Tel: 0032-10-472934 – Fax: 0032-10-474301 – Email: pierre.dehez@uclouvain.be

** CEREC, Facultés universitaires Saint-Louis, Boulevard du jardin botanique 38, 1000 Bruxelles, Belgium.
Tel: 0032-2-2117939 – Fax: 0032-2-2117873 – Email: tellone@fusl.ac.be

1. Introduction

Imagine the following situation. A group of firms consider collaborating on a project which requires a combination of elements which are owned by some of them. These elements are nonrival but excludable goods i.e. public goods with exclusion like for instance knowledge, data or informations, patents or copyrights.[1] The question is not to share the cost of these goods because they are already available. Their costs are sunk. The question is instead to possibly compensate the firms who own these goods, knowing that any *additional* cost would be shared independently. This problem can be framed as a cost sharing game in which the value of the grand coalition is zero. These are *compensation games* to which standard cost allocation rules can be applied, taking possibly into account the relative size of the players by assigning weights to each player e.g. firms' market shares or production volumes.

In what follows we shall use the term "data" for expository reasons and talk about "data (sharing) games". Data games are defined on the basis of the *replacement cost* of the goods involved e.g. the present cost of acquiring the data or developing alternative technologies.

The origin of the present paper is the cost sharing problem faced by the EU chemical industry, following the requirements imposed by the European Commission under the achronym "REACH". Each European firm must indeed submit by 2018 a detailed analysis on the substances it produces or imports, with a particular attention to toxicity. This is a huge program. Indeed there are about 30,000 substances and an average of 100 parameters for each substance! The European Commission is encouraging firms to cooperate, in particular by sharing the data they have gathered over the past.[2] Beyond the cost reducing motivation, the idea is also to avoid unecessary replications of analysis involving living beings.

To illustrate the problem, consider the simple case of three firms and a single data owned by one firm. How much should that firm receive as a compensation? Alternatively, how much should the other two firms pay? Asking them to cover the total value, say K, would not be fair. Instead they should be asked to contribute *one third* of K and the firm owning the data would then receive $2K/3$. If there were n firms and a single data owned by t firms, the $n–t$ firms without data should contribute K/n and the t firms owning the data would then receive

[1] To quote Drèze (1980, p.6): "Public goods with exclusion are public goods … the consumption of which by individuals can be controlled, measured and subjected to payment or other contractual limitations."

[2] Guidance for the implementation of REACH is provided by the *European Chemical Agency* (ECHA). For more details, see http://echa.europa.eu/reach_en.asp. As far as data sharing is concerned, only ad hoc methods are being proposed at this stage.

$K/t - K/n$: the cost of the complete dataset is uniformly allocated among all players and the cost of each data is uniformly redistributed to the players owning it. This turns out to be exactly the Shapley value of the associated cost game and it coincides with the nucleolus if $t = 1$. If instead $t \geq 2$, the nucleolus actually excludes any compensation, a property of the core.

Data games are essential, monotone decreasing and subadditive though they are generally not concave. They can be decomposed into a sum of "elementary" data games, one for each data. This is a useful feature – especially within a context involving a large number of data like REACH – for computing the Shapley value, possibly weighted, which is then the sum of the values attached to the elementary data games. Data games have a nonempty core as it always contains the no compensation allocation: no coalition can object when no one is asked to pay. In some situations the core is reduced to that trivial allocation, like for instance when each data is owned by at least two players.

Additional results are obtained when individual datasets form a *partition* of the complete dataset, a situation which fits perfectly the case of patents or copyrights.[3] Partition data games are concave and their core is a regular simplex. As a consequence, the nucleolus and the Shapley value coincide with the core centroid which is then simply defined as the average of its vertices. Surprisingly, simple accounting rules such as equal charge or separable costs remaining benefits (SCRB), when applied to partition data games, result in allocations which coincide with the Shapley value and thereby belong to the core.

A special attention is devoted to situations where there are reasons to treat players asymmetrically, independently of the initial distribution of data. Firms engaged into a joint project may have different size as measured for instance by their market shares. Such situations can be accomodated by using the asymmetric Shapley value based on weights assigned to the players. The case where some player are assigned a *zero* weight is of particular importance within a context of data sharing. Some players may indeed own data while not being otherwise part of the joint project. This is for instance the case within REACH where independent laboratories, like university laboratories, own relevant data on chemical substances but are not part of the submission process.

[3] The particular case where datasets are nested gives rise to a type of "reverse" airport game. See Dehez (2008).

The paper is organized as follows. Cost sharing games and their characterization in terms of elementary cost games are introduced in Section 2 together with the concept of marginal cost vector. Section 3 is devoted to the definition and properties of data games. Their core and Shapley value are analyzed in the subsequent sections, including the asymmetric (weighted) Shapley value. Accounting methods are defined and applied to data games in Section 7. Concluding remarks are offered in the last section.

2. Cost games

A set $N = \{1,\ldots,n\}$ of players, $n \geq 2$, have a common project and face the problem of dividing its cost. The cost of realizing the project to the exclusive benefit of any coalition is also known. This defines a real-valued function C on the subsets of N with $C(\emptyset) = 0$.[4] A pair (N,C) defines a *cost game* and the cost to be divided is $C(N)$. A *sharing rule* φ associates a cost allocation $y = \varphi(N,C)$ to any cost game (N,C) such that $\Sigma_i y_i = C(N)$. The dual (N,C^*) of the game (N,C) is defined by $C^*(S) = C(N) - C(N\backslash S)$. The *surplus* game (N,v) associated with the game (N,C) is defined by:

$$v(S) = \sum_{i \in S} C(i) - C(S)$$

Cost and surplus allocations y and x are linked by the n identities $y_i + x_i = C(i),\ i = 1,\ldots,n.$

Notation: The letters $n, s, t,\ldots$ will denote the size of the sets $N, S, T,\ldots$ For a vector y, $y(S)$ will denote the sum over S of its coordinates. Coalitions will be identified as $ijk\ldots$ instead of $\{i,j,k\}\ldots$ For any set S, $S\backslash i$ will denote the coalition from which player i has been removed.

The set $G(N)$ of functions defined on the subsets of some finite set N is a vector space and the collection of $2^n - 1$ *unanimity* games

$$\begin{aligned} u_T(S) &= 1 \quad \text{if } T \subset S \\ &= 0 \quad \text{if not} \end{aligned}$$

defined for all $T \subset N$, $T \neq \emptyset$, forms a basis of $G(N)$.[5] Here we shall use the basis formed by the collection of $2^n - 1$ games

$$\begin{aligned} e_T(S) &= 1 \quad \text{if } S \cap T \neq \emptyset \\ &= 0 \quad \text{if not} \end{aligned}$$

defined for all $T \subset N$, $T \neq \emptyset$.

[4] See for instance Young (1985b) or Moulin (1988, 2003).

[5] These games were introduced by Shapley in 1953 to prove existence and uniqueness of the value.

These games have been introduced by Kalai and Samet (1987) as *duals* of the unanimity games: $e_T = u_T*$.[6]

Given a coalition S and a player i in S, the *marginal cost* of player i to coalition S is defined by $C(S) – C(S\backslash i)$. Marginal costs play an central role in cost allocation. Let $\prod_n$ be the set of all players' permutations. To each permutation $\pi = (i_1,...,i_n) \in \prod_n$ we associate the vector of marginal costs $\mu(\pi)$ whose elements are given by:

$$\mu_{i_1}(\pi) = C(i_1) - C(\varnothing) = C(i_1)$$

$$\mu_{i_k}(\pi) = C(i_1,...,i_k) - C(i_1,...,i_{k-1}) \quad (k = 2,...,n)$$

It is easily seen that it defines a cost allocation: $\sum_{i \in N} \mu_i(\pi) = C(N)$.

A cost game (N,C) is *essential* if $C(N) < \sum_{i \in N} C(i)$. It is *subadditive* if $S \cap T = \varnothing$ implies $C(S \cup T) \leq C(S) + C(T)$. It is *concave* if $C(S \cup T) \leq C(S) + C(T) - C(S \cap T)$ for all S and T. Hence concavity implies subadditivity: concavity is a stronger form of economies of scale than subadditivity. Equivalently, a cost game (N,C) is concave if for all i, the marginal costs $C(S) – C(S\backslash i)$ are non increasing with respect to set inclusion.[7] The surplus game associated with a subadditive (resp. concave) cost game is superadditive (resp. convex) and the total surplus to be divided is positive if the cost game is essential.

The class of concave cost games is interesting because most solution concepts agree, as was proved by Shapley (1971) and Maschler, Peleg and Shapley (1972, 1979):

(i) the core is the unique stable set (in the sense of von Neumann and Morgentern);

(ii) the core coincides with the bargaining set (with respect to the grand coalition);

(iii) the kernel and the nucleolus coincide;

(iv) the Shapley value belongs to the core and is centrally located.[8]

3. Data games

We denote by M_i the set of data owned by player i and by M_S the set of data owned by coalition S:

$$M_S = \bigcup_{i \in S} M_i$$

[6] The elementary games e_T are normalized *fixed costs games*: coalitions containing players from T entail a fixed cost equal to 1. They are used in Dehez (2009) to characterize the weighted Shapley value in terms of the allocation of fixed costs, along the lines suggested by Shapley (1981b).

[7] See Shapley (1971).

[8] More precisely, the Shapley value is the center of gravity of core's vertices, accounting for multiplicity.

where $M_N = M$ is the complete dataset and $M_i \neq M$ for some $i \in N$. Hence players may own no data ($M_i = \varnothing$) or own the complete dataset ($M_i = M$). We denote by d_h the cost of *reproducing* data $h \in M$, with $d_h > 0$ for all $h \in M$. The cost associated with a non empty coalition is the cost of acquiring the missing data:

$$C(S) = \sum_{h \in M \setminus M_S} d_h = d_0 - \sum_{h \in M_S} d_h \text{ for all } S \neq \varnothing \tag{1}$$

where $d_0 = \Sigma_{h \in M} d_h$. This defines a class of cost games that we call "data games". Because $C(N) = 0$, data games are pure "compensation" games.

In what follows we shall consider two examples involving three players and three data, with a common cost vector. Only the distribution of data among players will change. Player 1 will however be assumed to own no data in both examples.

Example 1 The game associated with the datasets $M_1 = \varnothing$, $M_2 = \{1,2\}$ and $M_3 = \{2,3\}$, and replacement cost vector $d = (6, 9, 12)$, is given by:

$C(1) = d_1 + d_2 + d_3 = d_0 = 27$

$C(2) = C(12) = d_3 = 12$

$C(3) = C(13) = d_1 = 6$

$C(23) = C(123) = 0$

As a matter of illustration, the marginal costs vector associated with the permutation $\pi = (3,1,2)$ is given by $\mu(\pi) = (6,0,-6)$.

Lemma 1 Data games are essential and subadditive. Furthermore they are monotonically *decreasing*: $S \subset T$ implies $C(S) \geq C(T)$.

Proof $M \neq M_i$ for some i implies $\sum_{i \in N} C(i) = nd_0 - \sum_{i \in N} \sum_{h \in M_i} d_h > 0$. Essentiallity then follows from $C(N) = 0$. To verify subadditivity, assume $S \cap T = \varnothing$. We then have:

$$C(S) + C(T) = 2d_0 - \sum_{h \in M_S} d_h - \sum_{h \in M_T} d_h = C(S \cup T) + d_0 - \sum_{h \in M_S \cap M_T} d_h \geq C(S \cup T)$$

Finally, if $S \subset T$ we have $M_S \subset M_T$ and $C(T) - C(S) = \sum_{h \in M_S} d_h - \sum_{h \in M_T} d_h \leq 0$. •

Let us now consider the case where datasets form a *partition* of M:

$M_i \cap M_j = \varnothing$ for all $i \neq j$

If $k_i = \Sigma_{h \in M_i} d_h$ denotes the value of the data owned by player i, the value of the data owned by a coalition S can be written as

$$\sum_{h \in M_S} d_h = \sum_{i \in S} \sum_{h \in M_i} d_h = \sum_{i \in S} k_i$$

Using (1), a "partition" data game (N,C) is then simply defined by:

$$C(S) = d_0 - \sum_{i \in S} k_i \qquad (2)$$

Lemma 2 Partition data games are concave.

Proof For any subsets S and T, we have $C(S \cup T) + C(S \cap T) - C(S) - C(T) = 0$ if they have a nonempty intersection. Otherwise, $C(S \cap T) = 0$ and $C(S \cup T) - C(S) - C(T) = -d_0 < 0$. •

Looking at marginal costs, we observe that $C(S) - C(S \backslash i) = - k_i$ for all $i \in S$ and all $S \neq \{i\}$ while $C(i) - C(\varnothing) = d_0 - k_i > - k_i$. Hence marginal costs are negative. They are constant when associated with *proper* coalitions and decreasing when associated with *singletons*. This confirms concavity.[9] The surplus game (N,v) associated with a partition data game (N,C) is *symmetric*: $v(S) = (s-1)d_0$.

Example 2 The following datasets $M_1 = \varnothing$, $M_2 = \{1\}$ and $M_3 = \{2,3\}$ form a partition. The game associated with the cost vector d = (6,9,12) is given by:

$C(1) = d_0 = 27$

$C(2) = C(12) = d_2 + d_3 = 21$

$C(3) = C(13) = d_1 = 6$

$C(23) = C(123) = 0$

Let $T_h = \{i \in N \mid h \in M_i\}$ denote the subset of players owning data h. In the partition case, the T_h's are singletons. An "elementary" data game (N,C_h) can be associated with each data h:

$$\begin{aligned} C_h(S) &= 0 \quad \text{if } S \cap T_h \neq \varnothing \\ &= d_h \quad \text{if } S \cap T_h = \varnothing \end{aligned}$$

for all $S \subset N$, $S \neq \varnothing$. Clearly data games can be decomposed as sums of elementary data games:

$$\sum_{h \in M} C_h(S) = \sum_{h \in M \backslash M_S} d_h = C(S)$$

[9] Notice that partition data games are *not* constant sum games: $C(S) + C(N \backslash S) = d_0$ for all $S \neq \varnothing$.

and elementary data games can be written in terms of fixed cost games:

$$C_h(S) = (1 - e_{T_h}(S))\, d_h \tag{3}$$

4. The core

An *imputation* y is an individually rational cost allocation:

$$y(N) = C(N) \text{ and } y(i) \le C(i) \text{ for all } i \in N$$

Data games being essential, the set of imputations is nonempty. Actually there are cost allocations which are better than the stand-alone costs for all players. The *core* is the set of imputations y against which no coalition can object

$$y(S) \le C(S) \text{ for all } S \subset N \tag{4}$$

i.e. no coalition pays more that its stand-alone cost.[10] In general, the core is a *convex polyhedron*, possibly empty, whose dimension does not exceed n–1. Following Shapley (1971), the core of a concave cost game is the convex hull of its marginal cost vectors.[11]

Because it *always* contains the trivial allocation 0 = (0, 0,…, 0) defined by the absence of compensation, the core of a data game as defined by (1) is nonempty. Indeed, $C(S) \ge 0$ for all $S \subset N$. For the game defined in Example 1, the core is the set of allocations (y_1, y_2, y_3) such that $y_1 + y_2 + y_3 = 0$, $0 \le y_1 \le 18$, $-6 \le y_2 \le 12$ and $-12 \le y_3 \le 6$.

Proposition 1 If *one and only one* player owns the complete dataset M, there are core allocations different from 0 and only that player is possibly compensated. If *every* data is owned by at least two players, the core reduces to {0} and the nucleolus reduces to 0.

Proof Let y be a core allocation. If $M_n = M$, (4) implies $y(N/i) \le C(N/i) = 0$ for all $i \ne n$. Combining this with $y(N) = 0$, we get $y_i \ge 0$ for all $i \ne n$ and $y_n \le 0$. If each data is owned by at least two players, $C(N/i) = 0$ for all i. Combining (4) with $y(N) = 0$, we get $y_i \ge 0$ for all i. This is possible only if $y = 0$. The nucleolus being contained in the core, it reduces to 0. [12]

[10] Equivalently, an allocation y belongs to the core *if and only if* $y(S) \ge C(N) - C(N \backslash S)$ for all $S \subset N$. There is *no cross-subsidization* in the sense that every coalition pays at least its marginal cost. See Faulhaber (1975).

[11] The core of a concave cost game coincides with the *Weber set* which is the set of all *random order values*. See Weber (1988).

[12] The nucleolus is a single-valued solution introduced by Schmeidler (1969). It is always defined and belongs to the core if nonempty.

If $M_n = M$ and $M_i \neq M$ for all $i \neq n$, the allocation $y_b = (b,b,\ldots,(1\text{-}n)b)$ where $b = C(N/n)/n > 0$ belongs to the core. Indeed,

$$\begin{aligned} y_b(S) &= sb < nb = C(N \setminus n) \leq C(S) && \text{if } n \notin S \\ &= (s-n)b \leq 0 = C(S) && \text{if } n \in S \end{aligned}$$

for all coalition $S \subset N$. ●

Remark 1 The absence of compensation results from the competition between the players owning the same data. It is in particular the case if two players own the complete dataset.

In the partition case, the game is concave and consequently the core is the convex hull of the marginal cost vectors associated with the n! players' permutations. Actually there are n *distinct* marginal costs vectors and the core is a *regular* simplex i.e. an equilateral triangle for n = 3, a regular tetrahedron for n = 4,…

Proposition 2 The core of a partition data game (2) is a regular simplex of full dimension whose n vertices are $v^1 = (d_0 - k_1, -k_2,\ldots, -k_n)$, $v^2 = (-k_1, d_0 - k_2,\ldots, -k_n)$, … , $v^n = (-k_1, -k_2,\ldots, d_0 - k_n)$.

Proof If player i is first in a given permutation he/she pays his or her cost $d_0 - k_i$. Otherwise he/she saves the cost k_i of the data he/she owns. Hence there are n distinct marginal costs vectors each with a multiplicity equal to $(n–1)!$ The vector v^i associated with the permutations where player i is first is defined by $v_i^i = d_0 - k_i$ and $v_j^i = -k_i$ for all $j \neq i$. The n marginal costs vectors $v^1,\ldots, v^n$ are *affinely independent*. Indeed assume that

$$\sum_{i=1}^{n} \beta_i v^i = 0 \text{ and } \sum_{i=1}^{n} \beta_i = 0$$

for some $\beta_1,\ldots,\beta_n$. We then have $\sum_{i=1}^{n} \beta_i k_i = \beta_j d_0$ for all j implying $\beta_i = 0$ for all $i = 1,\ldots n$. This ensures that the core is a simplex. Its vertices are connected to each other by line segments of identical length $2^{1/2}d_0$. This implies its regularity. Positivity of d_0 ensures its full dimensionality. ●

Remark 2 Translating and normalizing (by adding the vector $(k_1, k_2,\ldots, k_n)$ and dividing by d_0) the core is transformed into the standard unit simplex. Applying this transformation to the partition data games defined by the vector $(k_1,\ldots,k_n)$, we obtain the (strategically) equivalent "constant" cost game defined by $C(S) = 1$ for all $S \neq \varnothing$. [13]

---

[13] Two games (N,C) and (N,C') are *strategically equivalent* if $C'(S) = aC(S) + b(S)$ for some $a > 0$ and $b \in \mathbb{R}^n$.

Remark 3 Looking at an elementary data game, there are two possibilities: either only one player owns the data and we fall in the partition case, or more than one player owns the data and the core reduces to 0.

Being a regular simplex, the core's centroid $\bar{y}$ is simply the average of its vertices:[14]

$$\bar{y}_i = \frac{d_0}{n} - k_i \quad (i = 1, \ldots, n) \tag{5}$$

and it coincides with least core and the nucleolus.[15] The only players to be compensated are those endowed with a dataset whose value exceeds the per capita cost of the complete data set. We shall see later that it also coincides with the Shapley value.

Marginal costs vectors in example 2 are given by v^1 = (27,–6,–21), v^2 = (0,21,–21) and v^3 = (0,–6,6). The resulting core's centroid is given by $\bar{y} = (9, 3, -12)$.

5. The Shapley value

The (symmetric) Shapley value of a cost game (N,C) is the average marginal cost vector:

$$\varphi(N, C) = \frac{1}{n!} \sum_{\pi \in \Pi} \mu(\pi) \tag{6}$$

It is the unique *additive* allocation rule which satisfies *symmetry* (players with identical marginal costs are *substitutes* and pay the same amount) and *dummy* (players with zero marginal costs are *dummies* and pay nothing). Additivity, efficiency, symmetry and dummy are the original axioms introduced by Shapley (1953, 1981a).[16]

The value defines an imputation for subadditive cost games and belongs to the core of concave cost games. Because data games can be written as sums of elementary games, computation of the value is straightforward as a result of additivity.

Proposition 3 The Shapley value of a data game (N,C) defined by the datasets ($M_1,\ldots,M_n$) and data costs ($d_1,\ldots,d_m$) is given by:

$$\varphi_i(N, C) = \frac{d_0}{n} - \sum_{h \in M_i} \frac{d_h}{t_h} \quad (i = 1, \ldots, n) \tag{7}$$

where t_h denotes the number of players owning data h.

[14] See Gonzales-Diaz and Sanchez Rodriguez (2007) for a definition of the core centroid (or barycenter).

[15] The nucleolus is a single-valued solution introduced by Schmeidler (1969). It is always defined and belongs to the core if nonempty. For a definition of the least core and nucleolus, see for instance Maschler et al. (1979) .

[16] There are alternative axiomatizations in particular the one proposed by Young (1985a). They are reviewed by Moulin (2003). The nucleolus satisfies symmetry and dummy but not additivity.

Proof The Shapley value of a fixed cost game (N,e_T) is given by:

$$\varphi_i(N,e_T) = \frac{1}{t} \quad \text{for all } i \in T$$
$$= 0 \quad \text{for all } i \notin T$$

Indeed players outside T are dummies and players in T are substitutes. The Shapley value is a linear operator. Using (3), the value of an elementary data game (N,C_h) is then given by:

$$\varphi_i(N,C_h) = \frac{d_h}{n} - \frac{d_h}{t_h} \quad \text{for all } i \in T_h$$
$$= \frac{d_h}{n} \quad \text{for all } i \notin T_h$$

A data games can be written as a sum of elementary data games. By additivity, the value of the data game (N,C) defined by ($M_1,\ldots,M_n$) and ($d_1,\ldots,d_m$) is given by (7). ●

Hence the cost of the complete dataset is uniformly allocated among all players and the cost of each data is uniformly redistributed to the players owning it. In Example 1, the Shapley compensation is given by (9,–1.5,–7.5) while the compensation derived from the nucleolus is given by (6,0,–6).

Remark 4 What a player receives increases with the cost of the data he/she she owns and decreases with the number of players owning the same data.

Proposition 4 In partition data games, the Shapley value and the nucleolus coincide.

Proof In the partition case, the associated surplus game is symmetric. The Shapley value and the nucleolus satisfying symmetry, they both divide the surplus equally:

$$y_i = C(i) - \frac{n-1}{n} d_0 = (d_0 - k_i) - \frac{n-1}{n} d_0 = \frac{d_0}{n} - k_i$$

It is the core centroid (5). Setting $t_h = 1$ for all h in (7) confirms the result. ●

6. The asymmetric (or weighted) Shapley value

The *weighted* Shapley value allows to take into account asymmetries between players.[17] Let ($w_1,\ldots w_n$) denote the weights assigned to the players. At this stage we assume that $w_i > 0$ for all $i \in N$. The case where some players are assigned a *zero weight* will be considered later.

[17] Weighted values were introduced in Shapley's Ph.D. dissertation and have been later axiomatized by himself (1981b) in a cost allocation context and by Kalai and Samet (1987). The set of all weighted values contains the core and a cost game is concave *if and only if* the set of weighted values and the core coincide. See Monderer, Samet and Shapley (1992) .

In a cost allocation context, w_i determines the share of player i in a fixed cost i.e.

$$\varphi_i(N,C) = \frac{w_i}{w(N)} F \quad (i = 1,...,n)$$

for the game (N,C) defined by $C(S) = F$ for all $S \subset N$, $S \neq \emptyset$. More generally, the value of a fixed cost game (N,e_T) is given by:

$$\begin{aligned} \varphi_i(N,e_T) &= \frac{w_i}{w(T)} \quad \text{for all } i \in T \\ &= 0 \qquad\quad \text{for all } i \notin T \end{aligned}$$

where $w(T)$ is the weight of coalition T. The symmetric case corresponds to $w_i = 1/n$ and $w(T) = t/n$. Using (3), the value of the elementary data game (N,C_h) associated with weights $(w_1,\ldots w_n)$ is given by:

$$\begin{aligned} \varphi_i(N,C_h) &= \frac{w_i}{w(N)} d_h - \frac{w_i}{w(T_h)} d_h \quad \text{for all } i \in T_h \\ &= \frac{w_i}{w(N)} d_h \qquad\qquad\qquad \text{for all } i \notin T_h \end{aligned}$$

We observe that, for a given data h, the ratio between what two players in T_h pay or receive is equal to their weight ratio, and the same applies to players outside T_h:

$$\frac{\varphi_i(N,C_h)}{\varphi_j(N,C_h)} = \frac{w_i}{w_j} \quad \text{for all } i, j \in T_h \text{ and for all } i, j \notin T_h$$

Using addivity, we obtain the following proposition which generalizes Proposition 3.

Proposition 5 Given *positive* weights $(w_1,\ldots w_n)$, the weighted Shapley value of the data game (N,C) defined by the datasets $(M_1,\ldots,M_n)$ and data costs $(d_1,\ldots,d_m)$ is given by:

$$\varphi_i(N,C) = \frac{w_i}{w(N)} d_0 - \sum_{h \in M_i} \frac{w_i}{w(T_h)} d_h \quad (i = 1,...,n) \tag{8}$$

The weighted value is not necessarily monotonic with respect to weights: what a player pays may well decrease while his or her weight increases. Monderer, Samet and Shapley (1992) have shown that concavity is a *necessary and sufficient* condition for monotonicity. This is well illustrated by partition data games. Indeed $w(T_h) = w_i$ for all $h \in M_i$ and (8) reduces to:

$$\varphi_i(N,C) = \frac{w_i}{w(N)} d_0 - k_i \quad (i = 1,...,n) \tag{9}$$

In the examples 1 and 2, the Shapley compensations associated with the weights (1, 1, 2) are given respectively by (6.75,–2.25,–4.5) and (6.75, 0.75,–7.5), to be compared to the Shapley compensations (9,–1.5,–7.5) and (9, 3,–12) under equal weights.

So far we have considered the case where weights are positive. A zero weight can be assigned to players who own data but are not interested in completing the dataset. Let us denote by $Z = \{i \in N \mid w_i = 0\}, Z \neq N,$ the set of zero weight players and by $U_h = T_h \cap Z$ the set of zero weight players owning data h. Consider the sequences (w^ν) defined by $w_i^\nu = w_i$ for all $i \in N \backslash Z$ and $w_i^\nu \to 0$ for all $i \in Z.$ Then players' permutation in which a zero weight player precedes a nonzero weight player has a zero limit probability.[18] As a consequence, we have the following proposition:

Proposition 6 Zero weight players are compensated for a data they own *if and only if* no positive weight player owns the same data.

In particular, if a data is owned by a single zero weight player, he/she receives the total value of his or her data. If a data is owned exclusively by several zero weight players, the way they share the value of the data is indeterminate. It there is no reason to discriminate among zero weight players, we may consider only sequences (w^ν) where $w_i^\nu = t^\nu \to 0$ for all $i \in Z$. The resulting value of an elementary game (N,C_h) is then unchanged for nonzero weight players while, for zero weight players, we get:

$$\begin{aligned} \varphi_i(N,C_h) &= -\frac{d_h}{u_h} \qquad \text{if } T_h \subset Z \\ &= 0 \qquad\quad \text{otherwise} \end{aligned}$$

where $u_h = |U_h|.$ Hence we have:

$$\varphi_i(N,C) = -\sum_{\substack{h \in M_i \\ T_h \subset Z}} \frac{1}{u_h} d_h \quad \text{for all } i \in Z$$

7. Accounting rules

There are various accounting rules for dividing joint costs based on players' marginal costs computed with respect to the grand coalition. They are typically of the form:

$$\phi_i(N,C) = MC_i + \alpha_i(C(N) - \sum_{j=1}^{n} MC_j) \quad (i = 1,...,n) \tag{10}$$

[18] See Dehez (2009) for more details.

where $MC_i = C(N) - C(N\backslash i)$ is the "separable cost" of player i the α_i's are weights satisfying $0 \leq \alpha_i \leq 1$ for all i and $\Sigma_i\, \alpha_i = 1$. There are two well known rules. The "*equal charge*" rule which uses equal weights and the "*separable costs remaining benefits*" rule (SCRB) whose weights are given by:[19]

$$\alpha_i = \frac{b_i}{b(N)}$$

where $b_i = C(i) - MC_i$ is the "remaining benefits" of player i and $b(N) = \Sigma_i\, b_i$.

When weights are independent of the cost game, ϕ is an *additive* (actually linear) sharing rule which satisfies the symmetry axiom but not the dummy axiom. The equal charge rule is additive while the SCRB rule is not. The latter is however scale invariant.

For an elementary data game (N,C_h), we observe that $C_h(N\backslash i) = 0$ for all i if $t_h \geq 2$ in which case $\phi(N, C_h) = 0$. Hence only data owned by a *single* player actually enter into (10). If $t_h = 1$, we have:

$$\begin{aligned} C_h(N \setminus i) &= d_h \quad \text{if } i \in T_h \\ &= 0 \quad \text{if } i \notin T_h \end{aligned}$$

and, consequently,

$$\begin{aligned} \phi_i(N, C_h) &= (\alpha_i - 1)\, d_h \quad \text{if } i \in T_h \\ &= \alpha_i\, d_h \qquad\quad \text{if } i \notin T_h \end{aligned}$$

Applied to a general data game, it leads to the following allocation:

$$\phi_i(N, C) \; = \alpha_i \sum_{h \in M^*} d_h - \sum_{h \in M_i^*} d_h \quad (i = 1, \ldots, n) \tag{11}$$

where M^* is the subset of data owned by a single player and $M_i^* = M_i \cap M^*$. The only players to be compensated are those who are alone to own some data. Like for core allocations, there is no compensation if every data is owned by at least two players.

In the partition case, $M^* = M$ and (11) reduces to

$$\phi_i(N, C) \; = \alpha_i d_0 - k_i \quad (i = 1, \ldots, n) \tag{12}$$

That equation defines an allocation which belongs to the core, *for any choice of weights*.

[19] See for instance Young (1985b).

Indeed we have:

$$C(S) - y(S) = d_0(1 - \sum_{i \in S} \alpha_i) \geq 0 \text{ for all } S \subset N$$

It coincides with the weighted Shapley value (9) whenever $w_i = \alpha_i$ for all i. In particular the equal share allocation coincides with the symmetric Shapley value which is also the allocation resulting from the SCRB method. Indeed we have:

$$b_i = (d_0 - k_i) - (-k_i) = d_0 \text{ for all } i \in N$$

and consequently $\alpha_i = 1/n$ for all $i \in N$.

8. Concluding remarks

The Shapley value is the natural allocation rule to be used in cost sharing as well as in the compensation framework considered here. The resulting allocation may not belong to the core because it involves cross subsidization. This should not be a reason to dismiss the Shapley value as a compensation mechanism because what the core suggests may be unacceptable. We have seen that the core excludes compensations when each data is held by two players or more. More generally if data are not sufficiently spread among players, compensations remain typically small. This appears forcefully in the situation where only two players own data, say players n and n-1, and the datasets they own differ only by a single data, say data 1:

$$M_n = \{1, \ldots, m\}, \; M_{n-1} = \{2, \ldots, m\} \text{ and } M_i = \varnothing \quad (i = 1, \ldots, n-1)$$

In this case, the core imposes that only player n may be compensated with an amount not exceeding d_1 – the cost of the missing data – while the other players may be asked to pay up to d_1, including player n-1. The nucleolus goes even further by imposing that the n-1 first players pay the same amount, namely d_1/n.

This is to be compared with the allocation derived from the Shapley value. Using (7) we get:

$$y_i = \frac{d_0}{n} \quad (i = 1, \ldots n-2)$$

$$y_{n-1} = \frac{d_0}{n} - \sum_{h=2}^{m} \frac{d_h}{2} = -\frac{n-2}{2n} d_0 + \frac{d_1}{2}$$

$$y_n = \frac{d_0}{n} - \sum_{h=2}^{m} \frac{d_h}{2} - d_1 = -\frac{n-2}{2n} d_0 - \frac{d_1}{2}$$

It is definitely more acceptable for the players: those without data pay the per capita cost of the complete dataset while players n and n-1 are both compensated, the difference between what they receive being precisely equal to the cost of the missing data.

In actual cost sharing problems, like the one faced by the European chemical industry, there must be an agreement on the compensation formula *and* on the costs parameters.[20] Reaching a consensus on the cost parameters is clearly the most difficult part in particular because, under the Shapley value, we know from Remark 4 that what a player pays decreases with the cost of the data he/she owns. One should however keep in mind that these cost parameters measure the present cost of *reproducing* the data and not the actual cost that has been sunk in the past.

References

Dehez, P. (2008), Reverse airport games, Revised CORE Discussion Paper 2007/94.

Dehez, P. (2009), The allocation of fixed costs and the weighted Shapley value, CORE Discussion Paper 2009/35.

Drèze, J.H. (1980), Public goods with exclusion, *Journal of Public Economics* 13,5-24.

Faulhaber, G. (1975), Cross-subsidization: Pricing in public enterprises, *American Economic Review* 65, 966-977.

Gonzáles-Díaz J. and Sánchez-Rodríguez E. (2007), A natural selection from the core of a TU game: the core-centre, *International Journal of Game Theory* 36, 27-46.

Kalai E. and D. Samet (1987), On weighted Shapley values, *International Journal of Game Theory* 16, 205-222.

Littlechild S. C. and Owen G. (1973), A simple expression for the Shapley value in a special case, *Management Science* 20, 370-372.

Maschler M., Peleg B. and Shapley L.S. (1972), The kernel and bargaining set for convex games, *International Journal of Game Theory* 1, 73-93.

Maschler M., Peleg B. and Shapley L.S. (1979), Geometric properties of the kernel, nucleolus and related solution concepts, *Mathematics of Operations Research* 4, 303-338.

Monderer D., D. Samet and L.S. Shapley (1992), Weighted values and the core, *International Journal of Game Theory* 21, 27-39.

[20] In that framework the firms are typically of different sizes and an agreement on weights must then also be reached. These are the weights that would be used to share the cost of *additionnal* data.

Moulin, H. (1988), *Axioms of cooperative decision making*, Cambridge University Press, Cambridge.

Moulin, H. (2003), *Fair division and collective welfare*, The MIT Press, Cambridge.

Roth A.E. (ed., 1988), *The Shapley value. Essays in honor of Lloyd Shapley*, Cambridge University Press, Cambridge.

Shapley, L.S. (1953), A value for n-person games, In Kuhn H. and Tucker A.W. (eds.), *Contributions to the theory of games* II, Princeton University Press, Princeton, 307-317. Reprinted in Roth (1988), 31-40.

Shapley, L. S. (1971), Cores of convex games, *International Journal of Game Theory* 1, 11-26.

Shapley, L.S. (1981a), Valuation of games, in Lucas W.F. (ed.), *Game Theory and its Applications*. Proceedings of Symposia in Applied Mathematics 24, American Mathematical Society, Providence, Rhode Island.

Shapley, L.S. (1981b), Discussion comments on "Equity considerations in traditional full cost allocation practices: An axiomatic approach", in Moriarity S. (ed.), *Joint cost allocation*, Proceeding of the University of Oklahoma Conference on Costs Allocations, April 1981, Center for Economic and Management Research, University of Oklahoma.

Schmeidler, D. (1969), The nucleolus of a characteristic function game, *SIAM Journal of Applied Mathematics* 17, 1163-1170.

Weber, R.J. (1988), Probabilistic values for games, in Roth (ed. 1988), 101-119.

Young H.P. (1985a), Monotonic solutions of cooperative games, *International Journal of Game Theory* 14, 65-72.

Young, H.P. (1985b), *Cost allocation: methods, principles, applications*, North-Holland, Amsterdam.