**2007/94**

■

Shapley compensation scheme

Pierre Dehez

# Shapley compensation scheme

Pierre DEHEZ[1]

## Abstract

We study a particular class of cost sharing games – "data games" – covering situations where some players own data which are useful for a project pursued by the set of all players. The problem is to set up compensations between players. Data games are subadditive but generally not concave, and have a nonempty core. We characterize the core and compute the compensation scheme derived from the Shapley value. We then compare it to the nucleolus. Although we use the term "data" our analysis actually applies to any good characterized by non rivalry and excludability.

**Keywords**: cost sharing, Shapley value, nucleolus.

**JEL Classification**: C71, D46, M41

**Introduction**

Imagine the following situation. A group of firms are willing to cooperate on a project requiring the collection of data and one firm already owns some data. The question is then the following: how should this firm be compensated? If there are $n$ firms and if $c$ is the cost of reproducing the existing data, a fair payment should be $c/n$. Each firm pays $c/n$ to the owner who collects $(n-1)c/n = c - c/n$. The owner gets $c$ but pays $c/n$ too. Applying the Shapley value to this cost game gives precisely the compensation described above. It is actually what any sensible compensation scheme should suggest in such a situation.

We can think of more complicated situations. There may be more than one "player" owning the same data and the data may be more or less complete. As we shall see, this gives rise to a cost sharing game which has an interesting and simple structure.

This story actually applies to any good characterized by non rivalry and excludability.[1] We shall however keep using the term "data" for expository reason.

The origin of the present paper is the cost sharing problem faced by the European chemical industry which must submit detailed analysis for about 30.000 substances that are being produced. This is a requirement imposed by EU under the acronym "REACH". Individual firms may submit these analyses but the European Commission encourages firms to get together, taking advantage of existing data, in order to reduce the overall cost of the program. In the official documents which are available, the cost sharing problem is described and hints are given for solving it through a few examples. No precise methodology is however proposed.

In what follows we set up a class of cost sharing games – "data games" – and analyze their properties. We show how they are related to airport games and observe that the core is nonempty as it always contains the no compensation allocation: no coalition can object when no one is asked to pay. We characterize the core which happens to have a simple and regular structure and show that in any core allocation only the player with the largest database is possibly compensated. We then define the compensation rule derived from the Shapley value and show that the resulting allocation does generally not belong to the core. We finally compare the Shapley value to the nucleolus, an other solution concept used in cost sharing. The nucleolus always belongs to the core and, applied to data games, all players except the one with the largest database, contribute the *same* amount.

**The cost sharing game**

We first define formally the notion of cost (sharing) game.[2]

A set $N = \{1,\dots,n\}$ of players have a common project and they face the problem of dividing its total cost. The cost of realizing the project to the exclusive benefit of the members of any subset of players ("coalition") is also known. This defines a real-valued function $C$ – a cost function – on the subsets of $N$ with $C(\varnothing) = 0$. A pair $(N,C)$ defines a cost game and the total cost to be divided is $C(N)$.

---

[1] This is what is sometime called a *public good with exclusion.*
[2] See for instance Young (1985) or Moulin (1988, 2003).

A sharing rule $\varphi$ associates a cost allocation $y = \varphi(N,C)$ to any cost game $(N,C)$ such that:[3]

$$\sum_{i=1}^{n} y_i = C(N)$$

**Notation:** The letters $n, s, t, \ldots$ will denote the size of the sets $N, S, T, \ldots$ For a vector $x$, $x(S)$ 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 of which player $i$ has been removed.

We now define *data games*. We denote by $k_i$ the replacement cost of the data owned by player $i$. Without loss of generality, we assume that

$$0 \leq k_1 \leq k_2 \leq \ldots \leq k_n \text{ and } k_n > 0 \tag{1}$$

where $k_i = 0$ means that player $i$ does not own data. The objective of the grand coalition $N$ is to share the data owned by player $n$. We denote by $c_i = k_n - k_i$ the cost of upgrading player $i$'s database to a level equivalent to player $n$'s database. The $c_i$'s satisfy the following inequalities:

$$c_1 \geq c_2 \geq \ldots \geq c_{n-1} \geq c_n = 0 \tag{2}$$

Assuming that if $i < j$, player $j$'s database includes player $i$'s database, the corresponding cost game $(N,C)$ is defined by:

$$C(S) = \text{Min}_{i \in S}(k_n - k_i) = \text{Min}_{i \in S} c_i \text{ for all } S \subset N, S \neq \varnothing \tag{3}$$

with $C(\varnothing) = 0$. Hence $C(i) = c_i$ and $C(N) = 0$. We shall denote by $(N,c)$ the data game with cost parameters $c_i = k_n - k_i$. Alternatively, the cost function $C$ can be written as:

$$C(S) = k_n - K(S) \tag{4}$$

where $K(S) = \text{Max}_{i \in S} k_i$. Given the inequalities (2), the cost function $K$ defines an *airport game*. It is well known that the cost functions associated to airport games are monotonically increasing, concave and thereby subadditive.[4] Here $k_n$ appears as a *fixed cost* which applies to all players and coalitions.[5]

In the case where a subset $D$ of players own the *same* data while the other players own *no* data, $k_i = a > 0$ if $i \in D$ and $k_i = 0$ otherwise. The associated cost function is given by:

$$C(S) = 0 \quad \text{for all } S \text{ such that } S \cap D \neq \varnothing$$

$$\tag{5}$$

$$C(S) = a \quad \text{for all } S \neq \varnothing \text{ such that } S \cap D = \varnothing$$

The case referred to in the introduction is the situation where only one player owns date.

---

[3] Equality is a requirement of efficiency.

[4] See Littlechild and Owen (1973) and Littlechild and Thomson (1977).

[5] The cost function $C$ can also be written as $C(S) = -C_0(S)$ where $C_0(S) = \text{Max}_{i \in S}(-c_i)$. However $C_0$ does not define a proper airport game because $C_0(S) \leq 0$ for all $S$. In particular, the cost function $C_0$ is superadditive.

It is clear from definition (3) that data games are *essential* (except in the uninteresting case where all $k_i$'s are equal and no compensations are then needed) and monotonically *decreasing*:

$$\sum_{i=1}^{n} C(i) > 0$$

$$S \subset T \implies C(S) \geq C(T)$$

It is also *subadditive*:

$$S \cap T = \varnothing \implies C(S) + C(T) \geq C(S \cup T)$$

Indeed if $S$ and $T$ be two disjoint subsets of $N$, $C(S \cup T) = \text{Min}\,(C(S), C(T))$. Assume $C(S) \geq C(T)$. Then $C(S) + C(T) - C(S \cup T) = C(S) \geq 0$.[6]

Subadditivity is a minimum form of scale economies. However the cost function $C$ is not concave[7] except in the particular case where data are owned by only one player. Indeed $C(S) + C(T) \geq C(S \cup T) + C(S \cap T)$ for all $S, T \subset N$ except if $S \cap T \neq \varnothing$, $S \cap D \neq \varnothing$, $T \cap D \neq \varnothing$ and $(S \cap T) \cap D = \varnothing$ in which cases the opposite inequality applies:

$$C(S) + C(T) = 0 < C(S \cup T) + C(S \cap T) = a$$

Such cases do not arise when $d \geq 2$. This does not imply that $C$ is a convex function because $C(S) + C(T) > C(S \cup T) + C(S \cap T)$ whenever $(S \cup T) \cap D = \varnothing$ and $S \cap T = \varnothing$.

It is sometimes convenient to consider the division of the surplus generated by the grand coalition. For any coalition $S \subset N$, we define $v(S)$ as the gain obtained by coalition $S$ if it forms:

$$v(S) = \sum_{i \in S} C(i) - C(S) \tag{6}$$

This defines the surplus (sharing) game $(N,v)$ associated to the cost game $(N,C)$. In particular $v(i) = 0$ and $v(N) > 0$ (for essential cost games).[8] Shares $z$ in the total surplus $v(N)$ and shares $y$ in the total cost $C(N)$ are related by the following identities:

$$z_i + y_i = C(i) \quad i = 1,\ldots,n \tag{7}$$

The surplus game $(N,v)$ associated to the data game $(N,c)$ is defined by the characteristic function

$$v(S) = \sum_{i \in S} c_i - \text{Min}_{i \in S} c_i$$

and $v(N) = \sum_{i=1}^{n} c_i$ is the total surplus to be divided.

---

[6] Subadditivity of the cost function $C$ also follows from the superadditivity of the cost function $C_0$.

[7] A set function $f$ is *concave* if $f(S) + f(T) \geq f(S \cup T) + f(S \cap T)$. Hence concavity implies subadditivity. Alternatively, a set function $f$ is concave if the marginal costs $f(S) - f(S\backslash i)$ are all decreasing.

[8] If the cost function $C$ is subadditive (resp. concave) then the surplus function $v$ is superadditive (resp. convex), and vice versa.

**The core**

Individual rationality is the minimal requirement to impose to a cost allocation $y$:

$$y_i \leq C(i) \quad \text{for all } i \in N$$

i.e. no player should pay more than his or her "stand alone" cost. This defines an *imputation*. Extending the argument to coalitions is a stronger requirement: the *core* is the set of allocations $y$ against which no coalition can object:

$$y(S) \leq C(S) \quad \text{for all } S \subset N \tag{8}$$

i.e. no coalition pays more that its stand alone cost. Equivalently, an allocation $y$ is in the core if and only if

$$y(S) \geq C(N) - C(N \backslash S) \quad \text{for all } S \subset N$$

i.e. there is no cross-subsidization: each coalition pays at least its marginal cost.[9]

Using (7) and (8), the core of the surplus game $(N,v)$ is defined by the allocations $z$ such that

$$z(S) \geq v(S) \quad \text{for all } S \subset N$$

The core of a data game as defined by (3) is *always* nonempty. Indeed, $C(S) \geq 0$ for all $S \subset N$, independently of the choice of the $c_i$'s. As a consequence, the trivial allocation defined by the absence of compensation $y_0 = (0,0,\ldots,0)$ belongs to the core for all $c_i$'s satisfying (2). Actually, the core happens to have a very simple structure which depends exclusively on $c_{n-1}$.

**Proposition 1**     The core of a data game $(N,c)$ is a *regular* convex polyhedron whose vertices are then $n$ vectors $(c_{n-1}, 0,\ldots, 0, -c_{n-1})$, $(0, c_{n-1}, 0,\ldots,0, -c_{n-1})$,$\ldots$, $(0, 0,\ldots, c_{n-1}, -c_{n-1})$ and $y_0 = (0,\ldots,0)$.[10]

**Proof**    We first observe that if $y$ belongs to the core, then $0 \leq y_i \leq c_i$ for all $i \neq n$ and $-c_{n-1} \leq y_n \leq 0$. Indeed, individual rationality requires $y_i \leq c_i$ for all $i$, with $c_n = 0$. For any given $i \neq n$, $y(N/i) \leq 0$ and $y(N) = 0$ imply $y_i \geq 0$. Finally $y(N/n) \leq c_{n-1}$ and $y(N) = 0$ imply $y_n \geq -c_{n-1}$.

We then show that $y$ belongs to the core *if and only if* there exist $\lambda_1,\ldots,\lambda_{n-1}$ in $[0,1]$ such that:

$$\sum_{i=1}^{n-1} \lambda_i \leq 1$$

$$y_i = \lambda_i c_{n-1} \quad i = 1,\ldots, n-1 \tag{9}$$

---

[9] See Faulhaber (1975).

[10] Hence the core is a polyhedron with $n$ congruent faces. It is an equilateral triangle for $n = 3$ and a tetrahedron for $n = 4$. All vertices are connected together by a line segment of length $2^{1/2} c_{n-1}$.

If $y$ is an allocation defined by (9), the following inequalities hold for all $S \subset N$

$$\sum_{i \in S} y_i = c_{n-1} \sum_{i \in S} \lambda_i \le c_{n-1} \le C(S) \quad \text{if } n \notin S$$

$$\sum_{i \in S} y_i = c_{n-1} \sum_{\substack{i \in S \\ i \ne n}} \lambda_i - c_{n-1} \sum_{i=1}^{n-1} \lambda_i = -c_{n-1} \sum_{\substack{i=1 \\ i \notin S}}^{n-1} \lambda_i \le 0 = C(S) \quad \text{if } n \in S$$

Hence $y$ belongs to the core.

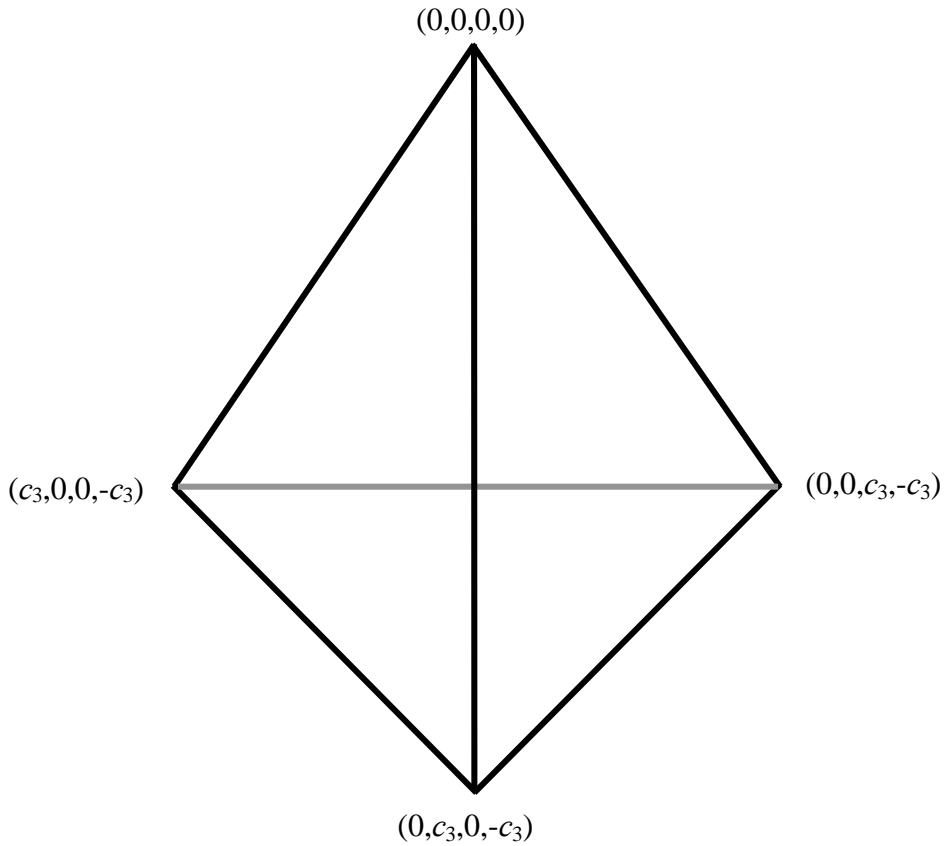If $y$ is an element of the core, we have successively:

$$0 \le y_{n-1} \le c_{n-1}$$
$$0 \le y_{n-2} \le y_{n-2} + y_{n-1} \le c_{n-1}$$
$$...$$
$$0 \le y_1 \le \sum_{i=1}^{n-1} y_i \le c_{n-1}$$

i.e. $0 \le y_i \le c_{n-1}$ for all $i = 1,\dots,n\text{-}1$. The $\lambda_i$'s defined by $\lambda_i = y_i/c_{n-1}$ then satisfy (9). $\ddot{\text{Y}}$

Figure 1 illustrates the core for the case $n = 4$.



**Figure 1**

From that proposition, it follows that the core reduces to $\{y_0\}$ *if and only if* more than one player owns the full database, independently of the other cost parameters. The size of the core depends only on $c_{n-1}$ – the largest compensation player $n$ can expect – and it has full dimension $(n-1)$ *if and only if* $c_{n-1} > 0$. Furthermore, the core being a regular polyhedron, its center of gravity[11] is simply defined by the average of its vertices, which is the point defined by (9) for $\lambda_i = 1/n$:

$$\hat{y} = \left( \frac{c_{n-1}}{n}, \frac{c_{n-1}}{n}, ..., \frac{c_{n-1}}{n}, -(n-1)\frac{c_{n-1}}{n} \right)$$

**The Shapley value**

To each permutation of the $n$ players we associate a vector $t$ of marginal costs. More precisely the element $i_k$ of the vector $t$ associated to the permutation $\pi = (i_1,...,i_n)$ is given by:

$$t_{i_1}(\pi) = C(i_1)$$
$$t_{i_k}(\pi) = C(i_1,...,i_k) - C(i_1,...,i_{k-1}) \text{ for } k = 2,...,n$$

There are $n!$ such vectors and the Shapley value is simply the average marginal cost vector:[12]

$$\varphi_i(N,C) = \frac{1}{n!} \sum_{\pi \in \Pi} t_i(\pi)$$

where $\Pi$ is the set of all permutations. The Shapley value is the unique *additive* sharing rule which satisfies *symmetry* and *dummy*.[13] There exist alternative axiomatizations of the Shapley value.[14] In the context of cost sharing it is shown that the Shapley sharing rule is the unique sharing rule which allocates fixed costs uniformly.[15] The Shapley value is individually rational for subadditive cost games and it belongs to the core for concave cost games.[16] Otherwise it does not necessarily belong to the core.

For the case $n = 3$, the marginal cost vectors associated to a data game $(N,c)$ are given by:

|  | 1 | 2 | 3 |
|---|---|---|---|
| 123 | $c_1$ | $c_2 - c_1$ | $- c_2$ |
| 132 | $c_1$ | $0$ | $- c_1$ |
| 213 | $0$ | $c_2$ | $- c_2$ |
| 231 | $0$ | $c_2$ | $- c_2$ |
| 312 | $0$ | $0$ | $0$ |
| 321 | $0$ | $0$ | $0$ |
| **Total** | $2c_1$ | $3c_2 - c_1$ | $- 3c_2 - c_1$ |

[11] See Gonzales-Diaz and Sanchez-Rodriguez (2007) for a general definition of the center of gravity of the core.
[12] The marginal cost vectors are distinct *if and only if* the cost function is *strictly* concave.
[13] These are the original axioms used by Shapley (1953, 1981): players with identical marginal costs pay the same amount (symmetry or "equal treatment of equals") and players with zero marginal costs pay nothing (dummy).
[14] See Moulin (2003).
[15] See Dehez (2007).
[16] The core is typically large and the Shapley value is located somewhere in its centre. See Shapley (1971).

The compensation derived from the Shapley value is obtained by dividing the totals by $n! = 6$. It is given by:

$$y = (\frac{c_1}{3}, \frac{c_2}{2} - \frac{c_1}{6}, -\frac{c_2}{2} - \frac{c_1}{6})$$

We notice that there are actually four distinct vectors if $c_1 > c_2$ and, in the case where $c_1 = c_2$, there are only three distinct vectors. The no compensation vector $y_0$ is always one of these vectors. It corresponds to the $(n-1)!$ permutations in which the last player comes first.

Using (7) the corresponding allocation of the total surplus $c_1 + c_2$ is given by:

$$z = c - y = (\frac{2c_1}{3}, \frac{c_2}{2} + \frac{c_1}{6}, \frac{c_2}{2} + \frac{c_1}{6})$$

It could have as well been derived from the marginal contribution vectors:[17]

|       | 1       | 2          | 3          |
|-------|---------|------------|------------|
| 123   | 0       | $c_1$      | $c_2$      |
| 132   | 0       | $c_2$      | $c_1$      |
| 213   | $c_1$   | 0          | $c_2$      |
| 231   | $c_1$   | 0          | $c_2$      |
| 312   | $c_1$   | $c_2$      | 0          |
| 321   | $c_1$   | $c_2$      | 0          |
| **Total** | **$4c_1$** | **$3c_2 + c_1$** | **$3c_2 + c_1$** |

Applied to the airport game $(N,K)$, the cost allocation derived from the Shapley value is well known and distributes the cost increments uniformly:

$$x_1 = \frac{k_1}{n}$$

$$x_2 = \frac{k_1}{n} + \frac{k_2 - k_1}{n-1}$$

$$x_2 = \frac{k_1}{n} + \frac{k_2 - k_1}{n-1} + \frac{k_3 - k_2}{n-2} \tag{10}$$

...

$$x_n = \frac{k_1}{n} + \frac{k_2 - k_1}{n-1} + ... + \frac{k_{n-1} - k_{n-2}}{2} + k_n - k_{n-1}$$

---

[17] Notice that 2 and 3 are substitute players. Actually the last two players are always substitutes in the surplus game associated to a data game: they both contribute $c_{n-1}$ to any coalition of which they are both members.

Using (4) and knowing that the Shapley value is a linear operator which satisfies symmetry, the compensation derived from the Shapley value is given by :

$$y_i = \frac{k_n}{n} - x_i \quad i = 1, ..., n$$

i.e.

$$y_1 = \frac{k_n - k_1}{n} = \frac{c_1}{n}$$

$$y_2 = \frac{k_n - k_1}{n} + \frac{k_1 - k_2}{n-1} = \frac{c_1}{n} + \frac{c_2 - c_1}{n-1} \quad\quad\quad (11)$$

...

$$y_n = \frac{k_n - k_1}{n} + ... + \frac{k_{n-2} - k_{n-1}}{2} + k_{n-1} - k_n = \frac{c_1}{n} + ... + \frac{c_{n-1} - c_{n-2}}{2} + c_n - c_{n-1}$$

We observe that the resulting allocation corresponds to the airport cost allocation as given by (10), replacing $k_i$ by $c_i$. This is consistent with the fact that $C(S) = - \text{Max}_{i \in S}(-c_i)$.

Formula (11) has a simple recursive structure and the Shapley compensation $y$ can be written simply as:

$$y = A.c$$

where $A$ is a $n \times n$ triangular matrix whose elements are defined by:

$$a_{ij} = \frac{-1}{(n-j+1)(n-j)} \quad \text{if} \quad j < i$$

$$a_{ii} = \frac{1}{n-i+1}$$

with $a_{ij} = 0$ otherwise. For $n = 4$, the matrix $A$ is given by:

$$A = \begin{bmatrix} 1/4 & 0 & 0 & 0 \\ -1/12 & 1/3 & 0 & 0 \\ -1/12 & -1/6 & 1/2 & 0 \\ -1/12 & -1/6 & -1/2 & 1 \end{bmatrix}$$

Notice that these matrices are overlapping, starting from the lower right element 1.[18] For instance, if $n = 5$ the first column starts with 1/5, followed by -1/20.[19]

---

[18] Actually the element $a_{nn}$ is arbitrary because $c_n = 0$. It is equal to 1 in the matrix defining the Shapley value of an airport game.

[19] The elements of any of the first $n$-1 columns sum up to 1 and the elements of any of the row sum up to $1/n$.

In the particular case where a subset $D$ of players own the same database while the other players own no data, the associated cost function is defined by (5) and the Shapley value is simply given by:

$$y_i = \frac{a}{n} \quad \text{if } i \notin D$$

$$y_i = -\frac{(n-d)a}{nd} = \frac{a}{n} - \frac{a}{d} \quad \text{if } i \in D$$

where $d$ is the size of $D$, $1 \le d \le n$. Players who do not own the data pay $c/n$ and the amount collected is redistributed uniformly among the players owning the data. We observe that the amount received by players in $D$ decreases with $d$. When $d = 1$, the game is concave and the Shapley value belongs to the core. In addition, for that particular game, the Shapley value coincides with the nucleolus as will be shown later.

We have seen that if more than one player own the full database the core reduces to the single allocation $y_0$ in which case the Shapley value does definitely not satisfy the core requirements. Actually no compensation does. The following examples show what may happen when there are differences in the databases.

**Example 1** The cost function associated to $c = (100, 40, 20, 0)$ is given by:

$C(1) = 100$
$C(2) = C(12) = 40$
$C(3) = C(13) = C(23) = C(123) = 20$

and $C(S) = 0$ for all the other coalitions (i.e. all coalitions including player 4). This game is not concave and the Shapley compensation $(25, 5, -5, -25)$ does *not* belong to the core. Indeed $y(123) = 25 > C(123)$.

**Example 2** The cost function associated to $c = (88, 76, 64, 0)$ is given by:

$C(1) = 88$
$C(2) = C(12) = 76$
$C(3) = C(13) = C(23) = C(123) = 64$

and $C(S) = 0$ for all the other coalitions. Although this game is not concave, the Shapley compensation $(22, 18, 12, -52)$ belongs to the core.

Using Proposition 1, it is easy to find conditions on the cost parameters $c_i$ such that the Shapley value belongs to the core. For instance, if $n = 3$, the Shapley compensation belongs to the core *if and only if*:

$$c_1 \le 3\, c_2 \tag{12}$$

When equality holds, the Shapley compensation is the allocation in the core in which player 1 pays the largest amount.

**Example 3** The vector $c = (90,30,0)$ satisfies (12) with equality. The Shapley compensation is given by $(30,0,-30)$ and the corresponding surplus allocation is $(60,30,30)$. Figure 2 illustrates the core of this three-player data game.
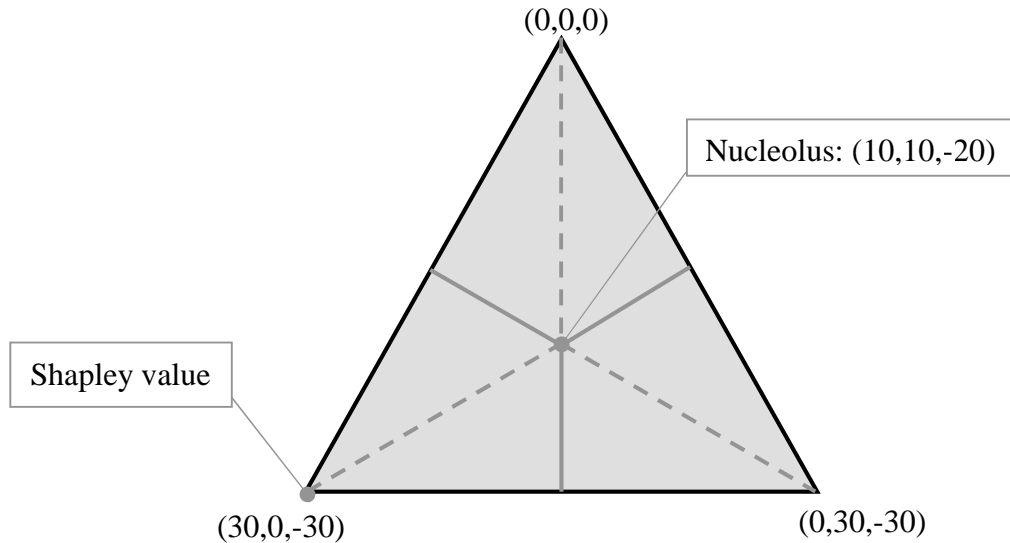


**Figure 2**

**The nucleolus**

The nucleolus is a solution concept introduced by Schmeidler (1969). Intuitively the idea is to minimize the loss incurred by coalitions suffering the highest loss – the loss (or excess) of a coalition being measured by the difference between the amount it pays and its cost. The nucleolus is always defined and belongs to the core if nonempty. It defines a sharing rule which satisfies symmetry and dummy but not additivity.[20]

In the examples 1 and 2, the compensations derived from the nucleolus are given respectively by $(5,5,5,-15)$ and $(16,16,16,-48)$. They differ significantly from the Shapley compensations. In particular, only the last player is compensated (this is a feature of the core) and all the other players contribute the same amount. This is actually a feature of the nucleolus when applied to data games.[21]

**Proposition 2** The nucleolus of a data game $(N,c)$ is the allocation $\hat{y}$ where

$$\hat{y}_i = \frac{c_{n-1}}{n} = \frac{k_n - k_{n-1}}{n} \text{ for all } i = 1 \dots n{-}1$$

---

[20] The nucleolus is a beautiful theoretical construct which however suffers from a number of limitations. The lack of monotonicity is probably the most serious one: if the cost of a coalition decreases, leaving the cost of all the other coalitions unchanged, it may be that some players pay more following that change.

[21] The equality between the nucleolus and the center of gravity of the core is a consequence of the regularity of the core. Indeed Maschler et al (1979) have shown that the nucleolus is the lexicographic center of the game.

**Proof** The excess function $e(y,S) = y(S) - \text{Min}_{i \in S}\, c_i$ is defined all $S \subset N$ ($S \neq \varnothing$ and $S \neq N$) over the set of individually rational cost allocations $y$. For each imputation $y$, the $2^n - 2$ excesses are ordered in a decreasing order. The vectors obtained are then compared lexicographically, retaining the allocations corresponding to the lowest sequence. Schmeidler (1969) proved that this procedure defines a single allocation, called the nucleolus, which belongs to the core in case it nonempty.[22]

The core being nonempty, we can use Proposition 1 to restrict the domain to the set

$$Y = \{y \in \tilde{N}^n \mid y(N) = 0,\ 0 \leq y_i \leq c_i,\ i = 1\ldots n{-}1,\ -c_{n-1} \leq y_n \leq 0\}$$

In view of the definition of the excess function, we only need to look at the coalitions $N/n$ and $N/i$ for $i = 1\ldots n{-}1$:

$$e(y,N/n) = y(N/n) - c_{n\text{-}1}$$

$$e(y,N/j) = y(N/j) \ \text{ for } i = 1\ldots n{-}1 \tag{13}$$

Indeed, for any given $y \in Y$, no excess associated to any other coalition can be higher. Minimizing over $Y$ then occurs at the allocation $y^*$ such that the $n$ excesses defined by (13) are all equal.

As a consequence

$$y_i = y_n + c_{n\text{-}1} \ \text{ for } i = 1\ldots n{-}1$$

Hence $y_i = w = y_n + c_{n\text{-}1}$ for $i = 1\ldots n{-}1$ and $y_n = -(n{-}1)w$. It then follows that $w = c_{n\text{-}1}/n$.     Ÿ

Whatever are the cost components $c_i$, the nucleolus only compensate the player with the largest database while all the other players contribute an equal amount. Furthermore the nucleolus is the no compensation allocation $y = y_0$ if and only if $c_{n\text{-}1} = 0$, irrespectively of the other cost parameters. The nucleolus is therefore clearly not an appropriate compensation scheme except in the extreme case where only one player owns data. In that case all the other player pay the same amount $w = k_n/n$ and the nucleolus coincide with the Shapley value.

In Example 3, the compensation derived from the nucleolus is $(10,10,-20)$ as shown in figure 2, to be compared to the Shapley compensation $(30,0,-30)$. The following example shows how the two rules may produce compensations which differ considerably in magnitude.

**Example 4** Consider the vector $k = (0,234,240)$ and the associated vector $c = (240,6,0)$. The compensations derived from the nucleolus and the Shapley value are given by $(2,2,-4)$ and $(80,-37,-43)$ respectively. In any case, the maximum compensation player $n$ can expect in the core is $c_{n\text{-}1} = 6$.

---

[22] The nucleolus actually belongs to the least core which, for data games, coincide with the nucleolus.

**Concluding remarks**

The cost sharing problem described in the present paper is highly stylized and its solution is straightforward. More realistic situations where databases are not necessarily interlocking lead to well defined cost games for which one can compute the Shapley compensation. These games do not however share the simple structure characterizing data games defined by (3).

If more than one project is involved, the problem can be written as a sum of cost games, one for each project, in which players who are not concerned are dummies. The Shapley value being additive, the solution to the global problem is then simply the sum of the Shapley values applied to the individual data games.

In an actual cost sharing problem, like the one faced by the European chemical industry, there must be an agreement on the compensation formula and on the value of the costs parameters. The fact that the compensation derived from the Shapley value may involve cross subsidization should not be a reason to dismiss it as a fair compensation mechanism. Indeed we have seen that core compensations are hardly acceptable because either they imply no compensation at all or only the last player is compensated.

Reaching a consensus on the cost parameters is clearly the most difficult part in particular because, under the Shapley value, what a player pays decreases with his or her cost parameter $k$, giving players an incentive to overvalue their cost parameter. 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. (2007), Fair division of fixed costs defines the Shapley value, Revised CORE Discussion Paper 2006/115.

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

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

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

Littlechild S. C. and G. F. Thomson  (1977), Aircraft landing fees: A game theory approach, *Bell Journal of Economics* 8,

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

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

Moulin, H. (2003), *Fair division and collective welfare*, The MIT 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 A.E. (ed. 1988), *The Shapley value. Essays in honor of Lloyd Shapley*, Cambridge University Press, Cambridge.

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

Shapley, L.S. (1981), 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.

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

Young, P.Y. (1985), Cost allocation, in Young P.Y. (ed.), *Fair allocation*, Proceedings of Symposia in Applied Mathematics 33, American Mathematical Society, Providence, Rhode Island.