Stable partitions in coalitional games

Krzysztof R. Apt* Tadeusz Radzik†

Abstract

We propose a notion of a stable partition in a coalitional game that is parametrized by the concept of a defection function. This function assigns to each partition of the grand coalition a set of different coalition arrangements for a group of defecting players. The alternatives are compared using their social welfare.

We characterize the stability of a partition for a number of most natural defection functions and investigate whether and how so defined stable partitions can be reached from any initial partition by means of simple transformations.

The approach is illustrated by analyzing an example in which a set of stores seeks an optimal transportation arrangement.

1 Introduction

The problem of coalition formation has become an important research direction in theoretical economics, notably game theory. It has been studied from many points of view beginning with [1], where the static situation of coalitional games in the presence of a given coalition structure (i.e., a partition) was considered and where the issue of coalition formation was briefly alluded to (on pages 233–234). The early research on the subject was discussed in [8].

More recently, the problem of formation of stable coalition structures was considered in [13] in the presence of externalities and in [11] in the presence of binding agreements. In both papers two-stage games are analyzed. In the first stage coalitions form and in the second stage the players engage in a non-cooperative game given the emerged coalition structure. In this context the question of stability of the coalition structure is then analyzed.

Much research on stable coalition structures focussed on hedonic games. These are games in which the payoff of a player depends exclusively on the members of the coalition he belongs to. In other words, a payoff of a player is a preference relation on the sets of players that include him. [4] considered four forms of stability in such games: core, Nash, individual stability and contractually individual stability. Each alternative captures the idea that no player, respectively, no group of playes has an incentive to change the existing coalition

*CWI, Amsterdam and University of Amsterdam, the Netherlands

†Institute of Mathematics, Wrocław University of Technology, Poland

structure. The problem of existence of (core, Nash, individually and contractually individually) stable coalitions was considered in this and other references, for example [12] and [5]. A potentially infinitely long coalition formation process in the context of hedonic games was studied in [2]. This leads to another notion of stability analogous to subgame perfect equilibrium.

Recently, [3] compared various notions of stability and equilibria in network formation games. These are games in which the players may be involved in a network relationship that, as a graph, may evolve. Other interaction structures which players can form were considered in [7], in which formation of hierarchies was studied, and [10] in which only bilateral agreements that follow a specific protocol were allowed.

Finally, the computer science perspective is illustrated by [6] in which an approach to coalition formation based on Bayesian reinforcement was considered and tested empirically.

In this paper we propose to study the existence and formation of stable coalition structures in the setting of coalitional games, by proposing and studying a concept that bears some similarity with the Nash equilibrium. We consider partitions of the grand coalition and view a partition stable if no group of players has a viable alternative to staying within the partition. The alternatives are provided as a set of different coalition arrangements for the group of defecting players and are compared using their social welfare.

The following example hopefully clarifies our approach. We shall return to it in the last section.

Example 1.1 Consider a set of stores located in a number of cities. Each store belongs to a chain. Suppose that each chain has a contract with a transportation company to deliver goods to all the stores belonging to the chain.

We can now envisage a situation in which a group of stores decides to leave this transportation arrangement and choose another one, for example the one in which the stores from the same city have a contract with one transportation company.

So from the viewpoint of the transportation logistics the stores from the 'defecting' group of stores are now partitioned not according to the chains they belong to but according to the cities they are located at. Such an alternative transportation arrangement is then a different, preferred, partition for the defecting group of stores. From our viewpoint the original transportation arrangement was then unstable. □

These alternatives to the existing partition of the grand coalition are formalized by means of a *defection function* that assigns to each partition a set of partitioned subsets of the grand coalition. By considering different defection functions we obtain different notions of stability. Two most natural defection functions are the one that allows formation of all partitions of all subsets and the one that allows formation of all partitions of the grand coalition.

We characterize these notions of stability and in the second part of the paper analyze the problem of whether and how so defined stable partitions can be reached from any initial partition by means of 'local' transformations.

2 Preliminary definitions

We begin by introducing the basic concepts. Let $N = \{1, 2, \ldots, n\}$ be a fixed set of players called the *grand coalition* and let (v, N) be a coalitional TU-game (in short a game). That is, v is a function from the powerset of N to $\mathcal{R}$. In what follows we assume that $v(\emptyset) = 0$. We call the elements of N *players* and non-empty subsets of N *coalitions*.

A game (v, N) is called

- *additive* if $v(A) + v(B) = v(A \cup B)$,
- *superadditive* if $v(A) + v(B) \leq v(A \cup B)$
- *strictly superadditive* if $v(A) + v(B) < v(A \cup B)$,

where in each case the condition holds for every two disjoint coalitions A and B of N.

A *collection* (in the grand coalition N) is any family $C := \{C_1, \ldots, C_l\}$ of mutually disjoint coalitions of N, and l is called its *size*. If additionally $\bigcup_{j=1}^{l} C_j = N$, the collection C is called a *partition* of N.

Given a collection $C := \{C_1, \ldots, C_l\}$ and a partition $P := \{P_1, \ldots, P_k\}$ we define

$$C[P] := \{P_1 \cap (\cup_{j=1}^{l} C_j), \ldots, P_k \cap (\cup_{j=1}^{l} C_j)\} \setminus \{\emptyset\}$$

and call $C[P]$ the *collection C in the frame of P*. (By removing the empty set we ensure that $C[P]$ is a collection.) To clarify this concept consider Figure 1. We depict in it a collection C, a partition P and C in the frame of P (together with P). Here C consists of three coalitions, while C in the frame of P consists of five coalitions.

Intuitively, given a subset S of N and a partition $C := \{C_1, \ldots, C_l\}$ of S, the collection C offers the players from S the 'benefits' resulting from the partition of S by C. However, if a partition P of N is 'in force', then the players from S enjoy instead the benefits resulting from the partition of S by $C[P]$, i.e., by C in the frame of P. Finally, note that if C is a partition of N, then $C[P]$ is simply P.

For a collection $C := \{C_1, \ldots, C_l\}$ we define now

$$sw(C) := \sum_{j=1}^{l} v(C_j)$$

and call $sw(C)$ the *social welfare* of C. So for a partition $P := \{P_1, \ldots, P_k\}$

$$sw(C[P]) = \sum_{i=1}^{k} v(P_i \cap (\cup_{j=1}^{l} C_j)).$$

We call $sw(C[P])$ the *P-modified social welfare* of the collection C. That is, the P-modified social welfare of C is the social welfare of C in the frame of P.

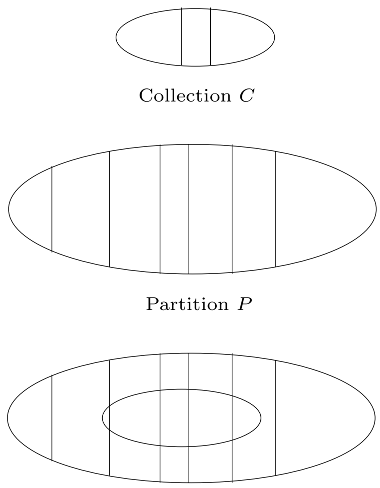

Figure 1: A collection C in the frame of a partition P

Given a partition $P := \{P_1, \ldots, P_k\}$ we call a coalition T of N P-*compatible* if for some $i \in \{1, \ldots, k\}$ we have $T \subseteq P_i$ and P-*incompatible* otherwise. Further, we call a partition $Q := \{Q_1, \ldots, Q_l\}$ P-*homogeneous* if for each $j \in \{1, \ldots, l\}$ some $i \in \{1, \ldots, k\}$ exists such that either $Q_j \subseteq P_i$ or $P_i \subseteq Q_j$. Equivalently, a partition Q is P-homogeneous if each Q_j is either P-compatible or equals $\cup_{i \in T} P_i$ for some $T \subseteq \{1, \ldots, k\}$. So any P-*homogeneous* partition arises from P by allowing each coalition either to split into smaller coalitions or to merge with other coalitions.

The crucial notion in our considerations is that of a *defection function*. It is a function $\mathbb{D}$ which associates with each partition P a family of collections in N. Intuitively we interpret $\mathbb{D}$ as follows. For a partition P the family $\mathbb{D}(P)$ consists of all the collections $C := \{C_1, \ldots, C_l\}$ whose players can leave the partition P by forming a new, separate, group of players $\cup_{j=1}^{l} C_j$ divided according to the collection C.

Let $\mathbb{D}$ be a defection function. A partition P of the grand coalition N is said to be $\mathbb{D}$-*stable* if

$$sw(C[P]) \geq sw(C) \tag{1}$$

for all collections $C \in \mathbb{D}(P)$.

If C is a partition, then, as already noted $C[P] = P$, so the above inequality reduces to the comparison of the social welfare of P and C:

$$sw(P) \geq sw(C). \tag{2}$$

This definition has the following natural interpretation. A partition P is $\mathbb{D}$-stable if no group of players is interested in leaving P when the players who wish to leave can only form collections allowed by the defection function $\mathbb{D}(P)$.

This is a consequence of the fact that the 'departing' players cannot improve upon their social welfare in comparison with their P-modified social welfare.

In what follows we shall consider three natural types of defection functions $\mathbb{D}$. Each will yield a different notion of a $\mathbb{D}$-stable partition.

3 $\mathbb{D}_c$-stability

We begin with the defection function $\mathbb{D}_c$, where for each partition P, $\mathbb{D}_c(P)$ is the family of all collections in N. So this defection function $\mathbb{D}_c$ allows any group of players to leave P and create an arbitrary collection in N. So by definition a partition P is $\mathbb{D}_c$-stable if and only if (1) holds for every collection C in N.

The following result shows that the $\mathbb{D}_c$-stability condition can be considerably simplified.

Theorem 3.1 *A partition $P := \{P_1, \ldots, P_k\}$ of N is $\mathbb{D}_c$-stable if and only if the following two conditions are satisfied:*

(i) for each $i \in \{1, \ldots, k\}$ and each pair of disjoint coalitions A and B such that $A \cup B \subseteq P_i$

$$v(A \cup B) \geq v(A) + v(B), \tag{3}$$

(ii) for each P-incompatible coalition $T \subseteq N$

$$\sum_{i=1}^{k} v(P_i \cap T) \geq v(T). \tag{4}$$

Proof.
($\Rightarrow$) Suppose that $A \cup B \subseteq P_i$ for some i and a pair A and B of disjoint coalitions. By taking the collection $C := \{A, B\}$ we get that (3) is an immediate consequence of (1) since by assumption $v(\emptyset) = 0$.

Now suppose that $T \subseteq N$ is a P-incompatible coalition. Then by taking the singleton collection $C := \{T\}$ we get that (4) is an immediate consequence of (1).

($\Leftarrow$) First note that (3) implies that for each $i \in \{1, \ldots, k\}$ and for each collection $C := \{C_1, \ldots, C_l\}$ with $l > 1$ and $\cup_{j=1}^{l} C_j \subseteq P_i$

$$v(\cup_{j=1}^{l} C_j) \geq \sum_{j=1}^{l} v(C_j). \tag{5}$$

Let now $C := \{C_1, \ldots, C_l\}$ be an arbitrary collection in N. Define $D^i := \{T \mid T \in C \text{ and } T \subseteq P_i\}$ for $i \in \{1, \ldots, k\}$. So D^i is the set of those elements of C that are subsets of P_i.

Further, let $E := C \setminus \cup_{i=1}^{k} D^i$. So $C \setminus E = \cup_{i=1}^{k} D^i$ and hence

$$\sum_{i=1}^{k} \sum_{T \in D^i} v(T) = \sum_{T \in C \setminus E} v(T).$$

Finally, let $E^i := \{T \cap P_i \mid T \in E\}$ for $i \in \{1, \dots, k\}$. By definition for every $i \in \{1, \dots, k\}$ and $T \in D^i \cup E^i$ we have $T \subseteq P_i$. Hence by (5) for $i \in \{1, \dots, k\}$

$$v(\cup_{T \in D^i \cup E^i} T) \geq \sum_{T \in D^i \cup E^i} v(T).$$

Moreover, for $i \in \{1, \dots, k\}$ we have

$$P_i \cap (\cup_{j=1}^l C_j) = \cup_{T \in D^i \cup E^i} T,$$

so

$$\sum_{i=1}^k v(P_i \cap (\cup_{j=1}^l C_j)) \geq \sum_{i=1}^k \sum_{T \in D^i \cup E^i} v(T).$$

Further, since $\cup_{i=1}^k D^i = C \setminus E$, using (4) we have the following chain of (in)equalities:

$$\sum_{i=1}^k \sum_{T \in D^i \cup E^i} v(T) = \sum_{T \in C \setminus E} v(T) + \sum_{i=1}^k \sum_{T \in E^i} v(T)$$

$$= \sum_{T \in C \setminus E} v(T) + \sum_{i=1}^k \sum_{T \in E} v(P_i \cap T) \geq \sum_{T \in C \setminus E} v(T) + \sum_{T \in E} v(T) = \sum_{j=1}^l v(C_j)\,.$$

So we have shown that conditions (3) and (4) imply (1). □

The following observation characterizes the games in which each partition is $\mathbb{D}_c$-stable.

Theorem 3.2 *A game (v, N) is additive if and only if each partition is $\mathbb{D}_c$-stable.*

Proof. (⇒) By Theorem 3.1.

(⇐) Take two disjoint coalitions A and B of N. First consider a partition of N which includes $A \cup B$ as an element. By item (i) of Theorem 3.1 we then get $v(A \cup B) \geq v(A) + v(B)$. Next, consider a partition P of N which includes A and B as elements. Then $A \cup B$ is P-incompatible so by item (ii) of Theorem 3.1 we get then $v(A) + v(B) \geq v(A \cup B)$. So (v, N) is additive. □

In turn, the following observation shows that only few partitions can be $\mathbb{D}_c$-stable.

Note 3.3 *If P is a $\mathbb{D}_c$-stable partition, then*

$$sw(P) = \max_Q sw(Q), \tag{6}$$

where the maximum is taken over all partitions Q in N.

Proof. For any partition C of N (1) reduces to (2). $\Box$

In fact, in general $\mathbb{D}_c$-stable partitions do not need to exist.

Example 3.4 Consider the game (v, N) with $N = \{1, 2, 3\}$ and v defined by:

$$v(S) := \begin{cases} 2 & \text{if } |S| = 1 \\ 5 & \text{if } |S| = 2 \\ 6 & \text{if } |S| = 3 \end{cases}$$

We now show that no partition of N is $\mathbb{D}_c$-stable. By Note 3.3 it suffices to check that no partition that maximizes the social welfare in the set of all partitions is $\mathbb{D}_c$-stable.

First note that the social welfare of a partition P is maximized when $|P| = 2$ —in that case $sw(P) = 7$. Then $P = \{P_1, P_2\}$ with $|P_1| = 2$ and $|P_2| = 1$. Suppose now P is $\mathbb{D}_c$-stable. We can assume that $P_1 = \{1, 2\}$ and $P_2 = \{3\}$. (The other subcases are symmetric since v is symmetric.) Now putting $T := \{2, 3\}$ in condition (4) of Theorem 3.1 we get $2+2 \geq 5$, which is a contradiction. $\Box$

For specific types of games the $\mathbb{D}_c$-stable partitions do exist, as the following two direct corollaries to Theorem 3.1 and Note 3.3 show.

Corollary 3.5 *The one element partition $P = \{N\}$ is $\mathbb{D}_c$-stable if and only if the game (v, N) is superadditive. Moreover, when (v, N) is strictly superadditive, then P is a unique $\mathbb{D}_c$-stable partition.*

Corollary 3.6 *The partition $P = \{\{1\}, \{2\}, \ldots, \{n\}\}$ is $\mathbb{D}_c$-stable if and only if the inequality $\sum_{i \in T} v(\{i\}) \geq v(T)$ holds for all $T \subseteq N$. Moreover, when all inequalities are strict, then P is a unique $\mathbb{D}_c$-stable partition.*

4 $\mathbb{D}_p$-stability

We now consider a weaker version of stability for which stable partitions are guaranteed to exist. To this end we consider the defection function $\mathbb{D}_p$, where for each partition P, $\mathbb{D}_p(P)$ is the family of all partitions of N. So the defection function $\mathbb{D}_p$ allows a group of players to leave P only as the group of all players. They can form then an arbitrary partition of N.

We have the following simple result.

Theorem 4.1 *A partition P is $\mathbb{D}_p$-stable if and only if (6) holds, where the maximum is taken over all partitions Q in N.*

Proof. For any partition C of N (1) reduces to (2). $\Box$

So the $\mathbb{D}_p$-stable partitions are exactly those that maximize the social welfare in the set of all partitions. Consequently the set of $\mathbb{D}_p$-stable partitions is non-empty. However, in specific situations the notion of a $\mathbb{D}_p$-stable partition may be inadequate.

As an example suppose there are k locations and we wish to associate 'optimally' each player with a location, where optimality means that the resulting social welfare is maximized. It may easily happen that in the considered game all $\mathbb{D}_p$-stable partitions have more than k coalitions, so an optimal partition in the above sense cannot be described as a $\mathbb{D}_p$-stable partition. To cope with such situations we modify this notion as follows.

Given $k \in \{1, \ldots, n\}$ let $\mathbb{D}_p^k$ be the defection function such that for each partition P, $\mathbb{D}_p^k(P)$ is the family of all partitions of size at most k. Then the following result holds.

Theorem 4.2 *For each $k \in \{1, \ldots, n\}$ there exists a partition P of size at most k which is $\mathbb{D}_p^k$-stable.*

Proof. Let P be any partition of size at most k and satisfying the equality

$$sw(P) = \max_{Q^k} sw(Q^k),$$

where the maximum is taken over all partitions Q^k of size at most k. Then P is the desired partition. □

5 $\mathbb{D}_{hp}$-stability

Finally, we focus on the defection function $\mathbb{D}_{hp}$, where for each partition P, $\mathbb{D}_{hp}(P)$ is the family of all P-homogeneous partitions in N. So the defection function $\mathbb{D}_{hp}$ allows the players to leave the partition P only by means of (possibly multiple) merges or splittings. The existence of a $\mathbb{D}_{hp}$-stable partition is then guaranteed by Theorem 4.1 since every $\mathbb{D}_p$-stable partition is also $\mathbb{D}_{hp}$-stable. Moreover, the following obvious analogue of Theorem 4.1 holds.

Theorem 5.1 *A partition P is $\mathbb{D}_{hp}$-stable if and only if (6) holds, where the maximum is taken over all P-homogeneous partitions Q in N.*

Also, this notion of stability admits the following characterization.

Theorem 5.2 *A partition $P := \{P_1, \ldots, P_k\}$ of N is $\mathbb{D}_{hp}$-stable if and only if the following two conditions are satisfied:*

(i) for each $i \in \{1, \ldots, k\}$ and for each partition $\{C_1, \ldots, C_l\}$ of the coalition P_i

$$v(P_i) \geq \sum_{j=1}^{l} v(C_j), \tag{7}$$

(ii) for each $T \subseteq \{1, \ldots, k\}$

$$\sum_{i \in T} v(P_i) \geq v(\cup_{i \in T} P_i). \tag{8}$$

Proof.
($\Rightarrow$) Let $C := \{C_1, \ldots, C_l\}$ be an arbitrary partition of some coalition P_i in P. Hence $\cup_{j=1}^{l} C_j = P_i$ and (7) is an immediate consequence of (1) applied to the P-homogenous partition $\{C_1, \ldots, C_l, P_1, \ldots, P_{i-1}, P_{i+1}, \ldots, P_n\}$.

Next, consider some $T \subseteq N$. Then by applying (1) to the partition $C := \{\cup_{i \in T} P_i\} \cup \{P_i \mid i \notin T\}$ we get the inequality (8).

($\Leftarrow$) Let $P := \{P_1, \ldots, P_k\}$ be any partition of N for which conditions (i) and (ii) hold, and let $C := \{C_1, \ldots, C_l\}$ be an arbitrary collection in $\mathbb{D}_{hp}(P)$. By definition C is a P-homogeneous partition. Therefore the coalitions of P can be divided into disjoint groups $\{G_1, \ldots, G_r\}$ of collections as follows: P_i and P_j belong to the same group if and only if $P_i \cup P_j \subseteq C_s$ for some s.

Let us define $L = \{t \mid |G_t| = 1\}$ and $M = \{t \mid |G_t| > 1\}$. By the P-homogeneity of partition C, for each $t \in L$ there is a set, say H_t, such that $P_t = \cup_{j \in H_t} C_j$. Similarly, for each $t \in M$ there is a coalition of C, say $C_{q(t)}$, such that $C_{q(t)} = \cup_{P_i \in G_t} P_i$. Hence, using (7) and (8), we get the following chain of (in)equalities:

$$\sum_{i=1}^{k} v(P_i \cap (\cup_{j=1}^{l} C_j)) = \sum_{t=1}^{k} v(P_t) = \sum_{t \in L} v(\cup_{j \in H_t} C_j) + \sum_{t \in M} \sum_{P_i \in G_t} v(P_i)$$

$$\geq \sum_{t \in L} \sum_{j \in H_t} v(C_j) + \sum_{t \in M} v(\cup_{P_i \in G_t} P_i) = \sum_{t \in L} \sum_{j \in H_t} v(C_j) + \sum_{t \in M} v(C_{q(t)}) = \sum_{j=1}^{l} v(C_j)\,.$$

So we have shown that conditions (7) and (8) imply (1). □

To summarize the relationship between the considered notions of stable partition, given a defection function $\mathbb{D}$ denote by $\mathcal{ST}(\mathbb{D})$ the set of $\mathbb{D}$-stable partitions. We have then the following obvious inclusions:

$$\mathcal{ST}(\mathbb{D}_c) \subseteq \mathcal{ST}(\mathbb{D}_p) \subseteq \mathcal{ST}(\mathbb{D}_{hp}).$$

With the possible exception of $\mathcal{ST}(\mathbb{D}_c)$ the considered sets of stable partitions are always non-empty.

6 Finding stable partitions

Next we consider the problem of finding stable partitions studied in the previous sections. To this end we introduce two rules that allow us to modify a partition of N.

merge

$$\{T_1, \ldots, T_k\} \cup P \rightarrow \{\cup_{j=1}^{k} T_j\} \cup P,$$

where $\sum_{j=1}^{k} v(T_j) < v(\cup_{j=1}^{k} T_j)$,

split

$$\{\cup_{j=1}^k T_j\} \cup P \to \{T_1, \ldots, T_k\} \cup P,$$

where $\{T_1, \ldots, T_k\}$ is a collection such that $v(\cup_{j=1}^k T_j) < \sum_{j=1}^k v(T_j)$,

The following observation holds.

Note 6.1 *Every iteration of the merge and split rules terminates.*

Proof. This is an immediate consequence of the fact that each rule application increases the social welfare. □

We now proceed with the characterizations of the introduced notions of stable partitions using the above rules. The cases of $\mathbb{D}_c$-stable partitions and $\mathbb{D}_p$-stable partitions are not so straightforward, so we begin with the notion of $\mathbb{D}_{hp}$-stability. The following result holds.

Theorem 6.2 *A partition is $\mathbb{D}_{hp}$-stable if and only if it is the outcome of iterating the merge and split rules.*

Proof. It is an immediate consequence of Theorem 5.2. □

So to find a $\mathbb{D}_{hp}$-stable partition it suffices to iterate the merge and split rules starting from any initial partition until one reaches a partition closed under the applications of these rules.

In general the outcome of various iterations of the merge and split rules does not need to be unique. Moreover, some of these outcomes do not have a maximal social welfare.

Example 6.3 Consider the following game (v, N). Let $N = \{1, 2, 3, 4\}$ and let v be defined as follows:

$$v(S) := \begin{cases} 1 & \text{if } S = \{1, 2\} \\ 2 & \text{if } S = \{1, 3\} \\ 0 & \text{otherwise} \end{cases}$$

Consider now the partition $\{\{1\}, \{2\}, \{3\}, \{4\}\}$ of N. By the merge rule we can transform it to $\{\{1, 2\}, \{3\}, \{4\}\}$ or to $\{\{1, 3\}, \{2\}, \{4\}\}$. The social welfare of these partitions is, respectively, 1 and 2. In each case we reached a partition to which neither merge nor split rule can be applied. □

We now proceed with an analysis of $\mathbb{D}_c$-stable partitions using the merge and split rules. First note the following observation.

Note 6.4 *Every $\mathbb{D}_c$-stable partition is closed under the applications of the merge and split rules.*

Proof. This is an immediate consequence of Note 3.3. □

Unfortunately it is not possible to characterize $\mathbb{D}_c$-stable partitions using the merge and split rules. First, not every partition closed under the applications of the merge and split rules is $\mathbb{D}_c$-stable. Indeed, take the game from Example 3.4 and consider a partition with the maximal social welfare. This partition is closed under the applications of the merge and split rules but we saw already that in this game no $\mathbb{D}_c$-stable partition exists.

Second, there are games in which $\mathbb{D}_c$-stable partitions exist, but some iterations of the merge and split rules may miss them.

Example 6.5 Let $N = \{1, 2, 3, 4\}$ and define v as follows:

$$v(S) := \begin{cases} 3 & \text{if } S = \{1, 2\} \\ |S| & \text{otherwise} \end{cases}$$

It is straightforward to show using Theorem 3.1 that $\{\{1, 2\}, \{3, 4\}\}$ is a $\mathbb{D}_c$-stable partition. Its social welfare is 5. Now, the partition $\{\{1, 3\}, \{2, 4\}\}$ is closed under the applications of the merge and split rules. But its social welfare is 4 so by Note 3.3 it is not $\mathbb{D}_c$-stable. □

Note that in the above example $\{\{1, 2\}, \{3, 4\}\}$ is also a unique $\mathbb{D}_p$-stable partition (since it is the only partition with the social welfare 5), so some iterations of the merge and split rules may also miss the $\mathbb{D}_p$-stable partitions.

Third, there are games in which all iterations of the merge and split rules have a unique outcome (which happens to be the unique partition that maximizes the social welfare in the set of all partitions), yet no $\mathbb{D}_c$-stable partition exists.

Example 6.6 Consider the following game. Let $N = \{1, 2, 3, 4\}$ and let v be defined as follows:

$$v(S) := \begin{cases} 6 & \text{if } S = \{1, 2, 3, 4\} \\ 4 & \text{if } S = \{1, 2\} \text{ or } S = \{3, 4\} \\ 3 & \text{if } S = \{1, 3\} \\ |S| & \text{otherwise} \end{cases}$$

Clearly neither merge nor split rule can be applied to the partition $\{\{1, 2\}, \{3, 4\}\}$. We now show that $\{\{1, 2\}, \{3, 4\}\}$ is a unique outcome of the iterations of the merge and split rules. Take a partition $\{T_1, \ldots, T_k\}$ different from $\{\{1, 2\}, \{3, 4\}\}$.

If for some i we have $T_i = \{1, 2\}$, then $\{T_1, \ldots, T_k\} = \{\{1, 2\}, \{3\}, \{4\}\}$ and consequently $\{\{1, 2\}, \{3\}, \{4\}\} \rightarrow \{\{1, 2\}, \{3, 4\}\}$ by the merge rule. If for some i we have $T_i = \{3, 4\}$, then the argument is symmetric. If for some i we have $T_i = \{1, 3\}$, then for $j \neq i$ we have $v(T_j) = |T_j|$. Consequently, by the merge and split rule

$$\{T_1, \ldots, T_k\} \rightarrow \{\{1, 2, 3, 4\}\} \rightarrow \{\{1, 2\}, \{3, 4\}\}.$$

If $\{T_1, \ldots, T_k\} = \{\{1,2,3,4\}\}$, then by the split rule

$$\{\{1,2,3,4\}\} \to \{\{1,2\},\{3,4\}\}.$$

If no T_i equals $\{1,2\}, \{3,4\}, \{1,3\}$ or $\{1,2,3,4\}$, then for all i we have $v(T_i) = |T_i|$ and consequently by the merge rule and by the split rule

$$\{T_1, \ldots, T_k\} \to \{\{1,2,3,4\}\} \to \{\{1,2\},\{3,4\}\}.$$

So in this game all iterations of the merge and split rules have a unique outcome, $P := \{\{1,2\},\{3,4\}\}$. By Note 6.4 P is the only possible $\mathbb{D}_c$-stable partition. However, for the P-incompatible set $\{1,3\}$ we have

$$v(\{1,3\}) > v(\{1\}) + v(\{3\}) = v(\{1,3\} \cap \{1,2\}) + v(\{1,3\} \cap \{3,4\}).$$

So by Theorem 3.1 P is not $\mathbb{D}_c$-stable and consequently this game has no $\mathbb{D}_c$-stable partitions. ☐

A natural question then arises whether some other simple rules exist using which we could characterize the $\mathbb{D}_c$-stable and $\mathbb{D}_p$-stable partitions.

This is very unlikely. To clarify the matters let us return to Example 6.5. We noted there that the partition $\{\{1,3\},\{2,4\}\}$ (with social welfare 4) is not $\mathbb{D}_c$-stable but is closed the applications of the merge and split rules.

So to transform $\{\{1,3\},\{2,4\}\}$ to $\{\{1,2\},\{3,4\}\}$, which is a unique $\mathbb{D}_c$-stable and $\mathbb{D}_p$-stable partition (and with social welfare 5), we need more powerful rules. As we limit ourselves to rules that lead to a strict increase of the social welfare, the only solution is to transform $\{\{1,3\},\{2,4\}\}$ to $\{\{1,2\},\{3,4\}\}$ directly, in one rule application. In particular, the following natural rule does not suffice:

transfer

$$\{T_1, T_2\} \cup P \to \{T_1 \setminus U,\ T_2 \cup U\} \cup P,$$

where $U \subset T_1$ and $v(T_1) + v(T_2) < v(T_1 \setminus U) + v(T_2 \cup U)$.

What we need is a rule that leads to a 'bidirectional transfer', for example

exchange

$$\{T_1, T_2\} \cup P \to \{T_1 \setminus U_1 \cup U_2,\ T_2 \setminus U_2 \cup U_1\} \cup P,$$

where $U_1 \subset T_1$, $U_2 \subset T_2$ and $v(T_1) + v(T_2) < v(T_1 \setminus U_1 \cup U_2) + v(T_2 \setminus U_2 \cup U_1)$.

This rule suffices here. But it is easy to construct an example where this rule does not suffice either. We just need to generalize appropriately Example 6.5.

Let $N := \{1, 2, \ldots, 2n$, where $n > 1$, and define v as follows:

$$v(S) := \begin{cases} \text{n+1} & \text{if } S = \{1, 3, \ldots, 2n-1\} \\ |S| & \text{otherwise} \end{cases}$$

Using Theorems 3.1 and 4.1 it is straightforward to check that $\{\{1, 3, \ldots, 2n-1\}, \{2, 4, \ldots, 2n\}\}$ is the unique $\mathbb{D}_c$-stable and unique $\mathbb{D}_p$-stable partition. Its social welfare is $2n+1$. Now the partition $\{P_1, \ldots, P_n\}$, where $P_i := \{2i-1, 2i\}$ for $i \in \{1, \ldots, n\}$ has the social welfare $2n$. So to transform it to $\{\{1, 3, \ldots, 2n-1\}, \{2, 4, \ldots, 2n\}\}$ we need a rule that can achieve it in one application. Of course such a rule exists: it suffices to group the odd numbers into one set and the even numbers into another. But it should be clear that any generic way of formulating this operation leads to a pretty complex rule.

Continuing this line it is easy to envisage a series of increasingly more complex examples which suggest that in the end the only rule using which we can characterize the $\mathbb{D}_p$-stable partitions seems to be the one that allows us to transform an arbitrary partition into another one when the social welfare increases. This defeats our purpose of finding a characterization by means of simple rules. The fact that the $\mathbb{D}_c$-stable partitions do not need to exist makes the task of characterizing them by means of simple rules even more unlikely.

7 Strictly stable partitions

As a way out of this dilemma we consider a more refined notion of a stable partition. Take a partition P of N. First note that for a collection C that consists of some coalitions of P, i.e., for $C \subseteq P$ we have $C[P] = C$ and consequently (1) then holds.

Given a defection function $\mathbb{D}$ we now say that a partition P of N is *strictly $\mathbb{D}$-stable* if

$$sw(C[P]) > sw(C)$$

for all collections $C \in \mathbb{D}(P)$ that are not subsets of P.

We now analyze strictly $\mathbb{D}_c$-stable partitions. The following analogue of Theorem 3.1 holds.

Theorem 7.1 *A partition $P := \{P_1, \ldots, P_k\}$ of N is strictly $\mathbb{D}_c$-stable if and only if the following two conditions are satisfied:*

- *for each $i \in \{1, \ldots, k\}$ and each pair of disjoint coalitions A and B such that $A \cup B \subseteq P_i$*

$$v(A \cup B) > v(A) + v(B), \tag{9}$$

- *for each P-incompatible coalition $T \subseteq N$*

$$\sum_{i=1}^{k} v(P_i \cap T) > v(T). \tag{10}$$

Note that (9) and (10) are simply the sharp counterparts of the inequalities (3) and (4).

Proof. The proof is a direct modification of the proof of Theorem 3.1. □

Next, we establish the following lemma.

Lemma 7.2 *Suppose that a strictly $\mathbb{D}_c$-stable partition P exists. Let P' be a partition which is closed under the applications of the merge and split rules. Then $P' = P$.*

Proof. Suppose $P = \{P_1, \ldots, P_k\}$ and $P' = \{T_1, \ldots, T_m\}$. Assume by contradiction that $\{P_1, \ldots, P_k\} \neq \{T_1, \ldots, T_m\}$. Then $\exists i_0 \in \{1, \ldots, k\}\ \forall j \in \{1, \ldots, m\}\ P_{i_0} \neq T_j$. Let $T_{j_1}, \ldots, T_{j_l}$ be the minimum cover of P_{i_0}.

Case 1. $P_{i_0} = \cup_{h=1}^{l} T_{j_h}$.

Then $l > 1$ and $\{T_{j_1}, \ldots, T_{j_l}\}$ is a partition of P_{i_0}. But P is strictly $\mathbb{D}_c$-stable, so by Theorem 7.1 and (9) $\sum_{h=1}^{l} v(T_{j_h}) < v(\cup_{h=1}^{l} T_{j_h})$. Consequently, the merge rule can be applied to $\{T_1, \ldots, T_m\}$, which is a contradiction.

Case 2. P_{i_0} is a proper subset of $\cup_{h=1}^{l} T_{j_h}$.

Then for some j_h the set $P_{i_0} \cap T_{j_h}$ is a proper non-empty subset of T_{j_h}. So T_{j_h} is a P-incompatible set. But P is strictly $\mathbb{D}_c$-stable, so by Theorem 7.1 and (10) $v(T_{j_h}) < \sum_{i=1}^{k} v(P_i \cap T_{j_h})$. Consequently, since $T_{j_h} = \cup_{i=1}^{k}(P_i \cap T_{j_h})$, the split rule can be applied to $\{T_1, \ldots, T_m\}$ which is a contradiction. □

This allows us to draw the following conclusions.

Theorem 7.3 *Suppose that P is a strictly $\mathbb{D}_c$-stable partition. Then*

(i) P is the outcome of every iteration of the merge and split rules.

(ii) P is a unique $\mathbb{D}_c$-stable partition.

(iii) P is a unique $\mathbb{D}_p$-stable partition.

(iv) P is a unique $\mathbb{D}_{hp}$-stable partition.

Proof.
(i) By Note 6.1 every iteration of the merge and split rules terminates, so the claim follows by Lemma 7.2.

(ii) Suppose that P' is a $\mathbb{D}_c$-stable partition. By Note 6.4 P' is closed under the applications of the merge and split rules, so by Lemma 7.2 $P' = P$.

(iii) By Theorem 4.1 a partition is $\mathbb{D}_p$-stable if and only if it maximizes the social welfare (in the set of all partitions). But each partition that maximizes the social welfare is closed under the applications of the merge and split rules, so the claim follows from (i).

(iv) By (i) and Theorem 6.2. □

Item (i) shows that if a strictly $\mathbb{D}_c$-stable partition exists, then we can reach it from any initial partition through an arbitrary iteration of the merge and split rules.

Example 6.5 shows that the concepts of unique and strictly $\mathbb{D}_c$-stable partitions do not coincide. Indeed, by Note 3.3 $\{\{1, 2\}, \{3, 4\}\}$ is there a unique

$\mathbb{D}_c$-stable partition. However, Theorem 7.1 implies that $\{\{1,2\},\{3,4\}\}$ is not strictly $\mathbb{D}_c$-stable. This is in contrast to the case of $\mathbb{D}_p$-stable partitions as the following characterization result shows.

Theorem 7.4 *A partition is strictly $\mathbb{D}_p$-stable if and only if it is a unique $\mathbb{D}_p$-stable partition.*

Proof.
($\Rightarrow$) Let P be a strictly $\mathbb{D}_p$-stable partition. By definition $sw(C[P]) > sw(C)$ for all partitions C different from P, or equivalently $sw(P) > sw(P')$ for all partitions P' different from P. Let P' be a $\mathbb{D}_p$-stable partition. By Theorem 4.1 both P and P' maximize the social welfare in the set of all partitions. So $sw(P) = sw(P')$ and consequently P and P' coincide.

($\Leftarrow$) Suppose P is a unique $\mathbb{D}_p$-stable partition. Let P' be a partition different from P. By Theorem 4.1 and uniqueness of P $sw(P) > sw(P')$. So P is a strictly $\mathbb{D}_p$-stable partition. $\Box$

The full analogue of Theorem 7.3 does not hold. Indeed, consider the following modification of Example 6.5.

Example 7.5 Let $N = \{1,2,3\}$ and define v as follows:

$$v(S) := \begin{cases} 3 & \text{if } S = \{1,2\} \\ |S| & \text{otherwise} \end{cases}$$

By Theorems 4.1 and 7.4 $\{\{1,2\},\{3\}\}$ is a strictly $\mathbb{D}_p$-stable partition. Its social welfare is 4. But the partition $\{\{1,3\},\{2\}\}$ is closed under the applications of the merge and split rules and hence, by Theorem 6.2, is $\mathbb{D}_{hp}$-stable. So $\{\{1,2\},\{3\}\}$ is not the outcome of every iteration of the merge and split rules and is not a unique $\mathbb{D}_{hp}$-stable partition. $\Box$

In turn, Example 6.6 shows that the existence of a strictly $\mathbb{D}_p$-stable partition does not imply the existence of a $\mathbb{D}_c$-stable partition.

Finally, the following result deals with the strictly $\mathbb{D}_{hp}$-stable partitions.

Theorem 7.6

(i) A partition is strictly $\mathbb{D}_{hp}$-stable if and only if it is a unique $\mathbb{D}_{hp}$-stable partition.

(ii) A partition is strictly $\mathbb{D}_{hp}$-stable if and only if it is strictly $\mathbb{D}_p$-stable.

Proof.
(i) The proof is the same as that of Theorem 7.4, relying on Theorem 5.1 instead of Theorem 4.1.

(ii)
($\Rightarrow$) Let P be a strictly $\mathbb{D}_{hp}$-stable partition. Take an arbitrary partition P' different from P and let P'' be an arbitrary closure of P' under the applications

of the merge and split rules. By Theorem 6.2 P'' is $\mathbb{D}_{hp}$-stable, so by the choice of P either $P = P''$ or $sw(P) > sw(P'')$.

If $P = P''$, then, by the choice of P', P' is different from P'', so $sw(P'') > sw(P')$ and consequently $sw(P) > sw(P')$. In turn, if $sw(P) > sw(P'')$, then, since $sw(P'') \geq sw(P')$, we get $sw(P) > sw(P')$, as well.

($\Leftarrow$) Directly by definition. □

8 Existence of stable partitions

We saw already that in general a $\mathbb{D}_c$-stable partition does not need to exist, so a strictly $\mathbb{D}_c$-stable partition does not need to exist either.

Under what condition a strictly $\mathbb{D}_c$-stable partition does exists? First note that by definition if the game is (strictly) superadditive, then $\{N\}$ is its (strictly) $\mathbb{D}_c$-stable partition. The following example introduces a natural class of non-superadditive games in which a $\mathbb{D}_c$-stable (respectively, a strictly $\mathbb{D}_c$-stable) partition exists.

Example 8.1 Consider a partition $P := \{P_1, \ldots, P_k\}$ of N. Let (v, N) be a game which is non-negative (that is, $v(A) \geq 0$ for all coalitions A of N), superadditive when limited to subsets of a coalition P_i (for all $i \in \{1, \ldots, n\}$) and such that $v(A) = 0$ for all P-incompatible sets. It is straightforward to check with the help of Theorem 3.1 that P is a (not necessarily unique —see Theorem 3.2) $\mathbb{D}_c$-stable partition.

But if we additionally stipulate that (v, N) is positive (that is, $v(A) > 0$ for all coalitions A of N) and strictly superadditive, in each case when limited to subsets of a coalition P_i (for all $i \in \{1, \ldots, n\}$), then P becomes a strictly $\mathbb{D}_c$-stable partition, and hence a unique partition with this property.

To see specific examples of such games choose a partition $\{P_1, \ldots, P_k\}$ of N and fix $m \geq 1$. Let

$$v(S) := \begin{cases} |S|^m & \text{if } S \subseteq P_i \text{ for some } i \\ 0 & \text{otherwise} \end{cases}$$

Then when $m = 1$ we get an example of a game in the first category and when $m > 1$ we get an example of a game in the second category. □

Next, Theorems 4.1 and 7.4 provide a simple criterion for the existence of a strictly $\mathbb{D}_p$-stable partition: such a partition exists if and only if exactly one partition maximizes the social welfare in the set of all partitions. By virtue of Theorem 7.6 the same existence criterion applies to strictly $\mathbb{D}_{hp}$-stable partitions.

Finally, let us return to our initial Example 1.1. Assume that:

- within each city the transportation costs per store decrease as the number of served stores increases (economy of scale),

- the transportation costs per store are always lower if the served stores are located in the same city,
- the stores aim at minimizing the transportation costs and that switching transportation companies incurs no costs.

To formally analyze this example assume that there are n stores and that $\{P_1, \ldots, P_k\}$ is the partition of the stores per city. Denote by $c(S)$ the total transportation costs to the set of stores S and let

$$v(S) := \sum_{i \in S} c(\{i\}) - c(S),$$

i.e., $v(S)$ is the cost saving for coalition S.

Then the first assumption states that the function $c(S)/|S|$, when limited to the sets of stores within one city, strictly decreases as the size of the set of stores S increases. This easily implies that the game (v, N) is strictly superadditive when limited to the set of stores within one city (see [9, pages 93-95]).[1]

In turn, the second assumption states that for each P-incompatible coalition $T \subseteq N$ (representing a set of stores from different cities) we have for $i \in \{1, \ldots, k\}$

$$c(P_i \cap T)/|P_i \cap T| < c(T)/|T|,$$

or equivalently $c(P_i \cap T) < c(T) \cdot |P_i \cap T|/|T|$. This implies $\sum_{i=1}^{k} c(P_i \cap T) < c(T)$, so $\sum_{i=1}^{k} v(P_i \cap T) > v(T)$ by the definition of v.

So using Theorem 7.1 we get that the partition $\{P_1, \ldots, P_k\}$ of the stores per city is strictly $\mathbb{D}_c$-stable. Using Theorem 7.3(i) we now conclude that the initial transportation arrangement, per chain, can be broken and will lead through an arbitrary sequence of splits and merges to the alternative transportation arrangement, per city.

References

[1] R.J. Aumann and J.H. Drèze. Cooperative games with coalition structures. *International Journal of Game Theory*, (3):217–237, 1974.

[2] F. Bloch and E. Diamantoudi. Noncooperative formation in coalitions in hedonic games, 2005. Working paper.

[1] The argument is as follows. Let S and T be disjoint coalitions, both subsets of some P_i. For some $\alpha, \beta \in (0, 1)$

$$c(S \cup T)/|S \cup T| = \alpha \cdot c(S)/|S| = \beta \cdot c(T)/|T|.$$

Let $\gamma := |S|/|S \cup T|$. Then $1 - \gamma = |T|/|S \cup T|$. Hence $\gamma \cdot c(S \cup T) = \alpha \cdot c(S)$ and $(1 - \gamma) \cdot c(S \cup T) = \beta \cdot c(T)$. Consequently

$$c(S \cup T) = \alpha \cdot c(S) + \beta \cdot c(T) < c(S) + c(T).$$

So by definition of v we get $v(S) + v(T) < v(S \cup T)$.

[3] F. Bloch and M. Jackson. Definitions of equilibrium in network formation, 2005. Working paper.

[4] A. Bogomolnaia and M. Jackson. The stability of hedonic coalition structures. *Games and Economic Behavior*, 38(2):201–230, 2002.

[5] N. Burani and W.S. Zwicker. Coalition formation games with separable preferences. *Mathematical Social Sciences*, 45(1):27–52, 2003.

[6] G. Chaldiakis and C. Boutilier. Bayesian reinforcement learning for coalition formation under uncertainty. *AAMAS '04* Conference, 2004.

[7] G. Demange. On group stability in hierarchies and networks. 2004. Working paper.

[8] J. Greenberg. Coalition structures. In R.J. Aumann and S. Hart, editors, *Handbook of Game Theory with Economic Applications*, volume 2 of *Handbook of Game Theory with Economic Applications*, chapter 37, pages 1305–1337. Elsevier, 1994.

[9] H. Moulin. Axioms of Cooperative Decision Making. Cambridge University Press, 1988.

[10] I. Macho-Stadler, D. Prez-Castrillo, and N. Porteiro. Sequential formation of coalitions through bilateral agreements, 2005. Working paper.

[11] D. Ray and R. Vohra. Equilibrium binding agreements. *Journal of Economic Theory*, (73):30–78, 1997.

[12] T. Sönmez, S. Banerjee, and H. Konishi. Core in a simple coalition formation game. *Social Choice and Welfare*, 18(1):135–153, 2001.

[13] S.S. Yi. Stable coalition structures with externalities. *Games and Economic Behavior*, 20:201–237, 1997.