A Generic Approach to Coalition Formation

Krzysztof R. Apt and Andreas Witzel

Abstract

We propose an abstract approach to coalition formation by focusing on partial preference relations between partitions of a grand coalition. Coalition formation is modelled by means of simple merge and split rules that transform partitions. We identify conditions under which every iteration of these rules yields a unique partition. The main conceptual tool is the notion of a stable partition. The results naturally apply to coalitional TU-games and to some classes of hedonic games.

1 Introduction

1.1 Background

Coalition formation has been a research topic of continuing interest in the area of coalitional games. It has been analyzed from several points of view, starting with [2], where the static situation of cooperative games in the presence of a given coalition structure (i.e., a partition) was considered. Early research on the subject is discussed in [10].

More recently, the problem of formation of stable coalition structures was considered in [15] in the presence of externalities and in [13] 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 focused 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. [5] considered four forms of stability in such games: core, Nash, individual and contractually individual stability. Each alternative captures the idea that no player, respectively, no group of players has an incentive to change the existing coalition structure. The problem of existence of (core, Nash, individually and contractually individually) stable coalitions was considered in this and other references, for example [14] and [6]. A potentially infinitely long coalition formation process in the context of hedonic games was studied in [3]. This leads to another notion of stability analogous to subgame perfect equilibrium.

Recently, [4] 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 [8], in which formation of hierarchies was studied, and [11] in which only bilateral agreements that follow a specific protocol were allowed. Various aspects of coalition formation are also discussed in the recent collection of articles [9].

In [1] we introduced the concept of a stable partition for coalitional TU-games and investigated whether and how so defined stable partitions can be reached from any initial partition by means of simple transformations. The underlying concept of 'quality' of a partition was defined there by means of social welfare, which is simply the summed value of the partition.

Finally, the computer science perspective is illustrated by [7] in which an approach to coalition formation based on Bayesian reinforcement was considered and tested empirically.

1.2 Approach

In this paper we generalize the approach of [1] and investigate the idea of coalition formation in an abstract setting. To this end we introduce an abstract preference relation $\triangleright$ between partitions of any subset of players. We then model coalition formation by means of simple transformations of partitions of the grand coalition through merges and splits that yield a 'local' improvement w.r.t. the $\triangleright$ preference relation.

We then turn to the question of identifying conditions to ensure that arbitrary sequences of merges and splits yield the same outcome. We provide an answer to this question by imposing natural conditions on the $\triangleright$ preference relation (namely transitivity and monotonicity) and by considering a parametrized concept of a stable partition.

The introduced notion of a stable partition focuses only on the way a group of players is partitioned. Intuitively, a partition P of the grand coalition is stable w.r.t. a class of partioned groups iff no such group gains advantage (modelled by an improvement w.r.t. $\triangleright$) by changing the way it is partitioned by P to its own partition.

This way we obtain a generic presentation that allows us to study the idea of coalition formation by focusing only on an abstract concept of the 'quality' of a partition. In particular this analysis does not take into account any allocations to individual players. Also, in our results no specific coalitional game is assumed.

In the setting of coalitional TU-games we obtain results for concrete preference relations induced by specific orders, some of which are discussed in [12], viz. the utilitarian, Nash, egalitarian and leximin orders. We also discuss applications to hedonic games.

In our future work we plan to incorporate into this analysis the concept of a network structure. In this context a *network* is an undirected graph on the set of players that makes explicit the direct links between players. In the presence of a network only coalitions formed by connected players are allowed.

The paper is organized as follows. In the next section we set the stage by introducing an abstract comparison relation between partitions of a group of players and the corresponding merge and split rules that act on such partitions. Then in Section 3 we discuss a number of natural comparison relations on partitions within the context of coalitional TU-games. Next, in Section 4, we introduce and study a parametrized concept of a stable partition and in Section 5 relate it to the merge and split rules. Finally, in Section 6 we explain how to apply the obtained results to the coalitional TU-games and some classes of hedonic games.

2 Comparing and transforming collections

Let $N = \{1, 2, \ldots, n\}$ be a fixed set of players called the *grand coalition*. Non-empty subsets of N are called *coalitions*. A *collection* (in the grand coalition N) is any family $C := \{C_1, \ldots, C_l\}$ of mutually disjoint coalitions, and l is called its *size*. If additionally $\bigcup_{j=1}^{l} C_j = N$, the collection C is called a *partition* of N. For $C = \{C_1, \ldots, C_k\}$, we define $\bigcup C := \bigcup_{i=1}^{k} C_i$.

In this article we are interested in comparing collections. In what follows we only compare collections A and B that are partitions of the same set, i.e., such that $\bigcup A = \bigcup B$. Intuitively, assuming a comparison relation $\rhd$, $A \rhd B$ means that the way A partitions K, where $K = \bigcup A = \bigcup B$, is preferable to the way B partitions K.

In specific examples we shall deal both with reflexive and non-reflexive transitive relations. So, to keep the presentation uniform we only assume that the relation $\rhd$ is transitive, i.e. for all collections A, B, C with $\bigcup A = \bigcup B = \bigcup C$,

$$A \rhd B \rhd C \text{ imply } A \rhd C, \tag{tr}$$

and that $\rhd$ is monotonic in the following two senses: for all collections A, B, C, D with $\bigcup A = \bigcup B$, $\bigcup C = \bigcup D$, and $\bigcup A \cap \bigcup C = \emptyset$,

$$A \rhd B \text{ and } C \rhd D \text{ imply } A \cup C \rhd B \cup D, \tag{m1}$$

and for all collections A, B, C with $\bigcup A = \bigcup B$ and $\bigcup A \cap \bigcup C = \emptyset$,

$$A \rhd B \text{ implies } A \cup C \rhd B \cup C. \tag{m2}$$

Of course, if $\rhd$ is reflexive (m2) follows from (m1).

The role of monotonicity will become clear in Section 4. If $\rhd$ is reflexive, we may denote it by $\succeq$ and if $\rhd$ is irreflexive, we may denote it by $\succ$.

Definition 2.1. By a *comparison relation* we mean a relation on collections that satisfies the conditions (tr), (m1) and (m2). □

In what follows we study coalition formation by focusing on the following two rules that allow us to transform partitions of the grand coalition:

merge: $\{T_1, \ldots, T_k\} \cup P \rightarrow \{\bigcup_{j=1}^{k} T_j\} \cup P$, where $\{\bigcup_{j=1}^{k} T_j\} \rhd \{T_1, \ldots, T_k\}$

split: $\{\bigcup_{j=1}^{k} T_j\} \cup P \rightarrow \{T_1, \ldots, T_k\} \cup P$, where $\{T_1, \ldots, T_k\} \rhd \{\bigcup_{j=1}^{k} T_j\}$

Note that both rules use the $\rhd$ comparison relation 'locally', by focusing on the coalitions that take part and result from the merge resp. split. In this paper we are interested in finding conditions that guarantee that arbitrary sequences of these two rules yield the same outcome. So, once these conditions hold, a specific *preferred* partition exists such that any initial partition can be transformed into it by applying the merge and split rules in an arbitrary order.

To start with, the following observation isolates the condition that guarantees the termination of the iterations of these two rules.

Note 2.2. *Suppose that $\rhd$ is an irreflexive comparison relation. Then every iteration of the merge and split rules terminates.*

Proof. Every iteration of these two rules produces by (m2) a sequence of partitions $P_1, P_2, \ldots$ with $P_{i+1} \rhd P_i$ for all $i \geq 1$. But the number of different partitions is finite. So by transitivity and irreflexivity of $\rhd$ such a sequence has to be finite. □

The analysis of the conditions guaranteeing the unique outcome of the iterations is now deferred to Section 5.

3 TU-games

To properly motivate the subsequent considerations and to clarify the status of the monotonicity conditions we now introduce some natural comparison relations on collections for coalitional TU-games. Recall that a *coalitional TU-game* is a pair (v, N), where $N = \{1, \ldots, n\}$ and v is a function from the powerset of N to the set of non-negative reals.[1] In what follows we assume that $v(\emptyset) = 0$.

For a coalitional TU-game (v, N) the comparison relations on collections are induced in a canonic way from the corresponding relations on the multisets of reals, by stipulating that for the collections A and B

$$A \rhd B \text{ iff } v(A) \rhd v(B),$$

where for a collection $A := \{A_1, \ldots, A_m\}$, $v(A) := \dot{\{} v(A_1), \ldots, v(A_m) \dot{\}}$, denoting the multisets using dotted braces.

To take into account payoffs to individual players we need to use the concept of a *value function* ϕ that given a coalition A assigns to each player $i \in A$ a real $\phi^A(i)$ such that $\sum_{i \in A} \phi^A(i) = v(A)$. Then for a collection $A := \{A_1, \ldots, A_m\}$ we put $v(A) := \dot{\{} \phi^{A_j}(i) \mid i \in A_j, j \in \{1, \ldots, m\} \dot{\}}$.

So first we introduce the appropriate relations on the multisets of non-negative reals. The corresponding definition of monotonicity for such a relation

[1] The assumption that the values of v are non-negative is non-standard and is needed only to accomodate for the Nash order, defined below.

$\rhd$ is that for all multisets a, b, c, d of reals

$$a \rhd b \text{ and } c \rhd d \text{ imply } a \dot{\cup} c \rhd b \dot{\cup} d$$

and

$$a \rhd b \text{ implies } a \dot{\cup} c \rhd b \dot{\cup} c,$$

where $\dot{\cup}$ denotes the multiset union.

Given two sequences $(a_1, \ldots, a_m)$ and $(b_1, \ldots, b_n)$ of real numbers we define the (extended) *lexicographic order* on them by putting

$$(a_1, \ldots, a_m) >_{lex} (b_1, \ldots, b_n)$$

iff

$$\exists i \leq min(m, n)\ (a_i > b_i \wedge \forall j < i\ a_j = b_j)$$

or

$$\forall i \leq min(m, n)\ a_i = b_i \wedge m > n.$$

Note that in this order we compare sequences of possibly different length. We have for example $(1, 1, 1, 0) >_{lex} (1, 1, 0)$ and $(1, 1, 0) >_{lex} (1, 1)$. It is straightforward to check that it is a linear order.

We assume below that $a = \dot{\{}a_1, \ldots, a_m\dot{\}}$ and $b = \dot{\{}b_1, \ldots, b_n\dot{\}}$ and that a^* is a sequence of the elements of a in decreasing order, and define

- the *utilitarian* order:

  $a \succ_{ut} b$ iff $\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j$,
- the *Nash* order:

  $a \succ_{Nash} b$ iff $\Pi_{i=1}^{m} a_i > \Pi_{j=1}^{n} b_j$,
- the *elitist* order:

  $a \succ_{el} b$ iff $max(a) > max(b)$,
- the *egalitarian* order:

  $a \succ_{eg} b$ iff $min(a) > min(b)$,
- the *leximin* order:

  $a \succ_{lex} b$ iff $a^* >_{lex} b^*$.

In [12] these orders were considered for the sequences of the same length. The intuition behind the Nash order is that when the sum $\sum_{i=1}^{m} a_i$ is fixed, the product $\Pi_{i=1}^{m} a_i$ is largest when all a_is are equal. So in a sense the Nash order favours an equal distribution.

For the first four relations, the corresponding reflexive counterparts are obtained by replacing $>$ by $\geq$. In turn, $\succeq_{lex}$, the reflexive version of $\succ_{lex}$, is obtained by additionally including all pairs of equal multisets. Note that all these preorders are in fact linear (i.e., total) preorders.

Note 3.1. *The above relations are all monotonic both in sense* (m1) *and* (m2).

Proof. The only relations for which the claim is not immediate are $\succ_{lex}$ and $\succeq_{lex}$. We will only prove (m1) for $\succ_{lex}$; the remaining proofs are analogous.

Let arbitrary multisets of non-negative reals a, b, c, d be given. We define, with e denoting any sequence or multiset of non-negative reals,

$$\begin{aligned} len(e) &:= \text{the number of elements in } e, \\ \mu &:= (a \,\dot\cup\, b \,\dot\cup\, c \,\dot\cup\, d)^* \text{ with all duplicates removed,} \\ \nu(x,e) &:= \text{the number of occurrences of } x \text{ in } e, \\ \beta &:= 1 + \max_{k=1}^{len(\mu)} \{\nu(\mu_k, a \,\dot\cup\, b \,\dot\cup\, c \,\dot\cup\, d)\}, \\ \#(e) &:= \sum_{k=1}^{len(\mu)} \nu(\mu_k, e) \cdot \beta^{-k}. \end{aligned}$$

So μ is the sequence of all distinct reals used in $a \,\dot\cup\, b \,\dot\cup\, c \,\dot\cup\, d$, arranged in a decreasing order. The function $\#(\cdot)$ injectively maps a multiset e to a real number y in such a way that in the floating point representation of y with base β, the kth digit after the point equals the number of occurrences of the kth biggest number μ_k in e. The base β is chosen in such a way that even if e is the union of some of the given multisets, the number $\nu(x, e)$ of occurrences of x in e never exceeds $\beta - 1$. Therefore, the following sequence of implications holds:

$$\begin{aligned} a^* >_{lex} b^* \text{ and } c^* >_{lex} d^* &\Rightarrow \#(a) > \#(b) \text{ and } \#(c) > \#(d) \\ &\Rightarrow \#(a) + \#(c) > \#(b) + \#(d) \\ &\Rightarrow \#(a \,\dot\cup\, c) > \#(b \,\dot\cup\, d) \\ &\Rightarrow (a \,\dot\cup\, c)^* >_{lex} (b \,\dot\cup\, d)^* \end{aligned}$$

□

As a natural example of a transitive relation that is not monotonic consider $\succeq_{av}$ defined by

$$a \succeq_{av} b \text{ iff } (\textstyle\sum_{i=1}^{m} a_i)/m \geq (\sum_{j=1}^{n} b_j)/n.$$

Note that for

$$a := \dot{\{}3\dot{\}}, b := \dot{\{}2,2,2,2\dot{\}}, c := \dot{\{}1,1,1,1\dot{\}}, d := \dot{\{}0\dot{\}}$$

we have both $a \succeq_{av} b$ and $c \succeq_{av} d$ but not $a \,\dot\cup\, c \succeq_{av} b \,\dot\cup\, d$ since $\dot{\{}3,1,1,1,1\dot{\}} \succeq_{av} \dot{\{}2,2,2,2,0\dot{\}}$ does not hold.

4 Stable partitions

We now return to our study of collections. One way to identify conditions guaranteeing the unique outcome of the iterations of the merge and split rules is through focusing on the properties of such a unique outcome. This brings us to a concept of a stable partition.

We follow here the approach of [1], although now no notion of a game is present. The introduced notion is parametrized by means of a *defection function* $\mathbb{D}$ that assigns to each partition some partitioned subsets of the grand coalition. Intuitively, given 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. Two most natural defection functions are $\mathbb{D}_p$, which allows formation of all partitions of the grand coalition, and $\mathbb{D}_c$, which allows formation of all collections in the grand coalition.

Next, given a collection C and a partition $P := \{P_1, \ldots, P_k\}$ we define

$$C[P] := \{P_1 \cap \bigcup C, \ldots, P_k \cap \bigcup C\} \setminus \{\emptyset\}$$

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.

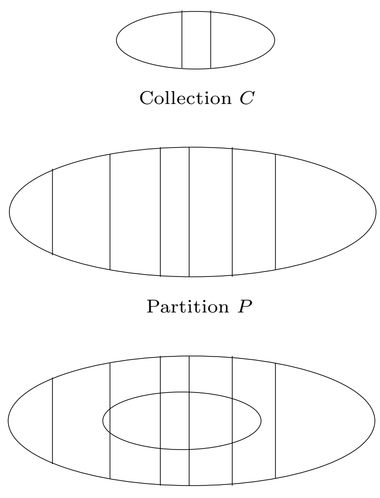

Figure 1: A collection C in the frame of a partition P

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., C in the frame of P.

To get familiar with the $C[P]$ notation note that

- if C is a singleton, say $C = \{T\}$, then $\{T\}[P] = \{P_1 \cap T, \ldots, P_k \cap T\} \setminus \{\emptyset\}$, where $P = \{P_1, \ldots, P_k\}$,
- if C is a partition of N, then $C[P] = P$,
- if $C \subseteq P$, that is C consists of some coalitions of P, then $C[P] = C$.

In general the following simple observation holds.

Note 4.1. *For a collection C and a partition P, $C[P] = C$ iff each element of C is a subset of a different element of P.* □

This brings us to the following notion.

Definition 4.2. Assume a defection function $\mathbb{D}$ and a comparison relation $\rhd$. We call a partition P $\mathbb{D}$-stable if $C[P] \rhd C$ for all $C \in \mathbb{D}(P)$ such that $C[P] \neq C$.

The last qualification, that is $C[P] \neq C$, requires some explanation. First note that if C is a partition of N, then $C[P] \neq C$ is equivalent to the statement $P \neq C$, since then $C[P] = P$. So in the case of the $\mathbb{D}_p$ defection function we have the following simpler definition.

Theorem 4.3. *A partition P is $\mathbb{D}_p$-stable iff for all partitions $P' \neq P$, $P \rhd P'$ holds.* □

Corollary 4.4. *Suppose that $\rhd$ is an irreflexive linear comparison relation. Then a $\mathbb{D}_p$-stable partition exists.* □

Next, if we deal with a reflexive comparison relation $\succeq$, then the qualification $C[P] \neq C$ can be dropped, as then $C[P] = C$ implies $C[P] \succeq C$. However, if we deal with an irreflexive comparison relation $\succ$, then this qualification is of course necessary. So using it we can deal with the irreflexive and reflexive case in a uniform way.

Intuitively, the condition $C[P] \neq C$ indicates that the players only care about the way they are partitioned. Indeed, if $C[P] = C$, then the partitions of $\bigcup C$ by means of P and by means of C coincide and are viewed as equally satisfactory for the players in $\bigcup C$. By disregarding the situations in which $C[P] = C$ we therefore adopt a limited viewpoint of cooperation according to which the players in C do not care about the presence of the players from outside of $\bigcup C$ in their coalitions.

The definition of $\mathbb{D}$-stability calls for checks involving (almost) all collections from $\mathbb{D}$(P). In the case of the $\mathbb{D}_c$ defection function, we can considerably simplify these checks as the following characterization results shows. Given a partition $P := \{P_1, \ldots, P_k\}$ we call here a coalition T *P-compatible* if for some $i \in \{1, \ldots, k\}$ we have $T \subseteq P_i$ and *P-incompatible* otherwise.

Theorem 4.5. *A partition $P = \{P_1, \ldots, P_k\}$ of N is $\mathbb{D}_c$-stable iff 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 $A \cup B \subseteq P_i$

$$\{A \cup B\} \triangleright \{A, B\}, \tag{1}$$

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

$$\{T\}[P] \triangleright \{T\}. \tag{2}$$

Proof. ($\Rightarrow$) Immediate.

($\Leftarrow$) Transitivity (tr), monotonicity (m2) and (1) imply by induction that for each $i \in \{1, \ldots, k\}$ and each collection $C = \{C_1, \ldots, C_l\}$ with $l > 1$ and $\bigcup C \subseteq P_i$,

$$\left\{\bigcup C\right\} \triangleright C. \tag{3}$$

Let now C be an arbitrary collection in N such that $C[P] \neq C$. We prove that $C[P] \triangleright C$. Define

$$D^i := \{T \in C \mid T \subseteq P_i\},$$
$$E := C \setminus \textstyle\bigcup_{i=1}^k D^i,$$
$$E^i := \{P_i \cap T \mid T \in E\} \setminus \{\emptyset\}.$$

Note that D^i is the set of P-compatible elements of C contained in P_i, E is the set of P-incompatible elements of C and E^i consists of the non-empty intersections of P-incompatible elements of C with P_i.

Suppose now that $\bigcup_{i=1}^k E^i \neq \emptyset$. Then $E \neq \emptyset$ and consequently

$$\bigcup_{i=1}^k E^i = \bigcup_{i=1}^k (\{P_i \cap T \mid T \in E\} \setminus \{\emptyset\}) = \bigcup_{T \in E} \{T\}[P] \overset{(m1),(2)}{\triangleright} E. \tag{4}$$

Consider now the following property:

$$|D^i \cup E^i| > 1. \tag{5}$$

Fix $i \in \{1, \ldots, k\}$. If (5) holds, then

$$\left\{P_i \cap \bigcup C\right\} = \left\{\bigcup (D^i \cup E^i)\right\} \overset{(3)}{\triangleright} D^i \cup E^i$$

and otherwise

$$\left\{P_i \cap \bigcup C\right\} = \left\{D^i \cup E^i\right\}.$$

Recall now that

$$C[P] = \bigcup_{i=1}^k \left\{P_i \cap \bigcup C\right\} \setminus \{\emptyset\}.$$

We distinguish two cases.

Case 1. (5) holds for some $i \in \{1, \ldots, k\}$.

Then by (m1) and (m2)

$$C[P] \rhd \bigcup_{i=1}^{k} (D^i \cup E^i) = (C \setminus E) \cup \bigcup_{i=1}^{k} E^i.$$

If $\bigcup_{i=1}^{k} E^i = \emptyset$, then also $E = \emptyset$ and we get $C[P] \rhd C$. Otherwise by (4), (tr) and (m2)

$$C[P] \rhd (C \setminus E) \cup E = C.$$

Case 2. (5) does not hold for any $i \in \{1, \ldots, k\}$.

Then

$$C[P] = \bigcup_{i=1}^{k} (D^i \cup E^i) = (C \setminus E) \cup \bigcup_{i=1}^{k} E^i.$$

Moreover, because $C[P] \neq C$, by Note 4.1 a P-incompatible element in C exists. So $\bigcup_{i=1}^{k} E^i \neq \emptyset$ and by (4) and (m2) we get as before

$$C[P] \rhd (C \setminus E) \cup E = C.$$

□

In [1] this theorem was proved for the coalitional TU-games and both the irreflexive and the reflexive utilitarian orders. The above result isolates the relevant conditions that the comparison relation, here $\rhd$, needs to satisfy.

In contrast to the case of the $\mathbb{D}_p$-stable partitions, as shown in [1], a $\mathbb{D}_c$-stable partition does not need to exist, even if $\rhd$ is irreflexive. In that paper a natural class of TU-games is defined for which $\mathbb{D}_c$-stable partitions are guaranteed to exist. In Section 6 we introduce a natural class of hedonic games for which $\mathbb{D}_c$-stable partitions exist.

5 Stable partitions and merge/split rules

We now resume our investigation of the conditions under which every iteration of the merge and split rules yields the same outcome. With this in mind we establish the following results concerned with the $\mathbb{D}_c$ defection function.

Note 5.1. *If $\rhd$ is an irreflexive comparison relation, then every $\mathbb{D}_c$-stable partition P is closed under the applications of the merge and split rules.*

Proof. To prove the closure under merge rule assume that for some $\{T_1, \ldots, T_k\} \subseteq P$ we have $\{\bigcup_{j=1}^{k} T_j\} \rhd \{T_1, \ldots, T_k\}$. $\mathbb{D}_c$-stability of P with $C := \{\bigcup_{j=1}^{k} T_j\}$ yields

$$\{T_1, \ldots, T_k\} = \{\bigcup_{j=1}^{k} T_j\}[P] \rhd \{\bigcup_{j=1}^{k} T_j\},$$

which is a contradiction by virtue of the transitivity and irreflexivity of $\rhd$.

The closure under the split rule is shown analogously. □

Lemma 5.2. *Assume that $\rhd$ is an irreflexive comparison relation and P is $\mathbb{D}_c$-stable. Let P' be closed under applications of merge and split rules. Then $P' = P$.*

Proof. Suppose $P = \{P_1, \ldots, P_k\}$, $P' = \{T_1, \ldots, T_m\}$. Assume $P \neq P'$. Then there is $i_0 \in \{1, \ldots, k\}$ such that for all $j \in \{1, \ldots, m\}$ we have $P_{i_0} \neq T_j$. Let $T_{j_1}, \ldots, T_{j_l}$ be the minimum cover of P_{i_0}. In the following case distinction we use Theorem 4.5.

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

Then $\{T_{j_1}, \ldots, T_{j_l}\}$ is a proper partition of P_{i_0}. But (1) (through its generalization to (3)) yields $P_{i_0} \rhd \{T_{j_1}, \ldots, T_{j_l}\}$, thus the merge rule is applicable to P'.

Case 2. $P_{i_0} \subsetneq \bigcup_{h=1}^{l} T_{j_h}$.

Then for some j_h we have $\emptyset \neq P_{i_0} \cap T_{j_h} \subsetneq T_{j_h}$, so T_{j_h} is P-incompatible. By (2) we have $\{T_{j_h}\}[P] \rhd \{T_{j_h}\}$, thus the split rule is applicable to P'. □

This allows us to conclude the following result that answers our initial question and clarifies the importance of the $\mathbb{D}_c$-stable partitions.

Theorem 5.3. *Suppose that $\rhd$ is an irreflexive comparison relation and P is a $\mathbb{D}_c$-stable partition. Then*

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

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

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

Proof. (*i*) By Note 2.2 every iteration of the merge and split rules terminates, so the claim follows by Lemma 5.2.

(*ii*) Since P is $\mathbb{D}_c$-stable, it is in particular $\mathbb{D}_p$-stable. Uniqueness follows from the transitivity and irreflexivity of $\rhd$ by virtue of Theorem 4.3.

(*iii*) Suppose that P' is a $\mathbb{D}_c$-stable partition. By Note 5.1 P' is closed under the applications of the merge and split rules, so by Lemma 5.2 $P' = P$. □

This generalizes the considerations of [1], where this result was established for the coalitional TU-games and the irreflexive utilitarian order. It was also shown there that there exist coalitional TU-games in which all iterations of the merge and split rules have a unique outcome which is not a $\mathbb{D}_c$-stable partition.

6 Hedonic games

Note that the results of the last two sections do not involve any notion of a game. Only by choosing the monotonic comparison relations introduced in Section 3 we obtain specific results that deal with coalitional TU-games.

These considerations also readily apply to NTU-games. However, one needs to be careful since the resulting notion of a stable coalition can be in some situations counterintuitive. To clarify the limitation of this approach we now focus on the hedonic games (see, e.g., [5]) that form a specific class of NTU-games. Recall that a *hedonic game* $(N, \succeq_1, \ldots, \succeq_n)$ consists of a set of players $N = \{1, \ldots, n\}$ and a sequence of linear preorders $\succeq_1, \ldots, \succeq_n$, where each $\succeq_i$ is the preference of player i over the subsets of N containing i. In what follows we shall not need the assumption that the $\succeq_i$ relations are linear.

Again, we let $\succ_i$ denote the associated irreflexive relation. Given a partition A of N and player i we denote by $A(i)$ the element of A to which i belongs and call it the set of *friends of i in A*. Given a hedonic game $(N, \succeq_1, \ldots, \succeq_n)$ a natural preference relation on the collections is given by:

$$A \succeq B \text{ iff } \neg\exists C \in B\, \forall i \in C.C \succ_i A(i), \tag{6}$$

where $\bigcup A = \bigcup B$.

It states that A is preferred over B unless B contains a coalition C such that each player in C strictly prefers C to his coalition in A. Clearly $\succeq$ is monotonic. The notion of $\mathbb{D}_p$-stability then coincides with the notion of core stability in [5].

However, the resulting notion of a $\mathbb{D}_c$-stable partition can contradict the intuition. To see this consider the following example.

Example 6.1. Suppose $N = \{1, 2, 3, 4\}$. Consider a hedonic game in which

$$\{2\} \succ_2 \{2, 3\} \succ_2 \{1, 2\}$$

and

$$\{3\} \succ_3 \{2, 3\} \succ_3 \{3, 4\}.$$

Now take $P = \{\{1, 2\}, \{3, 4\}\}$ and $C = \{\{2, 3\}\}$. Then $C[P] = \{\{2\}, \{3\}\}$. So both players 2 and 3 strictly prefer their coalition in $C[P]$ to the one in C and consequently P is 'stable' w.r.t. collection C. In fact, it is straightforward to extend the above ordering in such a way that P is $\mathbb{D}_c$-stable.

However, both players 2 and 3 favour the coalition $\{2, 3\}$ higher than their coalition within P, so intuitively P should not be stable. □

The difficulty in the above example arises from the fact that in players' preferences smaller coalitions can be preferred over the larger ones. Natural hedonic games in which this is not the case can be derived from arbitrary partitions of the set of players. Given a partition $P := \{P_1, \ldots, P_k\}$ of N we assume that each player

- prefers a larger set of friends over a smaller one,

- only 'cares' about the sets of his friends in P.

We formalize this order by putting for all sets of players that include i

$$S \succeq_i T \text{ iff } S \cap P(i) \supseteq T \cap P(i).$$

With this definition, all partitions which result from arbitrary (including no) applications of the merge rule to P are $\mathbb{D}_c$-stable w.r.t. the reflexive comparison relation $\succeq$ defined in (6).

Next, we provide an example of a hedonic game in which a $\mathbb{D}_c$-stable partition w.r.t. to a natural irreflexive comparison relation $\succ$ exists. To this end given a partition $P := \{P_1, \ldots, P_k\}$ of N we now assume that each player

- prefers a larger set of his friends in P over a smaller one,
- 'dislikes' coalitions that include a player who is not his friend in P.

We formalize this by putting for all sets of players that include i

$$S \succeq_i T \text{ iff } S \cup T \subseteq P(i) \text{ and } S \supseteq T,$$

and by extending this order to the coalitions that include player i and also a player from outside of $P(i)$ by assuming that they are the minimal elements in $\succeq_i$. So $S \succ_i T$ iff either $S \cup T \subseteq P(i)$ and $S \supset T$ or $S \subseteq P(i)$ and $\neg T \subseteq P(i)$.

We then define an irreflexive comparison relation on collections by

$$A \succ B \text{ iff for } i \in \{1, \ldots, n\}\ A(i) \succeq_i B(i) \text{ with at least one } \succeq_i \text{ being strict.}$$

It is straightforward to check that for this comparison relation the partition $\{P_1, \ldots, P_k\}$ satisfies the conditions (1) and (2) of Theorem 4.5. So by virtue of this theorem $\{P_1, \ldots, P_k\}$ is $\mathbb{D}_c$-stable. Further, by virtue of Theorem 5.3, $\{P_1, \ldots, P_k\}$ can be reached from any initial partition through an arbitrary sequence of the applications of the split and merge rules.

Acknowledgement

We thank Tadeusz Radzik for helpful comments.

References

[1] K. R. Apt and T. Radzik. Stable partitions in coalitional games, 2006. Available from http://arxiv.org/abs/cs.GT/0605132.

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

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

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

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

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

[7] G. Chalkiadakis and C. Boutilier. Bayesian reinforcement learning for coalition formation under uncertainty. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-04)*, pages 1090–1097, 2004.

[8] G. Demange. On group stability in hierarchies and networks. *Journal of Political Economy*, 112(4):754–778, 2004.

[9] G. Demange and M. Wooders, editors. *Group Formation in Economics*. Cambridge University Press, 2006.

[10] 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.

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

[12] H. Moulin. *Axioms of Cooperative Decision Making*. Cambridge University Press, 1998.

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

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

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

Krzysztof R. Apt
CWI
1098 SJ Amsterdam, The Netherlands
Email: K.R.Apt@cwi.nl

Andreas Witzel
ILLC, University of Amsterdam
1018 TV Amsterdam, The Netherlands
Email: awitzel@illc.uva.nl