CentER Discussion Paper

No. 2004–111

ENEMIES AND FRIENDS IN HEDONIC GAMES: INDIVIDUAL DEVIATIONS, STABILITY AND MANIPULATION

By D. Dimitrov, S.C. Sung

October 2004

ISSN 0924-7815

Enemies and friends in hedonic games: individual deviations, stability and manipulation*

Dinko Dimitrov
CentER and Department of Econometrics and Operations Research
Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

Shao Chin Sung
School of Information Science
Japan Advanced Institute of Science and Technology
1-1 Asahidai, Tatsunokuchi, Ishikawa 923-1292, Japan

October 2004

Abstract

We consider hedonic games with separable preferences, and explore the existence of stable coalition structures if only individual deviations are allowed. For two natural subdomains of separable preferences, namely preference domains based on (1) aversion to enemies and (2) appreciation of friends, we show that an individually stable coalition structure always exist, and a Nash stable coalition structure exists when mutuality is imposed. Moreover, we show that on the domain of separable preferences a contractual individually stable coalition structure can be obtained in polynomial time. Finally, we prove that, on

*Thanks are due to Hans Peters, Ton Storcken, and especially to Coralio Ballester and Bettina Klaus for very helpful comments. Dinko Dimitrov gratefully acknowledges financial support from a Marie Curie Research Fellowship of the European Community programme "Improving the Human Research Potential and the Socio-Economic Knowledge Base" under contract number HPMF-CT-2002-02121.

each of the two subdomains, the corresponding algorithm that we use for finding Nash stable and individually stable coalition structures turns out to be strategy-proof.

Journal of Economic Literature Classification Numbers: C71, A14, D20.

Keywords: additive separability, coalition formation, hedonic games, stability, strategy-proofness.

1 Introduction

The formal model of a hedonic coalition formation game as introduced by Banerjee, Konishi and Sönmez (2001) and Bogomolnaia and Jackson (2002) consists of two components, namely a finite set of players and a preference ranking for each player defined over the coalitions that player may belong to. The outcome of such a game is a partition of the society (the set of players) into coalitions, where a partition of the society into coalitions is called a *coalition structure*. This model explicitly takes into account the dependence of an agent's utility on the identity of the members of his or her coalition as recognized in the seminal paper of Drèze and Greenberg (1980).

In this paper we consider hedonic games with *separable preferences*, and study the existence of stable coalition structures if only individual deviations are allowed. A player's preference is separable if he or she views every other player either as a friend or as an enemy, and the division between friends and enemies guides the ordering of coalitions in the sense that adding a friend leads to a more preferable coalition, while adding an enemy leads to a less preferable coalition. More specifically, we concentrate on two natural subdomains of separable preferences, namely preference domains based on *aversion to enemies* and *appreciation of friends*.

The preference domain based on aversion to enemies corresponds to a situation in which, when comparing two coalitions, every player looks first at his or her enemies in either coalition. The coalition that contains less enemies is declared by the player as better than the other, and if the two coalitions have the same number of enemies, then the number of friends is decisive for the comparison. This kind of preference restriction can be illustrated by the formation of swimming teams, where a bad guy can affect the whole outcome of the coalition.

The preference domain based on appreciation of friends corresponds to a situation in which, when comparing two coalitions, every player pays attention first to his or her friends in either coalition. The coalition that contains more friends is declared by the player as better than the other, and if the two coalitions have the same number of friends, then the coalition with less enemies wins the comparison. This kind of preference restriction is appropriate in contexts where friendship collaboration is crucial for the payoff that a player obtains from a coalition, while the harm produced by the remaining players in the coalition is small for the player. An example may be the formation of research groups that compete for grants.

Notice that these two subdomains allow for indifferences in the corresponding rankings over coalitions. For the two classes of hedonic games corresponding to these two subdomains, Dimitrov, Borm, Hendrickx, and Sung (2004) study the existence of core stable coalition structures, and provide algorithms for generating such coalition structures. However, there are cases in which coalitional deviations are not possible and, hence, solution concepts that consider only individual deviations are warranted. We concentrate in this paper on Nash stability and individual stability. A coalition structure is Nash stable if no player wishes to migrate to another coalition in the same coalition structure. Individual stability, in addition, pays attention to the reaction of the welcoming coalition in the sense that no one in that coalition should be made worse off.

It turns out that individually stable coalition structures for the two classes of hedonic games always exist. The way we restrict players' preferences allow us also to present a positive result on the existence of Nash stable coalition structures when, in addition, mutuality on the preferences is imposed (i.e. the friendship among players is always mutual). It should be noted that, in contrast to Bogomolnaia and Jackson (2002), symmetry (requiring that the players have the same reciprocal values for each other and being stronger than mutuality) is no longer crucial for the existence proofs. For an excellent study of the role of symmetric additive separable preferences in hedonic games the reader is referred to Burani and Zwicker (2003).

When looking for individually stable or Nash stable coalition structures, we make use of the two algorithms proposed by Dimitrov, Borm, Hendrickx, and Sung (2004) for generating core stable coalition structures, and show that these core stable coalition structures are also individually stable or Nash stable. Moreover, on the separable preference domain a contractual individually stable coalition structure can be obtain in polynomial time by using one of the proposed algorithms.

The algorithms suggested by Dimitrov, Borm, Hendrickx, and Sung (2004) can be considered also as functions which assign a coalition structure to each hedonic game in the proposed domains. We show that these algorithms are strategy-proof on the corresponding domains, in the sense that no player can profitably misrepresent his or her preference to obtain a better outcome.

The rest of the paper is organized as follows. Section 2 presents the formal model of a hedonic game and different stability notions for this class of games already known in the literature; it introduces also the classes of preferences based on aversion to enemies and appreciation of friends, respectively. The algorithms used by Dimitrov, Borm, Hendrickx, and Sung (2004) for generating core stable coalition structures for these two classes of hedonic games are described and exemplified in Section 3. Section 4 presents our results on the existence of stable coalition structures on the corresponding domains when only individual deviations are allowed. Section 5 is devoted to the study of strategy-proofness of the proposed algorithms. We conclude in Section 6 with some final remarks.

2 Basic notions

Consider a finite set of players $N = \{1, 2, \ldots, n\}$. A *coalition* is a non-empty subset of N. For each player $i \in N$, we denote by $\mathcal{N}_i = \{X \subseteq N \mid i \in X\}$ the collection of all coalitions containing i. A collection $\mathcal{C}$ of coalitions is called a *coalition structure* if $\mathcal{C}$ is a partition of N, i.e. the coalitions in $\mathcal{C}$ are pairwise disjoint and $\bigcup_{X \in \mathcal{C}} X = N$. The set of all coalition structures on N is denoted by $\mathbf{C}^N$. For each coalition structure $\mathcal{C}$ and each player $i \in N$, by $\mathcal{C}(i)$ we denote the coalition in $\mathcal{C}$ containing i, i.e. $\{\mathcal{C}(i)\} = \mathcal{C} \cap \mathcal{N}_i$.

We assume that each player $i \in N$ is endowed with a preference $\succeq_i$ over $\mathcal{N}_i$, i.e. a binary relation over $\mathcal{N}_i$ which is reflexive, complete, and transitive. We denote by $\mathcal{P}_i$ the set of all player i's preferences, by $P = (\succeq_1, \succeq_2, \ldots, \succeq_n)$ a profile of preferences $\succeq_i$ for all $i \in N$, and by $\mathcal{P} = \mathcal{P}_1 \times \ldots \times \mathcal{P}_n$ the set of all preference profiles. Moreover, we assume that the preference of each player $i \in N$ over coalition structures is *purely hedonic*, i.e. it is completely characterized by $\succeq_i$ in such a way that, for each $\mathcal{C}$ and $\mathcal{C}'$, each player i weakly prefers $\mathcal{C}$ to $\mathcal{C}'$ if and only if $\mathcal{C}(i) \succeq_i \mathcal{C}'(i)$.

A *hedonic game* on a finite set N of players with a preference profile

$P \in \mathcal{P}$ is denoted by the pair (N, P). The set of all hedonic games will be denoted by $\mathcal{G}$.

Let $\mathcal{C}$ be a coalition structure. We say that

- $\mathcal{C}$ is *weak core stable* if there does not exist a nonempty coalition X such that $X \succ_i \mathcal{C}(i)$ for all $i \in X$;

- $\mathcal{C}$ is *strong core stable* if there does not exist a nonempty coalition X such that $X \succeq_i \mathcal{C}(i)$ for all $i \in X$, and $X \succ_j \mathcal{C}(j)$ for at least one $j \in X$;

- $\mathcal{C}$ is *Nash stable* if there do not exist $i \in N$ and a coalition $X \in \mathcal{C} \cup \{\emptyset\}$ such that $X \cup \{i\} \succ_i \mathcal{C}(i)$;

- $\mathcal{C}$ is *individually stable* if there do not exist $i \in N$ and a coalition $X \in \mathcal{C} \cup \{\emptyset\}$ such that $X \cup \{i\} \succ_i \mathcal{C}(i)$, and $X \cup \{i\} \succeq_j X$ for all $j \in X$;

- $\mathcal{C}$ is *contractual individually stable* if there do not exist $i \in N$ and a coalition $X \in \mathcal{C} \cup \{\emptyset\}$ such that $X \cup \{i\} \succ_i \mathcal{C}(i)$, $X \cup \{i\} \succeq_j X$ for all $j \in X$, and $\mathcal{C}(i) \setminus \{i\} \succeq_j \mathcal{C}(i)$ for all $j \in \mathcal{C}(i) \setminus \{i\}$.

Clearly, Nash stability implies individual stability that, in turn, implies contractual individual stability. Moreover, individual stability is implied by strong core stability as well.

We now specify the preference domains that will be considered. For each $i \in N$, we let $G_i = G(\succeq_i) = \{j \in N : \{i, j\} \succeq_i \{i\}\}$ be the set of friends of player i, and its complement $B_i = N \setminus G_i$ the set of enemies of player i. Notice that, from $\{i\} \succeq_i \{i\}$, we have $i \in G_i$ for each $i \in N$. The next definition suggests two natural ways of how each player i ranks the sets in $\mathcal{N}_i$ depending on the numbers of his or her friends and enemies.

Definition 1 *Let $P = (\succeq_1, \succeq_2, \ldots, \succeq_n) \in \mathcal{P}$ be a preference profile.*

- *We say that P is based on* ***aversion to enemies*** *if, for all $i \in N$ and all $X, Y \in \mathcal{N}_i$, $X \succeq_i Y$ if and only if* (1) $|X \cap B_i| < |Y \cap B_i|$ *or* (2) $|X \cap B_i| = |Y \cap B_i|$ *and* $|X \cap G_i| \geq |Y \cap G_i|$.

- *We say that P is based on* ***appreciation of friends*** *if, for all $i \in N$ and all $X, Y \in \mathcal{N}_i$, $X \succeq_i Y$ if and only if* (1) $|X \cap G_i| > |Y \cap G_i|$ *or* (2) $|X \cap G_i| = |Y \cap G_i|$ *and* $|X \cap B_i| \leq |Y \cap B_i|$.

Observe that if the preference profile is based on aversion to enemies, each player looks first at his or her enemies in the corresponding coalitions; if the preference profile is based on appreciation of friends, we have a priority for friends when comparing two coalitions. In the following, the set of all preference profiles based on aversion to enemies is denoted by $\mathcal{P}^e$, and the set of all preference profiles based on appreciation of friends is denoted by $\mathcal{P}^f$. The corresponding sets of hedonic games will be denoted by $\mathcal{G}^e$ and $\mathcal{G}^f$, respectively.

It is not difficult to see that if players' preferences are induced by either way suggested by Definition 1, then each player i will be equipped with a preference relation over $\mathcal{N}_i$ with G_i being its top and $B_i \cup \{i\}$ being its bottom. In fact, the preference profiles based on aversion to enemies and the preference profiles based on appreciation of friends belong to a more general class of preference profiles, namely the class of additive separable preferences.

A preference profile $P \in \mathcal{P}$ is *additive separable* if, for all $i \in N$, there exists a function $v_i : N \to \mathbb{R}$ such that for all $X, Y \in N_i$, $X \succeq_i Y$ if and only if $\sum_{j \in X} v_i(j) \geq \sum_{j \in Y} v_i(j)$. We denote the set of all additive separable preference profiles by $\mathcal{P}^{as}$, and the corresponding set of hedonic games by $\mathcal{G}^{as}$. For example, when $P \in \mathcal{P}^e$, one can take, for each $i \in N$, $v_i(j) = 1$

if $j \in G_i$, and $v_i(j) = -n$ otherwise; when $P \in \mathcal{P}^f$, one can take, for each $i \in N$, $v_i(j) = n$ if $j \in G_i$, and $v_i(j) = -1$ otherwise. Therefore, we have $(\mathcal{P}^e \cup \mathcal{P}^f) \subset \mathcal{P}^{as}$ and $(\mathcal{G}^e \cup \mathcal{G}^f) \subset \mathcal{G}^{as}$. All additive separable preference profiles are also separable.

Definition 2 *A preference profile $P \in \mathcal{P}$ is* ***separable*** *if, for every player $i \in N$, there is a partition (G_i, B_i) of N such that for every $j \in N$ and $X \in \mathcal{N}_i$ with $j \notin X$, we have $[X \cup \{j\} \succeq_i X \Leftrightarrow j \in G_i]$ and $[X \cup \{j\} \preceq_i X \Leftrightarrow j \in B_i]$.*

Since (G_i, B_i) is partition of N, each separable preference profile P is such that, for every $i, j \in N$ and $X \in \mathcal{N}_i$ with $j \notin X$, we have either $X \cup \{j\} \succeq_i X$ or $X \cup \{j\} \preceq_i X$ but not both. Hence, an equivalent definition of separability can be obtained by replacing $\succeq_i$ and $\preceq_i$ respectively by $\succ_i$ and $\prec_i$. We denote the set of all separable preference profiles by $\mathcal{P}^s$ and the corresponding set of games on player set N by $\mathcal{G}^s$. Clearly, we have $(\mathcal{P}^e \cup \mathcal{P}^f) \subset \mathcal{P}^{as} \subset \mathcal{P}^s \subset \mathcal{P}$ and $(\mathcal{G}^e \cup \mathcal{G}^f) \subset \mathcal{G}^{as} \subset \mathcal{G}^s \subset \mathcal{G}$.

3 Two algorithms

Let $H = (V, E)$ be a directed graph with set of vertices V and set of directed edges $E \subseteq V \times V$. A path $(k_1, k_2, \ldots, k_m)$ in H is a sequence of vertices $k_1, k_2, \ldots, k_m \in V$ for some positive integer m such that $(k_l, k_{l+1}) \in E$ for each $1 \leq l \leq m-1$, and we say that $(k_1, k_2, \ldots, k_m)$ is a path from k_1 to k_m. Let $X \subseteq V$. We say that X is *strongly connected* if, for every $i, j \in X$, there is a path from i to j which only contains vertices belonging to X. We say that X is a *strongly connected component* if X is strongly connected and, for all $Y \subseteq N$ which properly contains X, Y is not strongly connected. Moreover, we say that X is a *clique* in H if $(i, j) \in E$ for every $i, j \in X$.

For each $(N, P) \in \mathcal{G}^s$ and each nonempty $Y \subseteq N$, let $H_{(Y,P)} = (V, E)$ be a *directed graph* with $V = Y$ and $E = \{(i, j) \in Y \times Y \mid i \neq j,\ j \in G_i\}$. In the following, two ways of partitioning the player set N in terms of $H_{(Y,P)}$s are described.

Algorithm 1

- Set $Y := N$ and $\mathcal{C} := \emptyset$.
- Repeat the following until $Y = \emptyset$:
  - Find a clique $X \subseteq Y$ in $H_{(Y,P)}$ with the largest number of vertices.
  - Set $Y := Y \setminus X$ and $\mathcal{C} := \mathcal{C} \cup \{X\}$.
- Return $\mathcal{C}$.

Algorithm 1 partitions the set N into coalitions in such a way that it first subtracts the largest clique X in $H_{(N,P)}$ from N, then it subtracts the largest clique X' in $H_{(Y,P)}$ from $Y = N \setminus X$, and so on, until no vertex remains.

Algorithm 2

- Set $Y := N$ and $\mathcal{C} := \emptyset$.
- Repeat the following until $Y = \emptyset$:
  - Find a strongly connected component $X \subseteq Y$ in $H_{(Y,P)}$.
  - Set $Y := Y \setminus X$ and $\mathcal{C} := \mathcal{C} \cup \{X\}$.
- Return $\mathcal{C}$.

Algorithm 2 partitions the set N into coalitions each of which is a strongly connected component in $H_{(N,P)}$. That is, Algorithm 2 finds the strong decomposition of graph $H_{(N,P)}$, and it can be done in $O(|N|^2)$ time (see Tarjan

(1972)). Note that it is always possible to generate coalition structures by Algorithm 2 provided that players' preferences are separable. However, it is not guaranteed that all coalition structures generated by this algorithm will be for example core stable for a corresponding hedonic game with separable preferences: as exemplified by Banerjee, Konishi, and Sönmez (2001) the core of a hedonic game may be empty even when players' preferences are additive separable and symmetric.

Suppose now that we disallow for mutuality (and, hence, for symmetry) but restrict players' additive separable preferences to be based on aversion to enemies or on appreciation of friends. It turns out that in this case the above algorithms generate core stable coalition structures as follows.

Theorem 1 *Let $(N,P) \in \mathcal{G}^e$ and let $\mathcal{C}$ be a coalition structure generated by Algorithm* 1. *Then $\mathcal{C}$ is a weak core stable coalition structure for (N,P).*

Theorem 2 *Let $(N,P) \in \mathcal{G}^f$ and let $\mathcal{C}$ be a coalition structure generated by Algorithm* 2. *Then $\mathcal{C}$ is a strong core stable coalition structure for (N,P).*

For proofs of these statements the reader is referred to Dimitrov, Borm, Hendrickx, and Sung (2004).

Note that the largest clique in $H_{(N,P)}$ may not be unique, and in such a case, it is not clear from the description of Algorithm 1 which clique will be selected. In other words, a different selection of cliques may lead to a different outcome. A more precise description of Algorithm 1 will be given in Section 5 in order to obtain a unique outcome for each game $(N,P) \in \mathcal{G}^s$. In contrast, the outcome of Algorithm 2 is unique, because the strong decomposition of each directed graph is unique. This point is illustrated by the next example in which every player is indifferent among coalitions on the same row and, for each $i \in N$, the top row corresponds to G_i and the bottom row corresponds to $B_i \cup \{i\}$.

Example 1 *Let* $N = \{1,2,3,4\}$ *and* $G_1 = \{1,2,3,4\}$, $G_2 = \{2,3,4\}$, $G_3 = \{1,3\}$, *and* $G_4 = \{1,3,4\}$. *Then we have the following preferences of the players over the coalitions they may belong to.*
Aversion to Enemies:

1	**2**	**3**	**4**
134	234	1234	14
13, 14	23, 24	123, 134, 234	4
1	2	13, 23, 34	...
...	...	3	

Appreciation of Friends:

1	**2**	**3**	**4**
134	234	1234	14
1234	1234	123, 134, 234	124, 134
13, 14	23, 24	13, 23, 34	1234
123, 124	123, 124	3	4
1	2		...
...	...		

Algorithm 1 *selects as weak core stable coalition structure for the case of aversion to enemies either* $\mathcal{C} = \{\{1,3\},\{2\},\{4\}\}$ *or* $\mathcal{C}' = \{\{1,4\},\{2,3\}\}$. *Note that player* 2 *for example is not indifferent between* $\mathcal{C}$ *and* $\mathcal{C}'$ (*i.e. between* $\mathcal{C}(2)$ *and* $\mathcal{C}'(2)$). *Algorithm* 2 *selects* $\{1,2,3,4\}$ *as a strong core coalition structure for the case of appreciation of friends. In this case the coalition structures* $\{\{1,3,4\},\{2\}\}$ *and* $\{\{1,4\},\{2,3\}\}$ *are strong core stable as well.*

4 Individual deviations and stability

Having described the algorithms for generating core stable coalition structures for hedonic games with preference profiles belonging to $\mathcal{P}^e$ and $\mathcal{P}^f$, we redirect now our attention to the Nash stability, individual stability, and contractual individual stability for such games.

4.1 Nash stability

Although appreciation of friends and aversion of enemies are very strong restrictions, it turns out that they do not guarantee the existence of Nash stable coalition structures. The next example shows a hedonic game $(N, P) \in \mathcal{G}^e \cap \mathcal{G}^f$ for which there is no Nash stable coalition structure.

Example 2 *Let* $N = \{1, 2, 3\}$ *and* $G_1 = \{1\}$, $G_2 = \{2\}$, *and* $G_3 = \{1, 2, 3\}$. *Then we have the the following preferences of the players over the coalitions they may belong to:*

1	**2**	**3**
1	2	123
12, 13	12, 23	13, 23
123	123	3

Note that each of the players 1 *and* 2 *would prefer to stay alone, i.e. we have to check only the coalition structure* $\mathcal{C} = \{\{1\}, \{2\}, \{3\}\}$. *However,* $\{1, 3\} \succ_3 \{3\}$ (*and* $\{2, 3\} \succ_3 \{3\}$). *Hence, a Nash stable coalition structure does not exist.*

Remark 1 *For the game in Example* 2 *a* (*strong*) *core stable coalition structure still exists* ($\{\{1\}, \{2\}, \{3\}\}$), *and it is selected by both Algorithm* 1 *and Algorithm* 2.

As shown next, adding mutuality (for all $i, j \in N$, $i \in G_j$ if and only if $j \in G_i$) to enemy aversion always guarantees the existence of a Nash stable coalition structure.

Proposition 1 *Let $(N, P) \in \mathcal{G}^e$ satisfy mutuality. Then a Nash stable coalition structure exists.*

Proof. Let $\mathcal{C}$ be a coalition structure generated by Algorithm 1. By Theorem 1, $\mathcal{C}$ is weak core stable. We show that $\mathcal{C}$ is Nash stable as well. Suppose not. Then there is $i \in N$ and $X \in \mathcal{C} \cup \{\emptyset\}$ such that $X \cup \{i\} \succ_i \mathcal{C}(i)$. We distinguish the following two cases:

(1) $X = \emptyset$. In this case we have $\{i\} \succ_i \mathcal{C}(i)$ that contradicts the fact that $\mathcal{C}$ is weak core stable.

(2) $\emptyset \neq X \in \mathcal{C}$. Since $\mathcal{C}(i)$ is a clique in $H_{(N,P)}$, we have $\mathcal{C}(i) \subseteq G_i$. It follows from $X \cup \{i\} \succ_i \mathcal{C}(i)$ and $\mathcal{C}(i) \subseteq G_i$ that $X \cup \{i\} \subseteq G_i$, i.e., $j \in G_i$ for all $j \in X$. Then, by mutuality, we have $i \in G_j$ for all $j \in X$, and thus, $X \cup \{i\} \succ_j \mathcal{C}(j) = X$ for all $j \in X$. Therefore, $X \cup \{i\} \succ_j \mathcal{C}(j)$ for all $j \in X \cup \{i\}$, which contradicts again the weak core stability of $\mathcal{C}$. ■

Proposition 2 *Let $(N, P) \in \mathcal{G}^f$ satisfy mutuality. Then a Nash stable coalition structure exists.*

Proof. Let $\mathcal{C}$ be a coalition structure generated by Algorithm 2. By Theorem 2, $\mathcal{C}$ is strong core stable. We show that $\mathcal{C}$ is Nash stable as well.

Let $H_{(N,P)}$ be the directed graph that corresponds to (N, P), and let $i \in N$. Notice that each coalition in $\mathcal{C}$ is a strongly connected component of $H_{(N,P)}$. When $\mathcal{C}(i) \neq \{i\}$, there exists $j \in \mathcal{C}(i) \setminus \{i\}$ such that $j \in G_i$. Hence $\mathcal{C}(i) \succeq_i \{i\} = \emptyset \cup \{i\}$. Moreover, it follows by mutuality that there are no edges between players belonging to different strongly connected subgraphs, i.e., $X \subseteq B_i$ for each $X \in \mathcal{C} \setminus \{\mathcal{C}(i)\}$. Hence, we have $\mathcal{C}(i) \succeq_i \{i\} \succ_i X \cup \{i\}$

for each $X \in \mathcal{C} \setminus \{\mathcal{C}(i)\}$. Finally, it is obvious that $\mathcal{C}(i) \cup \{i\} \succeq_i \mathcal{C}(i)$. Therefore, each $i \in N$ satisfies $\mathcal{C}(i) \succeq_i X \cup \{i\}$ for all $X \in \mathcal{C} \cup \{\emptyset\}$, i.e. $\mathcal{C}$ is Nash stable. ■

Remark 2 *Note that mutuality is a crucial condition for proving the existence of a Nash stable coalition structure in Propositions* 1 *and* 2 *but symmetry is not. A proof for the existence of Nash stable coalition structures when player preferences are additive separable and symmetric is provided by Bogomolnaia and Jackson (2002).*

4.2 Individual stability

As exemplified above, restricting players' preferences in a hedonic game to be based either on aversion to enemies or on appreciation of friends does not guarantee the existence of a Nash stable coalition structure. However, these restrictions are sufficient for individual stability as shown next.

Proposition 3 *Let* $(N, P) \in \mathcal{G}^e$. *Then an individually stable coalition structure exists.*

Proof. Let $\mathcal{C}$ be a coalition structure generated by Algorithm 1. By Theorem 1, $\mathcal{C}$ is weak core stable. We show that $\mathcal{C}$ is individually stable as well. Suppose not. Then there is $i \in N$ and $X \in \mathcal{C} \cup \{\emptyset\}$ such that $X \cup \{i\} \succ_i \mathcal{C}(i)$ and $X \cup \{i\} \succeq_j X$ for all $j \in X$. We distinguish the following two cases:

(1) $X = \emptyset$. In this case we have $\{i\} \succ_i \mathcal{C}(i)$ that contradicts the fact that $\mathcal{C}$ is weak core stable.

(2) $\emptyset \neq X \in \mathcal{C}$. From $X \cup \{i\} \succeq_j X$ for all $j \in X$, we have $i \in G_j$ for all $j \in X$. Thus, in fact we have $X \cup \{i\} \succ_j X$ for all $j \in X$. Combining with $X \cup \{i\} \succ_i \mathcal{C}(i)$, we can conclude that $X \cup \{i\}$ is a strong deviation from $\mathcal{C}$, which contradicts again the weak core stability of $\mathcal{C}$. ■

Proposition 4 *Let $(N, P) \in \mathcal{G}^f$. Then an individually stable coalition structure exists.*

Proof. Let $\mathcal{C}$ be a coalition structure generated by Algorithm 2. By Theorem 2, $\mathcal{C}$ is strong core stable. Because strong core stability implies individual stability we are done. ■

4.3 Contractual individual stability

Since individual stability implies contractual individual stability, it follows that the coalition structures generated by Algorithm 1 and Algorithm 2 are also contractual individually stable on the domains based on aversion to enemies and appreciation of friends, respectively. We ask now the question whether these algorithms always generate a contractual individually stable coalition structure on a *larger* preference domain.

Indeed, as suggested by Ballester (2004), a contractual individually stable coalition structure on any preference domain can be obtained by an algorithm which starts with an arbitrary coalition structure $\mathcal{C}$ (say, a coalition structure consisting of singletons only), and repeats the following operation until the resulting coalition structure becomes contractual individually stable: If some player i wishes to migrate to another coalition $X \neq \mathcal{C}(i)$ in $\mathcal{C}$ and no one in X and $\mathcal{C}(i)$ is worse off (i.e., $\mathcal{C}$ is not contractual individually stable), then remove i from $\mathcal{C}(i)$ and put i in X. Observe that, by applying this operation, no one is worse off and at least one player is strictly better off. Hence, the algorithm halts after a finite number of applications of the operation. As a straightforward estimation, the running time of this algorithm is exponential of $|N|$.

In the following, we show that, when the domain of separable preferences is under consideration, a contractual individually stable coalition structure can be obtained by Algorithm 2, whose running time is $O(|N|^2)$, a polynomial of $|N|$. Observe that Algorithm 2 only requires a partial description of each separable preference $\succeq_i$, namely the set G_i for each $i \in N$, while a complete description of a preference may have length exponential of $|N|$ even if players' preferences are separable.

Proposition 5 *Let* $(N, P) \in \mathcal{G}^s$. *Then a contractual individually stable coalition structure can be obtained by Algorithm 2.*

Proof. Let $(N, P) \in \mathcal{G}^s$. Recall that, for every $i, j \in N$ and $X \in \mathcal{N}_i$ with $j \notin X$, we have $[X \cup \{j\} \succ_i X \Leftrightarrow j \in G_i]$ and $[X \cup \{j\} \prec_i X \Leftrightarrow j \in B_i]$. Let $\mathcal{C}$ be the coalition structure constructed by Algorithm 2, i.e. $\mathcal{C}$ is the strong decomposition of the directed graph $H_{(N,P)}$. In the following, we show that $\mathcal{C}$ is contractual individually stable. Let $i \in N$, and consider the following two cases:

(1) $\mathcal{C}(i) = \{i\}$. Since each $X \in \mathcal{C}$ is a strongly connected component in $H_{(N,P)}$, for all $X \in \mathcal{C} \setminus \{\mathcal{C}(i)\}$, $i \notin G_j$ for all $j \in X$ if $j \in G_i$ for some $j \in X$. Thus, for all $X \in \mathcal{C}$ with $X \neq \mathcal{C}(i)$, we have $\mathcal{C}(i) = \{i\} \succ_i X \cup \{i\}$ or $X \succ_j X \cup \{i\}$ for all $j \in X$.

(2) $\mathcal{C}(i) \neq \{i\}$. Again, since each $X \in \mathcal{C}$ is a strongly connected component in $H_{(N,P)}$, there exists $j \in \mathcal{C}(i) \setminus \{i\}$ such that $i \in G_j$. Thus, $\mathcal{C}(i) \succ_j \mathcal{C}(i) \setminus \{i\}$ for some $j \in \mathcal{C}(i) \setminus \{i\}$.

Therefore, we can conclude that $\mathcal{C}$ is contractual individually stable. ■

Example 3 *In order to show that there are cases in which Algorithm* 1 *does not deliver a contractual individually stable coalition structure on the class of separable games, let us consider the following game: Let* $N = \{1, 2, 3\}$ *and let the players have the following preferences over the coalitions they may* *belong to:*

1	**2**	**3**
12	123	123
123	12, 23	13, 23
1	123	3
13		

Suppose now that Algorithm 1 *selects* $\{\{1\}, \{2, 3\}\}$*. Notice that this coalition structure is not contractual individually stable because all players are better off in the coalition structure* $\{1, 2, 3\}$*.*

5 Strategy-proofness

In this section, we study the question whether the algorithms presented in Section 3 prevent players from strategic behavior. In other words, we ask whether there is a player who can obtain a more preferred outcome by submitting to the proposed algorithms a preference different from his or her true preference.

Let $\varphi : \mathcal{G} \to \mathbf{C}^N$ be a rule that associates a coalition structure to each hedonic game on player set N. Then, for each $(N, P) \in \mathcal{G}$, $\varphi(N, P)$ denotes the coalition structure obtained by applying φ to (N, P). For each $i \in N$, we denote by $\varphi_i(N, P)$ the coalition in $\varphi(N, P)$ to which player i belongs. For every $P = (\succeq_1, \succeq_2, \ldots, \succeq_n), P' = (\succeq'_1, \succeq'_2, \ldots, \succeq'_n) \in \mathcal{P}$, we denote by $\delta(P, P') = \{i \in N \mid \succeq_i \not\equiv \succeq'_i\}$ the set of players whose preferences in P and P' are different.

Definition 3 *Let* $\bar{\mathcal{P}} \subseteq \mathcal{P}$*. We say that* $\varphi : \mathcal{G} \to \mathbf{C}^N$ *is* ***strategy-proof*** *on* $\bar{\mathcal{P}}$ *if* $\varphi_i(N, P) \succeq_i \varphi_i(N, P')$ *for each* $i \in N$ *and every* $P, P' \in \bar{\mathcal{P}}$ *with* $\delta(P, P') = \{i\}$*, where* $\succeq_i$ *is the preference of player* $i \in N$ *in preference* *profile* P.

In what follows in this section we concentrate on the rule $\varphi^1 : \mathcal{G} \to \mathbf{C}^N$ that associates a coalition structure to each $(N, P) \in \mathcal{G}$ according to Algorithm 1, and on the rule $\varphi^2 : \mathcal{G} \to \mathbf{C}^N$ that associates a coalition structure to each $(N, P) \in \mathcal{G}$ according to Algorithm 2. However, as mentioned in Section 3, the outcome of Algorithm 1 may not be unique when the largest clique in $H_{(N,P)}$ is not unique. In the following, a more precise description of Algorithm 1 is given.

For each $X \subseteq N$, let $e^X = (e_1^X, e_2^X, \ldots, e_n^X)$ be the n-dimensional vector with $e_i^X = 1$ if $i \in X$, and $e_i^X = 0$ otherwise. Then, for all $X, Y \subseteq N$, we write $X \trianglerighteq Y$ if and only if (1) $e^X = e^Y$ or (2) there is $k \in \{1, \ldots, n\}$ such that $e_i^X = e_i^Y$ for all $i < k$ and $e_k^X < e_k^Y$. Notice that $\trianglerighteq$ is a lexicographic order over all subsets of N. We formulate now Algorithm 1*.

Algorithm 1*

- Set $Y := N$ and $\mathcal{C} := \emptyset$.
- Repeat the following until $Y = \emptyset$:
  - Find all cliques $X \subseteq Y$ in $H_{(Y,P)}$ with the largest number of vertices, and collect them in the set $\mathcal{X}$.
  - Find the clique $X \in \mathcal{X}$ such that $X \trianglerighteq X'$ for all $X' \in \mathcal{X}$.
  - Set $Y := Y \setminus X$ and $\mathcal{C} := \mathcal{C} \cup \{X\}$.
- Return $\mathcal{C}$.

Notice that, in Algorithm 1^*, the selection of cliques is guided by the lexicographic order $\trianglerighteq$. We denote by φ^{1^*} the rule that associates a coalition structure to each $(N, P) \in \mathcal{G}^e$ according to Algorithm 1*, and we consider φ^{1^*} instead of φ^1.

Proposition 6 *The rule φ^{1^*} is strategy-proof on $\mathcal{P}^e$.*

Proof. Let $i \in N$, and let $P, P' \in \mathcal{P}^e$ with $\delta(P, P') = \{i\}$. For each $j \in N$, we denote by G_j and G'_j the set of friends of player j respectively in P and P'. From $\delta(P, P') = \{i\}$, we have $G_i \neq G'_i$ and $G_j = G'_j$ for each $j \in N \setminus \{i\}$. We distinguish the following two cases:

(1) $G'_i \subset G_i$. Keeping in mind that φ^{1^*} selects cliques and players' preferences are based on aversion to enemies, we conclude that it will be always the case that $\left|\varphi_i^{1^*}(N, P')\right| \leq \left|\varphi_i^{1^*}(N, P)\right|$, and therefore, $\varphi_i^{1^*}(N, P) \succeq_i \varphi_i^{1^*}(N, P')$. In other words, player i has no incentive to report a smaller set of friends under aversion to enemies.

(2) $G'_i \cap B_i \neq \emptyset$. Since φ^{1^*} selects cliques and players' preferences are based on aversion to enemies, we have $\varphi_i^{1^*}(N, P') \succ_i \varphi_i^{1^*}(N, P)$ only if $\left|\varphi_i^{1^*}(N, P')\right| > \left|\varphi_i^{1^*}(N, P)\right|$. Notice that this can happen only if there is at least one player $k \in \varphi_i^{1^*}(N, P')$ with $i \in G_k$ and $k \in G'_i \setminus G_i$. But this means that $k \in N \setminus G_i = B_i$, and therefore, $\varphi_i^{1^*}(N, P) \succ_i \varphi_i^{1^*}(N, P')$. Hence, player i has no incentive to declare a player k as his friend if $k \in B_i$.

Therefore, player i has no incentive to misrepresent his preference, i.e., φ^{1^*} is a strategy-proof on $\mathcal{P}^e$. ■

We turn now to the rule φ^2 that associates a coalition structure to each game $(N, P) \in \mathcal{G}$ according to Algorithm 2.

Proposition 7 *The rule φ^2 is strategy-proof on $\mathcal{P}^f$.*

Proof. Let $i \in N$, and let $P, P' \in \mathcal{P}^f$ with $\delta(P, P') = \{i\}$. For each $j \in N$, we denote by G_j and G'_j the set of friends of player j respectively in P and P'. From $\delta(P, P') = \{i\}$, we have $G_i \neq G'_i$ and $G_j = G'_j$ for each $j \in N \setminus \{i\}$.

We first show that $(\varphi_i^2(N, P') \cap G_i) \subseteq (\varphi_i^2(N, P) \cap G_i)$. Let $j \in (\varphi_i^2(N, P') \cap G_i)$. Since $\varphi_i^2(N, P')$ is a strongly connected component in $H_{(N,P')}$, there exists a path from j to i. Without loss of generality we assume that i appears exactly once on such a path (because a shorter path from j to i can be obtained from a path from j to i on which i appears more than once). Then, i appears only as the last vertex on such a path. It follows that such a path is also in $H_{(N,P)}$, because $G_j = G'_j$ for each $j \in N \setminus \{i\}$. From $j \in G_i$, there exists a path from i to j in $H_{(N,P)}$, because edge (i, j) is contained in $H_{(N,P)}$. Therefore, $j \in \varphi_i^2(N, P)$, because $\varphi_i^2(N, P)$ is a strongly connected component in $H_{(N,P)}$.

Suppose $(\varphi_i^2(N, P') \cap G_i)$ is a proper subset of $(\varphi_i^2(N, P) \cap G_i)$. Then, we have $\varphi_i^2(N, P') \prec_i \varphi_i^2(N, P)$, because player's preferences are based on appreciation of friends. Thus, player i has no incentive to misrepresent G_i by G'_i, and we are done.

Now suppose $(\varphi_i^2(N, P') \cap G_i) = (\varphi_i^2(N, P) \cap G_i)$, and we show that $(\varphi_i^2(N, P) \cap B_i) \subseteq (\varphi_i^2(N, P') \cap B_i)$. Let $j \in (\varphi_i^2(N, P) \cap B_i)$. Since $\varphi_i^2(N, P)$ is a strongly connected component in $H_{(N,P)}$, there exists a path p from i to j in $H_{(N,P)}$, and there also exists a path p' from j to i in $H_{(N,P)}$. Again, without loss of generality, we assume that i appears exactly once on each of p and p'. It follows that all vertices which appear on p belong to $\varphi_i^2(N, P)$. Let k be the vertex appearing on path p immediately after i. Observe that $k \in G_i$, and thus, $k \in (\varphi_i^2(N, P) \cap G_i)$. By assumption, we have $k \in (\varphi_i^2(N, P') \cap G_i)$, and thus, there is a path from i to k in $H_{(N,P')}$.

Let p'' be the subpath of p from k to j. Observe that i does not appear on p'', and thus, p'' is also a path (from k to j) in $H_{(N,P')}$. Moreover, observe that p' is also a path (from j to i) in $H_{(N,P')}$, because i appears only as the last vertex on p'. Therefore, we have $j \in \varphi_i^2(N, P')$, because $\varphi_i^2(N, P')$ is a strongly connected component in $H_{(N,P')}$, and there are paths from i to k, from k to j, and from j to i.

Now we can conclude that $\varphi_i^2(N, P') \preceq_i \varphi_i^2(N, P)$. Therefore, each player has no incentive to misrepresent his or her preference, i.e., φ^2 is a strategy-proof on $\mathcal{P}^f$. ■

6 Conclusion

We have shown in this paper that for a separable hedonic game two simple algorithms play an important role: the first algorithm (Algorithm 1) partitions the set of players into coalitions each of which is a clique in the directed graph corresponding to the game; the second algorithm (Algorithm 2) partitions the set of players into coalitions each of which is a strongly connected component of the directed graph corresponding to the game. The importance of these algorithms is due to the following observations: (1) Algorithm 2 delivers a contractual individually stable coalition structure in polynomial time when players' preferences are separable; (2) Algorithm 1 and Algorithm 2 generate core stable and individually stable coalition structures when players' preferences are based on aversion to enemies and appreciation of friends, respectively; (3) adding mutuality to aversion to enemies and to appreciation of friends guarantees that the coalition structures generated by Algorithm 1 and Algorithm 2 are Nash stable on the corresponding preference domains as well; (4) both algorithms (with a slight modification of Algorithm 1) prevent players from strategic behavior on the corresponding domains.

References

[1] Ballester, C. (2004): NP-completeness in hedonic games, Games and Economic Behavior 49, 1-30.

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

[3] Bogomolnaia, A. and M. O. Jackson (2002): The stability of hedonic coalition structures, Games and Economic Behavior 38, 201-230.

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

[5] Dimitrov, D., P. Borm, R. Hendrickx, and S.-C. Sung (2004): Simple priorities and core stability in hedonic games, CentER Discussion Paper 2004-5, Tilburg University.

[6] Dréze, J. and J. Greenberg (1980): Hedonic coalitions: optimality and stability, Econometrica 48, 987-1003.

[7] Tarjan, R. E. (1972): Depth-first search and linear graph algorithms, SIAM Journal of Computing 1(2), 146-160.