



ELSEVIER

Mathematical Social Sciences 45 (2003) 185–203

www.elsevier.com/locate/econbase

mathematical
social
sciences

On coalition formation: durable coalition structures

Salvador Barberà^a, Anke Gerber^{b,*}

^a*Departament d'Economia i d'Història Econòmica and CODE, Universitat Autònoma de Barcelona,
08193 Bellaterra, Barcelona, Spain*

^b*Institute for Empirical Research in Economics, Blümlisalpstrasse 10, CH-8006 Zurich, Switzerland*

Received 12 June 2002; accepted 30 January 2003

Abstract

We define a solution to the problem of coalition formation that applies to purely hedonic games. Coalition structures satisfying our requirements are called durable, and we interpret them as much more likely to last than those coalition structures not satisfying the requirements, which we call transient. Durability results from a combination of foresight and extreme risk aversion on the part of agents, when considering to join others to disrupt an existing structure in search of higher gains. Agents' calculations are also constrained to satisfy a strong consistency requirement, which is reflected in the recursive structure of our definition. We prove that durable coalition structures always exist, and we provide examples of hedonic games where our solutions apply nicely.

© 2003 Elsevier Science B.V. All rights reserved.

Keywords: Hedonic games; Foresight; Risk aversion

JEL classification: C71

1. Introduction

Given the interests of agents in a society, and their possibilities to cooperate, will durable social arrangements exist? If so, how durable will they be? What is it that keeps the members of groups together and apart from other groups? The answers to these questions and the predictions we obtain from them are relevant for many economic and social problems, in particular for the study of local public goods such as, for example, in Guesnerie and Oddou (1981), Greenberg and Weber (1986, 1993) and Demange (1994). Despite its relevance, the issue of coalition formation has been neglected until very

*Corresponding author. Tel.: +41-1-634-3708; fax: +41-1-634-4907.

E-mail address: agerber@iew.unizh.ch (A. Gerber).

recently. In the context of characteristic function games, where each coalition can choose from a set of feasible utility allocations, the focus was on the question of the choice of a utility allocation for some exogenously given coalition structure.¹ Among the notable exceptions that also address the issue of coalition formation are Shenoy (1979), Hart and Kurz (1983), Bennett and Zame (1988), Zhou (1994) and Gerber (2000).

In this paper we concentrate on the aspect of coalition formation and define a solution concept that applies to hedonic coalition formation games: that is, to situations where agents derive utility from belonging to a group, and do not care about the arrangements among people outside their own group. Unlike Drèze and Greenberg (1980), who introduced the ‘hedonic aspect’ of coalition formation, we restrict ourselves to situations that are purely hedonic in the sense that, within a coalition, there is no distributional issue to solve. Rather, an agent’s utility only depends on the composition of the group he belongs to. Such hedonic games have recently been studied by Bogomolnaia and Jackson (2002) and Banerjee et al. (2001). Both papers provide sufficient conditions for the nonemptiness of the core, which is by no means guaranteed in this context. Alcalde and Revilla (2001) give another condition ensuring that the core is not empty, and also that the direct mechanisms that assign participants in coalition formation games to coalitional structures in the core will be strategy-proof. Pápai (2000) provides an even more stringent condition on families of coalition formation games, which guarantees the existence of a unique core allocation in all games within any family meeting her single-lapping requirement. Bogomolnaia and Jackson (2002) also propose the weaker notions of individual and Nash stability and analyze the existence of coalition structures satisfying these requirements.

We propose a new solution concept which captures the notion of ‘maxmin’ behavior on the part of the players and we prove that our solution selects a nonempty set of what we call *durable* coalition structures for any purely hedonic game. When the core of the coalition game is nonempty, all coalition structures in the core will be declared to be durable. What makes a coalition structure durable is a combination of foresight and risk aversion. Agents will not disrupt a coalition structure if they can foresee that they may end up in a situation that makes them worse off. Our definition qualifies this statement and makes it precise. Suppose that a coalition R is considering to form, thus disrupting a coalition structure \mathcal{S} where the agents of R do not form a group. What coalition structure will this change lead to? We consider that it will lead to any of the coalition structures \mathcal{S}' having the properties

1. R is one of the coalitions in \mathcal{S}' ,
2. coalitions in \mathcal{S} not affected by the formation of R remain in \mathcal{S}' , and
3. agents who were previously associated with members of R , but who are not part of R , are associated in a coalition substructure that would be durable if they were alone in society.

Requirement 3 introduces a recursive element into our definition of durability.

We shall define what coalition structures are durable on the basis of a construction

¹It is often assumed that a grand coalition is formed.

which already assumes what coalition structures would be durable would a smaller society form coalitions of its own. This recursive element incorporates a notion of consistency in the calculations of agents: when judging whether certain coalitions might form, they apply the same equilibrium concept to subsocieties as that resulting for society as a whole from their own calculations.

Joint deviations from \mathcal{S} by several coalitions will directly lead to another coalition structure \mathcal{S}' by similar considerations. We define the transitive closure of this relation, and say that \mathcal{S} leads to \mathcal{S}' if the two stand in this new relation. \mathcal{S} will be durable if any path away from it eventually leads to some coalition structure \mathcal{S}' where one of the agents disrupting \mathcal{S} would be worse off.

We are aware that our notion of durability incorporates some asymmetry in the degree of rationality that each agent attributes to himself, as compared to that he attributes to others. We elaborate on this point after the formal definition of durability in Section 2, where we emphasize the interpretation and the arguments that lead us to propose and defend our concept. Its merits should also be judged against the background of the alternative proposals in the currently growing literature on coalition formation.

There are a few papers that incorporate farsightedness in a context where coalitions can induce certain alternatives from others. Chwe (1994) proposes the notion of a consistent set, which shares the idea of maxmin behavior with our notion of durability. According to Chwe, a set of outcomes (coalition structures in our context) is consistent if any deviation from an outcome in this set is deterred, because it may lead to another outcome in the set that makes some deviator worse off. In the same framework, Xue (1998) utilizes Greenberg's (1990) theory of social situations and introduces the notion of a stable standard of behavior. Konishi and Ray (1999) model coalition formation as a dynamic process and analyse 'equilibrium deviation schemes', thereby capturing perfect foresight on the part of the players. Bhattacharya (2002) introduces a further requirement in Chwe's analysis, by imposing a credibility constraint over the type of outcomes that can dominate others.

None of these concepts, though, is directly applicable to a hedonic game, or, more generally, to a characteristic function game. The reason for this is that these models take as given an 'effectiveness relation' which describes the set of outcomes a deviating coalition can induce from a given alternative. However, in a context where the utility of a player depends on the group he belongs to, there is no naturally given effectiveness relation, and it is too naive to assume that those players who face a coalitional deviation by others (the 'residuals') either stay together or dissolve into singletons. It is even more critical to assume that a deviating coalition can impose any coalition structure on the residuals. Hence we believe that one important aspect of our solution concept is the development of a consistent theory of where coalitional deviations lead to. Another difference between our solution and the concepts mentioned above is that our notion of durability can be referred directly to each possible coalition structure, whereas, for example, Chwe's consistency is not the property of any individual coalition structure, but refers to sets of such objects.² Diamantoudi and Xue (2000) extend several of the

²This difference is analogous to that between the core and the von Neumann–Morgenstern stable set (von Neumann and Morgenstern, 1944). While being a core member is a property of each imputation, the von Neumann–Morgenstern stable set refers to a set taken as a whole.

pre-existing solution concepts by including foresight in the calculations of potential colluders in hedonic games, along the lines proposed by Greenberg (1990) and Chwe (1994). They have to address, as we do, the issue of potential coalition formation among agents who are not part of deviating coalitions. They assume that such agents stay put and take no immediate action. This is in contrast to our consistent treatment of coalition formation among residual members who have lost some of their partners.

The idea of recursive consistency that is inherent in our notion of durability can be found in several game theoretic solution concepts. In the context of strategic form games it underlies the definition of a *coalition-proof Nash equilibrium* in Bernheim et al. (1987). For characteristic function games, Ray (1989) proposes a modification of the core, where the proposals of blocking coalitions are tested for stability in the same way as is the original utility allocation. It turns out that the core and the modified core are identical. Dutta et al. (1989) introduce the *consistent bargaining set*, which also requires a test of counterobjections for validity, and this process is taken to the limit. For games in which the utility of a player depends on the whole coalition structure, Ray and Vohra (1997) have proposed the concept of an *equilibrium binding agreement*.³ The recursive element of their definition is similar to ours. If a coalition considers disrupting some given coalition structure it is neither assumed to be very pessimistic nor to be very optimistic concerning the partition that it expects to arise after the deviation. Rather, the resulting partition is expected to be consistent with the equilibrium notion one is going to define. Unlike Ray and Vohra (1997), who only consider deviations that lead to refinements of a given coalition structure, we allow for arbitrary coalitional deviations. A drawback of their definition when applied to purely hedonic games is that the set of equilibrium binding agreements coincides with the core: this leaves us with the usual problem of emptiness.

To conclude the literature review let us remark that our notion of durability has an interesting parallel in voting models, for which Rubinstein (1980) has introduced the ‘stability set’.⁴ The stability set captures ‘prudent’ behavior on the part of the voters: a voter i will not vote for alternative y over alternative x if y is then beaten by some alternative z which is worse for i than alternative x . Hence, as in our definition of durability, voters are assumed to behave conservatively. However, their farsightedness is limited since they only look one step ahead while we assume the players to look arbitrarily far ahead.

Summarizing, we are aware of some of the shortcomings of our proposal, but we feel that it fares well relative to other solutions which are well established in the literature. In particular, we value the fact that it incorporates a limited but well-defined degree of rationality and foresight for conservative players, and that it allows us to classify every single coalition structure, per se, as being durable or transient.

The paper proceeds as follows. After preliminaries, in Section 2 we provide a definition of durability followed by motivational remarks. We provide an existence proof

³In fact, the framework in Ray and Vohra (1997) is even more general, since they analyze games in strategic form, where binding agreements can be written between members of a coalition but not across coalitions.

⁴See also Martin and Merlin (2002) for an analysis of the relationship of the stability set with other social choice correspondences and its normative properties.

and prove some interesting properties of durable structures. In Section 3 we present examples to illustrate the potential of our proposed definition for special games or classes of games. Finally, Section 4 concludes.

2. Durable coalition structures

We consider a *hedonic game*, $(N; (\succeq_i)_{i \in N})$, where N is the finite set of *players*, and \succeq_i is a complete and transitive preference relation on $S_i(N) = \{S \subset N \mid i \in S\}$ for all $i \in N$. Strict preference and indifference will be denoted by $>_i$ and \sim_i , respectively. A set $\emptyset \neq S \subset N$ is called a *coalition*. If $(N; (\succeq_i)_{i \in N})$ is a hedonic game and T is a coalition, then the restriction of $(N; (\succeq_i)_{i \in N})$ to the player set T is again a hedonic game and is given by $(T; (\succeq_i^T)_{i \in T})$, where \succeq_i^T is the restriction of \succeq_i to $S_i(T)$. A *coalition structure* \mathcal{S} on N is a partition of N into disjoint coalitions. We denote the set of all coalition structures on N by \mathcal{P} . Let $S(i, \mathcal{S})$ be the coalition in $\mathcal{S} \in \mathcal{P}$ that contains player i .

We will propose the notion of a *durable* coalition structure as motivated in the Introduction. In order to give the basic idea, let us assume that we have explained how the players will partition themselves after some coalitions have deviated from a given coalition structure. Knowing that their deviation might be the starting point of further deviations by others, that is, knowing the graph on the set of coalition structures that is defined by all potential coalitional deviations, will these coalitions actually deviate? Our assumption is that they will only deviate if no matter which coalition structure is reached later, they will never be worse off than in their present coalition structure. Thus, we will consider as potentially ‘unstable’ those coalition structures which can be disrupted by some coalitions without the risk of a future ‘loss’ for any of the deviating players. Formally, we recursively define the notion of a *durable* coalition structure as follows.

Let $(N; (\succeq_i)_{i \in N})$ be a hedonic game and assume that the notion of a durable coalition structure has already been defined for all hedonic games with strictly less than $|N|$ players.⁵ Then, a coalition structure \mathcal{S} on N *directly leads* to a coalition structure \mathcal{S}' via coalitions T^1, T^2, \dots, T^m ($m \geq 1$), and we write

$$\mathcal{S}^{T^1, \dots, T^m} \rightarrow \mathcal{S}'$$

if

1. $T^h \in \mathcal{S}'$ and $T^h >_i S(i, \mathcal{S})$ for all $i \in T^h$ and for all $h = 1, \dots, m$,
2. $\mathcal{B} = \{T \mid T \in \mathcal{S} \text{ and } T \cap T^h = \emptyset \text{ for all } h = 1, \dots, m\} \subset \mathcal{S}'$,
3. $\mathcal{S}' \setminus (\cup_{h=1}^m \{T^h\} \cup \cup_{S \in \mathcal{B}} \{S\})$ is either empty or durable in the hedonic game obtained by restricting $(N; (\succeq_i)_{i \in N})$ to the player set $N \setminus (\cup_{h=1}^m T^h \cup \cup_{S \in \mathcal{B}} S)$.

⁵By $|A|$ we denote the cardinality of a set A .

A coalition structure \mathcal{S} directly leads to a coalition structure \mathcal{S}' and we write $\mathcal{S} \rightarrow \mathcal{S}'$ if there exist coalitions T^1, \dots, T^m such that

$$\mathcal{S}^{T^1, \dots, T^m} \rightarrow \mathcal{S}'.$$

Finally, we define \rightarrow to be the transitive closure of the relation \rightarrow , i.e. $\mathcal{S} \rightarrow \mathcal{S}'$, if there exist coalition structures $\mathcal{S}^0, \dots, \mathcal{S}^t$ such that $\mathcal{S}, \mathcal{S}^t = \mathcal{S}'$ and $\mathcal{S}^j \rightarrow \mathcal{S}^{j+1}$ for $j = 0, 1, \dots, t-1$. We then say that \mathcal{S} leads to \mathcal{S}' .

Definition 2.1. A coalition structure \mathcal{S} on N is *transient* if there exist coalitions T^1, \dots, T^m , such that for all coalition structures \mathcal{S}' with

$$\mathcal{S}^{T^1, \dots, T^m} \rightarrow \mathcal{S}',$$

and for all coalition structures \mathcal{S}'' with $\mathcal{S}' \rightarrow \mathcal{S}''$ it is true that

$$S(i, \mathcal{S}'') \geq_i S(i, \mathcal{S}') \text{ for all } i \in T^h \text{ and for all } h = 1, \dots, m.$$

A coalition structure \mathcal{S} is *durable* if it is not transient.

Observe that the core of the hedonic game $(N; (\geq_i)_{i \in N})$ is always a subset of the set of durable coalition structures.⁶

Let us reflect upon the definition of durability for a moment. First observe that the notion of a durable coalition structure is well defined even if the set of durable coalition structures were empty for some of the restricted hedonic games appearing in the definition of the relation ‘directly leading’. However, in the following we will show that there always exist durable coalition structures.

It is not straightforward to predict the way in which players partition themselves into coalitions after some coalition structure has been disrupted. But if we were to predict which coalition structure would arise, then it should be consistent with our theory. Therefore, we assume that the players in coalitions who are not affected by a deviation stay together and that the members of coalitions that are broken and who themselves are not part of any deviating coalition (if any) form a coalition structure that is durable in their ‘subsociety’. This defines the binary relation ‘directly leading’. Then, coalitions will hesitate to deviate from a durable coalition structure since this may trigger further deviations and finally lead to a partition in which some of the originally deviating players are worse off. On the other hand, coalition structures that can be safely disrupted by some coalitions will only be transient.

Now that we have presented our definitions of durable coalition structures and before turning to the proof of existence, let us elaborate on their interpretation and on some of the criticisms they are open to. First, we comment on the use of the maxmin criterion and its proper interpretation. Maxmin behavior is an expression of extreme risk aversion, with agents attaching not only very low utilities to their least preferred outcomes, but also attributing to them some probability of choice, however small. Agents compute the

⁶For any hedonic game $(N; (\geq_i)_{i \in N})$ the *core* is the set of all coalition structures $\mathcal{S} \in P$ such that there exists no coalition T with $T >_i S(i, \mathcal{S})$ for all $i \in T$.

min over outcomes which are possible, then choose actions that maximize this min. In order to be consistent with this interpretation, it is best to think of durability as a matter of degree, not as an all-or-nothing property. Durable coalition structures are likely to stay, for all the reasons we provide. Non-durable coalition structures are likely to be ephemeral, at best, but still possible. This relativistic view is more agreeable to us than any deterministic, predictive interpretation. We are not claiming that durable structures will happen for sure, and last forever.

Next, we address a criticism of our definition which has been advanced by Debraj Ray. Suppose I am a player who is considering to cooperate in disrupting a given coalition structure. Suppose I am deterred by the threat that this deviation ‘leads’ to another I dislike, after a few steps. These steps may be taken by agents who, when looking ahead, would also be deterred by a similar threat, if they used exactly the same reasoning. Yet, our definition does not sever the link between the current coalition structure and my threat, even when the intermediate deviants might in fact feel threatened when taking the actions we attribute to them.

This is not a trivial objection. Yet, the criticism would be especially bothersome if we were thinking of durability as the basis for a zero–one classification between what can happen and what cannot. Under such a radical interpretation, non-durable structures should not play any role in defining durable ones, other than being their complement. We leave this as an open puzzle, one that is far from our motivation. Under our more relativistic view we can actually defend our notion of durability. If agents are ‘maxminers’ they will attach positive probability, however small, to any utility-enhancing move that leads away from a durable coalition structure. Thus, we deliberately do not sever any links in the original graph defined by the relation ‘directly leading’.

We now turn attention to the question of the existence of durable coalition structures. We begin by a remark.

Remark 2.1.

- (a) If $N = \{1\}$, then $\{N\}$ is durable. This is because $\{N\}$ is the unique coalition structure on N which, by definition, does not lead to any coalition structure on N .
- (b) If $N = \{1, 2\}$, then we get the following three cases.
 - (b1) If $N >_i \{i\}$ for $i = 1, 2$, then $\{N\}$ is the only durable coalition structure.
 - (b2) If there exists $i \in N$ such that $\{i\} >_i N$, then $\{\{1\}, \{2\}\}$ is the only durable coalition structure.
 - (b3) If $N \succeq_i \{i\}$ for $i = 1, 2$ and $N \sim_i \{i\}$ for at least one $i \in N$, then both coalition structures $\{N\}$ and $\{\{1\}, \{2\}\}$ are durable.

The above remark proves that the set of durable coalition structures is non-empty, meaningful and easy to compute when there are one or two agents. Before proving existence for an arbitrary number of agents we provide a useful lemma. In words, it states the following: consider a cycle of coalition structures in the graph defined by the relation \rightarrow . Then there exists at least one agent participating in a deviation from some coalition structure \mathcal{S} in the cycle who loses utility compared to \mathcal{S} at some other coalition structure in the cycle. More formally:

Lemma 2.1. Let $(N; (\succeq_i)_{i \in N})$ be a hedonic game and let $\mathcal{S}^1, \dots, \mathcal{S}^J$ be a finite sequence of coalition structures on N such that the following condition is satisfied. For all $j = 1, \dots, J$, there exist coalitions $T_1^j, \dots, T_{m(j)}^j$, with

1.

$$\mathcal{S}^j \xrightarrow{T_1^j, \dots, T_{m(j)}^j} \mathcal{S}^{j+1},$$

2. for all coalition structures \mathcal{S}' such that $\mathcal{S}^{j+1} \twoheadrightarrow \mathcal{S}'$, it is true that $S(i, \mathcal{S}') \succeq_i S(i, \mathcal{S}^j)$ for all $i \in T_h^j$ and for all $h = 1, \dots, m(j)$.

Then $\mathcal{S}^1 \neq \mathcal{S}^J$.

Proof. Let $(N; (\succeq_i)_{i \in N})$ be a hedonic game and assume by way of contradiction that there exists a finite sequence of coalition structures $\mathcal{S}^1, \dots, \mathcal{S}^J$ which fulfills the conditions in the statement of the lemma but for which $\mathcal{S}^1 = \mathcal{S}^J$. Consider one of the coalitions T_h^1 for some $h \in \{1, \dots, m(1)\}$. Clearly, $T_h^1 \notin \mathcal{S}^1 = \mathcal{S}^T$ and therefore there exists a maximal $j \in \{2, \dots, T-1\}$ such that $T_h^1 \in \mathcal{S}^j$. Since $T_h^1 \notin \mathcal{S}^{j+1}$, by definition there exists $h' \in \{1, \dots, m(j)\}$ and $i \in T_h^1 \cap T_{h'}^j$. Hence, condition 2 implies that $S(i, \mathcal{S}^1) \succeq_i S(i, \mathcal{S}^j) = T_h^1$. However, this leads to a contradiction since $T_h^1 \succ_i S(i, \mathcal{S}^1)$. \square

Theorem 2.1. Let $(N; (\succeq_i)_{i \in N})$ be a hedonic game. If a coalition structure $\tilde{\mathcal{S}}$ is transient, then there exists a durable coalition structure \mathcal{S} such that $\tilde{\mathcal{S}} \twoheadrightarrow \mathcal{S}$.

Proof. Let $(N; (\succeq_i)_{i \in N})$ be a hedonic game and let $\tilde{\mathcal{S}}$ be a transient coalition structure. Define

$$\mathcal{B} = \{\tilde{\mathcal{S}}\} \cup \{\mathcal{S} \mid \tilde{\mathcal{S}} \twoheadrightarrow \mathcal{S}\}.$$

Observe that $\mathcal{B} \setminus \{\tilde{\mathcal{S}}\} \neq \emptyset$ since a necessary condition for $\tilde{\mathcal{S}}$ to be transient is that there exists a coalition structure that $\tilde{\mathcal{S}}$ directly leads to.

Suppose the claim is false, so that \mathcal{S} is transient for all $\mathcal{S} \in \mathcal{B}$. Then, for any $\mathcal{S} \in \mathcal{B}$ there exist coalitions T^1, \dots, T^m and some $\mathcal{S}' \in \mathcal{B}$ such that

$$\mathcal{S} \xrightarrow{T^1, \dots, T^m} \mathcal{S}',$$

and $S(i, \mathcal{S}'') \succeq_i S(i, \mathcal{S})$ for all $i \in T^h$, for all $h = 1, \dots, m$, and for all \mathcal{S}'' such that $\mathcal{S}' \twoheadrightarrow \mathcal{S}''$. Since this is true for all $\mathcal{S} \in \mathcal{B}$ and since \mathcal{B} is finite, there exists a finite sequence of coalition structures $\mathcal{S}^1, \dots, \mathcal{S}^T$ with $\mathcal{S}^1 = \mathcal{S}^T$, which fulfills conditions 1 and 2 in Lemma 2.1. This contradicts Lemma 2.1. \square

Corollary 2.1. For any hedonic game $(N; (\succeq_i)_{i \in N})$ the set of durable coalition structures is nonempty.

Proof. The claim immediately follows from Theorem 2.1. \square

We now provide a first simple test that our definitions are sensible, by showing that coalition structures that violate individual rationality can never be durable.

Definition 2.2. A coalition structure \mathcal{S} is said to be *individually rational* if $S(i, \mathcal{S}) \succeq_i \{i\}$ for all $i \in N$.

Theorem 2.2. Let $(N; (\succeq_i)_{i \in N})$ be a hedonic game. Then, any individually rational coalition structure leads to individually rational coalition structures only. Moreover, if \mathcal{S} is durable then \mathcal{S} is individually rational.

Proof. We will show inductively over $|N|$ that if \mathcal{S} is an individually rational coalition structure for the hedonic game $(N; (\succeq_i)_{i \in N})$, then it leads to individually rational coalition structures only, and if \mathcal{S} is durable, then \mathcal{S} is individually rational.

If $|N| = 1$, then nothing has to be proved. Let $|N| > 1$ and assume that, for all hedonic games $(T; (\succeq_i)_{i \in T})$ with $|T| < |N|$, the claim has been proved. Let \mathcal{S} be an individually rational coalition structure on N and assume by way of contradiction that there exists $i \in N$ and a coalition structure \mathcal{S}' on N such that

$$\mathcal{S}^{T^1, \dots, T^m} \rightarrow \mathcal{S}'$$

and $\{i\} >_i S(i, \mathcal{S}')$. Then $i \notin T^h$ for all $h \in \{1, \dots, m\}$. Therefore, either $S(i, \mathcal{S}) \in \mathcal{S}'$, which gives an immediate contradiction, or, for $\mathcal{B} = \{T \mid T \in \mathcal{S} \text{ and } T \cap T^h = \emptyset \text{ for all } h = 1, \dots, m\}$ we find that

$$S(i, \mathcal{S}') \in \mathcal{S}' \setminus (\cup_{h=1}^m \{T^h\} \cup \cup_{S \in \mathcal{B}} \{S\}),$$

where the latter is a durable coalition structure in the restriction of the hedonic game $(N; (\succeq_i)_{i \in N})$ to the player set $N \setminus (\cup_{h=1}^m T^h \cup \cup_{S \in \mathcal{B}} S)$. By the induction hypothesis, this implies that $S(i, \mathcal{S}') \succeq_i \{i\}$, which is a contradiction. Therefore, any individually rational coalition structure \mathcal{S} on N leads to individually rational coalition structures only.

Let \mathcal{S} be a non-individually rational coalition structure on N and let $J = \{i \in N \mid \{i\} >_i S(i, \mathcal{S})\}$. Then \mathcal{S} directly leads to some individually rational coalition structure \mathcal{S}' via the coalitions $\{i\}$, $i \in J$. We have shown above that \mathcal{S}' leads to individually rational coalition structures \mathcal{S}'' only, i.e. $S(i, \mathcal{S}'') \succeq_i \{i\} >_i S(i, \mathcal{S})$ for all $i \in J$. Therefore, \mathcal{S} is transient. \square

A direct consequence of Theorem 2.2 is the following corollary which proves to be very useful for computing durable coalition structures.

Corollary 2.2. Let $(N; (\succeq_i)_{i \in N})$ be a hedonic game. Then there does not exist a durable coalition structure \mathcal{S} and a coalition T such that $\{i\} \in \mathcal{S}$ and $T >_i \{i\}$ for all $i \in T$.

Proof. Let $(N; (\succeq_i)_{i \in N})$ be a hedonic game. Assume by way of contradiction that there exists a durable coalition structure \mathcal{S} and a coalition T such that $\{i\} \in \mathcal{S}$ and $T >_i \{i\}$ for all $i \in T$. Then, there exists some coalition structure \mathcal{S}' such that

$$\mathcal{S} \xrightarrow{T} \mathcal{S}'.$$

By Theorem 2.2 the coalition structures \mathcal{S} and \mathcal{S}' are individually rational and \mathcal{S}' leads to individually rational coalition structures only. Hence, $S(i, \mathcal{S}'') \succeq_i \{i\} = S(i, \mathcal{S})$ for all $i \in T$ and all coalition structures \mathcal{S}'' such that $\mathcal{S}' \twoheadrightarrow \mathcal{S}''$. Hence, \mathcal{S} is transient, which is a contradiction. \square

3. Further properties and examples

In this section we provide further results on durable coalition structures and present different examples of hedonic games which illustrate the versatility of the solution concept we have proposed. In the following we will frequently use a simplified notation for coalition structures. For example, we write $[12 \mid 3 \mid 456]$ for the coalition structure $\{\{1,2\}, \{3\}, \{4,5,6\}\}$ on the player set $N = \{1,2,\dots,6\}$. Also, for any hedonic game $(N; (\succeq_i)_{i \in N})$ we will only list the preferences of the players over individually rational coalitions, where a coalition S is individually rational if $S \succeq_i \{i\}$ for all $i \in S$. We begin by studying the relationship between the set of durable coalition structures and the core, which, as we have already seen, is always a subset of the set of durable coalition structures.

Example 3.1. This is a simple example of a roommate problem taken from Gale and Shapley (1962). There are $N = \{1,2,3,4\}$ students who can divide into pairs of roommates. Everyone prefers to share a room with someone over being alone.

Coalitions of size greater than two are unfeasible. We can represent the students' preferences over roommates by preferences over coalitions if we define these preference relations such that unfeasible coalitions are not individually rational. The preferences we are going to study are as follows:⁷

$$\begin{aligned} \{1,2\} &>_1 \{1,3\} >_1 \{1,4\} >_1 \{1\}, \\ \{2,3\} &>_2 \{1,2\} >_2 \{2,4\} >_2 \{2\}, \\ \{1,3\} &>_3 \{2,3\} >_3 \{3,4\} >_3 \{3\}, \\ \{2,4\} &>_4 \{3,4\} >_4 \{1,4\} >_4 \{4\}. \end{aligned}$$

By Theorem 2.2 and Corollary 2.2 the only candidates for durable coalition structures are $[12 \mid 34]$, $[13 \mid 24]$ and $[14 \mid 23]$. According to Remark 2.1 the coalition structure $\{\{ij\}\}$ is durable relative to $\{i,j\}$ for all $i,j \in \{1,2,3,4\}$, $i \neq j$. Hence we get the cycle illustrated in Fig. 1. While the core of this game is obviously empty, we can immediately see that all coalition structures in this cycle are durable since there is always one deviator who is deterred since she may end up with roommate 4 after a further deviation.

Example 3.2. This example illustrates that the core, even if nonempty, may be a strict

⁷Actually, all we need in this example is that 1's most preferred roommate is 2, that 2's most preferred roommate is 3, that 3's most preferred roommate is 1 and that 1, 2, 3 rank 4 last.

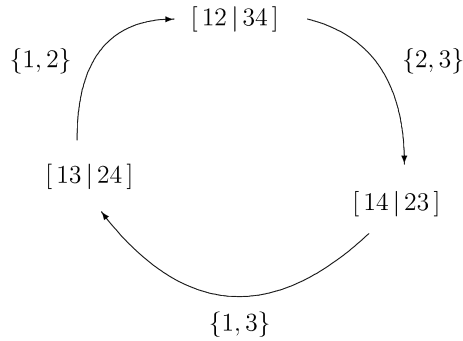


Fig. 1. Example 3.1.

subset of the set of durable coalition structures. Let $N = \{1, 2, 3\}$ and let the preferences \succeq_i be given as follows:

$$\begin{aligned} \{1, 3\} &>_1 \{1, 2, 3\} >_1 \{1, 2\} >_1 \{1\}, \\ \{1, 2\} &>_2 \{1, 2, 3\} >_2 \{2, 3\} >_2 \{2\}, \\ \{1, 2, 3\} &>_3 \{2, 3\} >_3 \{1, 3\} >_3 \{3\}. \end{aligned}$$

By Theorem 2.2 and Corollary 2.2 we can concentrate on the coalition structures $[12 | 3]$, $[13 | 2]$, $[1 | 23]$ and $[123]$. According to Remark 2.1 the coalition structure $\{\{i\}\}$ is durable relative to $\{i\}$. Therefore, we get the graph presented in Fig. 2. The core is given by the coalition structure $[123]$ while the set of durable coalition structures is given by

$$\{[12 | 3], [13 | 2], [123]\}.$$

To see this, note that, for example, $[12 | 3]$ is durable since it directly leads to $[13 | 2]$ via $\{1, 3\}$ and this directly leads to $[1 | 23]$ which is worse for player 1 than $[12 | 3]$.

In both examples the set of durable coalition structures contains more than one element and is a strict superset of the core. A straightforward question then is whether

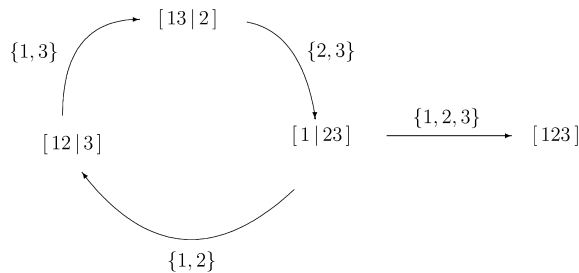


Fig. 2. Example 3.2.

one can construct an example with a unique durable coalition structure and an empty core. The next theorem shows that this is impossible.

Theorem 3.1. *Let $(N; (\succeq_i)_{i \in N})$ be a hedonic game. If there exists a unique durable coalition structure \mathcal{S} , then the core is nonempty and contains as its unique element the coalition structure \mathcal{S} .*

Proof. Let $(N; (\succeq_i)_{i \in N})$ be a hedonic game and let \mathcal{S} be its unique durable coalition structure. Assume that the core is empty. Since the core is always a subset of the set of durable coalition structures the theorem is proved once we have established that this assumption leads to a contradiction. If the core is empty any coalition structure \mathcal{S}' directly leads to another coalition structure \mathcal{S}'' . Since \mathcal{S}' is transient for all $\mathcal{S}' \neq \mathcal{S}$ we can construct a path $\mathcal{S}^1, \mathcal{S}^2, \dots$ as follows. Since \mathcal{S} is not in the core there exists a coalition T and a coalition structure \mathcal{S}' with

$$\mathcal{S} \xrightarrow{T} \mathcal{S}'.$$

Let $\mathcal{S}^1 = \mathcal{S}$ and $\mathcal{S}^2 = \mathcal{S}'$. For $j \geq 2$ and $\mathcal{S}^j \neq \mathcal{S}$ let $T_1^j, \dots, T_{m(j)}^j$, and \mathcal{S}^{j+1} be such that

$$\mathcal{S}^j \xrightarrow{T_1^j, \dots, T_{m(j)}^j} \mathcal{S}^{j+1}$$

and $\mathcal{S}^j \succeq_i S(i, \mathcal{S}^j)$ for all $i \in T_h^j$, for all $h = 1, \dots, m(j)$, and for all $\hat{\mathcal{S}}$ with $\mathcal{S}^{j+1} \rightarrow \hat{\mathcal{S}}$. If $\mathcal{S}^j = \mathcal{S}$ for some $j \geq 2$, let $\mathcal{S}^{j+1} = \mathcal{S}^2 = \mathcal{S}'$.

Since the core is empty and the set of coalition structures is finite, the path defined above necessarily ends up in a cycle. To simplify the notation let the coalition structures in this cycle again be numbered $\mathcal{S}^1, \mathcal{S}^2, \dots, \mathcal{S}^J$, with $J \geq 1$. Then

$$\mathcal{S}^j \xrightarrow{T_1^j, \dots, T_{m(j)}^j} \mathcal{S}^{j+1}$$

if $\mathcal{S}^j \neq \mathcal{S}$ and $\mathcal{S}^j \xrightarrow{T} \mathcal{S}^{j+1}$ if $\mathcal{S}^j = \mathcal{S}$. By Lemma 2.1 there exists $1 \leq j \leq J-1$ such that $\mathcal{S}^j = \mathcal{S}$. W.l.o.g. let $j = 1$. Since

$$\mathcal{S}^1 \xrightarrow{T} \mathcal{S}^2$$

it is true that $T \succ_i S(i, \mathcal{S}^1)$ for all $i \in T$. Moreover, since $T \notin \mathcal{S}^1$ there exists a maximal $j \in \{2, \dots, J-1\}$ such that $T \in \mathcal{S}^j$. Therefore, there exists $h \in \{1, \dots, m(j)\}$ and a player $i \in T_h \cap T$. By construction this implies $S(i, \mathcal{S}^1) \succeq_i S(i, \mathcal{S}^j) = T$, which is a contradiction. \square

Another way of stating this result is to say that there exist at least two durable coalition structures whenever the core is empty. One case in which there exists a unique durable coalition structure and hence the core and the set of durable coalition structures coincide is the one where the hedonic game satisfies the *top-coalition property* introduced by Banerjee et al. (2001).

Definition 3.1. Let $(N; (\succeq_i)_{i \in N})$ be a hedonic game and let $V \subset N$ be a nonempty set of

players. A nonempty subset $S \subset V$ is a *top-coalition* of V , if for any $i \in S$ and any coalition $T \subset V$ with $i \in T$, we have $S \succeq_i T$. The hedonic game satisfies the *top-coalition property* if for any nonempty set of players $V \subset N$, there exists a top-coalition of V .

Observe that if $(N; (\succeq_i)_{i \in N})$ is a hedonic game that satisfies the top-coalition property, then, for any $T \subset N$ the game $(T; (\succeq_i^T)_{i \in T})$ obtained by restricting the player set to T satisfies the top-coalition property as well. Banerjee et al. (2001) show that if a hedonic game satisfies the top-coalition property and if preferences are strict, then the core is nonempty and contains as its unique element the coalition structure \mathcal{S}^* defined as follows.⁸ For any $V \subset N$ let $\text{TOP}(V) = \{S \subset V \mid S \text{ is a top-coalition of } V\}$. Define $V_0 = N$ and let $\text{TOP}(V_0) = \{S_1^1, \dots, S_{m(1)}^1\}$. Define $V_1 = N \setminus \bigcup_{j=1}^{m(1)} S_j^1$ and let $\text{TOP}(V_1) = \{S_1^2, \dots, S_{m(2)}^2\}$. In the same way define $V_k = N \setminus \bigcup_{j=1}^{m(k)} S_j^k$ and let $\text{TOP}(V_k) = \{S_1^{k+1}, \dots, S_{m(k+1)}^{k+1}\}$ for each k . Since N is finite there exists K with $V_{K-1} \neq \emptyset$ and $V_K = \emptyset$. Let

$$\mathcal{S}^* = \{S_1^1, \dots, S_{m(1)}^1, S_1^2, \dots, S_{m(2)}^2, \dots, S_1^K, \dots, S_{m(K)}^K\}. \quad (1)$$

As the following theorem shows, \mathcal{S}^* is the unique durable coalition structure. Hence, in a case where there is a unique reasonable outcome of the coalition formation game the core and the notion of durability give the same prediction.

Theorem 3.2. *Suppose $(N; (\succeq_i)_{i \in N})$ is a hedonic game that satisfies the top-coalition property, and that preferences are strict. Then there exists a unique durable coalition structure and it is the unique coalition structure in the core.*

Proof. We prove the theorem by showing that a durable coalition structure must consist of the set of coalitions contained in \mathcal{S}^* as defined in expression (1). The proof is by induction over the number of players. Let $(N; (\succeq_i)_{i \in N})$ be a hedonic game that satisfies the top-coalition property. If $|N| = 1$, then obviously the single person coalition is the only element in the core and it is also the only durable coalition structure. Let $|N| = n$ and assume that the claim has been proved for all hedonic games with strictly less than n players. Let \mathcal{S} be a durable coalition structure in the hedonic game $(N; (\succeq_i)_{i \in N})$. If $S_j^1 \notin \mathcal{S}$ for some $j \in \{1, \dots, m(1)\}$, then, since preferences are strict, there exists \mathcal{S}^1 with

$$\mathcal{S} \xrightarrow{S_j^1} \mathcal{S}^1.$$

Since S_j^1 is a top-coalition of N , for no $i \in S_j^1$ there exists a strictly improving deviation. Moreover, by the induction hypothesis, for any $T \subsetneq N$ with $S_j^1 \subset T$ it is true that $S_j^1 \in \mathcal{S}_T^*$, where \mathcal{S}_T^* is the unique durable coalition structure in the hedonic game $(T; (\succeq_i^T)_{i \in T})$. Hence, $S_j^1 \in \mathcal{S}'$ for all \mathcal{S}' with $\mathcal{S}^1 \twoheadrightarrow \mathcal{S}'$. This implies that \mathcal{S} is transient,

⁸Pápai (2000) introduces the *single-lapping property* which also guarantees uniqueness of the core. Since this property implies the top-coalition property we do not treat it separately here.

which is a contradiction. Therefore, $S_j^1 \in \mathcal{S}$ for all $j = 1, \dots, m(1)$, and by the same argument as above $\bigcup_{h=1}^{m(1)} \{S_h^1\} \subset \mathcal{S}'$ for all \mathcal{S}' with $\mathcal{S} \twoheadrightarrow \mathcal{S}'$.

If $S_j^2 \notin \mathcal{S}$ for some $j \in \{1, \dots, m(2)\}$, then by the previous argument $S(i, \mathcal{S}) \subset N \setminus \bigcup_{h=1}^{m(1)} S_h^1$ for all $i \in S_j^2$. Hence, since preferences are strict and S_j^2 is a top-coalition of $N \setminus \bigcup_{h=1}^{m(1)} S_h^1$, there exists \mathcal{S}^2 with

$$\mathcal{S} \xrightarrow{S_j^2} \mathcal{S}^2.$$

We have shown above that $\bigcup_{h=1}^{m(1)} \{S_h^1\} \subset \mathcal{S}'$ for all \mathcal{S}' with $\mathcal{S} \twoheadrightarrow \mathcal{S}'$. Moreover, for no $i \in S_j^2$ there exists a strictly improving deviation involving only players in $N \setminus \bigcup_{h=1}^{m(1)} S_h^1$. Finally, from the induction hypothesis it follows that $S_j^2 \in \mathcal{S}_T^*$ for any T with $S_j^2 \subset T \subset N \setminus \bigcup_{h=1}^{m(1)} S_h^1$, where \mathcal{S}_T^* is the unique durable coalition structure in the hedonic game $(T; (\succeq_i)_{i \in T})$. Therefore, $S_j^2 \in \mathcal{S}'$ for all \mathcal{S}' with $\mathcal{S}^2 \twoheadrightarrow \mathcal{S}'$, from which it follows that \mathcal{S} is transient, which is a contradiction. Hence, $S_j^2 \in \mathcal{S}$ for all $j = 1, \dots, m(2)$, and the same argument shows that $\bigcup_{h=1}^{m(2)} \{S_h^2\} \subset \mathcal{S}'$ for all \mathcal{S}' with $\mathcal{S} \twoheadrightarrow \mathcal{S}'$.

In the same way, one proves that $S_j^k \in \mathcal{S}$ for all $j = 1, \dots, m(k)$, and all $k = 1, \dots, K$. Hence, $\mathcal{S} = \mathcal{S}^*$, which proves the theorem. \square

The following three examples consider matching problems which can be interpreted as hedonic games by defining the preference relations such that unfeasible coalitions are not individually rational. Stable matchings then correspond to coalition structures in the core. As Gale and Shapley (1962) have shown, the core is nonempty for any one-to-one matching problem. By contrast, there exist many-to-one matchings problems for which no stable matching exists. Our examples will illustrate that the notion of durability gives rise to reasonable predictions for both types of matching problems.

Example 3.3. This is an example of a one-to-one matching problem. Let $N = W \cup M$, where $W = \{w_1, w_2, w_3, w_4\}$ is the set of women and $M = \{m_1, m_2, m_3, m_4\}$ is the set of men. Each $w \in W$ has a strict preference ordering over $M \cup \{w\}$ which we represent by an ordered list $P(w)$:

$$P(w_1) = m_4 m_2 m_1 w_1,$$

$$P(w_2) = m_2 m_3 m_1 w_2,$$

$$P(w_3) = m_3 w_3,$$

$$P(w_4) = m_1 m_4 w_4.$$

The interpretation of, for example, $P(w_3)$ is that w_3 prefers m_3 over being alone, while all other men are strictly worse for w_3 than being alone. Since unacceptable mates do not play any role we omitted them from the list. Similarly, the strict preference ordering of each $m \in M$ over $W \cup \{m\}$ is represented by an ordered list $P(m)$:

$$P(m_1) = w_1 w_4 w_2 m_1,$$

$$P(m_2) = w_1 w_2 m_2,$$

$$P(m_3) = w_2 w_3 m_3,$$

$$P(m_4) = w_4 w_1 m_4.$$

As mentioned above it is straightforward to formulate this matching problem as a hedonic game if one defines the preference relations \succeq_i over $S_i(N)$ such that unfeasible coalitions, i.e. coalitions where more than one player belongs to either W or M , are not individually rational.

A matching μ is a one to-one function from $W \cup M$ onto itself such that $\mu^2(i) = i$ for all $i \in W \cup M$. Any matching defines a coalition structure on $W \cup M$ in an obvious way. A matching μ is *stable* if there is no $i \in W \cup M$ who prefers being unmatched over being matched with his mate at μ and if there is no $w \in W$ and $m \in M$ that are not matched to one another at μ but would each prefer one another to their mates at μ . It is immediate to see that a matching μ is stable if and only if the corresponding coalition structure is in the core of the hedonic game defined by the matching market.

In this example there is a unique stable matching which corresponds to the coalition structure $[w_1 m_4 \mid w_2 m_2 \mid w_3 m_3 \mid w_4 m_1]$. However, this is not the unique durable coalition structure. As Fig. 3 shows, the coalition structure $\mathcal{S} = [w_1 m_1 \mid w_2 m_2 \mid w_3 m_3 \mid w_4 m_4]$ is also durable since the only possible deviation is by coalition $\{w_1, m_2\}$ and it leads to

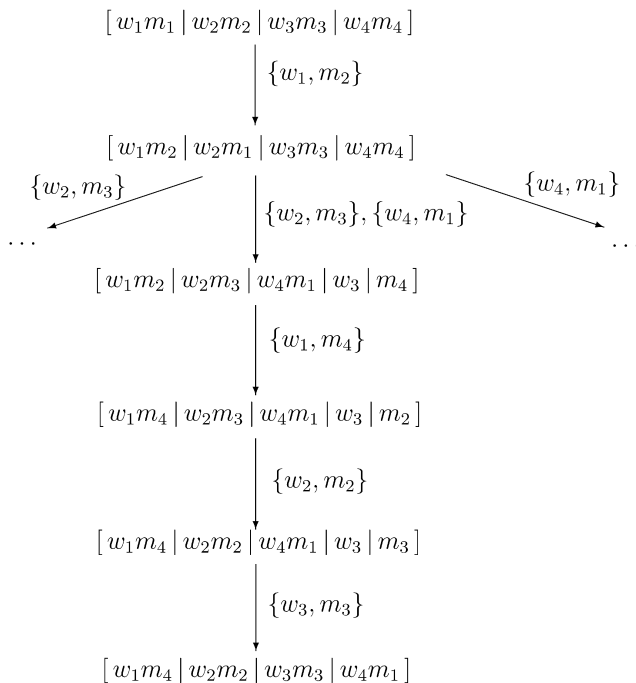


Fig. 3. Example 3.3.

$[w_1 m_4 \mid w_2 m_3 \mid w_4 m_1 \mid w_3 \mid m_2]$ which is strictly worse for m_2 than \mathcal{S} . Hence the core can be a strict subset of the set of durable coalition structures (cf. Example 3.2) also for the special case of a matching problem.

Example 3.4. The following example of a many-to-one-matching problem for which there exists no stable matching is taken from Dutta and Massó (1997). Let there be two firms F_1, F_2 , and four workers w_1, w_2, w_3, w_4 , so that $N = \mathcal{F} \cup \mathcal{W}$ where $\mathcal{F} = \{F_1, F_2\}$ and $\mathcal{W} = \{w_1, w_2, w_3, w_4\}$. Each firm F_j has a preference ordering over $\{F_j\} \cup \{W \mid \emptyset \neq W \subset \mathcal{W}\}$ and each worker w_i has a preference ordering over $(\mathcal{F} \times \mathcal{W}_i) \cup \{w_i\}$, where $\mathcal{W}_i = \{S \mid S \subset \mathcal{W}, w_i \in S\}$.

In this example we assume that workers' preferences are *worker-lexicographic*, i.e. w_i 's ranking over colleagues determines w_i 's preference ordering over firm-co-worker pairs in which the set of co-workers is distinct. Hence, for our purposes it is enough to list workers' preferences over colleagues only. The preferences of firms and workers are given in Table 1, where elements are ranked in descending order of preference and again only acceptable partners are listed.

Again one can represent this matching market as a hedonic game so that the stable matchings correspond to coalition structures in the core. As Dutta and Massó (1997) argue, the only candidates for stable matchings are

$$[F_1 w_3 w_4 \mid F_2 w_1 w_2],$$

$$[F_1 w_2 w_3 \mid F_2 w_1 w_4],$$

$$[F_1 w_2 w_3 w_4 \mid F_2 w_1].$$

However, all three matchings are unstable. More precisely, we obtain the cycle in Fig. 4.

It is straightforward to see that all coalition structures in this cycle are durable since all deviations are deterred. For example, $[F_1 w_3 w_4 \mid F_2 w_1 w_2]$ is durable since w_2 is not willing to join F_1, w_3, w_4 and deviate to $[F_1 w_2 w_3 w_4 \mid F_2 w_1]$ knowing that this may lead to $[F_1 w_2 w_3 \mid F_2 w_1 w_4]$ which is worse for w_2 than $[F_1 w_3 w_4 \mid F_2 w_1 w_2]$.

Table 1
Example 3.4

| F_1 | F_2 | w_1 | w_2 | w_3 | w_4 |
|---------------------|----------------|----------------|---------------------|---------------------|---------------------|
| $\{w_2, w_3, w_4\}$ | $\{w_1, w_2\}$ | $\{w_1, w_2\}$ | $\{w_1, w_2, w_3\}$ | $\{w_1, w_2, w_3\}$ | $\{w_1, w_4\}$ |
| $\{w_2, w_3\}$ | $\{w_1, w_4\}$ | $\{w_1, w_3\}$ | $\{w_1, w_2, w_4\}$ | $\{w_1, w_3, w_4\}$ | $\{w_2, w_4\}$ |
| $\{w_3, w_4\}$ | $\{w_1\}$ | $\{w_1, w_4\}$ | $\{w_2, w_3, w_4\}$ | $\{w_2, w_3, w_4\}$ | $\{w_1, w_3, w_4\}$ |
| $\{w_2, w_4\}$ | $\{w_2\}$ | $\{w_1\}$ | $\{w_1, w_2\}$ | $\{w_1, w_3\}$ | $\{w_2, w_3, w_4\}$ |
| $\{w_1, w_4\}$ | $\{w_3\}$ | | $\{w_2, w_3\}$ | $\{w_2, w_3\}$ | $\{w_3, w_4\}$ |
| $\{w_2\}$ | $\{w_4\}$ | | $\{w_2, w_4\}$ | $\{w_3, w_4\}$ | $\{w_4\}$ |
| $\{w_4\}$ | | | $\{w_2\}$ | $\{w_3\}$ | |
| $\{w_3\}$ | | | | | |
| $\{w_1\}$ | | | | | |

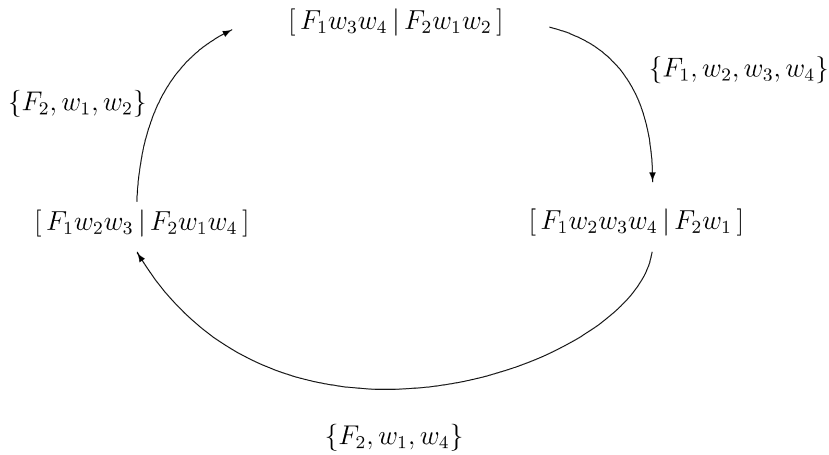


Fig. 4. Example 3.4.

Example 3.5. The final example is again a many-to-one matching problem with firms and workers. As before there are two firms F_1, F_2 and four workers w_1, w_2, w_3, w_4 , of whom (w_1, w_2) is a couple and w_3 and w_4 are singles. The firms are located in towns far away from each other, so the couple strictly prefers being unemployed over working for different firms. Firms' and workers' preferences are given in Table 2, where we again only list individually rational coalitions in descending order of preference.

Each firm has a most preferred worker and there are no returns to employ an additional worker once this worker has been appointed. There is a unique stable matching which corresponds to the coalition structure $[F_1w_3 | F_2w_4 | w_1 | w_2]$ in which the couple is unemployed. According to Theorem 2.2 and Corollary 2.2 there are only five candidates for durable coalition structures. In Fig. 5 we present the graph on these coalition structures that is defined by the relation 'directly leading.' It is straightforward to see that, apart from the core structure, there exists one additional durable coalition structure in which the couple is matched to F_1 and the singles are matched to F_2 . Clearly, the couple will prefer the suboptimal firm F_1 and refuse to change to F_2 when this leads to a matching where they are both unemployed. The full employment in the durable coalition structure $[F_1w_1w_2 | F_2w_3w_4]$ seems to be a much more appealing outcome of the matching problem than the stable matching.

Table 2
Example 3.5

| F_1 | F_2 | $w_i (i = 1, 2)$ | w_3 | w_4 |
|---------------------|---------------------|---------------------|---------------------|---------------------|
| $\{F_1, w_3\}$ | $\{F_2, w_4\}$ | $\{F_2, w_1, w_2\}$ | $\{F_2, w_3, w_4\}$ | $\{F_2, w_3, w_4\}$ |
| $\{F_1, w_1, w_2\}$ | $\{F_2, w_1, w_2\}$ | $\{F_1, w_1, w_2\}$ | $\{F_1, w_3\}$ | $\{F_1, w_3, w_4\}$ |
| $\{F_1, w_3, w_4\}$ | $\{F_2, w_3, w_4\}$ | $\{w_i\}$ | $\{F_1, w_3, w_4\}$ | $\{F_2, w_4\}$ |
| $\{F_1\}$ | $\{F_2\}$ | | $\{w_3\}$ | $\{w_4\}$ |

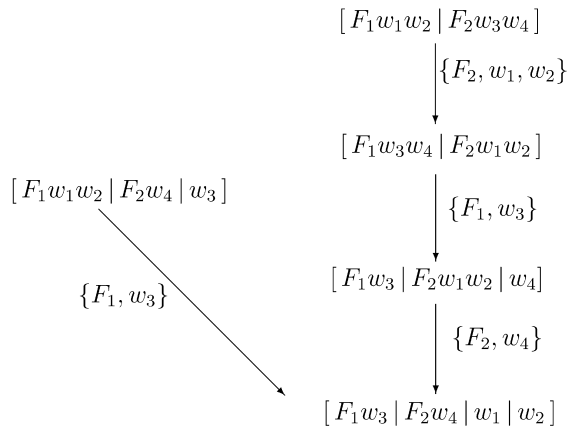


Fig. 5. Example 3.5.

4. Conclusion

We have proposed a solution concept for hedonic coalition formation games which assumes some form of maxmin behavior on the part of players. Players are farsighted in that they take into account that their initial deviation may trigger further deviations and they are extremely pessimistic in that they attribute to each coalition structure a positive probability that it may last forever. Our definition of durability provides an implicit theory about which coalition structures may arise following a coalitional deviation. Our assumption about the rearrangement of members of disrupted coalitions is consistent with our theory and gives rise to a recursive definition of our solution concept.

We have not concealed the shortcomings of our solution concept which are mainly due to the asymmetry in the degree of foresight and rationality that players attribute to others and to themselves. But the proposal has many attractive features. It always selects a nonempty set of coalition structures as being durable, it satisfies a very strong consistency notion and it does well in simple applications. It always selects coalition structures in the core when they exist. It chooses core allocations alone in simple cases where this choice is very compelling. In other simple examples it also selects as durable some additional coalition structures not in the core, and when this happens the retained structures always turn out to be plausible. We hope to have provided a sensible and well grounded tool for the analysis of possible outcomes of coalition formation games. The challenge of computing our solution (or any other) for large, complex games, still lies ahead.

Acknowledgements

We are grateful to the participants of the 1999 Manresa Conference on Networks, Groups and Coalitions for their comments, and very specially to our discussant there,

Debraj Ray. We also wish to thank an anonymous referee for useful suggestions. This research was started through the stimulating influence of Michael Maschler.

References

- Alcalde, J., Revilla, P., 2001. Tops responsiveness, strategy-proofness and coalition formation problems. Mimeo.
- Banerjee, S., Konishi, H., Sönmez, T., 2001. Core in a simple coalition formation game. *Social Choice and Welfare* 18, 135–153.
- Bennett, E., Zame, W.R., 1988. Bargaining in cooperative games. *International Journal of Game Theory* 17, 279–300.
- Bernheim, D., Peleg, B., Whinston, M., 1987. Coalition-proof Nash equilibria. I. Concepts. *Journal of Economic Theory* 42, 1–12.
- Bhattacharya, A., 2002. Coalitional stability with a credibility constraint. *Mathematical Social Sciences* 43, 27–44.
- Bogomolnaia, A., Jackson, M.O., 2002. The stability of hedonic coalition structures. *Games and Economic Behavior* 38, 201–230.
- Chwe, M.S.-Y., 1994. Farsighted coalitional stability. *Journal of Economic Theory* 63, 299–325.
- Demange, G., 1994. Intermediate preferences and stable coalition structures. *Journal of Mathematical Economics* 23, 45–58.
- Diamantoudi, E., Xue, L., 2000. Farsighted stability in hedonic games. Working Paper No. 2000-12, University of Aarhus, Denmark.
- Drèze, J., Greenberg, J., 1980. Hedonic coalitions: optimality and stability. *Econometrica* 48, 987–1003.
- Dutta, B., Massó, J., 1997. Stability of matchings when individuals have preferences over colleagues. *Journal of Economic Theory* 75, 464–475.
- Dutta, B., Ray, D., Sengupta, K., Vohra, R., 1989. A consistent bargaining set. *Journal of Economic Theory* 49, 93–112.
- Gale, D., Shapley, L.S., 1962. College admissions and the stability of marriage. *American Mathematical Monthly* 69, 9–15.
- Gerber, A., 2000. Coalition formation in general NTU games. *Review of Economic Design* 5, 149–175.
- Greenberg, J., 1990. *The Theory of Social Situations*. Cambridge University Press, Cambridge.
- Greenberg, J., Weber, S., 1986. Strong tiebout equilibrium under restricted preferences domain. *Journal of Economic Theory* 38, 101–117.
- Greenberg, J., Weber, S., 1993. Stable coalition structures with a unidimensional set of alternatives. *Journal of Economic Theory* 60, 62–82.
- Guesnerie, R., Oddou, C., 1981. Second best taxation as a game. *Journal of Economic Theory* 25, 67–91.
- Hart, S., Kurz, M., 1983. Endogenous formation of coalitions. *Econometrica* 51, 1047–1064.
- Konishi, H., Ray, D., 1999. Coalitional deviations as a dynamic process. Mimeo.
- Martin, M., Merlin, V., 2002. The stability set as a social choice correspondence. *Mathematical Social Sciences* 44, 91–113.
- Pápai, S., 2000. Unique stability in simple coalition formation games. Mimeo.
- Ray, D., 1989. Credible coalitions and the core. *International Journal of Game Theory* 18, 185–187.
- Ray, D., Vohra, R., 1997. Equilibrium binding agreements. *Journal of Economic Theory* 73, 30–78.
- Rubinstein, A., 1980. Stability of decision systems under majority rule. *Journal of Economic Theory* 23, 150–159.
- Shenoy, P.P., 1979. On coalition formation: a game theoretical approach. *International Journal of Game Theory* 8, 133–164.
- von Neumann, J., Morgenstern, O., 1944. *Theory of Games and Economic Behavior*, 3rd Edition, 1953. Princeton University Press, Princeton.
- Xue, L., 1998. Coalitional stability under perfect foresight. *Economic Theory* 11, 603–627.
- Zhou, L., 1994. A new bargaining set of an N -person game and endogenous coalition formation. *Games and Economic Behavior* 6, 512–526.