Pareto Optimality in Coalition Formation

Haris Aziz, Felix Brandt*, Paul Harrenstein

Institut für Informatik, Technische Universität München, 80538 München, Germany

Abstract

A minimal requirement on allocative efficiency in the social sciences is Pareto optimality. In this paper, we identify a close structural connection between Pareto optimality and perfection that has various algorithmic consequences for coalition formation. Based on this insight, we formulate the *Preference Refinement Algorithm (PRA)* which computes an individually rational and Pareto optimal outcome in hedonic coalition formation games or any other discrete allocation setting. Our approach also leads to various results for specific classes of hedonic games. In particular, we show that computing and verifying Pareto optimal partitions in general hedonic games, anonymous games, three-cyclic games, room-roommate games and B-hedonic games is intractable while both problems are tractable for roommate games, W-hedonic games, and house allocation with existing tenants.

Keywords: Coalition Formation, Hedonic Games, Pareto Optimality, Computational Complexity
JEL: C63, C70, C71, and C78

1. Introduction

Ever since the publication of von Neumann and Morgenstern's *Theory of Games and Economic Behavior* in 1944, coalitions have played a central role within game theory. The crucial questions in coalitional game theory are which coalitions can be expected to form and how the members of coalitions should divide the proceeds of their cooperation. Traditionally the focus has been on the latter issue, which led to the formulation and analysis of concepts such as the core, the Shapley value, or the bargaining set. Which coalitions are likely to form is commonly assumed to be settled exogenously, either by explicitly specifying the coalition structure, a partition of the players in disjoint coalitions, or, implicitly, by assuming that larger coalitions can invariably guarantee better outcomes to its members than smaller ones and that, as a consequence, the grand

*Corresponding author
Email addresses: `aziz@in.tum.de` (Haris Aziz), `brandtf@in.tum.de` (Felix Brandt), `harrenst@in.tum.de` (Paul Harrenstein)

coalition of all players will eventually form. The two questions, however, are clearly interdependent: the individual players' payoffs depend on the coalitions that form just as much as the formation of coalitions depends on how the payoffs are distributed.

Coalition formation games, which were first formalized by Drèze and Greenberg [15], model coalition formation in settings in which utility is *non-transferable*. In many such situations it is natural to assume that a player's appreciation of a coalition structure only depends on the coalition he is a member of and not on how the remaining players are grouped. Initiated by Banerjee et al. [7] and Bogomolnaia and Jackson [9], much of the work on coalition formation now concentrates on these so-called *hedonic games*. In this paper, we focus on Pareto optimality and individual rationality in this rich class of coalition formation games.

The main question in coalition formation games is which coalitions one may reasonably expect to form. To get a proper formal grasp of this issue, a number of stability concepts have been proposed for hedonic games—such as the core or Nash stability—and much research concentrates on conditions for existence, the structure, and computation of stable and efficient partitions. *Pareto optimality*—which holds if no coalition structure is strictly better for some player without being strictly worse for another—and *individual rationality*—which holds if every player is satisfied in the sense that no player would rather be on his own—are commonly considered minimal requirements for any reasonable partition.[1]

Another reason to investigate Pareto optimal partitions algorithmically is that, in contrast to other stability concepts like the core, they are guaranteed to exist. This even holds if we additionally require individual rationality. Moreover, even though the *Gale-Shapley algorithm* returns a core stable matching for marriage games, it is already NP-hard to check whether the core is empty in almost any generalization, such as roommate games [24], general hedonic games [6], and games with $\mathscr{B}$- and $\mathscr{W}$-preferences [12, 13]. Interestingly, when the status-quo partition cannot be changed without the mutual consent of all players, Pareto optimality can be seen as a stability notion [22].

When, there are indifferences in the preferences, a core stable outcome is not necessarily Pareto optimal. Thus, Pareto optimality can serve as a refinement of core stable outcomes. An outcome which is Pareto optimal and a Pareto improvement over a core stable outcome is called *Pareto-stable*. This notion further motivates the need for algorithms to compute Pareto improvements of given outcomes. Sotomayor and Özak [29] note that *"the study of the discrete two-sided matching models with non-necessarily strict preferences and the search for algorithms to produce the Pareto-stable matchings is a new and interesting line of investigation."*

---

[1]For example, in the context of TU coalitional games, Aumann [5] states that "the requirement that a feasible outcome y be undominated via one-person coalitions (individual rationality) and via the all-person coalition (efficiency or Pareto optimality) is thus quite compelling". His point can easily be seen to extend to hedonic games as well.

We investigate both the problem of finding a Pareto optimal and individually rational partition and the problem of deciding whether a given partition is Pareto optimal. In particular, our results concern *general hedonic games*, *B-hedonic* and *W-hedonic games* (two classes of games in which each player's preferences over coalitions are based on his most preferred and least preferred player in his coalition, respectively), *roommate games*, *house allocation with existing tenants*, *three-cyclic games*, *room-roommate games*, and *anonymous games*.

Many of our results, both positive and negative, rely on the concept of *perfection* and how it relates to Pareto optimality. A *perfect* partition is one that is most desirable for every player. We find (a) that under extremely mild conditions, NP-hardness of finding a perfect partition implies NP-hardness of finding a Pareto optimal partition (Lemma 1), and (b) that under stronger but equally well-specified circumstances, feasibility of finding a perfect partition implies feasibility of finding a Pareto optimal partition (Lemma 2). The latter we show via a Turing reduction to the problem of computing a perfect partition. At the heart of this algorithm, which we refer to as the *Preference Refinement Algorithm (PRA)*, lies a fundamental insight of how perfection and Pareto optimality are related. It turns out that a partition is Pareto optimal for a particular preference profile if and only if the partition is perfect for another but related profile (Theorem 1). In this way PRA is also applicable to any other discrete allocation setting.

For general allocation problems, *serial dictatorship*—which subsequently chooses the most preferred allocation for a player given a fixed ranking of all players—is well-established as a procedure for finding Pareto optimal solutions [see, e.g., 1]. However, it is only guaranteed to do so if the players' preferences over outcomes are strict, which is not feasible in many compact representations. Moreover, when applied to coalition formation games, there can be Pareto optimal partitions that serial dictatorship is unable to find, which may have serious repercussions if also other considerations, like fairness, are taken into account. By contrast, PRA handles weak preferences well and is complete in the sense that it may return any Pareto optimal partition, provided that the subroutine that computes perfect partitions can compute any perfect partition (Theorem 3).

2. Preliminaries

In this section, we review the terminology and notation used in this paper.

Hedonic games. Let N be a set of n players. A *coalition* is any non-empty subset of N. By $\mathscr{N}_i$ we denote the set of coalitions player i belongs to, i.e., $\mathscr{N}_i = \{S \subseteq N : i \in S\}$. A *coalition structure*, or simply a *partition*, is a partition π of the players N into coalitions, where $\pi(i)$ is the coalition player i belongs to.

A *hedonic game* is a pair (N, R), where $R = (R_1, \dots, R_n)$ is a *preference profile* specifying the preferences of each player i as a binary, complete, reflexive, and transitive *preference relation* R_i over $\mathscr{N}_i$. If R_i is also anti-symmetric we say that i's preferences are *strict*. We adopt the conventions of social choice theory by writing $S\,P_i\,T$ if $S\,R_i\,T$ but not $T\,R_i\,S$—i.e., if i *strictly prefers* S to T—and $S\,I_i\,T$ if both $S\,R_i\,T$ and $T\,R_i\,S$—i.e., if i is *indifferent* between S and T.

For a player i, a coalition S in $\mathscr{N}_i$ is *acceptable* if for i being in S is at least preferable as being alone—i.e., if $S\,R_i\,\{i\}$—and *unacceptable* otherwise.

In a similar fashion, for X a subset of $\mathscr{N}_i$, a coalition S in X is said to be *most preferred in X by i* if $S\,R_i\,T$ for all $T \in X$ and *least preferred in X by i* if $T\,R_i\,S$ for all $T \in X$. In case $X = \mathscr{N}_i$ we generally omit the reference to X. The sets of most and least preferred coalitions in X by i, we denote by $\max_{R_i}(X)$ and $\min_{R_i}(X)$, respectively.

In hedonic games, players are only interested in the coalition they are in. Accordingly, preferences over coalitions naturally extend to preferences over partitions and we write $\pi\,R_i\,\pi'$ if $\pi(i)\,R_i\,\pi'(i)$. We also say that partition π is *acceptable* or *unacceptable* to a player i according to whether $\pi(i)$ is acceptable or unacceptable to i, respectively. Moreover, π is *individually rational* if π is acceptable to all players. If there is a pre-existing individually rational partition π^*, then a mechanism returning a partition π is *individually rational* if $\pi(i)\,R_i\,\pi^*(i)$ for all $i \in N$.

Given a preference profile R, partition π *Pareto dominates* another partition π' if $\pi\,R_j\,\pi'$ for all players j and $\pi'\,P_i\,\pi$ for at least one player i. A partition π is *Pareto optimal for R* if there is no partition π' that Pareto dominates it given R. Partition π is, moreover, said to be *weakly Pareto optimal for R_i* if there is no π' with $\pi'\,P_i\,\pi$ for all players i.

Classes of hedonic games. The number of potential coalitions grows exponentially in the number of players. In this sense, hedonic games are relatively large objects and for algorithmic purposes it is often useful to look at classes of games that allow for concise representations (see Figure 1 for a schematic overview of the logical interrelationships between the different classes).

For *general hedonic games*, we will assume that each player expresses his preferences only over his acceptable coalitions. This representation is also known as *Representation by Individually Rational Lists of Coalitions* (*RIRLC*) [6].

We now describe classes of hedonic games in which the players' preferences over coalitions are induced by their preferences over the other players. For R_i such preferences of player i over players, we say that a player j is *acceptable* to i if $j\,R_i\,i$ and *unacceptable* otherwise. Any coalition containing an unacceptable player is unacceptable to player i.

B-hedonic and W-hedonic games. For a subset J of players, we denote by $\max_{R_i}(J)$ and $\min_{R_i}(J)$ the sets of the most and least preferred players in J by i, respectively. We will assume that $\max_{R_i}(\emptyset) = \min_{R_i}(\emptyset) = \{i\}$. In a *B-hedonic game* the preferences R_i of a player i over players extend to preferences over coalitions in such a way that, for all coalitions S and T in $\mathscr{N}_i$, we have $S\,R_i\,T$ if and only if either some player in T is unacceptable to i or all players in S are acceptable to i and $j\,R_i\,k$ for all $j \in \max_{R_i}(S \setminus \{i\})$ and $k \in \max_{R_i}(T \setminus \{i\})$. Analogously, in a *W-hedonic game* (N, R), we have $S\,R_i\,T$ if and only if either some player in T is unacceptable to i or $j \, R_i \, k$ for all $j \in \min_{R_i}(S \setminus \{i\})$ and $k \in \min_{R_i}(T \setminus \{i\})$.[2]

Roommate games. The class of *roommate games*, which are well-known from the literature on matching theory, can be defined as those hedonic games in which only coalitions of size one or two are acceptable and preferences R_i over other players are extended naturally over preferences over coalitions in the following way: $\{i, j\} \, R_i \{i, k\}$ if and only if $j \, R_i \, k$ for all $j, k \in N$.

Marriage games. A *marriage game* is a roommate game in which the set N of players can be partitioned into two sets *male* and *female* and a player finds a member of the same sex unacceptable. Moreover, marriage games can also be seen as B-hedonic or W-hedonic games in which for each player all other players of the same sex are unacceptable.

Anonymous hedonic games. Anonymous games are a subclass of hedonic games in which a player's preferences over coalitions only depends on the coalition sizes.

Three-cyclic games. A *three-cyclic game* is a hedonic game in which the set of players is divided into men, women, and dogs and only kind of acceptable coalitions are man-woman-dog triplets. Furthermore, men only care about women, women only care about dogs and dogs only care about men.

Room-roommate games. Room-rooommate games are hedonic games in which the set of players is partitioned into a set A of rooms and a set T of tenants, the rooms are completely indifferent among all outcomes, and for each tenant $i \in T$ the only coalitions other than $\{i\}$ that are acceptable are among $\{\{i, j, r\} : i, j \in T \text{ and } r \in A\}$.

Exchange economies of discrete goods. An exchange economy of discrete goods consists of agents and discrete goods such that agents have complete preferences over bundles of goods. Special cases include *house allocation* and *housing markets* in which each agent can be allocated at most one good. Exchange economies of discrete goods are special cases of the general hedonic games in which the goods can be treated as agents which are completely indifferent between outcomes.

Computational Complexity. We will assume familiarity with fundamental concepts in computational complexity: exponential time, polynomial time, polynomial-time reductions, worst-case time complexity analysis of an algorithm, NP-, and coNP-completeness. For an accessible overview of these concepts addressed at economists, we refer the reader to the excellent introduction by Roughgarden [26]. Another short description of these concepts can be found in Section 1.1 of [6].

[2] W-hedonic games are equivalent to hedonic games with $\mathscr{W}$-preferences if individually rational outcomes are assumed. Unlike hedonic games with $\mathscr{B}$-preferences, B-hedonic games are defined in analogy to W-hedonic games and the preferences are not based on coalition sizes [cf. 12].

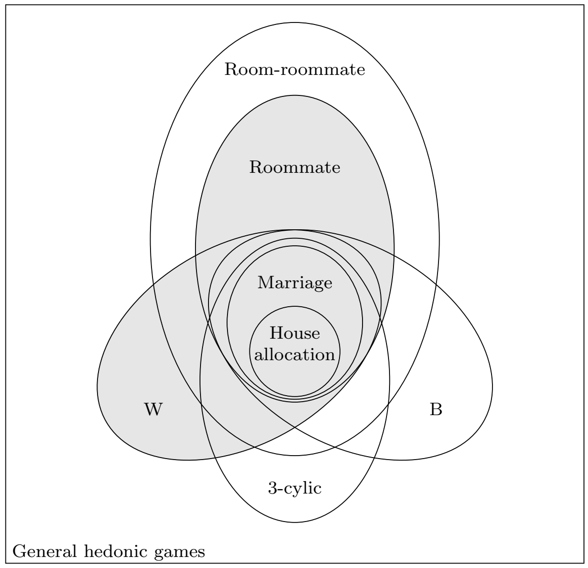

Figure 1: Classes of hedonic games: the classes within the grey area admit polynomial-time algorithms to compute and verify Pareto optimal partitions.

3. Perfection and Pareto Optimality

Pareto optimality constitutes a rather minimal efficiency requirement on partitions. A much stronger condition is that of *perfection*. We say that a partition π is *perfect* if $\pi(i)$ is a most preferred coalition for all players i. Thus, every perfect partition is Pareto optimal but not necessarily the other way round. Perfect partitions are obviously very desirable, but, in contrast to Pareto optimal ones, they are unfortunately not guaranteed to exist. Nevertheless, there exists a strong structural connection between the two concepts, which we exploit in our algorithm for finding Pareto optimal partitions in Section 4.

The problem of finding a perfect partition (`PerfectPartition`) we formally specify as follows.

`PerfectPartition`
Instance: A preference profile R
Question: Find a perfect partition for R.
If no perfect partition exists, output $\emptyset$.

We will later see that the complexity of `PerfectPartition` depends on the specific class of hedonic games that is being considered. By contrast, the related problem of *checking* whether a partition is perfect is an almost trivial problem for virtually all reasonable classes of games. If perfect partitions exist, they clearly coincide with the Pareto optimal ones. Hence, an oracle to compute a Pareto optimal partition can be used to solve `PerfectPartition`. If this Pareto optimal partition is perfect we are done, if it is not, no perfect partitions exist.

Thus, we obtain the following lemma, which we will invoke in our hardness proofs for computing Pareto optimal partitions.

Lemma 1. *For every class of hedonic games for which it can be checked in polynomial time whether a given partition is perfect, NP-hardness of* `PerfectPartition` *implies NP-hardness of computing a Pareto optimal partition.*

It might be less obvious that a procedure solving `PerfectPartition` can also be deployed as an oracle for an algorithm to compute Pareto optimal partitions. To do so, we first give a characterization of Pareto optimal partitions in terms of perfect partitions, which forms the mathematical heart of the Preference Refinement Algorithm to be presented in the next section.

The connection between perfection and Pareto optimality can intuitively be explained as follows. If all players are indifferent among all coalitions, every partition is perfect. It follows that the players can always relax their preferences up to a point where perfect partitions are possible. We find that, if a partition is perfect for a minimally relaxed preference profile—in the sense that, if any one player relaxes his preferences only slightly less, no perfect partition is possible anymore—, this partition is Pareto optimal for the original unrelaxed preference profile. To see this, assume π is perfect in some minimally relaxed preference profile and that some player i reasserts some strict preferences he had previously relaxed, thus rendering π no longer perfect. Now, π does not rank among i's most preferred partitions anymore. By assumption, none of i's most preferred partitions is also most preferred by all other players. Hence, it is impossible to find a partition π' that is better for i than π, without some other player strictly preferring π to π'. It follows that π is Pareto optimal.

To make this argumentation precise, we introduce the concept of a coarsening of a preference profile and the lattices these coarsenings define. Let $R = (R_1, \ldots, R_n)$ and $R' = (R'_1, \ldots, R'_n)$ be preference profiles over a set X and let i be a player. We write $R_i \leq_i R'_i$ if

$$R_i|_{\{x,y\}} = R'_i|_{\{x,y\}} \text{ for all } x \in X \text{ and all } y \in X \setminus \max_{R_i}(X).$$

Accordingly, R_i is exactly like R'_i, except that in R'_i player i may have strict preferences among some of his most preferred coalitions in R_i. Thus, R'_i is finer than R_i. It can easily be established that $\leq_i$ is a linear order for each player i.

We say that a preference profile $R = (R_1, \ldots, R_n)$ is a *coarsening of* or *coarsens* another preference profile $R' = (R'_1, \ldots, R'_n)$ whenever $R_i \leq_i R'_i$ for every player i. In that case we also say that R' *refines* R and write $R \leq R'$. Moreover, we write $R<R'$ if $R\leq R'$ but not $R'\leq R$. Thus, if R' refines R, i.e., if $R \leq R'$, then for each i and all coalitions S and T we have that $S\, R'_i\, T$ implies $S\, R_i\, T$, but not necessarily the other way round. It is worth observing that, if a partition is perfect for some preference profile R, then it is also perfect for any coarsening of R. The same holds for Pareto optimal partitions.

For preference profiles R and R' with $R \leq R'$, let $[R, R']$ denote the set $\{R'' : R \leq R'' \leq R'\}$, i.e., the set of all coarsenings of R' that also refine R.

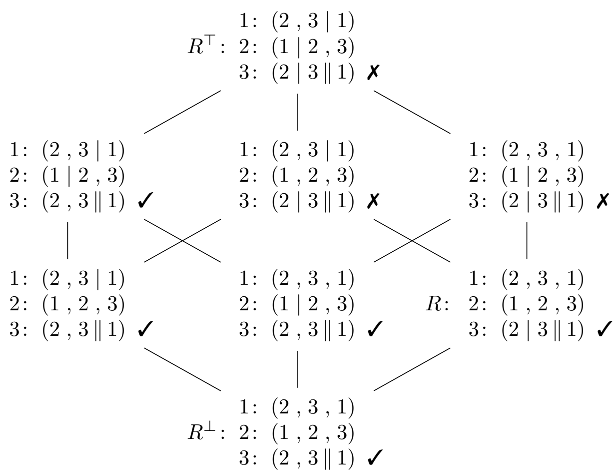

Figure 2: Lattice $([R^\top, R^\perp], \leq)$ for the W-hedonic games $(N, R^\top)$ and $(N, R^\perp)$ where $N = \{1, 2, 3\}$. For better readability, we denote indifference I_i by a comma, strict preferences P_i by a vertical bar, and unacceptability by a double bar. Thus, e.g., $R_3^\top$ represents the W-preferences of player 3 over coalitions such that $\{2,3\}\, P_3^\top \{3\}\, P_3^\top \{1,2,3\}\, I_3^\top \{1,3\}$. A checkmark indicates that a perfect partition exists for the respective preference profile, a cross that this is not the case. Thus, we find that partition $\{\{1\},\{2,3\}\}$ is perfect for preference profile R. Moreover, for preference profiles R' with $R < R' \leq R^\top$, no perfect partitions exist. By virtue of Theorem 1, we may conclude that $\{\{1\},\{2,3\}\}$ is Pareto optimal for $R^\top$.

$([R, R'], \leq)$ is a complete lattice with R and R' as bottom and top element, respectively (see Figure 2 for an example). We say that R *covers* R' if R is a minimal refinement of R' with $R' \neq R$, i.e., if $R' < R$ and there is no R'' such that $R' < R'' < R$. Observe that, if R covers R', R and R' coincide for all but one player, say i, for whom R_i is the unique minimal refinement of R'_i such that $R'_i \neq R_i$. We also denote R_i by $\texttt{Cover}(R'_i)$.

We are now in a position to prove the following theorem, which characterizes Pareto optimal partitions for a preference profile R as those that are perfect for coarsenings R' of R such that for no preference profile R'' with $R' < R'' \leq R$ perfect partitions exist. Figure 2 provides an example that illustrates Theorem 1.

Theorem 1. *Let $(N, R^\top)$ and $(N, R^\perp)$ be hedonic games such that $R^\perp \leq R^\top$ and π a perfect partition for $R^\perp$. Then, π is Pareto optimal for $R^\top$ if and only if there is some $R \in [R^\perp, R^\top]$ such that*

(*i*) *π is perfect for R, and*

(*ii*) *no partition is perfect for any R' with $R < R' \leq R^\top$.*

Proof. For the if-direction, assume there is some $R \in [R^\perp, R^\top]$ such that π is perfect for R and there is no perfect partition for any $R' \in [R^\perp, R^\top]$ with $R \leq R'$. For contradiction, also assume π is not Pareto optimal for $R^\top$. Then, there is some π' such that $\pi' \, R_j^\top \, \pi$ for all j and $\pi' \, P_i^\top \, \pi$ for some i. By $R \leq R^\top$ and π being perfect for R, it follows that π' is a perfect partition for R as well. Hence, $\pi' \, I_i \, \pi$. Let $R'_i = R_i^\top \cup \{(X,Y) : X \, R_i^\top \, \pi' \text{ and } Y \, R_i^\top \, \pi'\}$ and $R' = (R_1, \dots, R_{i-1}, R'_i, R_{i+1}, \dots, R_n)$. Observe that π' is perfect for R' but, as $\pi' \, P'_i \, \pi$, π is not. It can moreover easily be checked that $R < R' \leq R^\top$, and a contradiction follows.

For the only-if direction assume that π is Pareto optimal for $R^\top$. Let R be the finest coarsening of $R^\top$ in $[R^\top, R^\perp]$ for which π is perfect. Observe that $R = (R_1, \dots, R_n)$ can be defined such that $R_i = R_i^\top \cup \{(X,Y) : X \, R_i^\top \, \pi \text{ and } Y \, R_i^\top \, \pi\}$ for all i. Since π is perfect for $R^\perp$, we have $R^\perp \leq R$. If $R = R^\top$, we are done immediately. Otherwise, consider an arbitrary $R' \in [R^\perp, R^\top]$ with $R < R'$ and assume for contradiction that some perfect partition π' exists for R'. Then, in particular, $\pi' \, R'_j \, \pi$ for all j. As $R \leq R'$, it follows that π' is also perfect for R and, by choice of R as the *finest* coarsening of $R^\top$ for which π is perfect, $\pi' \, R_j^\top \, \pi$ for no j, i.e., $\pi \, R_j^\top \, \pi'$ for all j. Also by choice of π and $R < R'$, however, π is not perfect for R'. Hence, $\pi' \, P'_i \, \pi$ and thus $\pi' \, P_i^\top \, \pi$ for some i. This, however, contradicts that π is Pareto optimal for $R^\top$. $\square$

4. The Preference Refinement Algorithm (PRA)

In this section, we present the *Preference Refinement Algorithm (PRA)*, a general algorithm to compute Pareto optimal and individually rational partitions. We prove that the algorithm satisfies a number of desirable properties and show compare it to *serial dictatorship*.

4.1. Formal Description of the Preference Refinement Algorithm

The Preference Refinement Algorithm (PRA) invokes an oracle solving the problem `PerfectPartition` and is based on the formal connection between Pareto optimality and perfection made explicit in Theorem 1.

The idea underlying the algorithm is as follows. To compute a Pareto optimal and individually rational partition for a hedonic game (N, R), first find that coarsening R' of R in which each player is indifferent among all his acceptable coalitions and his preferences among unacceptable coalitions are as in R. In this coarsening, a perfect and individually rational partition is guaranteed to exist. Then, we search the lattice $([R', R], \leq)$ for a preference profile that allows for a perfect partition but none of the profiles refining it do. By virtue of Theorem 1, every perfect partition for such a preference profile will be a Pareto optimal partition for R. By only refining the preferences of one player at a time, we can use divide-and-conquer to conduct the search, thus achieving efficiency for a number of classes of hedonic games. By employing different search techniques, other desirable properties—such as *completeness*, *egalitarian fairness*, or *strategyproofness*—can be attained.

A formal specification of PRA is given in Algorithm 1, where `Choose`$(\{j \in N : Q_j^\perp \neq Q_j^\top\})$ returns a player in the set $\{j \in N : Q_j^\perp \neq Q_j^\top\}$ and `Refine`$(Q_i^\perp, Q_i^\top)$ a refinement Q'_i with $Q_i^\perp < Q'_i \leq Q_i^\top$. A run of PRA is depicted in Figure 3.

Algorithm 1 Preference Refinement Algorithm (PRA)

Input: Hedonic game (N, R)
Output: Pareto optimal and individually rational partition

1 $Q_i^\top \leftarrow R_i$, for each $i \in N$
2 $Q_i^\perp \leftarrow R_i \cup \{(X,Y) : X\ R_i\ \{i\} \text{ and } Y\ R_i\ \{i\}\}$, for each $i \in N$
3 **while** $Q_i^\perp \neq Q_i^\top$ for some $i \in N$ **do**
4     $i \leftarrow \texttt{Choose}(\{j \in N : Q_j^\perp \neq Q_j^\top\})$
5   $Q_i' \leftarrow \texttt{Refine}(Q_i^\perp, Q_i^\top)$
6   **if** $\texttt{PerfectPartition}(N, (Q_1^\perp, \dots, Q_{i-1}^\perp, Q_i', Q_{i+1}^\perp, \dots, Q_n^\perp)) = \emptyset$ **then**
7     $Q_i^\top \leftarrow Q_i''$ where $\texttt{Cover}(Q_i'') = Q_i'$
8   **else**
9     $Q_i^\perp \leftarrow Q_i'$
10   **end if**
11 **end while**
12 **return** $\texttt{PerfectPartition}(N, Q^\perp)$

The exact definition of the procedures `Choose` and `Refine` may vary, giving rise to multiple possible settings.[3] Still, for every setting, PRA will be sound in the sense that it invariably returns a Pareto optimal and individually rational partition for the hedonic game it gets as input. This result is captured by the following theorem.

Theorem 2. *For every hedonic game (N, R), PRA returns an individually rational and Pareto optimal partition.*

Proof. Observe that in the course of running PRA two preference profiles $Q^\top$ and $Q^\perp$ are updated. Below we show that

(*a*) for every assignment of $Q^\perp$ a perfect partition exists,

(*b*) PRA terminates, and

(*c*) that at termination there is no preference profile R' with $Q^\perp < R' \leq R$ for which a perfect partition exists.

PRA then outputs a perfect partition for the final assignment of $Q^\perp$, which, by virtue of Theorem 1, is Pareto optimal for R.

To see that (a) holds, observe that in steps 1 and 2, $Q^\top$ and $Q^\perp$ are initialized. For each $i \in N$, $Q_i^\top$ is set to R_i and $Q_i^\perp$ is set to the finest coarsening of R_i in which i is indifferent between all coalitions that are acceptable in R_i. Observe that for this first assignment the singleton partition $\{\{i\} : i \in N\}$ is perfect for $Q^\perp$. Due to the condition in step 6, moreover, $Q^\perp$ is only updated to preference profiles (step 9) for which a perfect partitions exist. This proves (a).

[3]Thus, PRA is not a single algorithm but rather a class of algorithms. We will slightly abuse terminology and speak of PRA and subclasses of PRA as single algorithms and implicitly quantify over the different settings.

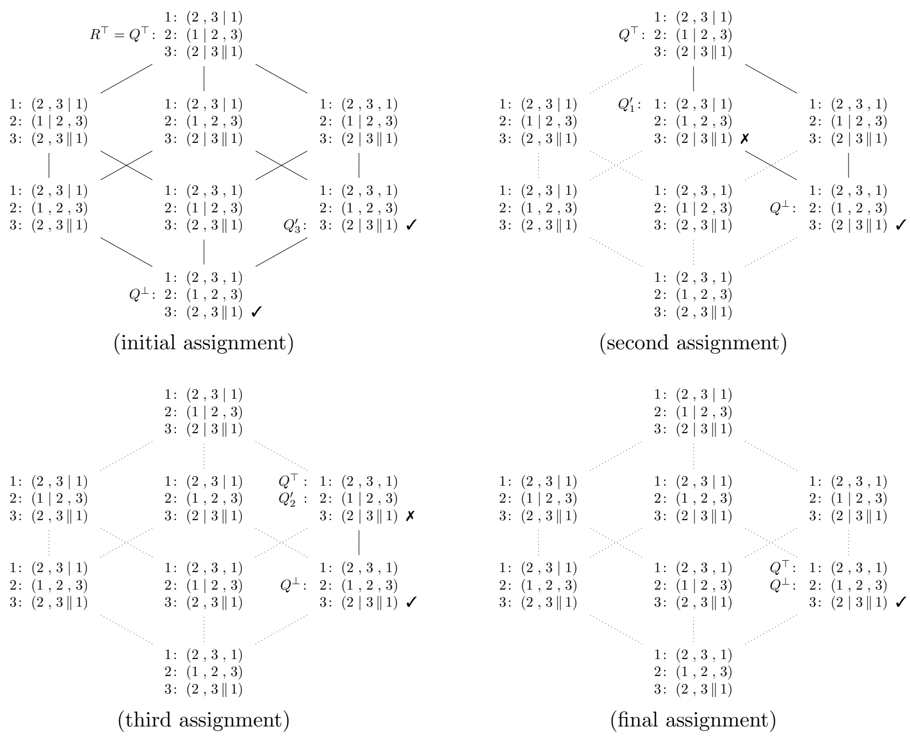

Figure 3: A run of the Preference Refinement Algorithm (any setting) for the W-hedonic game $(N, R^\top)$ as in Figure 2. As before, we denote indifference I_i by a comma, strict preferences P_i by a vertical bar, and unacceptability by a double bar. The figure depicts the subsequent preference profiles assigned to $Q^\top$ and $Q^\perp$. For each for these we have Q'_i denote $\texttt{Refine}(Q^\top_i, Q^\perp_i)$. A cross indicates that $\texttt{PerfectPartition}(Q^\perp_1, \ldots, Q^\perp_{i-1}, Q'_i, Q^\perp_{i+1}, \ldots, Q^\perp_n) = \emptyset$ and a checkmark that this is not the case. In this particular example $\texttt{Choose}$ subsequently selects players 3, 1, and 2. At termination $Q^\top_i = Q^\perp_i$ for each player i and PRA outputs a perfect partition in $\texttt{PerfectPartition}(Q^\perp)$. In this case, this is $\{\{1\}, \{2, 3\}\}$, which is Pareto optimal for $(N, R^\top)$.

For (b), observe that in each iteration of the while-loop a preference profile in $[Q^\perp, Q^\top]$ is inspected and either $Q^\top$ is (strictly) coarsened (step 7) or $Q^\perp$ (strictly) refined. Hence, after a finite number of steps, $Q^\perp_i = Q^\top_i$ for all players i and the condition of the while-loop is falsified (step 3). The algorithm then terminates one step later, which proves (b).

Finally, assume for contradiction that (c) does not hold for the final assignment of $Q^\perp$. As at termination $Q^\perp = Q^\top$, it follows that there is some R' with $Q^\top < R' \leq R$ for which there is a perfect partition. Obviously, for the initial assignment R of $Q^\top$, there is no R' with $Q^\top < R' \leq R$, let alone one for which a perfect partition exists. Let Q^* and Q^{**} be the last assignments of $Q^\perp$ and $Q^\top$, respectively, with the property that there is no R' with $Q^{**} < R' \leq R$ that also allows for a perfect partition. Then, in the next iteration of the while-loop, the condition of the if-clause in step 6 is satisfied i.e., for the player i chosen in step 4 and Q'_i selected in step 5, no perfect partition for $(Q^{**}_1, \ldots, Q^{**}_{i-1}, Q'_i, Q^{**}_{i+1}, \ldots, Q^{**}_n)$ exists. (If this were not the case, $Q^\perp$ would be updated in step 9 and Q^* and Q^{**} are not the last assignments of $Q^\perp$ and $Q^\top$ with the above property.) As can easily be appreciated, then no perfect partition exists for any R' with $(Q^{**}_1, \ldots, Q^{**}_{i-1}, Q'_i, Q^{**}_{i+1}, \ldots, Q^{**}_n) \leq R' \leq R$ either. Let Q'_i cover Q''_i. By linearity of the refinement relation on individual preferences, Q''_i is unique. Hence, the next assignment of $Q^\top$ equals $(Q^{**}_1, \ldots, Q^{**}_{i-1}, Q''_i, Q^{**}_{i+1}, \ldots, Q^{**}_n)$ (step 7). It moreover follows that there is no perfect partition for any R' with $(Q^{**}_1, \ldots, Q^{**}_{i-1}, Q''_i, Q^{**}_{i+1}, \ldots, Q^{**}_n) < R' \leq R$. Hence, Q^{**} was not the last assignment of $Q^\top$ with this property, a contradiction. □

In other respects, however, the behavior of PRA may depend on the specific settings of `Choose` and `Refine`. Among the many possibilities, we distinguish the following types of setting.

- *Divide-and-conquer* (PRA$_{DC}$). `Choose` is defined arbitrarily. `Refine` is specified such that the number of refinements from $Q^\perp_i$ to $\texttt{Refine}(Q^\perp_i, Q^\top_i)$ is the smallest integer larger than half the number of refinements from $Q^\perp_i$ to $Q^\top_i$. For the remainder of the paper, we assume PRA$_{DC}$ to be the default setting of PRA and frequently omit the subscript when the context is clear.

- *Serial dictatorship* (PRA$_{SD}$). `Choose` selects players according to a fixed order of the players and `Refine` returns a player's finest preference relation, i.e., generally $\texttt{Refine}(Q^\perp_i, Q^\top_i) = Q^\top_i$.

- *Conservative* (PRA$_{Cons}$). `Choose` selects players non-deterministically and `Refine` is defined such that $\texttt{Refine}(Q^\perp_i, Q^\top_i) = \texttt{Cover}(Q^\perp_i)$.

- *Egalitarian* (PRA$_{Egal}$). `Choose` selects a player that has been selected the fewest number of times during the execution or PRA. `Refine` is defined such that $\texttt{Refine}(Q^\perp_i, Q^\top_i) = \texttt{Cover}(Q^\perp_i)$.

Completeness—in the sense that for every individually rational and Pareto optimal partition π of a hedonic game (N, R), there is an execution of PRA that returns π on input (N, R)—cannot be achieved for all settings of `Choose` and `Refine`. However, for PRA$_{Cons}$, which proceeds cautiously by refining preferences only minimally in each iteration, we have the following result.

Theorem 3. *For every hedonic game (N, R) and every partition π that is individually rational and Pareto optimal for R, there is an execution of PRA$_{Cons}$ on input (N, R) that returns a partition π' such that $\pi \, I_i \, \pi'$ for all i in N.*

Proof. By virtue of Theorem 1, for each Pareto optimal and individually rational partition π for a preference profile R there is some coarsening Q^* of R where π is perfect and no perfect partitions exist for any R' with $Q^* < R' \leq R$. By individual rationality of π, it follows that Q^* is a refinement of the initial assignment of $Q^\perp$. Observe that perfect partitions exist for all R'' with $Q^\perp \leq R'' \leq R^*$. Hence, if `Choose` selects the appropriate players an appropriate number of times, the final assignment of $Q^\perp$ is Q^*. It follows that the perfect partition π' returned by `PerfectPartition` on input (N, Q^*), and therewith eventually by PRA$_{Cons}$, is such that $\pi \, I_i \, \pi'$ for all i in N. □

Whether PRA runs in polynomial time depends on a number of factors, in particular on whether the various subroutines `Choose`, `Refine`, `Cover`, and `PerfectPartition` as well as the initial assignment to $Q^\perp$ can be executed efficiently. Among these tractability of `PerfectPartition` is the most crucial and usually the least trivial to prove. The polynomial-time computability of the other routines can easily be appreciated to hold for all the classes of hedonic games considered in this paper. The best computational results are achieved by PRA$_{DC}$, the divide-and-conquer setting of PRA.

Lemma 2. *For any class of hedonic games for which any coarsening and* `PerfectPartition` *can be computed in polynomial time, PRA$_{DC}$ runs in polynomial time.*

Furthermore, if for any given preference profile R and partition π, the coarsening $Q^\perp$ of R such that $Q_i^\perp = R_i \cup \{(X,Y) : X \, R_i \, \pi(i)$ and $Y \, R_i \, \pi(i)\}$ can be computed in polynomial time, then the it can also be verified in polynomial time whether a given partition is Pareto optimal. For partitions that are not Pareto optimal, PRA then yields a partition that Pareto dominates it.

Proof. Under the given conditions, we prove that PRA runs in polynomial time. We first prove that the while-loop in PRA iterates a polynomial number of times. In each iteration of the while-loop, for a given player i, $Q_i^\top$ is lowered or $Q_i^\perp$ is raised. Due to the divide-and-conquer definition of `Refine`, $Q_i^\top$ coincides with $Q_i^\perp$ after a polynomial number of refinements. After this, player i is not considered for preference refinement. Therefore, even if the representation of (N, R) may be such that each player differentiates between an exponential number of coalitions, divide-and-conquer ensures that the while loop in PRA iterates a polynomial number of times. Having assumed that the crucial subroutine `PerfectPartition` takes polynomial time, PRA runs in polynomial time as well.

For the second part of the lemma, we run PRA to find a Pareto optimal partition that Pareto dominates π if there is any. We therefore modify Step 2 by setting $Q_i^\perp$ to the coarsening of R' of R in which for all $S \in \mathscr{N}_i$ such that $S \, R_i \, \pi(i)$, it is the case that $S \, I_i' \, \pi(i)$. It is clear that π is a perfect partition for R'. Since such a coarsening can be computed in polynomial time as stated by the condition in the lemma, Step 2 takes polynomial time. Since an initial perfect partition exists for $Q_i^\perp$, we run PRA as usual after Step 2. □

Observe that PRA not only applies to the canonical representation of general hedonic games, but also to many natural classes of hedonic games in which the preferences over coalitions (with possibly exponentially many indifference classes) for each player are defined implicitly. For example, in W-hedonic games,

$\max_{R_i}(N)$ specifies the set of favorite players of player i, but can also implicitly represent all those coalitions S such that the least preferred player in S is also a favorite player for i. In fact, if `PerfectPartition` can be solved efficiently, PRA_{DC} runs in polynomial time even if there is an exponential number of equivalence classes. Note that the lattice $([R^{\perp}, R^{\top}], \leq)$ inspected by PRA can be of exponential height and doubly-exponential width in the input (N, R). By virtue of divide-and-conquer, PRA_{DC} also traverses through the lattice in an orderly and efficient fashion to compute a Pareto optimal partition.

For all the specific classes of hedonic games we consider in this paper, each player differentiates between a number of equivalence classes that is polynomial in the input. Therefore, for any of these classes the other settings of PRA also run in polynomial time, provided that `PerfectPartition` can be solved in polynomial time. Hedonic games in which players can possibly distinguish between a number of equivalences exponential in the input, however, do exist. For an example consider the class of hedonic games in which each player has preferences R_i over the other players. These are then used to define a *leximin* ordering over coalitions i is a member of defined inductively such that $S\, R_i\, T$ if and only if either $\min_{R_i}(S)\, P_i \min_{R_i}(T)$ or $\min_{R_i}(S)\, I_i \min_{R_i}(T)$ and $S \setminus \min_{R_i}(S)\, R_i\, T \setminus \min_{R_i}(T)$. Contingent on the tractability of `PerfectPartition` in such classes, PRA_{DC} will run in polynomial time, whereas this is not the case for SD.

4.2. Advantages of PRA over Serial Dictatorship

Serial dictatorship (SD) is a well-studied mechanism in resource allocation, in which an arbitrary player is chosen as the 'dictator' who is then assigned his most favored allocation and the process is repeated until all players or resources have been dealt with. In the context of coalition formation, SD is well-defined only if in every iteration, the dictator has a *unique* most preferred coalition. In addition individual rationality can be achieved by giving the dictator his most preferred allocation that is also acceptable to the other players.

Proposition 1. *For general hedonic games, W-hedonic games, and roommate games, an individually stable and Pareto optimal partition can be computed in polynomial time when preferences are strict.*

Proposition 1 follows from the application of SD to hedonic games with strict preferences over the coalitions. If the preferences over coalitions are not strict, then the decision to assign one of the favorite coalitions to the dictator may be sub-optimal. Even if players express strict preferences over other players, SD may not work if the preferences induced over coalitions admit ties. Hence, SD is not applicable to B-hedonic games.

If SD works properly and efficiently in some setting, then so does PRA_{SD}, the serial dictatorship setting of PRA. As in each iteration in Algorithm 1, the same player i is chosen in step 4 until it cannot be chosen any more and $Q_i^{\top}$ is each time chosen in step 5, PRA_{SD} emulates serial dictatorship. Note that PRA_{SD} is a *generalization* of SD that can handle both ties and initial endowments.

Instead of forming the most preferred coalition of a dictator, PRA_{SD} considers whether a perfect partition can be formed in which the dictator gets one of his most preferred coalitions while ensuring that the other players also get one of their most preferred coalitions (in the current coarsened preference profile). Therefore, in contrast to SD, PRA_{SD} can compute a Pareto optimal partition that is a Pareto improvement over a given partition (also compare Lemma 2), even when preferences over coalitions are strict.

Abdulkadiroğlu and Sönmez [1] showed that in the case of strict preferences and house allocation settings, every Pareto optimal allocation can be found by SD. In the case of coalition formation, however, it is easy to construct a four-player hedonic game with strict preferences for which there is a Pareto optimal partition that SD cannot return. Consider, for instance, four players with strict preferences who are primarily interested in the size of the coalition they are in, preferring coalitions of size three to those of size two, and those of size two to being alone or being all together in the grand coalition. Then, any partition in which players are grouped in pairs is Pareto optimal and individually rational but not most preferred by any of the players. In this case, PRA_{Cons} will yield better results.

Finally, PRA can be used to compute the most egalitarian Pareto optimal partition, in which the satisfaction level of the worst off player is maximized. Such a partition is achieved by PRA_{Egal}.

We note that the Top Trading Cycle (TTC) [27] algorithm is another algorithm that achieves Pareto optimality and individual rationality for the restricted setting of housing markets with strict preferences. TTC can be generalized to achieve Pareto optimality and individual rationality in the presence of ties [see e.g., 4]. However TTC and its generalizations are designed for the exchange of a single type of good and cannot handle coalition formation, network formation, and general exchange economies of discrete goods.

4.3. Strategyproofness

In this subsection, we make some remarks concerning *strategyproofness* along with Pareto optimality and individual rationality in coalition formation mechanisms. A coalition formation mechanism is *strategyproof* if for all players i and all preference profiles R and $R' = (R_1, \dots, R_i', \dots R_n)$ it is the case that $f(R)\ R_i\ f(R')$, i.e, to express one's true preferences is a dominant strategy. The following theorem dashes hope for the existence of strategyproof mechanisms for hedonic games.

Theorem 4 (Alcalde and Barberà [3], Roth [25])**.** *There exists no strategyproof mechanism that returns an individually rational and Pareto optimal partition for marriage games if preferences are strict but allow unacceptability and there are at least four players.*

Theorem 4, of course, also applies to superclasses of marriage games. As a consequence, SD is not strategyproof for general hedonic games, marriage games, roommate games and W-hedonic game with strict but incomplete preferences (unacceptability can be expressed).

In contrast to the negative statement of Theorem 4, strategyproofness can be achieved if players are unable to express unacceptability of coalitions.

Theorem 5. *For general hedonic games with no unacceptability, PRA_{SD} is strategyproof.*

The reason is that if a player is not allowed to express any coalition as unacceptable then he cannot prevent a dictator from forming a most preferred coalition in which the player is also present. Theorem 5 applies to all subclasses of general hedonic games. As a corollary we also obtain the well-known statement that for house allocation with strict preferences, serial dictatorship is Pareto optimal and strategyproof.

5. Specific Computational Results

In this section, we consider the problems VERIFICATION (verifying whether a given partition is Pareto optimal) and COMPUTATION (computing a Pareto optimal partition) for various classes and representations of hedonic games. For our positive results we show that `PerfectPartition` can be computed in polynomial time. Tractability of COMPUTATION and VERIFICATION then generally follow as a corollary to the first and second part of Lemma 2, respectively. To show intractability of COMPUTATION we show that `PerfectPartition` cannot be solved in polynomial time for the respective class of games. Lemma 1 then gives the result. Hardness of `PerfectPartition` we usually establish by showing that every instance I of a particular intractable problem can in polynomial time be reduced to an equivalent instance (N, R) of `PerfectPartition`. [4]

5.1. General hedonic games

As shown in Proposition 1, Pareto optimal partitions can be found efficiently for general hedonic games with strict preferences. If preferences are not strict, the problem turns out to be NP-hard. We prove this statement by utilizing Lemma 1 and showing that `PerfectPartition` is NP-hard.

Theorem 6. *For a general hedonic game, computing a Pareto optimal partition is NP-hard even when each player has a maximum of four acceptable coalitions and the maximum size of each coalition is three.*

Interestingly, *verifying* Pareto optimality is coNP-complete even for strict preferences. This result contrasts with the general observation by Cechlárová that "*in the area of matching theory usually ties are 'responsible' for NP-completeness*" [10].

[4] For a number of classes of games, we find that I is also equivalent to verifying whether a particular partition π obtainable from I in polynomial time is not *weakly* Pareto optimal. If, moreover, this π is always identical to the grand coalition $\{N\}$ and no player is indifferent between N and any other coalition—in that case $\{N\}$ is Pareto optimal if and only if it is weakly Pareto optimal—we may conclude that VERIFICATION is computationally hard as well.

Theorem 7. *For a general hedonic game, verifying whether a given partition π is Pareto optimal and whether π is weakly Pareto optimal is coNP-complete even when preferences are strict and π consists of the grand coalition of all players.*

5.2. Roommate games

For the class of roommate games, we obtain more positive results.

Theorem 8. *For roommate games, an individually rational and Pareto optimal coalition can be computed in polynomial time.*

The statement follows from Lemma 2 and the fact that `PerfectPartition` can be solved in polynomial time for W-hedonic games. The latter is proved by a polynomial-time reduction of `PerfectPartition` to *maximum-weight matching.*

By utilizing the second part of Lemma 2, it can be seen that there exists an algorithm to compute a Pareto optimal improvement of a given roommate matching which takes time $O(n^3) \cdot O(n \log(n)) = O(n^4 \log(n))$. As a corollary we get the following.

Theorem 9. *For roommate games, it can be checked in polynomial time whether a given partition is Pareto optimal.*

The issue of computing a Pareto optimal improvement of a given 'status-quo' roommate matching has already enjoyed some attention in the literature [29, 22]. Morrill [22] examined roommate games with *strict* preferences and proposed an algorithm which, for any given matching, finds a Pareto optimal matching that Pareto dominates the original one. We devise a tailor-made algorithm for roommate games which finds a Pareto optimal Pareto improvement of a given matching in $O(n^3)$—the same asymptotic complexity required by Morrill's algorithm for the restricted case of strict preferences.

Theorem 10. *For roommate games, there exists an algorithm that finds a Pareto optimal Pareto improvement of a given matching in $O(n^3)$, even if preferences contain ties.*

5.3. W-hedonic games

We now turn to Pareto optimality in W-hedonic games. Recall that W-hedonic games may not admit a core stable partition and in fact even checking whether a core stable partition exists is NP-hard [13]. These negative existence and computational results in the literature contrast with the following positive result.

Theorem 11. *For W-hedonic games, a partition that is both individually rational and Pareto optimal can be computed in polynomial time.*

The statement follows from Lemma 2 and the fact that `PerfectPartition` can be solved in polynomial time for W-hedonic games. The latter is proved by a polynomial-time reduction of `PerfectPartition` to a polynomial-time solvable problem called *clique packing.*

Due to the second part of Lemma 2, the following is evident.

Theorem 12. *For W-hedonic games, it can be checked in polynomial time whether a given partition is Pareto optimal or weakly Pareto optimal.*

Our positive results for W-hedonic games also apply to hedonic games with $\mathscr{W}$-preferences as proposed by Cechlárová and Hajduková [12, 13].

5.4. B-hedonic games

For W-hedonic games, a Pareto optimal partition can be computed efficiently, even in the presence of unacceptable players. In the absence of unacceptable players, computing such a partition for B-hedonic games is trivial, as the partition consisting of the grand coalition is always a solution.

Interestingly, if preferences do allow for unacceptable players, the same problem becomes NP-hard. The statement is shown by using Lemma 1 and a reduction from SAT.

Theorem 13. *For B-hedonic games, computing a Pareto optimal partition is NP-hard.*

Using similar techniques, we also have the following.

Theorem 14. *For B-hedonic games, verifying whether a given partition is weakly Pareto optimal is coNP-complete.*

We expect the result above to hold for (strong) Pareto optimality as well.

5.5. Exchange of multiple types of goods

An *exchange economy with multiple types of goods* consists of agents who have complete preferences over bundles of discrete goods. The goods can be classified into multiple types—e.g., houses, cars etc.— and each agent can be allocated up to a certain number goods of a certain type [see e.g., 21]. Each good is either initially owned by some agent or is a social endowment not owned by any agent. The goal is to achieve an individually rational and Pareto optimal allocation. The problem can be construed as a coalition formation problem in which coalitions with two or more agents are unacceptable and goods are completely indifferent.

We also notice that if there is only type of good, say houses, and each agent can own a maximum of one house, then the problem reduces to '*house allocation with existing tenants*' [2, 4]. House allocation with existing tenants itself is a hybrid generalization of two classic settings: *house allocation problem* and *housing markets*.[5]

For house allocation with existing tenants, PRA can be used to compute individually rational and Pareto optimal outcomes in polynomial time via a reduction to finding perfect matchings.

[5] The *house allocation problem* (also known as the *assignment problem*) in which agents have strict preferences over houses, and each agent is allocated a maximum of one house. As mentioned earlier, serial dictatorship can be used to compute a Pareto optimal allocation for the classic house allocation setting (with strict preferences). In the classic *housing market* setting, each agent owns a house and has strict preferences over all the houses in the market.

Theorem 15. *For house allocation with existing tenants, an individually rational and Pareto optimal outcome can be computed in polynomial time, even if preferences contain ties.*

On the other hand, the problem of computing a Pareto optimal outcome becomes NP-hard if there is more than one type of goods. Even two types of goods suffice for this to hold.

Theorem 16. *For exchange economies with multiple types of goods, computing a Pareto optimal outcome is NP-hard even if there is no unacceptability, there are only two types of goods, and each agent can own only one good of each type.*

As a corollary of the proof idea of Theorem 16 we obtain the following theorem, which has previously been shown by Cechlárová [Theorem 3, 11].

Theorem 17 (Cechlárová [11])**.** *For exchange economies with multiple types of goods, verifying whether a given outcome is weakly Pareto optimal is coNP-complete, even if there are two types of goods, preference are strict, and there is no unacceptability.*

We briefly comment on the strategic aspects of the restricted setting of exchange economies of discrete goods. Even if there are no ties in the preferences and there are only two types of goods, it is known that there exists no strategyproof mechanism that yields an individually rational and Pareto optimal outcome for exchange economies with multiple types of goods (see, Konishi et al. [Proposition 4.1, 21]). On the other hand, if unacceptability is not expressed, PRA_{SD} is strategyproof even if there are ties in the preferences. We get the following as a corollary of Theorem 5.

Corollary 1. *For exchange economies with multiple types of goods but with no unacceptability, PRA_{SD} is strategyproof.*

Recently, it has been shown that there exists a rather elaborate strategyproof mechanism that yields a Pareto optimal and individually rational outcome for the house allocation with existing tenants problem, even when allowing for ties in the preferences [2, 4].

5.6. Three-cyclic games

Knuth [20] proposed a three-dimensional extension of marriage games in which there are three sets of agents: men, women and dogs. A feasible matching is a set of disjoint families, i.e., triples of the form (man, woman, dog). If each agent has preferences over all pairs from the other two sets, then we call the game a *three-dimensional matching game*. Based on Knuth's idea, Ng and Hirschberg [23] formalized the *cyclic* three-dimensional matching setting (which we will refer to as three-cyclic games) in which men only care about women, women only care about dogs and dogs only care about men. Recently, Biró and McDermid [8] proved that for three-cyclic games, checking the existence of a core stable or a strict core stable matching is NP-complete. We show that even computing a Pareto optimal outcome is NP-hard for three-cyclic games.

Theorem 18. *For three-cyclic games, computing a Pareto optimal outcome is NP-hard even if there is no unacceptability.*

The corresponding verification problem turns out to be coNP-complete.

Theorem 19. *For three-cyclic games, verifying whether a given partition is weakly Pareto optimal is coNP-complete even for strict preferences.*

The intractability results for three-cyclic games carry over to three-dimensional matching games, the latter being a generalization of the former.

5.7. Room-roommate games

In house allocation, non-sharable items (e.g, jobs or houses) are divided among the agents. *Room-roommate games* are a generalization of both house allocation and roommate games. We see that neither our positive algorithmic results for house allocation nor those for roommate games extend to room-roommate games.

Theorem 20. *For room-roommate games, computing a Pareto optimal outcome is NP-hard even if no unacceptability is expressed in the preferences.*

Theorem 21. *For room-roommate games, verifying whether a given partition is weakly Pareto optimal is coNP-complete, even for strict preferences.*

In the absence of ties, Computation is easier. The individually rational version of serial dictatorship gives the result.

Theorem 22. *For room-roommate games with strict preferences, a Pareto optimal and individually rational partition can be computed in polynomial time.*

5.8. Anonymous games

Anonymous hedonic games are a subclass of hedonic games in which the players' preferences over coalitions only depend on coalition sizes. Therefore, anonymous hedonic games can be represented compactly by preferences lists over the integers $1, \ldots, n$. Anonymous hedonic games were first considered by Bogomolnaia and Jackson [9]. Ballester [6] later examined the complexity of checking the existence of stable partitions of anonymous games. We show that both computing and verifying Pareto optimal outcomes is intractable for anonymous games.

Theorem 23. *For anonymous games, computing a Pareto optimal partition is NP-hard.*

Theorem 24. *For anonymous games, verifying whether a given partition is Pareto optimal is coNP-complete, even if the partition in question consists of the grand coalition.*

These results exploit discontinuities in the preferences. Bogomolnaia and Jackson [9] proved that for anonymous games with single-peaked preferences, an individually stable partition exists and can be computed efficiently. It will be interesting to see if the structure of single-peakedness in anonymous games can be exploited in order to obtain a positive algorithmic result.

6. Conclusions

Pareto optimality and individual rationality are important requirements for desirable partitions in coalition formation. In this paper, we examined computational, structural, and strategic issues related to Pareto optimality in various classes of hedonic games (see Table 1 and Table 2). The relations between the classes of hedonic games considered is depicted in Figure 1.

At the basis of most of our computational results lies an intimate conceptual connection between Pareto optimality and the notion of perfection. Exploiting this connection, we formulated the *Preference Refinement Algorithm* (PRA) for computing Pareto optimal and individually rational outcomes. A quality of PRA that distinguishes it from a well-known procedure like serial dictatorship, is that it handles both weak preferences and unacceptability well. This feature is especially relevant when dealing with classes of hedonic games in which the players preferences over (possibly exponentially many) coalitions are concisely represented by, e.g., preferences over (polynomially many) players—as in W- and B-hedonic games—or over numbers—as in the case of anonymous games. Even if the preferences in the concise representation are strict, the corresponding preferences over coalitions are bound to contain indifferences.

We saw that unacceptability and ties are a major source of intractability when computing Pareto optimal outcomes. In some cases, checking whether a given partition is Pareto optimal can be significantly harder than finding one. Furthermore, we observed that unacceptability (rather than ties) is the main obstacle in devising strategyproof mechanisms to find Pareto optimal and individually rational partitions.

It should be noted that most of our insights gained into Pareto optimality, perfection, and the resulting algorithmic techniques—especially those presented in Section 3 and Section 4—do not only apply to coalition formation but also to many other settings, such as school choice, discrete allocation and exchange [see e.g., 28], and network formation [see e.g., 19].

Acknowledgements

This material is based on work supported by the Deutsche Forschungsgemeinschaft under grants BR-2312/6-1 (within the European Science Foundation's EUROCORES program LogICCC), BR 2312/7-1, and BR 2312/9-1.

References

[1] Abdulkadiroğlu, A., Sönmez, T., 1998. Random serial dictatorship and the core from random endowments in house allocation problems. Econometrica 66 (3), 689–702.

[2] Abdulkadiroğlu, A., Sönmez, T., 1999. House allocation with existing tenants. Journal of Economic Theory 88 (2), 233–260.

Game	Verification	Computation
General	coNP-C (Th. 7)	NP-hard (Th. 6)
General (strict prefs)	coNP-C (Th. 7)	in P (Prop. 1)
Roommate	in P (Th. 9)	in P (Th. 8)
B-hedonic	coNP-C (Th. 14, weak PO)	NP-hard (Th. 13)
W-hedonic	in P (Th. 12)	in P (Th. 11)
Anonymous	coNP-C (Th. 24)	NP-hard (Th. 23)
Room-roommate	coNP-C (Th. 21, weak PO)	NP-hard (Th. 20)
Three-cyclic	coNP-C (Th. 19, weak PO)	NP-hard (Th. 18)
House allocation (w/ existing tenants)	in P (Th. 15)	in P (Th. 15)
General exchange economy	coNP-C (Th. 17, weak PO [11])	NP-hard (Th. 16)

Table 1: Complexity of Pareto optimality in hedonic games: positive results hold for both Pareto optimality and individual rationality.

Game	Preference Restriction	Strategyproof Mechanism
General, Marriage, Roommate W, B, 3-cyclic, Room-roommate	No unacceptability	✓ (Th. 5)
House allocation	Strict, No unacceptability	✓ (SD)
House allocation (w. existing tenants)	Strict	✓ [2]
House allocation (w. existing tenants)		✓ [4]
General exchange economy	No unacceptability	✓ (Cor. 1)
General exchange economy		✗ [Prop. 4.1, 21]
General, Marriage, Roommate, W, B, 3-cyclic, Room-roommate		✗ (Th. 4)

Table 2: Existence of strategyproof mechanisms for hedonic games that yield a Pareto optimal and individually rational partition

[3] Alcalde, J., Barberà, S., 1994. Top dominance and the possibility of strategy-proof stable solutions to matching problems. Economic Theory 4 (3), 417–435.

[4] Alcalde-Unzu, J., Molis, E., 2011. Exchange of indivisible goods and indifferences: The top trading absorbing sets mechanisms. Games and Economic Behavior 73 (1), 1–16.

[5] Aumann, R. J., 1987. Game theory. In: Eatwell, J., Milgate, M., Newman, P. (Eds.), The New Palgrave: A Dictionary of Economics. Vol. 2. MacMillan, pp. 460–482.

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

[7] Banerjee, S., Konishi, H., Sönmez, T., 2001. Core in a simple coalition formation game. Social Choice and Welfare 18, 135–153.

[8] Biró, P., McDermid, E., 2010. Three-sided stable matchings with cyclic preferences. Algorithmica 58 (1), 5–18.

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

[10] Cechlárová, K., 2002. On the complexity of exchange-stable roommates. Discrete Applied Mathematics 116, 279–287.

[11] Cechlárová, K., 2009. On the complexity of the shapley-scarf economy with several types of goods. Kybernetika 45 (5), 689–700.

[12] Cechlárová, K., Hajduková, J., 2004. Stability of partitions under $\mathscr{BW}$-preferences and $\mathscr{WB}$-preferences. International Journal of Information Technology and Decision Making 3 (4), 605–618.

[13] Cechlárová, K., Hajduková, J., 2004. Stable partitions with $\mathscr{W}$-preferences. Discrete Applied Mathematics 138 (3), 333–347.

[14] Cornuéjols, G., Hartvigsen, D., Pulleyblank, W., 1982. Packing subgraphs in a graph. Operations Research Letters 1 (4), 139–143.

[15] Drèze, J. H., Greenberg, J., 1980. Hedonic coalitions: Optimality and stability. Econometrica 48 (4), 987–1003.

[16] Edmonds, J., 1965. Paths, trees and flowers. Canadian Journal of Mathematics 17, 449–467.

[17] Gabow, H. N., Tarjan, R., 1991. Faster scaling algorithms for general graph matching problems. Journal of the ACM 38 (4), 815–853.

[18] Hell, P., Kirkpatrick, D. G., 1984. Packings by cliques and by finite families of graphs. Discrete Mathematics 49 (1), 45–59.

[19] Jackson, M. O., Wolinsky, A., 1996. A strategic model of social and economic networks. Journal of Economic Theory 71 (1), 44–74.

[20] Knuth, D. E., 1976. Mariages stables. Les Presses de l'Université de Montréal.

[21] Konishi, H., Quint, T., Wako, J., 2001. On the Shapley-Scarf economy: the case of multiple types of indivisible goods. Journal of Mathematical Economics 35 (1), 1–15.

[22] Morrill, T., 2010. The roommates problem revisited. Journal of Economic Theory 145 (5), 1739–1756.

[23] Ng, C., Hirschberg, D. S., 1991. Three-dimensional stable matching problems. SIAM Journal on Discrete Mathematics 4 (2), 245–252.

[24] Ronn, E., 1990. NP-complete stable matching problems. Journal of Algorithms 11, 285–304.

[25] Roth, A. E., 1982. The economics of matching: Stability and incentives. Mathematics of Operations Research 7 (4), 617–628.

[26] Roughgarden, T., 2010. Computing equilibria: a computational complexity perspective. Economic Theory 42 (1), 193–236.

[27] Shapley, L. S., Scarf, H., 1974. On cores and indivisibility. Journal of Mathematical Economics 1 (1), 23–37.

[28] Sönmez, T., Ünver, M. U., 2011. Matching, allocation, and exchange of discrete resources. In: Benhabib, J., Jackson, M. O., Bisin, A. (Eds.), Handbook of Social Economics. Vol. 1 of Handbooks in Economics. Elsevier, Ch. 17, pp. 781–852, forthcoming.

[29] Sotomayor, M. A. O., Özak, O., 2009. Two-sided matching models. In: Meyers, R. A. (Ed.), Encyclopedia of Complexity and Systems Science. Springer, pp. 9654–9678.

[30] von Neumann, J., Morgenstern, O., 1947. Theory of Games and Economic Behavior, 2nd Edition. Princeton University Press.

Appendices

Appendix A: Proofs

Proof of Theorem 6.

We can prove the statement by utilizing Lemma 1 and showing that `PerfectPartition` is NP-hard by a reduction from X3C. Recall that an instance of X3C is a pair (X, C) such that $X = \{1, \ldots, 3m\}$ is a set, m some positive integer, and C is a collection of subsets of size 3. The question is whether there is a sub-collection $C' \subseteq C$ that partitions X. The problem remains NP-hard even if $|\{c \in C : x \in c\}| \leq 3$ for all $x \in X$.

For an instance (X, C) of X3C, define a general hedonic game (N, R) such that $N = X$ and each player i is indifferent among coalitions in $\{c \in C : i \in c\}$, strictly prefers them to $\{i\}$, and finds every other coalition unacceptable. Then, a perfect partition exists if and only if (X, C) is a 'yes' instance.

Assume that there exists $C' \subseteq C$ such that C' is a partition of X. Then there exists a partition $\pi' = C'$ in which each player gets one of his most desirable coalition. Therefore, (N, R) is a 'yes' instance of `PerfectPartition`.

Assume that there exists no $C' \subseteq C$ such that C' is a partition of X. Then there exists no partition π' such that each player $i \in N$ gets one of his most desirable coalition. Therefore, (N, R) is a 'no' instance of `PerfectPartition`. □

Proof of Theorem 7.

VERIFICATION is clearly in coNP. A partition π' which (weakly) Pareto dominates π is a polynomial time certificate that π is not (weakly) Pareto optimal. Similarly, a partition π' which weakly Pareto dominates π is a polynomial time certificate that π is not weakly Pareto optimal.

We now prove that the problem is coNP-hard by a reduction from X3C. An instance (X, C) of X3C can be reduced to an instance $((N, R), \pi)$ of VERIFICATION, where $\pi = \{N\}$ and (N, R) is a general hedonic game with $N = X$ and $c_1 \, P_i \cdots P_i \, c_k \, P_i \, N \, P_i \, \{i\}$ for each $i \in N$, where $c_1, \ldots, c_k$ is any linear ordering of $\{c \in C : i \in c\}$ (all other coalitions are unacceptable to i). Then, $\{N\}$ is Pareto dominated by another partition if and only if there is a partition $C' \subseteq C$ of X.

Assume that there exists $C' \subseteq C$ such that C' is a partition of X. Then there exists a partition $\pi' = C'$ which (weakly) Pareto dominates π. Therefore π is not weakly Pareto optimal or Pareto optimal. Assume that there exists no $S' \subseteq S$ such that S' is a partition of R. Then there exists no partition π' which Pareto dominates and therefore weakly Pareto dominates π. □

Proof of Theorem 8.

We utilize Lemma 2. It is sufficient to show that `PerfectPartition` can be solved in time $O(n^3)$. We do so by a linear time reduction the polynomial-time solvable problem of checking whether a graph admits a perfect matching [16].

For notational convenience we write $F(i)$ for $\max_{R_i}(N)$. Construct an undirected graph $G = (V, E)$ where $V = N \cup (N \times \{0\})$, $E = \{\{i, j\} : i \neq j, i \in F(j) \text{ and } j \in F(j)\} \cup \{\{i, (i, 0)\} : i \in F(i)\}$.

We claim that there exists a perfect partition for (N, R) if and only if there exists a matching of size n in graph G. It is clear that in a matching of size n, each $v \in N$ is matched. If there exists a perfect partition, then each player in N is matched to a player $j \neq i$ such that $j \in F(i)$ or i is unmatched but $i \in F(i)$. In either case there exists a matching in which each i is matched. In the first case, i is matched to j in G. In the second case, i is matched to $(i, 0)$.

Now assume that there exists a matching of size n in G. Then, each $i \in N$ is matched to $j \neq i$ or $(i, 0)$. If i is matched to j, then we know $\{i, j\} \in E$ and therefore $j \in F(i)$. If i is matched to $(i, 0)$, then we know $\{i, (i, 0)\} \in E$ and therefore $i \in F(i)$. Thus, there exists a perfect partition. □

Proof of Theorem 9.

By observing the proof of Theorem 8 and utilizing the second part of Lemma 2, it can be seen that there exists an algorithm to compute a Pareto optimal improvement of a given roommate matching which takes time $O(n^3) \cdot O(n \log(n)) = O(n^4 \log(n))$. As a corollary we get the theorem. □

Proof of Theorem 10.

We reduce the problem to the polynomial-time solvable problem of computing a maximum weight matching of a graph.

For a given roommate game (N, R), let π be the partition we want to check for Pareto optimality. Since π contains coalitions of size one or two, we can construct an undirected graph $G = (V, E)$ where $V = N \cup (N \times \{0\})$, $E = V \times V \setminus (\{\{i, j\} : \pi(i) \, P_i \, \{i\}\} \cup \{\{i, (i, 0)\} : \pi(i) \, P_i \, \{i\}\})$. For graph (V, E), consider the matching $M = \{S \in \pi : |S| = 2\} \cup \{\{i, (i, 0)\} : \{i\} \in \pi\}$.

We now define a weight function such that for all $i \in V$, $w_i : E \to \mathbb{R}^+$ where w_i is defined inductively in the following way: $w_{(i,0)}(e) = 0$ for all e such that $(i, 0) \in e \in E$ and $i \in N$; $w_i(\pi(i)) = n$ if $\pi(i) \neq \{i\}$ and $\pi(i) = \{i, j\}$; $w_i(\{i, (i, 0)\}) = n$ if $\pi(i) = \{i\}$; $w_i(S) = -n$ if $i \notin S$; $w_i(T) = w_i(S) + 1/n$ if there is a coalition T such that $i \in T$, $T \, P_i \, S$, and there exists no coalition T' such that $T \, P_i \, T' \, P_i \, S$; and $w_i(T) = w_i(S)$ if $S \, R_i \, \pi(i)$ and T is coalition such that $T \, I_i \, S$. Define a weight function $w' : E \to \mathbb{R}^+$ such that for any $S = \{i, j\} \in E$, $w'(S) = w_i(S) + w_j(S)$. For $E'' \subseteq E$, denote by $w'(E'')$, the value $\sum_{e \in E''} w'(e)$. We can then prove that π is Pareto optimal if and only if π is the maximum weight matching of $G^{w'}$, the graph G, weighted by weight function w'. Since we have a linear-time reduction to maximum weight matching [17], the complexity of the algorithm is $O(n^3)$.

We first prove that if π is a maximum weight matching, then π is Pareto optimal. Assume π is the maximum matching. A partition which Pareto dominates π has a bigger weight than π. Since π is the maximum weight matching, there exists no matching which Pareto dominates π.

We now prove the converse. Assume π is Pareto optimal but π is not a maximum weight matching. Then there exists another matching π' such that

$w'(\pi') > w'(\pi)$. Since π is Pareto optimal, it cannot be that π' Pareto dominates π. Therefore, there exists a player $i \in N$ such that the utility of i in π is less than in π'. But this is only possible if $w_i(\pi(i)) = n$ and $w_i(\pi'(i)) = 0$. Since $w'(\pi') > w'(\pi)$, then $\sum_{j \in N \setminus \{i\}} w_j(\pi'(j)) - \sum_{j \in N \setminus \{i\}} w_j(\pi(j)) > n$. But we, know that for each player $j \in N \setminus \{i\}$, the increase in w_i can be at most $(n-1)/n$. Therefore $\sum_{j \in N \setminus \{i\}} w_j(\pi'(j)) - \sum_{j \in N \setminus \{i\}} w_j(\pi(j)) < \frac{n-1}{n} \times n = n - 1$. This is a contradiction.

Therefore, checking whether π is Pareto optimal reduces to checking whether π is a maximum weight matching of $G^{w'}$. This can be done as follows. Compute a maximum weight matching M of G^w in polynomial time and check whether $w'(M) = w'(\pi)$. □

Proof of Theorem 11.

The statement follows from Lemma 2 and the fact that `PerfectPartition` can be solved in polynomial time for W-hedonic games. The latter is proved by a polynomial-time reduction of `PerfectPartition` to a polynomial-time solvable problem called *clique packing*.

We first introduce the more general notion of graph packing. Let $\mathscr{F}$ be a set of undirected graphs. An $\mathscr{F}$-packing of a graph G is a subgraph H such that each component of H is (isomorphic to) a member of $\mathscr{F}$. The size of $\mathscr{F}$-packing H is $|V(H)|$. We will informally say that vertex i is *matched* by $\mathscr{F}$-packing H if i is in a connected component in H. Then, a maximum $\mathscr{F}$-packing of a graph G is one that matches the maximum number of vertices. It is easy to see that computing a maximum $\{K_2\}$-packing of a graph is equivalent to maximum cardinality matching. Hell and Kirkpatrick [18] and Cornuéjols et al. [14] independently proved that there is a polynomial-time algorithm to compute a maximum $\{K_2, \ldots, K_n\}$-packing of a graph. Cornuéjols et al. [14] note that finding a $\{K_2, \ldots, K_n\}$-packing can be reduced to finding a $\{K_2, K_3\}$-packing.

We are now in a position to reduce `PerfectPartition` for W-hedonic games to computing a maximum $\{K_2, K_3\}$-packing. For a W-hedonic game (N, R), construct a graph $G = (N \cup (N \times \{0, 1\}), E)$ such that $\{(i, 0), (i, 1)\} \in E$ for all $i \in N$; $\{i, j\} \in E$ if and only if $i \in \max_{R_j}(N)$ and $j \in \max_{R_i}(N)$ for $i, j \in N$ such that $i \neq j$; and $\{i, (i, 0)\}, \{i, (i, 1)\} \in E$ if and only if $i \in \max_{R_i}(N)$ for all $i \in N$. Let H be a maximum $\{K_2, K_3\}$-packing of G.

It can then be proved that there exists a perfect partition of N according to R if and only if $|V(H)| = 3|N|$. This is equivalent to saying that there exists no perfect partition of N according to R if and only if $|V(H)| < 3|N|$. We first prove the left to right implication of the last statement. Assume that there exists no perfect partition of N according to R. This implies that for every partition π, there exists some player i which is not matched with a subset of his favorite players and also does not consider himself a favorite player. Therefore, $i \notin V(H)$. Thus, $|V(H)| < 3|N|$.

We now prove the implication from right to left. Consider the case that $|V(H)| < 3|N|$ in the the corresponding graph G of (N, R) for R. We will first show that there exists an $i \in N$ such that $i \notin V(H)$. Assume for contradiction that $N \subseteq V(H)$ and assume without loss of generality there is a vertex $(i, 0)$

such that $(i,0) \notin V(H)$. If $(i,1)$ is not matched with i, then $(i,0)$ can be matched with $(i,1)$ because $\{(i,0),(i,1)\} \in E(G)$ so that $(i,0) \in V(H)$. If $(i,1)$ is matched with i, then this implies that $\{i,(i,1)\} \in E(G)$ and also $\{i,(i,0)\} \in E(G)$. Therefore $i,(i,0)$ and $(i,1)$ can be matched into a K_3 by H so that $(i,0) \in V(H)$. Since, $N \times \{0,1\} \subseteq V(H)$, it be must that there exists an $i \notin V(H)$. Therefore i is not in an acceptable (and favorite) coalition.

Since `PerfectPartition` for W-hedonic games reduces to checking whether graph G can be packed perfectly by elements in $\mathscr{F} = \{K_2, K_3\}$, we have a polynomial-time algorithm to solve `PerfectPartition` for W-hedonic games. Denote by $CC(H)$ the set of connected components of graph H. If $|V(H)| = 3|N|$ and a perfect partition does exist, then $\{V(S) \cap N : S \in CC(H)\} \setminus \emptyset$ is a perfect partition. □

Proof of Theorem 12.

The statement follows from second part of Lemma 2 and the proof of Theorem 11 in which it is shown that `PerfectPartition` is polynomial-time solvable for W-hedonic games.

□

Proof of Theorem 13.

It can be checked in polynomial time whether a partition is perfect in a B-hedonic game. Hence, by Lemma 1, it suffices to show that `PerfectPartition` is NP-hard. We do so by a reduction from SAT. Let $\varphi = X_1 \wedge \cdots \wedge X_k$ a Boolean formula in conjunctive normal form in which the Boolean variables $p_1, \ldots, p_m$ occur. Now define the B-hedonic game (N, R), where $N = \{X_1, \ldots, X_k\} \cup \{p_1, \neg p_1, \ldots, p_m, \neg p_m\} \cup \{0,1\}$ and the preferences for each literal p or $\neg p$, and each clause $X = (x_1 \vee \cdots \vee x_\ell)$ are as follows, where commas denote indifference, vertical bars strict preference P_i, and double bars unacceptability:

$$\begin{aligned}
p:&\quad (0,1 \mid N \setminus \{0,1,\neg p\} \parallel \neg p)\\
\neg p:&\quad (0,1 \mid N \setminus \{0,1,p\} \parallel p)\\
X:&\quad (x_1,\ldots,x_\ell \mid N \setminus \{0,x_1,\ldots,x_\ell\} \parallel 0)\\
0:&\quad (N \setminus \{0,1\} \mid 0 \parallel 1)\\
1:&\quad (N \setminus \{0,1\} \mid 1 \parallel 0)
\end{aligned}$$

We prove that φ is satisfiable if and only if a perfect (and individually rational) partition for (N, R) exists. To this end, first assume that v is a valuation that satisfies φ. Then, define the partition π such that

$$\pi = \{\{1, x'_1, \ldots, x'_{\ell'}, X_1, \ldots, X_k\}, \{0, x''_1, \ldots, x''_{\ell''}\}\}.$$

where $x'_1, \ldots, x'_{\ell'}$ are the literals rendered true by v and $x''_1, \ldots, x''_{\ell''}$ those that are rendered false. It can then be shown that π is a perfect partition.

Obviously, π is a perfect partition for both 0 and 1 as well as for every "literal" player x. As v is a satisfying valuation, for every "clause" player $X = (x_1, \ldots, x_\ell)$ at least one of the literals $x_1, \ldots, x_\ell$ is set to true by v, which will then be in the same coalition as X. It follows that $\pi(X)$ is a favorite coalition for X as well.

For the opposite direction, assume that π is a perfect partition. Now define the valuation v such that every literal x is to true if and only if $x \in \pi(1)$.

It remains to be shown that v is properly defined and that v satisfies φ. Observe that π consists of exactly two coalitions. The players 0 and 1 have to be in different coalitions, i.e., $\pi(0) \neq \pi(1)$, because they are unacceptable to one another. Moreover, any two "literal" players p and $\neg p$ have to be in either $\pi(0)$ or $\pi(1)$, otherwise they are not in a favorite coalition. Finally, every "clause" player $X = (x_1, \ldots, x_\ell)$ has to be in a coalition to which π assigns at least of the "literal players" $x_1, \ldots, x_\ell$, for very much the same reason. Because, player 0 is unacceptable to any "clause" player X, each of them has to be in $\pi(1)$.

As for every Boolean variable p, precisely one of p and $\neg p$ is assigned to $\pi(1)$, v is well-defined. Moreover, because every "clause" player $X = (x_1, \ldots, x_\ell)$ is in $\pi(1)$ together with at least one of $x_1, \ldots, x_\ell$, v is readily seen to satisfy formula φ. $\square$

Proof of Theorem 14.

The reduction is the same as in the proof of Theorem 13. The partition π to be considered is one in which each player in $N \setminus \{0, 1\}$ is a singleton and 0 and 1 are together. Then π is weakly Pareto optimal if and only if φ is not satisfiable. $\square$

Proof of Theorem 16.

We can prove the statement by utilizing Lemma 1 and showing that `PerfectPartition` is NP-hard by a reduction from 3-DimensionalMatching (3-DM). Let X, Y, and Z be finite, disjoint sets, and let T be a subset of $X \times Y \times Z$. Now $M \subseteq T$ is a *three-dimensional matching* if the following holds: for any two distinct triples $(x_1, y_1, z_1) \in M$ and $(x_2, y_2, z_2) \in M$, we have $x_1 \neq x_2$, $y_1 \neq y_2$, and $z_1 \neq z_2$. The decision question in problem 3-DM is whether for given X, Y, Z and $T \subseteq X \times Y \times Z$, there exists an $M \subseteq T$ such that $|M| = |X| = |Y| = |Z|$.

Based on an instance of 3-DM, we form an exchange economy with two kinds of goods as follows. Let N be equal to X, Y be the set of goods of type one, and Z be the goods of type two. Then, for any triple $(x, y, z) \in T$, set the preferences in N, in a way such that $\{y, z\}$ is one of the favorite allocations of agent x. Then, there exists a three-dimensional matching which matches each element in $X \cup Y \cup Z$ if and only if there exist a perfect allocation in which each player gets one of his favorite allocations. $\square$

Proof of Theorem 17.

The proof is by a reduction from 3-DM. Based on an instance of 3-DM, form an exchange economy with two kinds of goods in a similar way as in the proof of Theorem 16. The main difference is that each player has strict preferences and arbitrarily breaks ties among the allocations which were his favorite in the proof of Theorem 16. Secondly, all other allocations are unacceptable. Consider the partition π consisting of singletons, i.e, an allocation in which each agent has no initial endowment. Then, π is not weakly Pareto optimal if and only if there exists a partition in which each player is in a coalition strictly more preferred than the singleton coalition. The problem is equivalent to 3-DM. $\square$

Proof of Theorem 18.

We can prove the statement by utilizing Lemma 1 and showing that `PerfectPartition` is NP-hard by a reduction from 3-DM.

Based on an instance of 3-DM, we form a 3-cyclic game as follows. Let $N = X \cup Y \cup Z$ where X, Y, Z are the set of men, women, and dogs respectively. Then, for any triple $(x, y, z) \in T$, set the preferences in N, in a way such that y is one the favorite women of man x, z is one the favorite dogs of woman y, and x is one of the favorite men of dog z. Then, there exists a three-dimensional matching which matches each element in $X \cup Y \cup Z$ if and only if there exist a perfect partition of the 3-cyclic game. $\square$

Proof of Theorem 19.

Based on an instance of 3-DM, form a 3-cyclic game in a similar way as in the proof of Theorem 18. The main difference is that each player has strict preferences among those players that were his favorite in the proof of Theorem 18. Furthermore, all other players are considered unacceptable.

Consider the partition π consisting of singletons. Then, π is not weakly Pareto optimal if and only if there exists a partition in which each player is in a coalition strictly more preferred than the singleton coalition. The problem is equivalent to 3-DM. $\square$

Proof of Theorem 20.

We can prove the statement by utilizing Lemma 1 and showing that `PerfectPartition` is NP-hard by a reduction from 3-DM.

Based on instance of 3-DM, let $N = X \cup Y$, $R = Z$, and the preferences of players in N are as follows. For each $x \in X$, x most prefers coalitions $\{x, y, z\}$ if $(x, y, z) \in T$. Similarly, each $y \in Y$ most prefers coalitions $\{x, y, z\}$ if $(x, y, z) \in T$. Then, the claim is that we have a 'yes' instance of 3-DM if and only if we have a 'yes' instance of `PerfectPartition`.

If there exists a perfect three-dimensional matching, then the same matching also gives us a perfect partition. Now assume that there exists a perfect partition. Then, it is easy to see that there also exists a perfect three-dimensional matching. $\square$

Proof of Theorem 21.

The reduction is similar to the reduction in the proof of Theorem 20. The first difference is that each player has strict preferences among those allocations that were his favorite in the proof of Theorem 20. Secondly, all other allocations are considered unacceptable.

The partition π in question consists of singletons. Then, π is not weakly Pareto optimal if and only if there exists a partition in which each player is in a coalition strictly more preferred than the singleton coalition. The problem can be shown to be equivalent to 3-DM. □

Proof of Theorem 23.

We prove the statement by utilizing Lemma 1 and showing NP-hardness of `PerfectPartition` for anonymous games by a reduction from X3C.

For an instance (X, C) of X3C, order the elements in C in any arbitrary order $c_1, \ldots, c_{|C|}$. Define function f such that $f(c_i) = 2 + i$. The function f is clearly a one-to-one function. We now present a reduction which reduces an instance (X, C) to an anonymous game (N, R) in the following way.

- $N = X \cup \{y_1^i, \ldots, y_{i-1}^i : i \in \{1, \ldots, |C|\}\}$
- For each $x \in X$, consider the set $\bigcup_{c \in C}\{f(c) : x \in c\}$. Then player x has the elements in $\bigcup_{c \in C}\{f(c) : x \in c\}$ as his first choice coalition sizes, then size 1 as his second most preferred coalition size, and finds all other coalitions sizes unacceptable.
- For each $i \in \{1, \ldots, |C|\}$, the preferences of players $y_1^i, \ldots, y_{i-1}^i$ are identical. They equally prefer coalitions of size 1 and size $i + 2$ and all other coalitions sizes are unacceptable.

Observe that this is a polynomial-time reduction. The total number of players is $|C| + (|C - 1| \times |C - 2|)/2$ which is polynomial in the input size of (X, C). Furthermore, each player has a preference list of length less than or equal to four. Now the claim is that a favorite partition exists if and only if (X, C) is a 'yes' instance of X3C.

Assume that there exists a exists a perfect partition π for game (N, R). Then, π is such that each player is in a coalition of one of his most preferred sizes. This means that each player in X is in a coalition with a unique size (of three or more). Then, $C' = \{N \cap s : s \in \pi \ \wedge \ |S| \geq 3\}$ is a proper partition of X such that $C' \subseteq C$. Therefore (X, C) is a 'yes' instance of X3C.

We now show that if (X, C) is a 'yes' instance of X3C, then there exists a perfect partition. Let $C' \subseteq C$ be a proper partition of X. Then, consider the following partition π of N:

$$\pi = \{c_i \cup \{y_1^i, \ldots, y_{i-1}^i\} : c_i \in C'\} \cup \{\{y_1^i\}, \ldots, \{y_{i-1}^i\} : c_i \notin C'\}.$$

It is clear that π is a perfect partition for (N, R) in which each player in the set of coalitions $\{\{c_i \cup \{y_1^i, \ldots, y_{i-1}^i\}\} : c_i \in C'\}$ gets one of his favorite (non-singleton) coalitions and the players $\{\{y_1^i\}, \ldots, \{y_{i-1}^i\} : c_i \notin C'\}$ are in a singleton but nonetheless a favorite-sized coalition. This completes the proof. □

Proof of Theorem 24.

The proof idea is similar to that of Theorem 23 but with subtle differences. We take an instance of X3C and construct an anonymous hedonic game and a partition π consisting of the grand coalition. The partition π is not Pareto optimal if and only the X3C instance is a 'yes' instance.

For an instance (X, C) of X3C, order the elements in C in any arbitrary order $c_1, \ldots, c_{|C|}$. Define function f such that $f(c_i) = 2 + i$. The function f is clearly a one-to-one function. We now present a reduction which reduces instance (X, C) to an anonymous game (N, R) in the following way.

- $N = X \cup \{y_1^i, \ldots, y_{i-1}^i : i \in \{1, \ldots, |C|\}\}$
- For each $x \in X$, consider the set $\bigcup_{c \in C}\{f(c) : x \in c\}$ Then player x has the elements in $\bigcup_{c \in C}\{f(c) : x \in c\}$ as his first choice coalition sizes, then considers $|N|$ as the second choice, size 1 as the third choice, and finds all other coalitions sizes unacceptable.
- For each $i \in \{1, \ldots, |C|\}$, the preferences of players $y_1^i, \ldots, y_{i-1}^i$ are identical. They equally prefer coalitions of size 1, size $i + 2$ and size $|N|$, and find all other coalition sizes unacceptable.

The claim is that $\{N\}$ is not Pareto optimal if and only if (X, C) is a 'yes' instance of X3C.

Assume that $\{N\}$ is not Pareto optimal. Then, there exists a partition π such that each player is either in the grand coalition or a more preferred coalition. Since one player is certainly in a more preferred coalition, then the other players cannot remain in the grand coalition. This means that each player in X is in a coalition of a unique size (of three or more). Then, $C' = \{N \cap s : s \in \pi \wedge |S| \geq 3\}$ is a proper partition of X such that $C' \subseteq C$. Therefore (X, C) is a 'yes' instance of X3C.

We now show that if (X, C) is a 'yes' instance of X3C, then $\{N\}$ is not Pareto optimal. Let $C' \subseteq C$ be a proper partition of X. Then, consider the following partition π of N:

$$\pi = \{c_i \cup \{y_1^i, \ldots, y_{i-1}^i\} : c_i \in C'\} \cup \{\{y_1^i\}, \ldots, \{y_{i-1}^i\} : c_i \notin C'\}.$$

It is clear that π is a Pareto improvement over $\{N\}$ in which each players in the set of coalitions $\{\{c_i \cup \{y_1^i, \ldots, y_{i-1}^i\}\} : c_i \in C'\}$ gets a more preferred coalition size than $|N|$ and the players $\{\{y_1^i\}, \ldots, \{y_{i-1}^i\} : c_i \notin C'\}$ are in a singleton but nonetheless a favorite-sized coalition. Therefore, we know that $\{N\}$ is not Pareto optimal. This completes the proof. □