Stable Roommate Problem with Diversity Preferences [*]

Niclas Boehmer[1][†] , **Edith Elkind**[2]

[1]TU Berlin, Berlin, Germany
[2]University of Oxford, Oxford, UK
niclas.boehmer@tu-berlin.de, elkind@cs.ox.ac.uk

Abstract

In the multidimensional stable roommate problem, agents have to be allocated to rooms and have preferences over sets of potential roommates. We study the complexity of finding good allocations of agents to rooms under the assumption that agents have diversity preferences [Bredereck *et al.*, 2019]: each agent belongs to one of the two types (e.g., juniors and seniors, artists and engineers), and agents' preferences over rooms depend solely on the fraction of agents of their own type among their potential roommates. We consider various solution concepts for this setting, such as core and exchange stability, Pareto optimality and envy-freeness. On the negative side, we prove that envy-free, core stable or (strongly) exchange stable outcomes may fail to exist and that the associated decision problems are NP-complete. On the positive side, we show that these problems are in FPT with respect to the room size, which is not the case for the general stable roommate problem.

1 Introduction

Alice and Bob are planning their wedding. They have agreed on the gift registry and the music to be played, but they still need to decide on the seating plan for the wedding reception. They expect 120 guests, and the reception venue has 20 tables, with each table seating 6 guests. However, this task is far from being easy: e.g., Alice's great-aunt does not get along with Bob's family and prefers not to share the table with any of them; on the contrary, Bob's younger brother is keen to meet Alice's family and would be upset if he were stuck with his relatives. After spending an evening trying to find a seating plan that would keep everyone happy, Alice and Bob are on the brink of canceling the wedding altogether.

Bob's friend Charlie wonders if the hapless couple may benefit from consulting the literature on the *stable roommate problem*. In this problem, the goal is to find a stable assignment of $2n$ agents into n rooms of size 2, where every agent

---

has a preference relation over her possible roommates. The most popular notion of stability in this context is *core stability*: no two agents should strictly prefer each other to their current roommate. Another relevant notion is *exchange stability*: no two agents should want to swap their places. However, for the stable roommate problem, neither core stable nor exchange stable outcomes are guaranteed to exist. Further, while Irving [1985] proved that it is possible to decide in time linear in the size of the input if an instance of the roommate problem with strict preferences admits a core stable outcome, many other algorithmic problems for core and exchange stability are computationally hard [Cechlárová and Manlove, 2005; Ronn, 1990]. For the *s-dimensional stable roommate problem*, where each room has size s and agents have preferences over all $s-1$-subsets of agents as their potential roommates, even the core non-emptiness problem for strict preferences is NP-complete for $s \geq 3$ [Huang, 2007; Ng and Hirschberg, 1991].

However, Charlie then notes that Alice and Bob's problem has additional structure: the invitees can be classified as bride's family or groom's family, and it appears that all constraints on seating arrangements can be expressed in terms of this classification: each person only has preferences over the ratio of groom's relatives and bride's relatives at her table. Thus, the problem in question is closely related to *hedonic diversity games*, recently introduced by Bredereck *et al.* [2019]. These are coalition formation games where agents have *diversity preferences*, i.e., they are partitioned into two groups (say, *red* and *blue*), and every agent is indifferent among coalitions with the same ratio of red and blue agents. However, positive results for hedonic diversity games are not directly applicable to the roommate setting: in hedonic games, agents form groups of varying sizes, while the wedding guests have to be split into groups of six.

In this paper, we investigate the multidimensional stable roommate problem (for arbitrary room size s) with diversity preferences; we refer to the resulting problem as the *roommate diversity problem*. This model captures important aspects of several real-world group formation scenarios, such as flat-sharing, splitting students into teams for group projects and seating arrangements at important events. We consider common solution concepts from the literature on the stable roommate problem; for each solution concept, we analyze the complexity of checking if a given outcome is a valid solution,

whether the set of solutions is guaranteed to be non-empty, and, if not, how hard it is to check if an instance admits a solution as well as to compute a solution if it exists.

1.1 Our Contribution

We show that for room size two, every instance admits an outcome that is core stable, exchange stable and Pareto optimal. For $s > 2$, we provide counterexamples showing that core stable, exchange stable or envy-free outcomes may fail to exist. We also prove that for core stability, strong exchange stability and envy-freeness the existence questions are computationally hard; for Pareto optimality, we show that it is not only hard to find a Pareto optimal outcome, but also to verify whether a given outcome is Pareto optimal. We provide an overview of our results in Table 1.

On the positive side, we show that all existence questions we consider are in FPT with respect to the room size. To the best of our knowledge, apart from some work on the multidimensional stable roommate problem with cyclic preferences [Hofbauer, 2016], this is one of the first positive results for the multidimensional stable roommate problem. Thus, the roommate diversity problem offers an attractive combination of expressive power and computational tractability.

1.2 Related Work

The stable roommate problem was proposed by Gale and Shapley [1962]. While Irving [1985] proved that it is possible to check in time linear in the input whether a roommate problem admits a core stable outcome if the preferences are restricted to be strict, Ronn [1990] showed that this problem becomes NP-complete if ties in the preference relations are allowed. As in practice a group deviation usually requires some regrouping, Alcalde [1994] initiated the study of local stability notions that do not require reallocating non-deviating agents, by introducing the notion of exchange stability. Subsequently, Cechlárová and Manlove [2005] proved that it is NP-complete to decide whether an instance of the stable roommate problem with strict preferences admits an exchange stable outcome. Another concept that is relevant for the roommate problem is Pareto optimality [Abraham *et al.*, 2005; Cseh *et al.*, 2019; Sotomayor, 2011]. Morrill [2010] proved that for room size two it is possible to check if a given outcome is Pareto optimal, and to find a Pareto improvement if it exists, in time quadratic in the number of agents. This implies that a Pareto optimal outcome can be found in polynomial time. Researchers have also considered various notions of fairness in the context of the roommate problem [Abdulkadiroğlu and Sönmez, 2003; Aziz and Klaus, 2019]. One such notion is envy-freeness: an outcome is said to be *envy-free* if no agent wants to take the place of another agent.

While much of the work on the stable roommate problem focuses on the case $s = 2$, there are a few papers that consider the three-dimensional stable roommate problem [Huang, 2007; Ng and Hirschberg, 1991]. For different possible definitions of the agents' preferences, it is NP-complete to decide whether an instance with strict preferences admits a core stable outcome [Huang, 2007; Iwama *et al.*, 2007; Ng and Hirschberg, 1991]. For this reason, despite its great practical relevance, the multidimensional version of the roommate problem has not attracted much attention yet.

One possibility to circumvent these negative results is to search for reasonable subclasses of the stable roommate problem, i.e., to identify realistic restrictions on the agents' preferences for which the associated computational problems become tractable. This approach has been successful in the study of the two-dimensional stable roommate problem [Abraham *et al.*, 2007; Bartholdi III and Trick, 1986; Bredereck *et al.*, 2017; Chung, 2000; Cseh and Juhos, 2019], as well as in the context of hedonic games [Aziz *et al.*, 2019; Banerjee *et al.*, 2001; Bogomolnaia and Jackson, 2002]. In particular, Bredereck *et al.* [2019] and Boehmer and Elkind [2020a] analyzed the complexity of finding stable outcomes in hedonic diversity games for several notions of stability, such as Nash stability, individual stability and core stability. However, these results do not directly translate to our model: first, with the exception of core stability, the solution concepts we consider are different from those considered in hedonic diversity games, and second, the hard constraint on the room sizes in the roommate problem changes both the set of feasible allocations and the set of possible deviations.

Finally, we note that Bredereck *et al.* [2019] and Boehmer and Elkind [2020a] related hedonic games with diversity preferences to *anonymous hedonic games* [Bogomolnaia and Jackson, 2002], where the agents' preferences over coalitions only depend on coalition sizes. This connection proves useful in our setting as well: e.g., in our hardness reductions we use the fact that it is NP-complete to decide whether an anonymous hedonic game admits a Nash or core stable outcome [Ballester, 2004].

Diversity-related questions have also been studied in the context of stable matching problems [Huang, 2010; Kamada and Kojima, 2015]. However, our approach is fundamentally different: in our model, types are used to define agents' preferences rather than distributional constraints on the outcome.

Full version The full version of the paper is available on arXiv [Boehmer and Elkind, 2020b]. It contains all missing proofs and some preliminary results on extensions of our model, such as non-uniform room sizes and more than two types of agents.

2 Preliminaries

For every positive integer t, we denote the set $\{1, \ldots, t\}$ by $[t]$, and we write $[0, t]$ to denote $\{0\} \cup [t]$.

Definition 1. *A* roommate diversity problem *with room size s is a quadruple $G = (R, B, s, (\succsim_i)_{i \in R \cup B})$ with $N = R \cup B$ and $|N| = k \cdot s$ for some $k \in \mathbb{N}$. The preference relation $\succsim_i$ of each agent $i \in N$ is a weak order over the set $D = \{\frac{j}{s} : j \in [0, s]\}$.*

In the following, we call all agents in R *red* agents and all agents in B *blue* agents, and write $r = |R|$, $b = |B|$ and $n = |R \cup B|$. We refer to size-s subsets of N as *coalitions* or *rooms*; the quantity $k = \frac{|N|}{s}$ is then the number of rooms. For each $i \in N$, let $\mathcal{N}_i = \{S \subseteq N : |S| = s, i \in S\}$ denote all size-s subsets of N containing i, i.e., all possible rooms that i can be part of.

	unrestricted			strict			dichotomous		
	Gu.	Ex.	Co.	Gu.	Ex.	Co.	Gu.	Ex.	Co.
Core	✗ (2)	NPc (3)	NPh (3)	✗ (2)	NPc (3)	NPh (3)	✓ (4)	-	P (4)
Strong Core	✗ (2)	NPc (5)	NPh (5)	✗ (2)	NPc (5)	NPh (5)	✗ (5)	NPc (5)	NPh (5)
Same Ex.	✓ (6)	-	P (6)	✓ (6)	-	P (6)	✓ (6)	-	P (6)
Exch.	✗ (7)	?	?	✗ (7)	?	?	?	?	?
Strong Ex.	✗ (7)	NPc (8)	NPh (8)	✗ (7)	?	?	✗ (7)	NPc (8)	NPh (8)
Pareto	✓	-	NPh (9)	✓	-	?	✓	-	NPh (9)
Envy-free	✗ (10)	NPc (12)	NPh (12)	✗ (10)	NPc (12)	NPh (12)	✗ (10)	NPc (12)	NPh (12)

Table 1: Overview of complexity results for different solution concepts and restrictions on the preference relations. For each solution concept and restriction, we indicate whether every instance satisfying this restriction is guaranteed to admit an outcome with the respective property (Gu.), the complexity of deciding if an instance admits such an outcome (Ex.) and of finding one if it exists (Co.). The number in brackets is the number of the respective theorem. For all solution concepts, the problem of verifying whether a given outcome has the desired property is in P except for Pareto optimality, for which this problem is coNP-complete. We prove that all existence problems in this table are in FPT with respect to the room size.

An *outcome* of G is a partition of all agents into k rooms $\pi = \{C_1, \ldots, C_k\}$ such that $|C_i| = s$ for all $i \in [k]$. Let $\pi(i)$ denote i's coalition in π. Given a coalition $C \subseteq N$, let $\theta(C)$ denote the fraction of red agents in C, i.e., $\theta(C) = \frac{|C \cap R|}{|C|}$: we say that C *is of fraction* $\theta(C)$. A coalition $C \subseteq N$ is called *pure* if $C \subseteq R$ or $C \subseteq B$ and *mixed* otherwise.

For each agent $i \in N$, we interpret her preference relation $\succsim_i$ over D as her preferences over the fraction of red agents in her coalition; for instance, $\frac{2}{5} \succsim_i \frac{3}{5}$ means that agent i prefers a room where two out of five agents are red to a room where three out of five agents are red.[1] Thereby, $\succsim_i$ induces agent i's preferences over all possible rooms she can be part of.[2] Given two rooms $S, T \in \mathcal{N}_i$, overloading notation, we write $S \succ_i T$ and say that i *strictly prefers* S to T if $\theta(S) \succsim_i \theta(T)$ and $\theta(T) \not\succsim_i \theta(S)$. Further, we write $S \succsim_i T$ and say that i *weakly prefers* S to T if $\theta(T) = \theta(S)$ or $\theta(T) \succsim_i \theta(S)$. If i weakly prefers S to T and T to S, we write $S \sim_i T$ and say that i is *indifferent* between S and T.

The preference relation of agent i is said to be *single-peaked* if there exists a peak $p_i \in D$ such that for all $\alpha, \beta \in D$ such that $p_i \leq \alpha < \beta$ or $\beta < \alpha \leq p_i$ it holds that $\alpha \succsim_i \beta$. The preference relation of i is said to be *dichotomous* if it is possible to partition D into two sets D^+ and D^- so that for all $d \in D^+, d' \in D^-$ it holds that $d \succ_i d'$, for all $d, d' \in D^-$ it holds that $d \sim_i d'$ and for all $d, d' \in D^+$ it holds that $d \sim_i d'$. We say that i *approves* all fractions in D^+ and *disapproves* all fractions in D^-.

Generalizing the definition of Gale and Shapley [1962], we say that a coalition $S \subseteq N$ with $|S| = s$ *blocks* an outcome

π if for all $i \in S$ it holds that $\theta(S) \succ_i \theta(\pi(i))$; we say that S *weakly blocks* an outcome π if for all $i \in S$ it holds that $\theta(S) \succsim_i \theta(\pi(i))$ and there exists an $i \in S$ such that $\theta(S) \succ_i \theta(\pi(i))$. An outcome π is called *(strongly) core stable* if no coalition (weakly) blocks it.

In an outcome π, a pair of agents $i, j \in N$ with $\pi(i) \neq \pi(j)$ has an *exchange deviation* if they would like to exchange places, i.e., $\theta\big((\pi(j) \setminus \{j\}) \cup \{i\}\big) \succ_i \theta\big(\pi(i)\big)$ and $\theta\big((\pi(i) \setminus \{i\}) \cup \{j\}\big) \succ_j \theta\big(\pi(j)\big)$. Further, a pair of agents $i, j \in N$ with $\pi(i) \neq \pi(j)$ has a *weak exchange deviation* if $\theta\big((\pi(j) \setminus \{j\}) \cup \{i\}\big) \succ_i \theta\big(\pi(i)\big)$ and $\theta\big((\pi(i) \setminus \{i\}) \cup \{j\}\big) \succsim_j \theta\big(\pi(j)\big)$. An outcome is called *(strongly) exchange stable* if no pair of agents has a (weak) exchange deviation [Alcalde, 1994].

An outcome π is called *Pareto optimal* if there is no other outcome that makes all agents weakly better off and some agents strictly better off, i.e., there is no outcome π' such that $\theta(\pi'(i)) \succsim_i \theta(\pi(i))$ for all $i \in N$ and $\theta(\pi'(i)) \succ_i \theta(\pi(i))$ for some $i \in N$.

An outcome π is called *envy-free* if there does not exist a pair of agents $i, j \in N$ with $\pi(i) \neq \pi(j)$ such that i envies j's place, i.e., $\theta\big((\pi(j) \setminus \{j\}) \cup \{i\}\big) \succ_i \theta\big(\pi(i)\big)$.

Due to space constraints, we defer several proofs to the full version of the paper.

3 Roommate Diversity Problem With Room Size Two

For $s = 2$, the roommate diversity problem becomes a special case of the classical stable roommate problem. Our first observation is that diversity preferences make the classical problem significantly easier: we can efficiently find an outcome that is Pareto optimal, core stable and exchange stable. This result motivates us to focus on the case $s > 2$ in the remainder of the paper.

Theorem 1. *Every instance of the roommate diversity problem with room size two admits an outcome that is Pareto optimal, core stable and exchange stable, even if we allow for indifferences in the preferences.*

Proof sketch. We say that a mixed pair is *happy* if both agents in the pair weakly prefer a mixed pair to a pure pair. The

---

[1] We could equivalently define the agents' preferences over the number of red agents in each room; we chose the ratio-based definition for consistency with the hedonic diversity games literature and to emphasize the room size.

[2] For succinctness and consistency with prior work, we assume that each agent has preferences over the entire set D, including 0 and 1, even though a red agent cannot be part of a room with ratio 0 and a blue agent cannot be part of a room with ratio 1. Allowing agents to have preferences over 'impossible' ratios has no impact on our results: even if, say, a blue agent ranks 1 highly, she cannot deviate to a coalition with this ratio. In all of our examples, the impossible ratios are ranked at the bottom of agent's preferences.

general idea of the algorithm is to first create as many happy mixed pairs as possible. The algorithm then attempts to put the remaining agents in pure pairs. If at this point the number of blue agents is odd, this is not possible. In this case, depending on the preferences of agents in mixed pairs, it either inserts an additional mixed pair or breaks up one of the mixed pairs to create two pure pairs: one red and one blue. □

4 (Strong) Core Stability

We have seen that for the roommate diversity problem with $s = 2$ a core stable outcome is guaranteed to exist. However, for larger values of s this is not the case.

Theorem 2. *An instance of the roommate diversity problem may fail to have a core stable outcome, even if no indifferences in the preferences are allowed.*

Proof. Let $G = (\{r_1, r_2, r_3, r_4\}, \{b_1, b_2, b_3, b_4\}, 4, (\succsim_i)_{i \in N})$ with

$$r_1, r_2, r_3 : \frac{2}{4} \succ \frac{4}{4} \succ \frac{3}{4} \succ \frac{1}{4} \succ \frac{0}{4}; r_4 : \frac{4}{4} \succ \frac{1}{4} \succ \frac{2}{4} \succ \frac{3}{4} \succ \frac{0}{4}$$

$$b_1, b_2, b_3, b_4 : \frac{1}{4} \succ \frac{2}{4} \succ \frac{3}{4} \succ \frac{0}{4} \succ \frac{4}{4}.$$

Assume for the sake of contradiction that G has a core stable outcome π. If π consists of two coalitions of fraction $\frac{2}{4}$, $\{r_4, b_1, b_2, b_3\}$ is blocking. If π consists of one coalition of fraction $\frac{3}{4}$ and one coalition of fraction $\frac{1}{4}$, $\{r_1, r_2, r_3, r_4\}$ is blocking. If π consists of a purely blue and a purely red coalition, $\{r_1, r_2, b_1, b_2\}$ is blocking. □

Further, if we assume that the room size is flexible and part of the input, the associated existence question becomes NP-complete.

Theorem 3. *It is* NP-*complete to decide whether a given instance of the roommate diversity problem admits a core stable outcome. The hardness result holds even if no indifferences in the preferences are allowed.*

Proof sketch. To check whether an outcome is core stable, we iterate over all $j \in [0, s]$ and check if there are j red agents and $s - j$ blue agents preferring $\frac{j}{s}$ to the fraction of their current coalition. To prove hardness, we reduce from the problem of deciding whether an anonymous hedonic game admits a core stable outcome [Ballester, 2004]. Given an anonymous hedonic game G' with n agents, we build an n-dimensional instance G of the roommate diversity problem. To this end, for each agent i in G', we introduce a red agent r_i whose preference relation is constructed by replacing each size $\ell \in [n]$ in the preferences of i with fraction $\frac{\ell}{n}$. Moreover, for all $\ell \in [n]$, we introduce n^2 blue agents preferring $\frac{\ell}{n}$ to every other fraction, while they rank all other fractions arbitrarily. To construct an outcome of G from an outcome π' of G', for each coalition $S \in \pi'$, we insert a coalition consisting of the corresponding red agents $\{r_i : i \in S\}$ and $n - |S|$ blue agents with $\frac{|S|}{n}$ as their top fraction. If a coalition T blocks π' in G', the coalition consisting of red agents $\{r_i : i \in T\}$ and $n - |T|$ blue agents with $\frac{|T|}{n}$ as their top fraction is a blocking coalition for the respective outcome of G. The other direction works analogously. □

Using the construction in the proof of Theorem 3, we can map the single-peaked anonymous hedonic game with empty core constructed by Banerjee *et al.* [2001] to a single-peaked instance of the roommate diversity problem with an empty core. We obtain the following corollary.

Corollary 1. *An instance of the roommate diversity problem may fail to have a core stable outcome, even if all agents' preferences are single-peaked.*

In contrast, if agents' preferences are dichotomous, the core is non-empty, and an outcome in the core can be computed efficiently: following the approach of Peters [2016], to construct a core stable outcome, we iterate over all fractions $\frac{\ell}{s}$ for $\ell \in [0, s]$ and add the maximum possible number of rooms consisting of ℓ red agents and $s - \ell$ blue agents who all approve $\frac{\ell}{s}$. The rest of the agents are split into the remaining rooms. We obtain the following result.

Theorem 4. *Every instance of the roommate diversity problem with dichotomous preferences admits a core stable outcome; moreover, an outcome in the core can be computed in polynomial time.*

However, this positive result does not extend to the more demanding notion of strong core stability.

Theorem 5. *It is* NP-*complete to decide whether a given roommate diversity problem admits a strongly core stable outcome, even if the preferences are restricted to be dichotomous and every agent approves at most four fractions.*

Proof sketch. Peters [2016] showed that the corresponding problem for anonymous hedonic games is NP-complete. By reducing from this problem, we can establish that our problem is NP-hard as well; the reduction is similar to the one used in the proof of Theorem 3. □

5 (Strong) Exchange Stability

As pointed out in the introduction, it is not always plausible to assume that agents are allowed to perform group deviations. Therefore, in this section, we focus on stability concepts that are defined in terms of agent swaps.

5.1 Same-Type Swaps

If the set of possible deviations is limited to agent swaps, it may be the case that only agents of the same type are allowed to swap their places: we call the resulting stability notion *same-type-exchange stability*. For example, if a professor forms fixed-size teams for a group project, she may want to fix the fraction of graduate students in each group (e.g., to ensure that the experienced students are equally distributed) but still allow for swaps between two undergraduate students or between two graduate students. Under this weaker version of exchange stability, the agents are guaranteed to eventually converge to a stable outcome.

Theorem 6. *Every instance of the roommate diversity problem has a (strongly) same-type-exchange stable outcome, and some such outcome can be computed in polynomial time.*

Proof. To compute a (strongly) same-type-exchange stable outcome, we start at an arbitrary outcome π and swap pairs

who have a (weak) same-type-exchange deviation until this is no longer possible. To see that this procedure always terminates, note that same-type exchange deviations do not change the fraction of red agents in any coalition. Thereby, nothing changes for agents who are not involved in the swap, while both agents involved in the swap weakly improve and at least one of them strictly improves. As every agent's preference relation is defined over $s + 1$ elements, the total number of swaps is at most ns. □

5.2 Unrestricted Swaps

Unfortunately, the existence guarantee for same-type swaps does not extend to unrestricted swaps.

Theorem 7. *An instance of the roommate diversity problem may fail to have an exchange stable outcome, even if the preferences are single-peaked and no indifferences in the preferences are allowed.*

Proof sketch. The following single-peaked roommate diversity game $G = (\{r_1, r_2, r_3, r_4, r_5\}, \{b_1, b_2, b_3, b_4\}, 3, (\succsim_i)_{i \in N})$ with:

$$r_1, r_2 : \frac{3}{3} \succ \frac{2}{3} \succ \frac{1}{3} \succ \frac{0}{3}; \ r_3, r_4, r_5 : \frac{2}{3} \succ \frac{1}{3} \succ \frac{3}{3} \succ \frac{0}{3};$$
$$b_1 : \frac{0}{3} \succ \frac{1}{3} \succ \frac{2}{3} \succ \frac{3}{3}; \ b_2, b_3, b_4 : \frac{1}{3} \succ \frac{2}{3} \succ \frac{0}{3} \succ \frac{3}{3}.$$

does not admit an exchange-stable outcome. □

Further, the associated existence problem for strong exchange stability is NP-complete.

Theorem 8. *It is NP-complete to decide whether a given instance of the roommate diversity problem admits a strongly exchange stable outcome. The hardness result still holds if the preferences are dichotomous and every agent approves at most four fractions.*

Proof sketch. To show hardness, we reduce from the problem of deciding whether an anonymous hedonic game admits a Nash stable outcome [Ballester, 2004; Peters, 2016]; recall that an outcome of a hedonic game is said to be *Nash stable* if no agent wants to move from her current coalition to another existing coalition or to form a singleton coalition. Our reduction is similar to the one in the proof of Theorem 3: we map an anonymous hedonic game G' with n agents to an n-dimensional instance of the roommate diversity problem G so that an agent i in G' is mapped to a red agent in G whose preference relation is obtained from that of i by replacing ℓ with $\frac{\ell}{n}$ for all $\ell \in [1, n]$; however, here, we introduce $n^2 - n$ blue agents who are indifferent among all fractions. □

We believe that deciding whether a roommate diversity problem admits an exchange stable outcome is NP-complete as well, but we were unable to extend the proof of Theorem 8 to show this.

6 Pareto Optimality

In the roommate problem, Pareto optimality emerges as a natural notion of stability. Indeed, an outcome is not Pareto optimal if and only if we can rearrange the agents so that all of them are weakly better off and at least some of them are strictly better off, i.e., there is a weakly improving deviation by the grand coalition [Elkind *et al.*, 2016; Morrill, 2010].

While for many other stability concepts we consider stable outcomes are not guaranteed to exist, by the definition of Pareto optimality, every instance of the roommate diversity problem admits a Pareto optimal outcome. Indeed, we can start at an arbitrary outcome and perform a sequence of at most ns Pareto improvements, i.e., rearrangements of the agents that make all agents weakly better off and some agents strictly better off. However, it is still computationally hard to compute a Pareto optimal outcome, as finding a Pareto improvement is difficult.

Theorem 9. *For the unrestricted roommate diversity problem, it is* coNP-*complete to decide whether a given outcome is Pareto optimal; moreover, we cannot compute a Pareto optimal outcome in polynomial time unless* P=NP. *These results hold even if preferences are dichotomous.*

Proof sketch. To show that an outcome is not Pareto optimal, it suffices to guess a Pareto improvement and verify that it indeed makes all agents weakly better off and some agents strictly better off; this establishes that our decision problem is in coNP. To prove hardness for the unrestricted case, we construct reductions from the related problems for anonymous hedonic games, which are NP-hard, as proven by Aziz *et al.* [2013]. The reductions are similar to the reduction in the proof of Theorem 8. □

7 Envy-Freeness

We can think of envy-freeness as a "one-sided" version of exchange stability. Thus, similarly to same-type-exchange stability, we can define same-type-envy-freeness by only considering envy among agents of the same type. Same-type-envy-freeness is a plausible variant of envy-freeness, as people tend to envy those who are similar to them [Salovey and Rodin, 1984]. Moreover, same-type-envy-freeness is also an appealing notion of fairness: if agent i and agent j are of the same type and i envies j, swapping i and j has no effect on other agents, so the decision which of these agents should get a better set of roommates is essentially arbitrary. Unfortunately, an outcome that is fair in this sense is not guaranteed to exist.

Theorem 10. *There exists an instance of the roommate diversity problem with room size two that has no same-type-envy-free outcome and thereby also no envy-free outcome. Moreover, in this instance the agents' preferences are single-peaked and dichotomous.*

Proof. Let $G = (\{r_1\}, \{b_1, b_2, b_3\}, 2, (\succsim_i)_{i \in N})$ with:

$$r_1 : \frac{2}{2} \succ \frac{1}{2} \succ \frac{0}{2}; \qquad b_1, b_2, b_3 : \frac{0}{2} \succ \frac{1}{2} \succ \frac{2}{2}.$$

This game is clearly single-peaked and can be transformed into a dichotomous game by putting the two bottom fractions

for each agent in the same equivalence class. As every outcome of this game consists of one mixed and one purely blue coalition, the blue agent in the mixed coalition always envies the two blue agents in the pure coalition. □

On the positive side, there exist two special cases where the existence of a same-type-envy-free outcome is guaranteed.

Theorem 11. *A same-type-envy-free outcome is guaranteed to exist if the number of red agents is divisible by s or by k.*

Proof. If s divides r, there exists an outcome consisting of pure coalitions only. If k divides r, there exists an outcome where the fraction of each coalition is $\frac{r}{k}$. In either case, all agents of the same type are in coalitions of the same fraction, so no agent envies another agent of her type. □

Nevertheless, it can be proven by a reduction from EXACT COVER BY 3-SETS that the general existence question for envy-freeness is NP-complete.

Theorem 12. *It is* NP-*complete to decide whether a given instance of the roommate diversity problem admits an envy-free outcome, even if no indifferences in the preference relations are allowed. This hardness result also holds if the agents' preferences are dichotomous and every agent is allowed to approve at most four fractions.*

We were not able to extend Theorem 12 to same-type-envy-freeness, but conjecture that the hardness result still holds. However, if preferences are strict, there exists a simple algorithm that solves this problem in time linear in n and single-exponential in s, i.e., this problem is in FPT with respect to s.

Theorem 13. *Given an instance of the roommate diversity problem with room size s and strict preferences, it is possible to check in time $\mathcal{O}(ns2^s)$ whether this instance admits a same-type-envy-free outcome and to find one if it exists.*

8 Parameterized Analysis

In Section 3, we saw that fixing the size of the rooms to $s = 2$ has a significant impact on the complexity of finding stable outcomes. Motivated by this result, as well as by Theorem 13, in this section, we study the parameterized complexity of the roommate diversity problem with respect to parameter s. This is a promising parameter, since in most of our hardness reductions we converted an anonymous hedonic game with n agents into an instance with room size n; it is also appealing because in practice the room size can be much smaller than the number of agents. Indeed, most of the algorithmic problems considered in this work turn out to be in FPT with respect to s. We start by considering (strong) core stability.

Theorem 14. *The problem of determining whether an instance of the roommate diversity problem admits a (strongly) core stable outcome is fixed-parameter tractable with respect to the room size.*

Proof sketch. As every agent is fully characterized by her preference relation and type, the number of distinct agents can be upper-bounded by $2 * 2^{(s+1)^2}$. For each such agent class, we introduce $s + 1$ variables denoting the number of

agents from this class that are put into a room with $i \in [0, s]$ red agents. Using these variables, we can check whether they induce a valid outcome and whether this outcome is (strongly) core stable by appropriately selected linear constraints. Thereby, it is possible to formalize this problem as an Integer Linear Program (ILP) and use the algorithm of Lenstra Jr [1983] that solves an ILP with ρ variables and input length L in time $\mathcal{O}(\rho^{2.5\rho+o(\rho)}L)$. □

This approach, with appropriate modifications, extends to (strong) exchange stability and envy-freeness.

Theorem 15. *The problem of determining whether an instance of the roommate diversity problem admits a (strongly) exchange stable outcome is fixed-parameter tractable with respect to the room size.*

Theorem 16. *The problem of determining whether an instance of the roommate diversity problem admits an envy-free outcome is fixed-parameter tractable with respect to the room size.*

Theorems 14–16 highlight a fundamental difference between the classical stable roommate problem and the roommate problem with diversity preferences: for the former all studied existence problems are NP-complete for $s = 3$ or even for $s = 2$, while for the latter many of these problems are polynomial-time solvable if the size of the rooms is fixed or bounded. Thus, assuming diversity preferences fundamentally changes the complexity of the roommate problem. We note, however, that we were unable to extend our ILP techniques to questions concerning Pareto optimality; it remains an open problem whether these questions are also in FPT with respect to the room size s.

9 Conclusions And Future Directions

In this paper, we have proposed the roommate diversity problem, which is an interesting special case of the multidimensional stable roommate problem. We have initiated the algorithmic study of this problem by considering various standard solution concepts and analyzing the existence of stable outcomes and the complexity of computing them (see Table 1). While we have answered many questions that arise in this context, the complexity of deciding whether a roommate diversity problem admits an exchange stable outcome remains open. Moreover, it is unclear if every instance with dichotomous preferences admits an exchange stable outcome. By parameterizing our computational problems by the size of the rooms, we showed that diversity preferences are a powerful restriction, as all studied existence problems lie in FPT with respect to this parameter. However, it would be desirable to obtain parameterized algorithms that are combinatorial rather than ILP-based, since ILP-based algorithms tend to be slow in practice.

Acknowledgements

References

Atila Abdulkadiroğlu and Tayfun Sönmez. School choice: A mechanism design approach. *American Economic Review*, 93(3):729–747, 2003.

David J. Abraham, Péter Biró, and David Manlove. "almost stable" matchings in the roommates problem. In *WAOA'05*, pages 1–14, 2005.

David J Abraham, Ariel Levavi, David F Manlove, and Gregg O'Malley. The stable roommates problem with globally-ranked pairs. In *WINE'07*, pages 431–444, 2007.

José Alcalde. Exchange-proofness or divorce-proofness? Stability in one-sided matching markets. *Economic Design*, 1(1):275–287, 1994.

Haris Aziz and Bettina Klaus. Random matching under priorities: Stability and no envy concepts. *Social Choice and Welfare*, 53(2):213–259, 2019.

Haris Aziz, Felix Brandt, and Paul Harrenstein. Pareto optimality in coalition formation. *Games and Economic Behavior*, 82:562–581, 2013.

Haris Aziz, Florian Brandl, Felix Brandt, Paul Harrenstein, Martin Olsen, and Dominik Peters. Fractional hedonic games. *ACM Transactions on Economics and Computation*, 7(2):6:1–6:29, 2019.

Coralio Ballester. NP-completeness in hedonic games. *Games and Economic Behavior*, 49:1–30, 2004.

Suryapratim Banerjee, Hideo Konishi, and Tayfun Sönmez. Core in a simple coalition formation game. *Social Choice and Welfare*, 18(1):135–153, 2001.

John Bartholdi III and Michael A Trick. Stable matching with preferences derived from a psychological model. *Operations Research Letters*, 5(4):165–169, 1986.

Niclas Boehmer and Edith Elkind. Individual-based stability in hedonic diversity games. In *AAAI'20*, 2020.

Niclas Boehmer and Edith Elkind. Stable roommate problem with diversity preferences, 2020. arXiv 2004.14640.

Anna Bogomolnaia and Matthew O Jackson. The stability of hedonic coalition structures. *Games and Economic Behavior*, 38(2):201–230, 2002.

Robert Bredereck, Jiehua Chen, Ugo Paavo Finnendahl, and Rolf Niedermeier. Stable roommate with narcissistic, single-peaked, and single-crossing preferences. In *ADT'17*, pages 315–330, 2017.

Robert Bredereck, Edith Elkind, and Ayumi Igarashi. Hedonic diversity games. In *AAMAS'19*, pages 565–573, 2019.

Katarína Cechlárová and David F Manlove. The exchange-stable marriage problem. *Discrete Applied Mathematics*, 152(1-3):109–122, 2005.

Kim-Sau Chung. On the existence of stable roommate matchings. *Games and Economic Behavior*, 33(2):206–230, 2000.

Ágnes Cseh and Attila Juhos. Pairwise preferences in the stable marriage problem. In *STACS'19*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

Ágnes Cseh, Tamás Fleiner, and Petra Harján. Pareto optimal coalitions of fixed size, 2019. arXiv 1901.06737.

Edith Elkind, Angelo Fanelli, and Michele Flammini. Price of Pareto optimality in hedonic games. In *AAAI'16*, pages 475–481, 2016.

David Gale and Lloyd S Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

Johannes Hofbauer. d-dimensional stable matching with cyclic preferences. *Mathematical Social Sciences*, 82:72–76, 2016.

Chien-Chung Huang. Two's company, three's a crowd: Stable family and threesome roommates problems. In *ESA'07*, pages 558–569, 2007.

Chien-Chung Huang. Classified stable matching. In *SODA'10*, pages 1235–1253, 2010.

Robert W Irving. An efficient algorithm for the "stable roommates" problem. *Journal of Algorithms*, 6(4):577–595, 1985.

Kazuo Iwama, Shuichi Miyazaki, and Kazuya Okamoto. Stable roommates problem with triple rooms. In *WAAC'07*, pages 105–112, 2007.

Yuichiro Kamada and Fuhito Kojima. Efficient matching under distributional constraints: Theory and applications. *American Economic Review*, 105(1):67–99, 2015.

Hendrik W. Lenstra Jr. Integer programming with a fixed number of variables. *Mathematics of Operations Research*, 8(4):538–548, 1983.

Thayer Morrill. The roommates problem revisited. *Journal of Economic Theory*, 145(5):1739–1756, 2010.

Cheng Ng and Daniel S Hirschberg. Three-dimensional stable matching problems. *SIAM Journal on Discrete Mathematics*, 4(2):245–252, 1991.

Dominik Peters. Complexity of hedonic games with dichotomous preferences. In *AAAI'16*, pages 579–585, 2016.

Eytan Ronn. NP-complete stable matching problems. *Journal of Algorithms*, 11(2):285–304, 1990.

Peter Salovey and Judith Rodin. Some antecedents and consequences of social-comparison jealousy. *Journal of Personality and Social Psychology*, 47(4):780, 1984.

Marilda Sotomayor. The Pareto-stability concept is a natural solution concept for discrete matching markets with indifferences. *International Journal of Game Theory*, 40(3):631–644, 2011.