Hedonic Diversity Games

Robert Bredereck
TU Berlin
Berlin, Germany
robert.bredereck@tu-berlin.de

Edith Elkind
University of Oxford
Oxford, U.K.
elkind@cs.ox.ac.uk

Ayumi Igarashi
Kyushu University
Fukuoka, Japan
igarashi@agent.inf.kyushu-u.ac.jp

ABSTRACT

We consider a coalition formation setting where each agent belongs to one of the two types, and agents' preferences over coalitions are determined by the fraction of the agents of their own type in each coalition. This setting differs from the well-studied Schelling's model in that some agents may prefer homogeneous coalitions, while others may prefer to be members of a diverse group, or a group that mostly consists of agents of the other type. We model this setting as a hedonic game and investigate the existence of stable outcomes using hedonic games solution concepts. We show that a core stable outcome may fail to exist and checking the existence of core stable outcomes is computationally hard. On the other hand, we propose an efficient algorithm to find an individually stable outcome under the natural assumption that agents' preferences over fractions of the agents of their own type are single-peaked.

KEYWORDS

Hedonic games; Schelling segregation; fractional hedonic games

ACM Reference Format:
Robert Bredereck, Edith Elkind, and Ayumi Igarashi. 2019. Hedonic Diversity Games. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 9 pages.

1 INTRODUCTION

At a conference dinner, researchers split into groups to chat over food. Some junior researchers prefer to stay in the company of other junior researchers, as they want to relax after a long day of talks. Some senior researchers prefer to chat with their friends, who also happen to be senior researchers. But there are also junior researchers who want to use the dinner as an opportunity to network with senior researchers, as well as senior researchers who are eager to make the newcomers feel welcome in the community, and therefore want to talk to as many junior people as possible.

This example can be viewed as an instance of a coalition formation problem. The agents belong to two types (senior and junior), and their preferences over coalitions are determined by the fraction of agents of each type in the coalition. This setting is reminiscent of the classic Schelling model [25], but there is an important difference: a standard assumption in the Schelling model is *homophily*, i.e., the agents are assumed to prefer to be surrounded by agents of their own type, though they can tolerate the presence of agents of the other type, as long as their fraction does not exceed a prespecified threshold. In contrast, in our example some agents have homophilic preferences, while others have *heterophilic preferences*, i.e., they seek out coalitions with agents who are not like them.

There is a very substantial body of research on homophily and heterophily in group formation. It is well-known that in a variety of contexts, ranging from residential location [25, 28, 29] to classroom activities and friendship relations [22, 23], individuals prefer to be together with those who are similar to them. There are also settings where agents prefer to be in a group with agents of the other type(s): for instance, in a coalition of buyers and sellers, a buyer prefers to be in a group with many other sellers and no other buyers, so as to maximize their negotiating power. Aziz et al. [3, 4] model this scenario as a *Bakers and Millers game*, where a baker wants to be in a coalition with many millers, whereas a miller wants to be in a coalition with many bakers. However, there are also real-life scenarios where agents can have different attitudes towards diversity: this includes, for instance, language learning by immersion (with types being learners' native languages), shared accommodation (with types being genders), primary and secondary education (with types being races and income groups), etc. In all these settings we expect the agents to display a broad range of preferences over ratios of different types in their group.

Our contribution The goal of our paper is to provide strategic foundations for the study of coalition formation scenarios where each agent may have a different degree of homophily. Specifically, we consider settings where agents are divided into two types (blue and red), and each agent has preferences regarding the fraction of the agents of her own type, which determines her preferences over coalitions. For most of our results, we assume that agents' preferences are single-peaked, i.e., each agent has a preferred ratio θ of agents of her own type, and prefers one fraction θ_1 to another fraction θ_2 if θ_1 is closer to θ than θ_2 is. Our model allows agents to express a variety of preferences including both complete homophily and complete heterophily.

We model this setting as a hedonic game, and investigate the existence of stable outcomes according to several established notions of stability, such as core stability, Nash stability and individual stability [6, 9]. We demonstrate that a core stable outcome may fail to exist, even when all agents have single-peaked preferences. Moreover, we show that deciding whether a core stable outcome exists is NP-complete. However, we identify several interesting special cases where the core is guaranteed to be non-empty.

We then consider stability notions that are defined in terms of individual deviations. While a Nash stable outcome may fail to exist, we propose an efficient algorithm to reach an individually stable outcome, i.e., an outcome where if some agent would like to deviate from her current coalition to another coalition, at least one agent in the target coalition would object to the move. Our proof employs a careful and non-trivial adaptation of the algorithm of Bogomolnaia and Jackson [9] for single-peaked anonymous games. Our algorithm

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada. © 2019 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

is *decentralized* in the sense that, by following a certain set of rules, the agents can form a stable partition by themselves.

Related work Our work is related to an established line of research that studies the impact of homophily on residential segregation. The seminal paper of Thomas Schelling [25] introduced a model of residential segregation in which two types of individuals are located on a line, and at each step a randomly chosen individual moves to a different location if the fraction of the like-minded agents in her neighborhood is below her tolerance ratio. With simple experiments using dimes and pennies, Schelling [25] found that any such dynamics almost always results in total segregation even if each agent only has a mild preference for her own type.

Following numerous papers empirically confirming Shelling's result [1, 13, 14, 17, 18, 21, 26], Young [27] was the first to provide a rigorous theoretical argument, showing that stability can only be achieved if agents are divided into homogeneous groups. In contrast, Brandt et al. [10] showed that with tolerance parameter being exactly $\frac{1}{2}$, the average size of the monochromatic community is independent of the size of the whole system; subsequently, Immorlica et al. [20] extended this analysis to the two-dimensional grid. Recent papers of Chauhan et al. [12] and Elkind et al. [16] consider game-theoretic variants of this model, which take into account both the fraction of the like-minded agents in the neighborhood and agents' preferences for being close to specific locations, and investigate existence and quality of Nash equilibria.

We note, however, that our model is fundamentally different from Shelling's model, for at least three reasons. First, we do not assume any underlying topology that restricts coalition formation. Second, the coalitions in an outcome of a hedonic game are pairwise disjoint, while the neighborhoods in the Schelling model may overlap. Finally, as argued above, our model does not assume homophilic preferences.

There is also a substantial literature on stability in hedonic games, starting with the early work of Bogomolnaia and Jackson [9]. Among the various classes of hedonic games, two classes are particularly relevant for our analysis: fractional hedonic games [3, 4] and anonymous hedonic games [9].

In fractional hedonic games, the agents are located on a social network, and they prefer a coalition C to a coalition C' if the fraction of their friends in C is higher than in C'. The Bakers and Millers game is an example of a fractional hedonic game, where it is assumed that each baker is a friend of each miller, but no two agents of the same type are friends. Aziz et al. [3, 4], and, subsequently, Bilò et al. [7, 8] identify several special cases of fractional hedonic games where the set of core stable outcomes is non-empty. In particular, Aziz et al. [3, 4] give a characterization of the set of strictly core stable outcomes in the Bakers and Millers game.

In anonymous hedonic games, agents' preferences over coalitions depend on the size of these coalitions only. Similarly to our setting, it is known that with single-peaked anonymous preferences, there is a natural decentralized process to reach individual stability; however, a core stable outcome may fail to exist [9], and deciding the existence of a core stable outcome is NP-complete [5].

There are also other subclasses of hedonic games where stable outcomes are guaranteed to exist, such as acyclic hedonic games [15, 19], dichotomous games [24], and top-responsive games [2].

Full version. The full version of the paper is available on arXiv [11]. It contains the proofs of Theorem 3.3, Proposition 3.4, and Proposition 3.5 which are omitted from this version due to space constraints.

2 OUR MODEL

For every positive integer s, we denote by $[s]$ the set $\{1, \ldots, s\}$. We start by defining the class of games that we are going to consider.

Definition 2.1. A *diversity game* is a triple $G = (R, B, (\succ_i)_{i \in R \cup B})$, where R and B are disjoint sets of *agents* and for each agent $i \in R \cup B$ it holds that $\succ_i$ is a linear order over the set

$$\Theta = \left\{ \frac{r}{s} \;\middle|\; r \in \{0, 1, \ldots, |R|\}, s \in \{1, \ldots, |R| + |B|\} \right\}.$$

We set $N = R \cup B$; the agents in R are called the *red agents*, and the agents in B are called the *blue agents*.

We refer to subsets of N as *coalitions*. For each $i \in N$, we denote by $\mathcal{N}_i$ the set of coalitions containing i. For each coalition $S \subseteq N$, we say that S is *mixed* if it contains both red and blue agents; a mixed coalition S is called a *mixed pair* if $|S| = 2$.

For each agent $i \in N$, we interpret the order $\succ_i$ as her preferences over the fraction of the red agents in a coalition; for instance, if $\frac{2}{3} \succ_i \frac{3}{5}$, this means that agent i prefers a coalition in which two thirds of the agents are red to a coalition in which three fifths of the agents are red.

For each coalition S, we denote by $\theta_R(S)$ the fraction of the red agents in S, i.e., $\theta_R(S) = \frac{|S \cap R|}{|S|}$; we refer to this fraction as the *red ratio of* S. For each $i \in N$ and $S, T \in \mathcal{N}_i$, we say that agent i *strictly prefers* S to T if $\theta_R(S) \succ_i \theta_R(T)$, and we say that i *weakly prefers* S to T if $\theta_R(S) = \theta_R(T)$ or $\theta_R(S) \succ_i \theta_R(T)$. A coalition S is said to be *individually rational* if every agent i in S weakly prefers S to $\{i\}$.

An *outcome* of a diversity game is a partition of agents in N into disjoint coalitions. Given a partition π of N and an agent $i \in N$, we write $\pi(i)$ to denote the unique coalition in π that contains i. A partition π of N is said to be *individually rational* if all coalitions in π are individually rational.

The *core* is the set of partitions that are resistant to *group deviations*. Formally, we say that a coalition $S \subseteq N$ *blocks* a partition π of N if every agent $i \in S$ strictly prefers S to her own coalition $\pi(i)$. A partition π of N is said to be *core stable*, or in the *core*, if no coalition $S \subseteq N$ blocks π.

We also consider outcomes that are immune to *individual deviations*. Consider an agent $i \in N$ and a pair of coalitions $S \notin \mathcal{N}_i$ and $T \in \mathcal{N}_i$. An agent $j \in S$ *accepts* a deviation of i to S if j weakly prefers $S \cup \{i\}$ to S. A deviation of i from T to S is said to be an *NS-deviation* if i prefers $S \cup \{i\}$ to T; and an *IS-deviation* if it is an NS-deviation and all agents in S accept it. A partition π is called *Nash stable* (NS) (respectively, *individually stable* (IS)) if no agent $i \in N$ has an NS-deviation (respectively, an IS-deviation) from $\pi(i)$ to another coalition $S \in \pi$ or to $\emptyset$.

We say that the preferences $\succ_i$ of an agent $i \in N$ are *single-peaked* if for every $i \in N$ there is a peak $p_i \in [0, 1]$ such that

$$\theta_1 < \theta_2 \leq p_i \text{ or } \theta_1 > \theta_2 \geq p_i \text{ implies that } \theta_2 \succ_i \theta_1.$$

In particular, if an agent has a strong preference for being in the majority, then her preferences are single-peaked, as illustrated in the following example.

Example 2.2 (Birds of a feather flock together). Suppose that all agents in R are smokers and all agents in B are non-smokers. Then we expect an agent to prefer groups with the maximum possible ratio of agents of her own type. Formally, for each $r \in R$ and each $\theta, \theta' \in \Theta$ we have $\theta \succ_r \theta'$ if and only if $\theta > \theta'$, and for each $b \in B$ and each $\theta, \theta' \in \Theta$ we have $\theta \succ_b \theta'$ if and only if $\theta < \theta'$. In this case, the partition in which each agent forms a singleton coalition is core stable and Nash stable (and hence also individually stable).

If an agent strongly prefers to be surrounded by agents of the other type, her preferences are single-peaked, too.

Example 2.3 (Bakers and Millers [4]). Suppose that each agent prefers the fraction of agents of the other type to be as high as possible. This holds, for instance, when individuals of the same type compete to trade with individuals of the other type. A *Bakers and Millers game* is a diversity game where for each $r \in R$ and each $\theta, \theta' \in \Theta$ we have $\theta \succ_r \theta'$ if and only if $\theta < \theta'$, and for each $b \in B$ and each $\theta, \theta' \in \Theta$ we have $\theta \succ_b \theta'$ if and only if $\theta > \theta'$. Note that if $|R| = |B|$, a partition into mixed pairs is core stable and Nash stable; indeed, Aziz et al. [4] prove that every Bakers and Millers game has a non-empty core.

3 CORE STABILITY

Examples 2.2 and 2.3 illustrate that if all agents have extreme homophilic or extreme heterophilic preferences, the core is guaranteed to be non-empty. However, we will now show that in the intermediate case the core may be empty, even if all agents have single-peaked preferences.

Example 3.1. Consider a diversity game G, where the set of agents is given by $R = \{r_1, r_2, r_3, r_4, r_5, r_6, r_7\}$ and $B = \{b_1, b_2\}$. Agents can be divided into the following three categories with essentially the same preferences: $X = \{r_1, r_2, r_3, r_4, b_1\}$, $Y = \{r_5\}$, and $Z = \{r_6, r_7, b_2\}$. We have $\Theta = \{0, \frac{1}{3}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{5}{7}, \frac{3}{4}, \frac{7}{9}, \frac{4}{5}, \frac{5}{6}, \frac{6}{7}, \frac{7}{8}, 1\}$. Each agent has the following single-peaked preferences over the ratios of red agents:

- $r_1, r_2, r_3, r_4 : \frac{6}{7} \succ \frac{5}{6} \succ \frac{4}{5} \succ \frac{7}{9} \succ \frac{3}{4} \succ \frac{7}{8} \succ 1 \succ \frac{5}{7} \succ \cdots$
- $b_1 : \frac{6}{7} \succ \frac{5}{6} \succ \frac{4}{5} \succ \frac{7}{9} \succ \frac{3}{4} \succ \frac{7}{8} \succ \frac{5}{7} \succ \cdots$
- $r_5 : \frac{5}{6} \succ \frac{4}{5} \succ \frac{7}{9} \succ \frac{3}{4} \succ \frac{6}{7} \succ \frac{7}{8} \succ 1 \succ \frac{5}{7} \succ \cdots$
- $b_2 : \frac{3}{4} \succ \frac{7}{9} \succ \frac{4}{5} \succ \frac{5}{6} \succ \frac{6}{7} \succ \frac{7}{8} \succ \frac{5}{7} \succ \cdots$
- $r_6, r_7 : \frac{3}{4} \succ \frac{7}{9} \succ \frac{4}{5} \succ \frac{5}{6} \succ \frac{6}{7} \succ \frac{7}{8} \succ 1 \succ \frac{5}{7} \succ \cdots$

Figure 1 illustrates the preferences of each preference category.

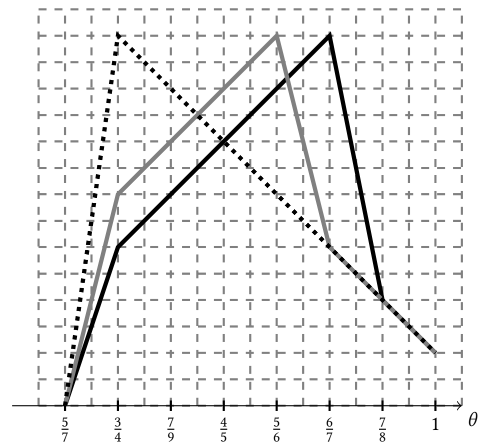


Figure 1: Single-peaked preferences over the ratios of red agents. The thick, grey, and dotted lines represent the preferences of agents in X, Y, and Z, respectively. We omit the ratios $\frac{1}{3}, \frac{1}{2}, \frac{2}{3}$ since there is no individually rational coalition with these fractions of red agents.

We will now argue that the game in Example 3.1 has empty core.

Proposition 3.2. *The game G in Example 3.1 has no core stable outcomes.*

Proof. Suppose towards a contradiction that there exists a core stable outcome π. We note that, by individual rationality, $\theta_R(S) \geq 3/4$ for every mixed coalition $S \in \pi$, as red agents in a coalition S with $\theta_R(S) \leqslant 5/7 < 3/4$ would strictly prefer to be alone. Also, π contains at least one mixed coalition, as otherwise $\{r_1, r_2, r_3, b_1\}$ would block π.

Let θ^* be the largest ratio of red agents in a mixed coalition in π. We have argued that $\theta^* \geq 3/4$. Also, if $3/4 \leq \theta^* \leq 4/5$, then $\theta_R(\pi(x)) \leq 4/5$ for all $x \in X$ and $3/4 \leq \theta_R(\pi(r_5)) \leq 4/5$ or $\theta_R(\pi(r_5)) = 1$; thus coalition $X \cup Y$ with $\theta_R(X \cup Y) = 5/6$ blocks π. Hence, $\theta^* \geq 5/6$. Now since π contains at least one mixed coalition with red ratio at least $5/6$, and all mixed coalitions in π must have red ratio at least $3/4$, if there is more than one mixed coalition, the number of agents would be at least 10, a contradiction. Thus, π contains exactly one mixed coalition of red ratio at least $5/6$. It follows that for each agent i it holds that either

- i belongs to a mixed coalition of red ratio at least $5/6$, i.e., $\theta_R(\pi(i)) \geq 5/6$; or
- i belongs to a completely homogeneous coalition, i.e., if $i \in Y$, then $\theta_R(\pi(i)) = 0$ and if $i \in R$, then $\theta_R(\pi(i)) = 1$.

In particular, this means that $\theta_R(\pi(z)) \neq 3/4$ for all $z \in Z$, and hence each $z \in Z$ prefers $3/4$ to $\theta_R(\pi(z))$.

Now suppose that r_5 does not belong to a coalition with his favorite red ratio, i.e., $5/6$. Then $\theta_R(\pi(r_5)) \geq 6/7$ and thus r_5 prefers $3/4$ to $\theta_R(\pi(r_5))$. Thus, the coalition $Y \cup Z$ of red ratio $3/4$ blocks π, a contradiction. Hence, r_5 belongs to a coalition of his favorite red ratio $5/6$, and thus $\theta^* = 5/6$. Further, if some agent $r \in X \cap R$ does not belong to a mixed coalition, then $\theta_R(\pi(r)) = 1$ and the coalition $\{r\} \cup Z$ would block π. Hence, the unique mixed coalition of red ratio $5/6$ contains both r_5 and all four red agents in X, which means that no red agent in Z belongs to the mixed coalition. Now we have:

- $b_1 \in X \cap B$ prefers $6/7$ to $\theta_R(\pi(b_1))$ since $\theta_R(\pi(b_1)) = 0$ or $\theta_R(\pi(b_1)) = 5/6$; and
- each $x \in X \cap R$ prefers $6/7$ to $\theta_R(\pi(x)) = 5/6$; and
- each $z \in Z \cap R$ prefers $6/7$ to $\theta_R(\pi(z)) = 1$.

Then coalition $X \cup (Z \cap R)$ of red ratio $6/7$ would block π, a contradiction. We conclude that G does not admit a core stable partition. □

Indeed, we can show that checking whether a given diversity game has a non-empty core is NP-complete.

Theorem 3.3. *The problem of checking whether a diversity game $G = (R, B, (\succ_i)_{i \in R \cup B})$ has a non-empty core is* NP*-complete.*

In Example 3.1 there are at least two agents of each type. In contrast, if one of the types is represented by a single agent, then the core is guaranteed to be non-empty.

Proposition 3.4. *Let $G = (R, B, (\succ_i)_{i \in R \cup B})$ be a diversity game with $|R| = 1$ or $|B| = 1$. Then the core of G is non-empty, and a partition in the core can be constructed in polynomial time.*

We also note that in the game in Example 3.1 agents' preferences belong to one of the three categories. The next proposition shows that if all agents have the same preferences, then there is a core stable outcome. We conjecture that with only two types of single-peaked preferences, the core is non-empty as well.

Proposition 3.5. *Let $G = (R, B, (\succ_i)_{i \in R \cup B})$ be a diversity game such that each agent has the same preference over $\Theta \setminus \{0, 1\}$. Then G has a non-empty core and a partition in the core can be constructed in polynomial time.*

4 NASH STABILITY AND INDIVIDUAL STABILITY

We have seen that core stability may be impossible to achieve. It is therefore natural to ask whether every diversity game has an outcome that is immune to individual deviations. It is easy to see that the answer is 'no' if we consider NS-deviations, even if we restrict ourselves to single-peaked preferences: for instance, the game where there is one red agent who prefers to be alone and one blue agent who prefers to be in a mixed coalition has no Nash stable outcomes. In contrast, each diversity game with single-peaked preferences admits an individually stable outcome. Moreover, such an outcome can be computed in polynomial time. In the remainder of this section, we present an algorithm that achieves this, and prove that it is correct.

Theorem 4.1. *Let $G = (R, B, \succ_{i \in R \cup B})$ be a diversity game with single-peaked preferences. Then an individually stable outcome exists and can be constructed in $O(|N|^4)$ time.*[1]

The algorithm will be divided into three parts:

(1) For agents with peaks greater than half, make mixed coalitions with red majority. For agents with peaks smaller than half, make mixed coalitions with blue majority.
(2) Make pairs from the remaining red agents and blue agents who are not in the mixed coalitions.
(3) Put all the remaining agents into singletons.

We will first show that one can construct a sequence of mixed coalitions with red majority that are immune to IS-deviations.

4.1 Create mixed coalitions with red majority

In making mixed coalitions, we will employ a technique that is similar to the algorithm for anonymous games proposed by Bogomolnaia and Jackson [9]. Intuitively, imagine that red agents and blue agents with peaks at least 1/2 form two lines, each of which is ordered from the highest peak to the lowest peak. The agents enter a room in this order, with a single blue agent entering first and red agents successively joining it as long as the fraction of red agents does not exceed the minimum peak of the agents already in the room. Once the fraction of the red agents reaches the minimum peak, a red agent who enters the room may deviate to a coalition that has been formed before. We alternate these two procedures as long as there is a red agent who can be added without exceeding the minimum peak. If no red agent can enter a room, then agents start entering another room and create a new mixed coalition in the same way. The algorithm terminates if either all red agents or all blue agents with their peaks at least half join a mixed coalition. Figure 2 illustrates this coalition formation process. We formalize this idea in Algorithm 1. We will create mixed coalitions containing exactly one blue agent, so we define the *virtual peak* q_i to be the favorite ratio of agent i among the ratios of red agents in coalitions containing exactly one blue agent.

In what follows, we assume that $S_0, S_1, \ldots, S_k$ are the final coalitions that have been obtained at the termination of the algorithm, and that π is the output of the algorithm. For each $t > 0$ and $i \in S_t$,

- i is called a *default agent* of S_t if i is a blue agent or i is a red agent who has joined S_t at Step 9;
- i is called a *new agent* of S_t if i is a red agent who has joined S_t in Step 14 or Step 19.

We denote by D_t the set of default agents of S_t, and we denote by N_t the set of new agents of S_t.

For each S_t with $t > 0$, each agent in S_t is either a default agent of a new agent. Notice that S_0 starts with the empty set and plays the role of the *last resort* option for red agents, i.e., red agents can always deviate to S_0 if they strictly prefer staying alone to the mixed coalition they have joined.

Deviations of blue agents We will first show that the algorithm constructs a sequence of coalitions such that no blue agent in the coalitions has an IS-deviation to other coalitions. We establish this by proving a sequence of claims. First, it is immediate that the red ratio of each coalition S_t is at least half, except for the last coalition, which may contain a single blue agent.

Lemma 4.2. *For each $S \in \pi$, the red ratio of S is at least half, i.e., $\theta_R(S) \geq \frac{1}{2}$.*

Proof. Take any $S \in \pi$. The claim is immediate when $S \subseteq S_0$. Suppose that $S = S_t$ for some $t > 0$. Then it is clear that S_t contains exactly one blue agent. If S_t contains no red agent, then this would mean that all red agents with their peaks at least half joined S_t, but deviated to smaller indexed-coalitions, meaning that S_t is the last coalition that has been formed, i.e., $t = k$; but this contradicts the construction in Step 21. Thus, S_t contains exactly one blue agent and at least one red agent, which implies that $\theta_R(S_t) \geq \frac{1}{2}$. □

This leads to the following lemma.

Lemma 4.3. *For each $r \in S_0$, r is a red agent and r strictly prefers $\{r\}$ to a mixed pair.*

Proof. By construction every agent in S_0 is a red agent. Suppose that before joining S_0, r has joined a mixed coalition S_t with $t > 0$ at Step 9 and then deviated to S_0 at Step 14. Just before r has left S_t, the ratio θ' of red agents in S_t is at least half, implying that $\theta' \succ_r \frac{1}{2}$ or $\theta' = \frac{1}{2}$. Since r joined S_0 at Step 14, we also have $1 \succ_r \theta'$.

[1] Indeed, a straightforward bound on the running time of the subsequent Algorithm 1 is $O(|B| \cdot |R| \cdot (|R| \cdot |B|))$. The running time of the subsequent Algorithm 2 is dominated by the running time of Algorithm 1. A detailed analysis is deferred to the full version.

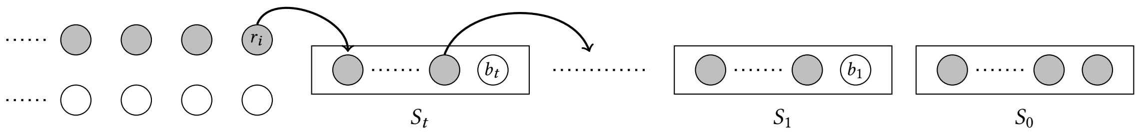


Figure 2: Illustration of how HALF$(R, B, (\succ_i)_{i \in R\cup B})$ creates coalitions with red majority. Gray circles correspond to red agents, while white circles correspond to blue agents.

Algorithm 1: HALF$(R, B, (\succ_i)_{i\in R\cup B})$

input : A single-peaked diversity game $(R, B, (\succ_i)_{i\in R\cup B})$
output: π

1 sort red and blue agents so that $q_{r_1} \geq q_{r_2} \geq \dots q_{r_x}$ and $q_{b_1} \geq q_{b_2} \geq \dots \geq q_{b_y}$;
2 initialize $i \leftarrow 1$, $k \leftarrow 1$ and $S_0 \leftarrow \emptyset$;
3 initialize $R' \leftarrow \{ r \in R \mid p_r \geq \frac{1}{2} \}$ and $B' \leftarrow \{ b \in B \mid p_b \geq \frac{1}{2} \}$;
4 **while** *$R' \neq \emptyset$ and $B' \neq \emptyset$* **do**
5 set $S_k \leftarrow \{b_k\}$;
6 **while** $\theta_R(S_k \cup \{r_i\}) \leq \min\{q_{r_i}, q_{b_k}\}$*, or there exist an agent $r \in S_k \cap R$ and $t < k$ such that r has an IS-deviation from S_k to S_t* **do**
7 `// add red agents to` S_k `as long as the ratio of red agents does not exceed the minimum virtual peak;`
8 **while** $\theta_R(S_k \cup \{r_i\}) \leq \min\{q_{r_i}, q_{b_k}\}$ **do**
9 set $S_k \leftarrow S_k \cup \{r_i\}$;
10 set $R' \leftarrow R' \setminus \{r_i\}$ and $i \leftarrow i + 1$;
11 `// let red agents in` S_k `deviate to smaller-indexed coalitions;`
12 **if** *there exists an agent $r \in S_k \cap R$ and $t < k$ such that r has an IS-deviation from S_k to S_t* **then**
13 choose r and S_t so that $\theta_R(S_t \cup \{r\})$ is r's most preferred ratio among the coalitions S_t satisfying the above;
14 set $S_t \leftarrow S_t \cup \{r\}$, $S_k \leftarrow S_t \setminus \{r\}$;
15 set $B' \leftarrow B' \setminus \{b_k\}$, and $k \leftarrow k + 1$;
16 `// let the remaining agents deviate to mixed coalitions as long as they prefer the ratio of the deviating coalition to half;`
17 **while** *there is an agent $r \in R'$ and a coalition S_t such that $t \geq 0$, $\theta_R(S_t \cup \{r\}) \succ_r \frac{1}{2}$, and all agents in S_t accept a deviation of r to S_t* **do**
18 choose r and S_t so that $\theta_R(S_t \cup \{r\})$ is r's most preferred ratio among the coalitions S_t satisfying the above;
19 $S_t \leftarrow S_t \cup \{r\}$ and $R' \leftarrow R' \setminus \{r\}$;
20 **if** *$R' = \emptyset$ and S_k consists of a single blue agent* **then**
21 return $\pi = \{ \{r\} \mid r \in S_0 \} \cup \{S_1, S_2, \dots, S_{k-1}\}$;
22 **else**
23 return $\pi = \{ \{r\} \mid r \in S_0 \} \cup \{S_1, S_2, \dots, S_k\}$;

Combining these yields $1 \succ_r \frac{1}{2}$. The claim is immediate when r joined S_0 at Step 19. □

We also observe that, by the construction of the algorithm, all default agents belong to the coalition whose red ratio is at most their virtual peak; see Figure 3 (a) for an illustration.

Lemma 4.4. *For every default agent i in S_t, where $S_t \in \pi$ with $t > 0$,*

(1) the red ratio of S_t is at most i's virtual peak, i.e., $\theta_R(S_t) \leq q_i$; and
(2) i weakly prefers S_t to $\{i\}$ and to a mixed pair.

Proof. Take any $t > 0$ and take any default agent i in S_t. Before the algorithm starts forming the next coalition S_{t+1}, the ratio $\theta_R(S_t)$ remains below q_i by the **while**-condition in Step 6. After S_{t+1} starts being formed, $\theta_R(S_t)$ can only increase by accepting red agents from larger-indexed coalitions. However, if a deviation of some red agent to S_t increases the fraction above q_i, this would mean that i would not be willing to accept such a deviation. Thus, $\theta_R(S_t) \leq q_i$. To show the second statement, recall that $p_i \geq \frac{1}{2}$ by construction of the algorithm and $\theta_R(S_t) \geq \frac{1}{2}$ by Lemma 4.2. If i is a blue agent, then by single-peakedness i weakly prefers S_t both to $\{i\}$ and to a mixed pair. If i is a red agent who has joined S_t at Step 9, then by single-peakedness and the fact that he has not deviated to S_0 at Step 14, i weakly prefers S_t both to his singleton and to a mixed pair. □

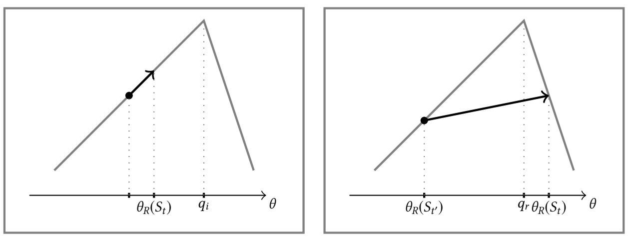
We are now ready to prove that no coalition in π admits a deviation of a blue agent.

Lemma 4.5. *For each $S \in \pi$, there is an agent in S who does not accept a deviation of a blue agent to S.*

Proof. If $S \subseteq S_0$ and $|S| = 1$, it follows from Lemma 4.3 that no $r \in S$ accepts a deviation of a blue agent. Suppose $S = S_t$ for some $t > 0$, and assume towards a contradiction that some blue agent $b \in B$ can be accepted by all agents in $S_t \in \pi$ for some $t > 0$. By Lemma 4.2, the ratio of red agents in S_t is at least $\frac{1}{2}$. If the fraction of red agents in S_t is at most the peak of some agent, i.e., $\theta_R(S_t) \leq p_i$ for some agent $i \in S_t$, this means that $\theta_R(S_t \cup \{b\}) < \theta_R(S_t) \leq p_i$, i.e., i would not accept b, a contradiction. Hence, suppose that the fraction of red agents in S_t is beyond the maximum peak, i.e., $\theta_R(S_t) > \max_{i \in S_t} p_i$. Take $i \in S_t \cap B$. By Lemma 4.4, $p_i < \theta_R(S_t) \leq q_i$. Since $p_i \geq \frac{1}{2}$ by construction of the algorithm, we have $\theta_R(S_t) > \frac{1}{2}$, which means that coalition S_t contains one blue agent and at least two red agents. Hence, even if one red agent leaves the coalition, its red ratio would be at least half, i.e.,

$$\frac{1}{2} \leq \frac{|S_t \cap R| - 1}{|S_t| - 1} (:= \theta^*) < p_i,$$

where the second inequality holds, since otherwise $p_i \leq \theta^* < \theta_R(S_t) \leq q_i$, which contradicts the fact that q_i is i's favorite ratio of the coalitions containing exactly one blue agent. Further, agent i prefers $\theta_R(S_t)$ to θ^* by single-peakedness and by the fact that $\theta^* < \theta_R(S_t) \leq q_i$. Since i accepts the deviation of b to S_t, after adding b to S_t, the red ratio should remain at least θ^*, implying that $\theta^* < \theta_R(S_t \cup \{b\})$. But this would mean that $\frac{|S_t \cap R|-1}{|S_t|-1} < \frac{|S_t \cap R|}{|S_t|+1}$, or, equivalently, $(|S_t|+1)(|S_t \cap R|-1) < |S_t \cap R|(|S_t|-1)$. This inequality can be simplified to $2|S_t \cap R| < |S_t|+1$, implying $\theta^* < \frac{1}{2}$, a contradiction. □



(a) Virtual peak of a default agent. (b) Virtual peak of a new agent.

Figure 3: The red ratio of a coalition is at most the virtual peak of a default agent, but exceeds the virtual peak of a new agent.

Deviations of red agents We now show that coalitions $S_1, \ldots, S_k$ do not admit an IS-deviation by red agents. To this end, we first observe that the red ratio of a coalition to which a new agent belongs exceeds the virtual peak of the new agent; see Figure 3 (b) for an illustration. Due to single-peakedness, this means that new agents do not accept further deviations of red agents.

Lemma 4.6. *Let S be the set of agents that belong to S_t with $t > 0$ after the* **while***-loop of Step 8 of Algorithm 1, and let r_i be the last agent in S to join S_t. Then no red agent r_j with $j > i$ can join S without exceeding the minimum peak, i.e., $\theta_R(S \cup \{r_j\}) > \min\{q_{b_t}, q_{r_j}\}$.*

Proof. Take any r_j with $j > i$. If $\theta_R(S \cup \{r_j\}) \leq \min\{q_{b_t}, q_{r_j}\}$, then this would mean that $\theta_R(S \cup \{r_{i+1}\}) \leq \min\{q_{b_t}, q_{r_{i+1}}\}$, since $q_{r_{i+1}} \geq q_{r_j}$. Hence, the algorithm would have added r_{i+1} to S, a contradiction. □

Lemma 4.7. *For all $S_t \in \pi$ ($t > 0$) and each new agent $r \in S_t$:*

(1) the red ratio of S_t exceeds r's virtual peak, i.e., $\theta_R(S_t) > q_r$;
(2) r is the unique new agent in S_t;
(3) r weakly prefers S_t to $\{r\}$ and to a mixed pair.

Proof. Let r be the first new agent who joined S_t at Step 14 or Step 19. When r joined S_t, his virtual peak q_r was at most that of any other red agent in S_t. By Lemma 4.6, he cannot join S_t without exceeding the virtual peak q_r, i.e., $\theta_R(D_t \cup \{r\}) < q_r$. Thus, by single-peakedness, he does not accept any further deviations by red agents. Hence, r is the unique new agent in S_t. To see that the third statement holds, note that if r joined S_t at Step 19, then r prefers $\theta_R(S_t)$ to $\frac{1}{2}$ by the **if**-condition of Step 17. Also, r chooses his favorite coalition among the coalitions to which he can deviate, and thus r prefers $\theta_R(S_t)$ to $\theta_R(S_0 \cup \{r\}) = 1$. Now suppose that r joined S_t from $S_{t'}$ with $t' > t$ at Step 14. At that point, the ratio θ' of red agents in $S_{t'}$ was at least half, and agent r weakly prefers $S_{t'}$ to a mixed pair since $\frac{1}{2} \leq \theta' \leq q_r$. Hence, since r prefers $\theta_R(S_t)$ to θ', by transitivity r prefers $\theta_R(S_t)$ to $\frac{1}{2}$. Also, by the **if**-condition in Step 12, r prefers S_t to being in a singleton coalition. □

We also observe that the red ratio of a higher-indexed coalition is smaller than or equal to that of a lower-indexed coalition.

Lemma 4.8. *Let $t' \geq t > 0$ and let S be the set of agents in S_t just after the* **while***-loop of Step 8. Then, the red ratio of $D_{t'}$ is at most the red ratio of S, i.e., $\theta_R(D_{t'}) \leq \theta_R(S)$; in particular, $\theta_R(D_{t'}) \leq \theta_R(D_t)$.*

Proof. Assume for the sake of contradiction that $\theta_R(D_{t'}) > \theta_R(S)$. Since S and $D_{t'}$ contain exactly one blue agent, this means that $D_{t'}$ contains more red agents than S does, i.e., $|S \cap R| + 1 \leq |D_{t'} \cap R|$. Thus, even if we add a red agent to S, its red ratio does not exceed $\theta_R(D_{t'})$, i.e., for each $r \in R \setminus S$ we have $\theta_R(S \cup \{r\}) \leq \theta_R(D_{t'})$. Now recall that by Lemma 4.4, the red ratio of $S_{t'}$ does not exceed the virtual peak of each default agent, which means that

$$\theta_R(D_{t'}) \leq \theta_R(S_{t'}) \leq \min_{x \in D_{t'}} q_x.$$

Combining these observations, for every $r \in R \setminus S$ we have

$$\theta_R(S \cup \{r\}) \leq \min_{x \in D_{t'}} q_x. \tag{1}$$

Moreover, since $t = t'$, or $S_{t'}$ has been created after S_t, there is a red agent $r_j \in D_{t'} \setminus S$ with $q_{r_j} \leq q_r$ for every $r \in S \cap R$ and $q_{b'} \leq q_b$, where b' and b are the unique blue agents in $D_{t'}$ and S, respectively. Combining this with inequality (1) yields

$$\theta_R(S \cup \{r_j\}) \leq \min_{x \in S \cup \{r_j\}} q_x,$$

which contradicts Lemma 4.6. □

Lemma 4.9. *Let $t' > t > 0$ with $\theta_R(D_{t'}) = \theta_R(D_t)$. Then some agent in $D_{t'}$ does not accept a deviation of a red agent to $S_{t'}$.*

Proof. Since $\theta_R(D_{t'}) = \theta_R(D_t)$, both D_t and $D_{t'}$ contain exactly one blue agent as well as the same number of red agents. By the description of the algorithm, the minimum virtual peak of agents in $D_{t'}$ is smaller than that of agents in D_t, i.e.,

$$\min_{x \in D_{t'}} q_x \leq \min_{x \in D_t} q_x. \tag{2}$$

Further, by Lemma 4.6, no red agent in $D_{t'}$ can join D_t without exceeding the minimum virtual peak, which implies that for every $r \in D_{t'}$ we have

$$\min_{x \in D_t \cup \{r\}} q_x < \theta_R(D_t \cup \{r\}) = \frac{|D_{t'} \cap R| + 1}{|D_{t'}| + 1}.$$

Combining this with the inequality (2) implies

$$\min_{x \in D_{t'}} q_x < \frac{|D_{t'} \cap R| + 1}{|D_{t'}| + 1}.$$

Thus, by single-peakedness, there is an agent in $D_{t'}$ who does not accept a deviation of a red agent to $D_{t'}$. □

We are now ready to show that no coalition in π admits a deviation of a red agent.

Lemma 4.10. *No agent $r \in S_t \cap R$ has an IS-deviation to another coalition $S_{t'}$ with $0 < t' < t$.*

Proof. Suppose towards a contradiction that there is such an agent $r \in S_t \cap R$ and a coalition $S_{t'}$ with $0 < t' < t$. Let t' be the largest index of coalition $S_{t'}$ to which r has an IS-deviation. Observe that if $S_{t'}$ contains a new agent, then the new agent does not accept a deviation of agent r by Lemma 4.7; thus, $S_{t'}$ has no new agent and $S_{t'} = D_{t'}$. If r is a default agent in S_t, then r weakly prefers S_t to D_t and hence r strictly prefers $D_{t'} \cup \{r\}$ to D_t by transitivity. Thus, agent r could have deviated to $S_{t'}$ from S_t at Step 14, a contradiction. If r is a new agent in S_t, then r joined S_t in Step 14 or Step 19, but this means that r could have deviated to $S_{t'}$ instead of S_t, as r strictly prefers $D_{t'} \cup \{r\}$ to $D_t \cup \{r\}$, a contradiction. □

Lemma 4.11. *No agent $r \in S_t \cap R$ has an IS-deviation to another coalition $S_{t'}$ with $t' > t$.*

Proof. Suppose towards a contradiction that there is such an agent $r \in S_t \cap R$ and a coalition $S_{t'}$. Let $t' > t$ be the smallest index of coalition $S_{t'}$ to which r has an IS-deviation. Again, if $S_{t'}$ contains a new agent, then r cannot deviate to $S_{t'}$; thus, $S_{t'} = D_{t'}$, and $\theta_R(S_{t'}) = \theta_R(D_{t'}) \leq \theta_R(D_t) \leq \theta_R(S_t)$ by Lemma 4.8.

First, consider the case where agent r is a default agent in S_t and $t > 0$. Then, by Lemma 4.4 and by the fact that $\theta_R(S_{t'}) \leq \theta_R(S_t)$, we have

$$\theta_R(S_{t'}) \leq \theta_R(S_t) \leq q_r.$$

Since r's preferences are single-peaked, r has an incentive to join $S_{t'}$ only if $\theta_R(S_{t'}) = \theta_R(S_t)$; however, by Lemma 4.9 this implies that there is an agent in $S_{t'}$ who is not willing to accept r's deviation, a contradiction.

Second, consider the case where agent r joined S_t at Step 14. Let S_ℓ be the coalition to which the red agent r belonged before joining S_t, and let S be the set of agents in S_ℓ just before r deviated from S_ℓ to S_t at Step 14. Note that S_ℓ is a coalition that has been created before $S_{t'}$ emerged, i.e., $\ell \leq t'$, since otherwise r would have deviated to $S_{t'}$ instead of S_t at Step 14. By Lemma 4.8, $\theta_R(D_{t'})$ is at most $\theta_R(S)$ and $\theta_R(S) \leqslant q_r$ by the description of the algorithm. Now we have $\theta_R(D_{t'}) \leq \theta_R(S) \leq q_r < \theta_R(S_t)$.

Recall that r strictly prefers S_t to S; thus, by single-peakedness, r has an incentive to deviate from S_t to $S_{t'}$ only if $\theta_R(D_{t'}) = \theta_R(S)$. Since $\theta_R(D_\ell) \leq \theta_R(S)$ by Lemma 4.8, this means that $\theta_R(D_{t'}) = \theta_R(D_\ell)$. Also, if $\ell < t'$, some agent in $D_{t'}$ does not accept the deviation of r by Lemma 4.9, and hence $\ell = t'$. Thus, agent r deviated from $S_{t'}$ to S_t and later the coalition $S_{t'}$ regained the same number of red agents as before. Now, if $\theta_R(S) = q_r$, then r would not have left the coalition S at Step 14; hence $\theta_R(S) < q_r$, which implies $\theta_R(D_{t'} \cup \{r\}) \leq q_r$. Recall that by Lemma 4.4, the red ratio of $D_{t'}$ is at most the minimum virtual peak, i.e., $\theta_R(D_{t'}) \leq \min_{x \in D_{t'}} q_x$. If adding r to $D_{t'}$ exceeds the minimum, some agent would not accept a deviation of r; thus, we have

$$\theta_R(D_{t'} \cup \{r\}) \leq \min_{x \in D_{t'} \cup \{r\}} q_x. \tag{3}$$

Now, since $\theta_R(S \setminus \{r\}) < \theta(D_{t'})$, there is at least one red agent $r_j \in D_{t'}$ who does not belong to S. By (3),

$$\theta(S \cup \{r_j\}) = \theta(D_{t'} \cup \{r\}) \leq \min_{x \in D_{t'} \cup \{r\}} q_x \leq \min_{x \in S \cup \{r_j\}} q_x,$$

contradicting Lemma 4.6.

Finally, consider the case where r joined S_t at Step 19. Since $S_{t'}$ does not contain any new agents, agent r could have joined $S_{t'}$ instead of S_t at Step 19, a contradiction. □

4.2 Algorithm for individual stability

We will now construct an individually stable outcome using the algorithm described in Section 4.1 as a subroutine. Suppose that we are given a diversity game $(R, B, (\succ_i)_{i \in R \cup B})$. For each $i \in R \cup B$, we denote by $\succ_i^B$ the preference over the ratios of the blue agents in each coalition: $\theta_1 \succ_i^B \theta_2$ if and only if $(1 - \theta_1) \succ_i (1 - \theta_2)$.

Algorithm 2: Algorithm for an individually stable outcome

input : A single-peaked diversity game $(R, B, (\succ_i)_{i \in R \cup B})$
output: π
1 make mixed coalitions with red majority, i.e., set $\pi_R = \mathrm{HALF}(R, B, (\succ_i)_{i \in R \cup B})$;
2 let $R_{\text{left}} \leftarrow R \setminus \bigcup_{S \in \pi_R} S$ and $B_{\text{left}} \leftarrow B \setminus \bigcup_{S \in \pi_R} S$;
3 make mixed coalitions with blue majority, i.e., set $\pi_B = \mathrm{HALF}(B_{\text{left}}, R_{\text{left}}, (\succ_i^B)_{i \in R_{\text{left}} \cup B_{\text{left}}})$;
4 set $R_{\text{left}} \leftarrow R_{\text{left}} \setminus \bigcup_{S \in \pi_B} S$, and $B_{\text{left}} \leftarrow B_{\text{left}} \setminus \bigcup_{S \in \pi_B} S$;
5 make mixed pairs from remaining agents in R_{left} and B_{left} who prefer a coalition of ratio $\frac{1}{2}$ to his or her own singleton; add them to π_{pair} ;
6 set $R_{\text{left}} \leftarrow R_{\text{left}} \setminus \bigcup_{S \in \pi_{\text{pair}}} S$, and $B_{\text{left}} \leftarrow B_{\text{left}} \setminus \bigcup_{S \in \pi_{\text{pair}}} S$;
7 `// let the remaining agents deviate to mixed coalitions as long as they prefer the deviating coalition to his or her singleton;`
8 **foreach** $A \in \{R, B\}$ **do**
9   **while** *there is an agent $i \in A_{\text{left}}$ and a mixed coalition $S \in \pi_A$ such that $S \cup \{i\}$ is individually rational and all agents in S accept a deviation of i to S* **do**
10   choose such pair i and S where $\theta_A(S \cup \{i\})$ is i's most preferred ratio among the coalitions S satisfying the above;
11     $\pi_A \leftarrow \pi_A \setminus \{S\} \cup \{S \cup \{i\}\}$, and $A_{\text{left}} \leftarrow A_{\text{left}} \setminus \{i\}$;
12 put all the remaining agents in R_{left} and B_{left} into singletons, and add them to π_{single};
13 return $\pi = \pi_R \cup \pi_B \cup \pi_{\text{pair}} \cup \pi_{\text{single}}$;

Now let π_R, π_B, π_{pair}, and π_{single} denote the partitions computed by Algorithm 2, and let $\pi = \pi_R \cup \pi_B \cup \pi_{\text{pair}} \cup \pi_{\text{single}}$. Arguing as in the previous section, we can establish the following properties.

Lemma 4.12. *For each $A \in \{R, B\}$ and each coalition $S \in \pi_A$ it holds that:*

(a) S is individually rational;
(b) if i joined S at Step 11, then i does not accept a further deviation of an agent in A to S;
(c) there is an agent who does not admit a deviation of an agent in $N \setminus A$;
(d) for each agent $i \in S$ who joined S before Step 11 of Algorithm 2, i weakly prefers S to a mixed pair.

Proof. Without loss of generality suppose that A is the set of red agents. Claim (a) (individual rationality) holds due to Lemmas 4.4 and 4.7 and the **while**-condition in Step 9 of Algorithm 2.

To show claim (b), suppose that a red agent r is the first agent who joined $S \in \pi_R$ at Step 11. By single-peakedness, the claim holds when $p_r \leq \frac{1}{2}$; so suppose $p_r > \frac{1}{2}$. Since r did not belong to any of the coalitions in π_R before, his virtual peak q_r is at most that of any other red agent in S. Since all agents in $S \setminus \{r\}$ accept r at Step 9 of Algorithm 2, S did not contain a new agent by Lemma 4.7, so the red ratio of $S \setminus \{r\}$ is at most the minimum virtual peak of agents in $S \setminus \{r\}$ by Lemma 4.4 and this remains true after accepting r. However, by Lemma 4.6, agent r cannot join S without exceeding either the virtual peak of the unique blue agent or the virtual peak q_r; thus, $\theta_R(S) < q_r$. By single-peakedness, we conclude that r does not accept any further deviation of red agents.

To prove claim (c), we can use the argument in the proof of Lemma 4.5. To establish claim (d), we recall that each agent $i \in S$ weakly prefers S to a mixed pair by Lemmas 4.4 and 4.7; by transitivity, this still holds after accepting the deviation of red agents at Step 11 of Algorithm 2. □

Lemma 4.13. *Mixed pairs created at Step 5 of Algorithm 2 do not admit an IS-deviation.*

Proof. Let R' and B' be the set of red and blue agents in R_{left} and B_{left} just before Step 5 of Algorithm 2, respectively. Note that both R' and B' are non-empty only if red agents (respectively, blue agents) are left when mixed coalitions with red majority have been formed, and blue agents (respectively, red agents) are left when mixed coalitions with blue majority have been formed. So, the red agents in R' and the blue agents in B' have opposite peaks, i.e., we have either

- $p_r \geq \frac{1}{2}$ for all $r \in R'$ and $p_b \leq \frac{1}{2}$ for all $b \in B'$; or
- $p_r \leq \frac{1}{2}$ for all $r \in R'$ and $p_b \geq \frac{1}{2}$ for all $b \in B'$.

Thus, for each agent who would like to join, one of the agents in the pair would not accept such a deviation. □

Lemma 4.14. *The partition π is individually stable.*

Proof. We first observe that π is individually rational. Indeed, all singletons in π are individually rational. Also, all mixed pairs are individually rational, since they prefer a coalition of ratio $\frac{1}{2}$ to being in a singleton coalition. All coalitions in π_R and π_B are individually rational by Lemma 4.12(a).

We will now show that π satisfies individual stability. Take any red agent $r \in R$. We note that by individual rationality, r has no incentive to deviate to red-only coalitions; also, by Lemma 4.13, r cannot deviate to a coalition in π_{pair}. Further, by Lemma 4.12(c), agent r has no IS-deviation to a coalition in π_B. Thus, it remains to check whether r has an IS-deviation to a coalition in π_R, or a blue singleton in π_{single}. Consider the following cases:

(1) Agent r joined $S \in \pi_R$ before Step 11 of Algorithm 2.
By Lemmas 4.4 and 4.7, agent r weakly prefers S to a mixed pair before accepting the deviation of red agents at Step 11 of Algorithm 2, and by transitivity, this remains true after Step 11 of Algorithm 2, meaning that r does not want to join a blue singleton. Also, r has no IS-deviation to a coalition in π_R by Lemmata 4.10 and 4.11.

(2) Agent r joined $S \in \pi_R$ at Step 11 of Algorithm 2.
By individual rationality, agent r weakly prefers his coalition to forming his own singleton. Hence, if r has an IS-deviation to a coalition of a single blue agent in π_{single}, then it means that both agent strictly prefer a mixed pair to their own singleton coalitions and hence they would have formed a pair at Step 5, a contradiction. Also, if agent r has an IS-deviation to some coalition $T \in \pi_R$, then it means that no red agent joined T at Step 11 of Algorithm 2 by Lemma 4.12(c), and thus agent r would have joined T instead of S, a contradiction.

(3) Agent r belongs to a coalition $S \in \pi_B$.
By construction of the algorithm, we have $p_r \leq \frac{1}{2}$, and agent r weakly prefers S to a mixed pair before accepting the deviation of blue agents at Step 11 of Algorithm 2. By transitivity, this remains true after Step 11 of Algorithm 2, which means that r prefers his coalition to a coalition with blue ratio at most $\frac{1}{2}$. Thus, r has no incentive to deviate to a coalition in π_R, or to a blue-only coalition.

(4) Agent r belongs to a coalition in π_{pair}.
Clearly, agent r has no incentive to deviate to singleton coalitions of blue agents. Also, if r has an IS-deviation to some coalition $T \in \pi_R$, r would have joined T at Step 19 of Algorithm 1, a contradiction.

(5) Agent r belongs to a coalition in π_{single}.
If r has an IS-deviation to some coalition $T \in \pi_R$, r could have joined T at Step 11 of Algorithm 2, a contradiction. If r has an IS-deviation to some coalition in π_{single} that consists of a single blue agent b, then r and b could have formed a pair at Step 5, a contradiction.

A symmetric argument applies to blue agents' deviations, and hence no blue agent has an IS-deviation to other coalitions. □

5 CONCLUSION

We have initiated the formal study of coalition formation games with varying degree of homophily and heterophily. Our results suggest several directions for future work.

First, while we have argued that Nash stable outcomes may fail to exist, the complexity of deciding whether a given diversity game admits a Nash stable outcome remains unknown. Also, we have obtained an existence result for individual stability under the assumption that agent's preferences are single-peaked, but it is not clear if the single-peakedness assumption is necessary; in fact, we do not have an example of a diversity game with no individually stable outcome. In a similar vein, it would be desirable to identify further special classes of diversity games that admit core stable outcomes.

More broadly, it would be interesting to extend our model to more than two agent types. Another possible extension is to consider the setting where agents are located on a social network, and each agent's preference over coalitions is determined by the fraction of her acquaintances in these coalitions; this model would capture both the setting considered in our work and fractional hedonic games.

Acknowledgements. We thank reviewers at AAMAS-19 for helpful feedback. We are also grateful to Ilan Nehama and Nathanaël Barrot for useful discussions. This work was supported by ERC grant 639945 (ACCORD) and by KAKENHI (Grant-in-Aid for JSPS Fellows, No. 18J00997) Japan.

REFERENCES

[1] R. Alba and J. Logan. 1993. Minority Proximity to Whites in Suburbs: An Individual-Level Analysis of Segregation. *Amer. J. Sociology* 98, 6 (1993), 1388–1427.

[2] J. Alcade and P. Revilla. 2004. Researching with Whom? Stability and Manipulation. *Journal of Mathematical Economics* 40, 8 (2004), 869 – 887.

[3] H. Aziz, F. Brandl, F. Brandt, P. Harrenstein, M. Olsen, and D. Peters. 2019. Fractional Hedonic Games. *ACM Transactions on Economics and Computation* (2019). forthcoming.

[4] H. Aziz, F. Brandt, and P. Harrenstein. 2014. Fractional Hedonic Games. In *Proceedings of the 13th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*. 5–12.

[5] C. Ballester. 2004. NP-competeness in Hedonic Games. *Games and Economic Behavior* 49 (2004), 1–30.

[6] S. Banerjee, H. Konishi, and T. Sönmez. 2001. Core in a Simple Coalition Formation Game. *Social Choice and Welfare* 18, 1 (2001), 135–153.

[7] V. Bilò, A. Fanelli, M. Flammini, G. Monaco, and L. Moscardelli. 2014. Nash Stability in Fractional Hedonic Games. In *Proceedings of the 10th Conference on Web and Internet Economics (WINE)*. 486–491.

[8] V. Bilò, A. Fanelli, M. Flammini, G. Monaco, and L. Moscardelli. 2015. On the Price of Stability of Fractional Hedonic Games. In *Proceedings of the 14th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*. 1239–1247.

[9] A. Bogomolnaia and M.O. Jackson. 2002. The Stability of Hedonic Coalition Structures. *Games and Economic Behavior* 38, 2 (2002), 201–230.

[10] C. Brandt, N. Immorlica, G. Kamath, and R. Kleinberg. 2012. An Analysis of One-dimensional Schelling Segregation. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing (STOC)*. 789–804.

[11] R. Bredereck, E. Elkind, and A. Igarashi. 2019. Hedonic Diversity Games. *arXiv preprint arXiv:1903.00303* (2019).

[12] A. Chauhan, P. Lenzner, and L. Molitor. 2018. Schelling Segregation with Strategic Agents. In *Proceedings of the 11th International Symposium on Algorithmic Game Theory (SAGT)*. 137–149.

[13] W. Clark. 2008. Geography, Space, and Science: Perspectives from Studies of Migration and Geographical Sorting. *Geographical Analysis* 40, 3 (2008), 258–275.

[14] W. Clark and M. Fossett. 2008. Understanding the Social Context of the Schelling Segregation Model. *Proceedings of the National Academy of Sciences* 105, 11 (2008), 4109–4114.

[15] G. Demange. 2004. On Group Stability in Hierarchies and Networks. *Journal of Political Economy* 112, 4 (2004), 754–778.

[16] E. Elkind, J. Gan, A. Igarashi, W. Suksompong, and A. Voudouris. 2019. Schelling Games on Graphs. *arXiv preprint arXiv:1902.07937* (2019).

[17] M. Emerson, K. Chai, and G. Yancey. 2001. Does Race Matter in Residential Segregation? Exploring the Preferences of White Americans. *American Sociological Review* 66, 6 (2001), 922–935.

[18] R. Farley, C. Steeh, T. Jackson, M. Krysan, and K. Reeves. 1993. Continued Racial Residential Segregation in Detroit: "Chocolate City, Vanilla Suburbs" Revisited. *Journal of Housing Research* 4, 1 (1993), 1–38.

[19] A. Igarashi and E. Elkind. 2016. Hedonic Games with Graph-restricted Communication. In *Proceedings of the 15th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*. 242–250.

[20] N. Immorlica, R. Kleinberg, B. Lucier, and M. Zadomighaddam. 2017. Exponential Segregation in a Two-dimensional Schelling Model with Tolerant Individuals. In *Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 984–993.

[21] D. S. Massey, A. B. Gross, and K. Shibuya. 1994. Migration, Segregation, and the Geographic Concentration of Poverty. *American Sociological Review* 59, 3 (1994), 425–445.

[22] J. Moody. 2001. Race, School Integration, and Friendship Segregation in America. *Amer. J. Sociology* 107, 3 (2001), 679–716.

[23] M. Newman. 2010. *Networks: An Introduction*. Oxford University Press, Inc., New York, NY, USA.

[24] D. Peters. 2016. Complexity of Hedonic Games with Dichotomous Preferences. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*. 579–585.

[25] T. C. Schelling. 1969. Models of Segregation. *The American Economic Review* 59, 2 (1969), 488–493.

[26] W. v. d. Laan Bouma-Doff. 2007. Involuntary Isolation: Ethnic Preferences and Residential Segregation. *Journal of Urban Affairs* 29, 3 (2007), 289–309.

[27] H. P. Young. 1998. Individual Strategy and Social Structure: an Evolutionary Theory of Institutions. *Princeton University Press*. (1998).

[28] J. Zhang. 2004. A Dynamic Model of Residential Segregation. *The Journal of Mathematical Sociology* 28, 3 (2004), 147–170.

[29] J. Zhang. 2004. Residential Segregation in an All-Integrationist World. *Journal of Economic Behavior and Organization* 54, 4 (2004), 533–550.