Individual-Based Stability in Hedonic Diversity Games

Niclas Boehmer[*]
TU Berlin
Berlin, Germany
niclas.boehmer@tu-berlin.de

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

Abstract

In *hedonic diversity games (HDGs)*, recently introduced by Bredereck, Elkind, and Igarashi (2019), each agent belongs to one of two classes (men and women, vegetarians and meat-eaters, junior and senior researchers), and agents' preferences over coalitions are determined by the fraction of agents from their class in each coalition. Bredereck et al. show that while an HDG may fail to have a Nash stable (NS) or a core stable (CS) outcome, every HDG in which all agents have single-peaked preferences admits an individually stable (IS) outcome, which can be computed in polynomial time. In this work, we extend and strengthen these results in several ways. First, we establish that the problem of deciding if an HDG has an NS outcome is NP-complete, but admits an XP algorithm with respect to the size of the smaller class. Second, we show that, in fact, all HDGs admit IS outcomes that can be computed in polynomial time; our algorithm for finding such outcomes is considerably simpler than that of Bredereck et al. We also consider two ways of generalizing the model of Bredereck et al. to $k \geq 2$ classes. We complement our theoretical results by empirical analysis, comparing the IS outcomes found by our algorithm, the algorithm of Bredereck et al. and a natural better-response dynamics.

1 Introduction

A number of French exchange students at a Spanish university have signed up for a game theory class. All students taking the class are advised to form study groups to discuss the material and to work on problem sheets. Now, some French students see this as an excellent opportunity to improve their Spanish and would like to join study groups where no one else speaks French. Other students have less confidence in their ability to communicate in Spanish and therefore want to be in a group where at least a few other students speak French; in fact, some of the students prefer to be in an exclusively French-speaking group. Spanish students, too, have different preferences over the fraction of French students in their groups: some are eager to meet new friends, while others are worried that language issues will affect their learning.

Many important aspects of the setting described in the previous paragraph can be captured by the recently intro-

duced framework of *hedonic diversity games (HDG)* (Bredereck, Elkind, and Igarashi 2019). In these games, agents can be split into two classes (say, *red* and *blue*), and each agent has preferences over the fraction of red agents in her group. The outcome of a game is a partition of agents into groups. This model is relevant for analyzing a variety of application scenarios, ranging from interdisciplinary collaborations to racial segregation.

In their work, Bredereck et al. aim to understand whether HDGs admit stable outcomes, for several common notions of stability for hedonic games, such as Nash stability, individual stability and core stability; the first two concepts are based on deviations by individual agents, while the third concept captures resilience against group deviations. Bredereck et al. show that an HDG may fail to have a Nash stable outcome or a core stable outcome and that deciding if an HDG has a core stable outcome is NP-complete. For individual stability, they get a positive result under the additional assumption that agents' preferences are single-peaked, i.e., each agent i has a preferred ratio of red agents in her group (say, ρ_i), and for any two ratios ρ, ρ' such that $\rho < \rho' \leq \rho_i$ or $\rho_i \leq \rho' < \rho$ she prefers a group with ratio ρ' to a group with ratio ρ. Specifically, Bredereck et al. show that every HDG with single-peaked preferences admits an individually stable outcome and describe a polynomial-time algorithm for finding some such outcome. Their work leaves open the question whether HDGs with non-single-peaked preferences always have an individually stable outcome.

Our Contribution   In this paper, we answer two open questions from the paper of Bredereck et al., as well as extend their model to an arbitrary number of agent classes.

First (Section 3), we show that deciding if an HDG has a Nash stable outcome is NP-complete. Our hardness result holds even if agents have dichotomous preferences, i.e., approve some ratios and disapprove the remaining ratios; in fact, it remains true if each agent approves at most 4 ratios. On the other hand, we show that the existence of a Nash stable outcome can be decided in polynomial time if the size of one of the classes can be bounded by a constant.

We then turn our attention to individual stability (Section 4). We describe a polynomial-time algorithm that finds an individually stable outcome of any HDG; our algorithm is significantly simpler than that of Bredereck et al. How-

---

[*]Most of this research was done when the first author was an MSc student at the University of Oxford.

ever, we show that neither algorithm Pareto-dominates the other: there are settings where the algorithm of Bredereck et al. produces a much better partition than our algorithm, and there are settings where the converse is true.

In Section 5, we consider two different ways of extending HDGs to settings with k agent classes, $k > 2$. First, we consider a very general model, where each agent may have arbitrary preferences over the ratios of different classes in her group. We show that our positive result for individual stability does not extend to this model: we describe a game with $k = 3$ that has no individually stable outcomes and prove that deciding the existence of such outcomes is NP-complete if $k \geq 5$. We then propose a more restrictive model, where an agent only cares about the fraction of agents that belong to her class. We show that this model encompasses both HDGs and another well-known class of hedonic games, namely, anonymous games, i.e., games where agents have preferences over the size of their group.

In Section 6, we empirically compare the outcomes produced (1) by our algorithm for finding individually stable outcomes, (2) by the algorithm of Bredereck et al. and (3) by a natural better-response dynamics with respect to several measures, such as the average social welfare and the diversity of resulting groups. We summarize our findings in Section 7.

Related Work Hedonic games were introduced by Drèze and Greenberg (1980) and have received a lot of attention in the computational social choice literature, as they offer a simple, but powerful formalism to study group formation in strategic settings; see, e.g., the survey by Aziz and Savani (2016). Hedonic diversity games share common features with two other well-studied classes of hedonic games, namely, anonymous games (Bogomolnaia and Jackson 2002) and fractional hedonic games (Aziz et al. 2019).

In anonymous games, agents cannot distinguish among other agents, and therefore the only feature of a group that matters to them is its size. HDGs may appear to be more general than anonymous games, since in HDGs each agent can distinguish between two classes of agents. Indeed, many proof techniques developed for anonymous games turn out to be relevant for HDGs. However, in a technical sense, anonymous games are not a subclass of HDGs, and some of the positive results for HDGs do not hold for anonymous games. In particular, while we prove that every HDG has an individually stable outcome, Bogomolnaia and Jackson (2002) show that this is not the case for anonymous games. Throughout the paper, we compare our results for HDGs to relevant results for anonymous games, and in Section 5 we propose a succinct representation formalism for hedonic games that captures both HDGs and anonymous games.

In fractional hedonic games, every agent assigns a numerical value to every other agent, and an agent's value for a group of size s that includes her is equal to the sum of the values she assigns to the group members, divided by s; each agent prefers coalitions with a higher value. Now, if agents are divided into two classes, so that each agent assigns the same value to all agents in each class, the resulting game is

a hedonic diversity game. However, not all hedonic diversity games can be obtained in this fashion; in particular, an HDG where each agent prefers groups that have the same number of agents from each class cannot be represented in this way. Conversely, there are fractional hedonic games that cannot be represented as HDGs.

Full version The full version of the paper is available on arXiv (Boehmer and Elkind 2019). It contains the proofs of Theorem 3.1, Theorem 3.3, Proposition 4.4 and Theorem 5.4.

2 Preliminaries

For every positive integer n, we write $[n]$ to denote the set $\{1, \ldots, n\}$.

A *hedonic game* is a pair $G = (N, (\succeq_i)_{i \in N})$, where $N = [n]$ is the set of *agents*, and for each $i \in N$ the relation $\succeq_i$ is a weak order over all subsets of N that contain i. The subsets of N are called *coalitions*; the set of all coalitions containing agent i is denoted by $\mathcal{N}(i)$. We refer to the set N as the *grand coalition*.

Given two coalitions $C, D \in \mathcal{N}(i)$, we write $C \sim_i D$ if $C \succeq_i D$ and $D \succeq_i C$; we write $C \succ_i D$ if $C \succeq_i D$ and $C \not\preceq_i D$. We say that i *weakly prefers* C to D if $C \succeq_i D$; if $C \succ_i D$, we say that i *strictly prefers* C to D, and if $C \sim_i D$, we say that i is *indifferent* between C and D. For succinctness, when describing agents' preferences, we often omit coalitions C with $\{i\} \succ_i C$.

An *outcome* of a hedonic game with the set of agents N is a partition $\pi = \{C_1, \ldots, C_k\}$ of N; we write π_i to denote the coalition in π that contains agent i. An agent i has an *NS-deviation* from an outcome π if there exists a coalition $C \in \pi \cup \{\varnothing\}$ such that $C \cup \{i\} \succ_i \pi_i$; i has an *IS-deviation* from an outcome π if there exists a coalition $C \in \pi \cup \{\varnothing\}$ such that $C \cup \{i\} \succ_i \pi_i$ and, additionally, $C \cup \{i\} \succeq_j C$ for each $j \in C$. An outcome π is *Nash stable (NS)* (respectively, *individually stable (IS)*) if no agent has an NS-deviation (respectively, an IS-deviation). An outcome π' *Pareto dominates* an outcome π if $\pi_i' \succeq \pi_i$ for each $i \in N$ and $\pi_j' \succ_j \pi_j$ for some $j \in N$; an outcome is *Pareto optimal* if it is not Pareto-dominated by another outcome.

Consider a hedonic game $G = (N, (\succeq_i)_{i \in N})$. We say that G is *dichotomous* if for every agent $i \in N$ there exist disjoint sets $\mathcal{N}^+(i)$ and $\mathcal{N}^-(i)$ such that $\mathcal{N}(i) = \mathcal{N}^+(i) \cup \mathcal{N}^-(i)$, and for every pair of coalitions $C, D \in \mathcal{N}(i)$, we have $C \sim_i D$ if $C, D \in \mathcal{N}^+(i)$ or $C, D \in \mathcal{N}^-(i)$ and $C \succ_i D$ if $C \in \mathcal{N}^+(i), D \in \mathcal{N}^-(i)$; we say that i *approves* coalitions in $\mathcal{N}^+(i)$ and *disapproves* coalitions in $\mathcal{N}^-(i)$. We say that G is *anonymous* if for every agent $i \in N$ there exists a weak order $\succeq_i^*$ on $[|N|]$ such that for every pair of coalitions C, D we have $C \succeq_i D$ if and only if $|C| \succeq_i^* |D|$.

In a *hedonic diversity game (HDG)*, the set of agents N is partitioned as $N = R \cup B$, and each agent i is indifferent between any two coalitions $C, D \in \mathcal{N}(i)$ that have the same fraction of agents in R; we refer to agents in R and B as *red* and *blue* agents, respectively. For each coalition $C \subseteq N$, we write $\theta(C) = \frac{|C \cap R|}{|C|}$; we refer to the quantity $\theta(C)$ as the *red ratio* of C. In hedonic diversity games, the preferences of every agent i can be described by a weak order $\succeq_i'$ over

the set $\Theta = \{\frac{j}{k} \mid 0 \le j \le |R|, j \le k \le |N|\}$: for all $C, D \in \mathcal{N}(i)$ we have $C \succeq_i D$ if and only if $\theta(C) \succeq'_i \theta(D)$. We say that a coalition C is *homogeneous* if $C \subseteq R$ or $C \subseteq B$, i.e., $\theta(C) \in \{0, 1\}$.

An anonymous game $(N, (\succeq_i)_{i \in N})$ is said to be *single-peaked* if for every agent $i \in N$ there exists a preferred size $s_i \in [n]$ such that for every pair of coalitions $C, D \in \mathcal{N}(i)$ with $|C| < |D| \le s_i$ or $s_i \le |D| < |C|$ it holds that $D \succeq_i C$. Similarly, a hedonic diversity game $(R \cup B, (\succeq'_i)_{i \in R \cup B})$ is said to be *single-peaked* if for every agent $i \in R \cup B$ there exists a preferred value $\rho_i \in \Theta$ such that for every $\rho, \rho' \in \Theta$ such that $\rho_i \le \rho < \rho'$ or $\rho' < \rho \le \rho_i$ it holds that $\rho \succeq'_i \rho'$.

3 Nash Stability

Bredereck, Elkind, and Igarashi (2019) show that a hedonic diversity game may fail to have a Nash stable outcome, even if agents' preferences are single-peaked: indeed, it is easy to see that an HDG with $|R| = |B| = 1$ where the red agent prefers to be alone and the blue agent prefers to be with the red agent has no NS outcome. However, Bredereck et al. do not consider the complexity of checking whether a given HDG admits a Nash stable outcome. We will now establish that this problem is NP-complete. We note that deciding whether an anonymous game has a Nash stable outcome is known to be NP-complete as well (Ballester 2004); Peters (2016) shows that the hardness result holds even for anonymous games with dichotomous preferences.

Theorem 3.1. *Given an HDG $G = (R \cup B, (\succeq'_i)_{i \in R \cup B})$, it is NP-complete to decide whether G has a Nash stable outcome. The hardness result holds even if G is dichotomous or if each relation $\succeq'_i$ is a strict order over Θ.*

To complement the hardness result of Theorem 3.1, we show that deciding the existence of a Nash stable outcome is in P if $\min\{|R|, |B|\}$ is bounded by a constant.

Theorem 3.2. *Given an HDG $G = (R \cup B, (\succeq'_i)_{i \in R \cup B})$ with $|R \cup B| = n$, $\min\{|R|, |B|\} = p$, we can decide if G has a Nash stable outcome in time $(np)^p \cdot \text{poly}(n)$.*

Proof. Assume without loss of generality that $|B| \le |R|$, so $p = |B|$. Our algorithm proceeds in three stages.

In the first stage, we guess a partition $\{C_1, \ldots, C_k\}$ of B, together with k positive integers $n_1, \ldots, n_k$ such that $n_i \ge |C_i|$ and $n_1 + \cdots + n_k \le n$; let $C_0 = \varnothing$ and $n_0 = n - (n_1 + \cdots + n_k)$. We will try to construct an NS partition $\{A_1, \ldots, A_k, D_1, \ldots, D_\ell\}$ for some $\ell \ge 0$ so that $C_i \subseteq A_i$, $|A_i| = n_i$ for each $i \in [k]$, and $D_i \subseteq R$ for $i \in [\ell]$. Note that we can enumerate all possible guesses in time $O(p^p \cdot n^p)$.

In the second stage, given a guess $\{C_1, \ldots, C_k\}$, $(n_1, \ldots, n_k)$, we construct an instance of the network flow problem as follows. We create a source, a sink, a node x_i for each $i \in R$ and a node y_j for each $j = 0, \ldots, k$. The source is connected to all nodes x_i, $i \in R$, by an edge of capacity 1, and each node y_j, $j = 0, \ldots, k$, is connected to the sink by an edge of capacity $n_j - |C_j|$. Further, there is an edge of capacity 1 from x_i to y_j, $j \in [k]$, if and only if $\frac{n_j - |C_j|}{n_j} \succeq'_i \frac{n_s - |C_s| + 1}{n_s + 1}$ for each $s \in [k] \setminus \{j\}$ and, moreover,

$\frac{n_j - |C_j|}{n_j} \succeq'_i 1$. Finally, there is an edge of capacity 1 from x_i to y_0 if $1 \succeq'_i \frac{n_s - |C_s| + 1}{n_s + 1}$ for all $s \in [k]$.

Intuitively, an edge from x_i to y_j, $j \in [k]$, indicates that i weakly prefers being in a coalition of size n_j with $|C_j|$ blue agents to deviating to a coalition of size n_s with $|C_s|$ blue agents, for $s \ne j$, or to a homogeneous red coalition; an edge from x_i to y_0 indicates that i weakly prefers a homogeneous red coalition to other available options. By construction, there is a flow of size $|R|$ in this network if and only if there is a partition $\{A_1, \ldots, A_k, D_0\}$ of $R \cup B$ such that $C_i \subseteq A_i$, $|A_i| = n_i$ for $i \in [k]$ and no red agent has an NS-deviation from this partition. Thus, if this instance of network flow does not admit a flow of size $|R|$, we reject the current guess, and otherwise we proceed to the next stage.

In the third stage, we split D_0 into $D_1, \ldots, D_\ell$ for some $\ell \ge 0$. Note that, no matter how we do this, if no red agent had an NS-deviation in $\{A_1, \ldots, A_k, D_0\}$, this will also be the case for the new partition. Thus, we can focus on the blue agents. Given a $t \in [n_0]$, we say that t is *safe for an agent $j \in B$* if j weakly prefers her current coalition in $\{A_1, \ldots, A_k, D_0\}$ to a coalition consisting of herself and t red agents; we say that t is *safe* if it is safe for each $j \in B$. Let $T \subseteq [n_0]$ be a collection of all safe integers. Then, we can subdivide D_0 so that in the resulting partition no blue agent has an NS-deviation if and only if n_0 can be represented as a sum of integers from T; the latter problem is a variant of KNAPSACK and can be solved by dynamic programming in time $O(n^2)$. $\square$

Theorem 3.2 puts the problem of finding a Nash stable outcome in the complexity class XP with respect to the parameter p; however, we do not know if this problem is fixed-parameter tractable (FPT) with respect to this parameter.

For HDGs with dichotomous preferences, another natural parameter is the number of ratios approved by each agent. However, our problem turns out to be para-NP-hard with respect to this parameter: the hardness proof in Theorem 3.1 goes through even if each agent only approves at most four ratios in Θ. Similarly, Peters (2016) shows that finding a Nash stable outcome in dichotomous anonymous games remains NP-hard if each agent approves at most four coalition sizes. Interestingly, we can prove that the latter problem becomes polynomial-time solvable if each agent approves at most one coalition size, but it is not clear how to extend this proof to dichotomous HDGs.

Theorem 3.3. *Given a dichotomous anonymous game $G = (N, (\succeq_i)_{i \in N})$ where for each agent $i \in N$ the set $\mathcal{N}^+(i)$ is of the form $\{C \in \mathcal{N}(i) : |C| = s_i\}$ for some $s_i \in \mathbb{N}$, we can decide in polynomial time whether G admits a Nash stable outcome.*

4 Individual Stability

Bredereck, Elkind, and Igarashi (2019) describe an algorithm that, given an HDG with single-peaked preferences, outputs an individually stable outcome in polynomial time. This algorithm is fairly complex: its description and analysis take up almost four pages of the AAMAS'19 paper. It is similar in spirit to the algorithm of Bogomolnaia and

Algorithm 1 Computing an individually stable outcome

Require: HDG $G = (R \cup B, (\succeq'_i)_{i \in R \cup B})$
Ensure: Individually stable outcome π of game G
 1: Let $B^* = \{b^*_1, \ldots, b^*_k\} = \{b \in B : \frac{1}{2} \succeq'_b 0\}$;
 2: Let $R^* = \{r^*_1, \ldots, r^*_\ell\} = \{r \in R : \frac{1}{2} \succeq'_r 1\}$;
 3: Let $C_0 = \{r^*_1, \ldots, r^*_{\min\{k,\ell\}}\} \cup \{b^*_1, \ldots, b^*_{\min\{k,\ell\}}\}$;
 4: Let $C = C_0$;
 5: **repeat**
 6:     **for** $i \in N \setminus C$ **do**
 7:         **if** i has an IS-deviation from $\{i\}$ to C **then**
 8:             $C = C \cup \{i\}$;
 9: **until** C has not changed in the previous iteration
10: Let $C_1 = C$;
11: **return** $\pi = \{\{i\} \mid i \in N \setminus C_1\} \cup \{C_1\}$;

Jackson (2002) that finds an IS outcome of an anonymous game with single-peaked preferences in polynomial time. The main contribution of this section is a much simpler polynomial-time algorithm that can find an individually stable outcome of *any* HDG; this result is particularly surprising, because it is known that not every anonymous game admits an IS outcome (Bogomolnaia and Jackson 2002).

Theorem 4.1. *Given an HDG* $G = (R \cup B, (\succeq'_i)_{i \in R \cup B})$, *we can compute an individually stable outcome of* G *in polynomial time.*

Proof. In what follows, we say that a coalition $C \subseteq R \cup B$ is *balanced* if $|C \cap R| = |C \cap B|$.

We claim that Algorithm 1 outputs an IS outcome of G. The first phase of this algorithm creates a maximum-size balanced coalition C such that all agents in C prefer C to being alone; all other agents are placed in singleton coalitions. In the second phase, the algorithm checks if any of the remaining agents has an IS-deviation to C; if yes, some such agent is invited to join C. This step is repeated until none of the remaining agents has an IS-deviation to C. To avoid ambiguity, we use C_0 and C_1 to denote the 'large' coalition obtained at the end of the first phase and at the end of the second phase, respectively.

Consider the sets B^* and R^* defined by our algorithm, and assume without loss of generality that $|B^*| \geq |R^*|$. By construction, C_0 contains all red agents who weakly prefer being in a balanced coalition to being in a homogeneous coalition. In particular, this means that no red agent in $N \setminus C_1$ would allow a blue agent to join her singleton coalition.

To prove that the partition computed by our algorithm is individually stable, we consider three classes of agents:

Agents in C_0: By deviating, these agents can form a homogeneous coalition or a balanced coalition. They weakly prefer C_0 to a homogeneous coalition, and, since they approve all subsequent changes to C, they weakly prefer C_1 to C_0 (which is balanced). Thus, they have no IS-deviation.

Agents in $N \setminus C_1$: By construction, these agents do not have an IS-deviation to C_1, and they are indifferent between being alone and joining another agent of the same

color. Further, a deviation that results in a two-agent balanced coalition is not an IS-deviation: all red agents in $N \setminus C_1$ strictly prefer being alone to being in a balanced coalition. Thus, agents in $N \setminus C_1$ have no IS-deviation.

Agents in $C_1 \setminus C_0$: Joining C, these agents strictly preferred C to being in a homogeneous coalition, and they have approved all changes to C since then. Thus, they strictly prefer being in C_1 to being in a homogeneous coalition. Further, a blue agent in $C_1 \setminus C_0$ cannot join a singleton red coalition, since the red agent strictly prefers to be left alone. On the other hand, every red agent in $C_1 \setminus C_0$ strictly prefers being alone to being in a balanced coalition, so by transitivity she strictly prefers staying in C_1 to joining a singleton blue coalition.

$\square$

The following example illustrates the execution of our algorithm.

Example 4.2. Consider an HDG with $B = \{1, 2, 3\}$, $R = \{4, 5\}$. The agents' preferences over Θ are given by

$$\frac{2}{3} \sim'_1 \frac{1}{2} \succ'_1 0, \qquad \frac{2}{3} \succ'_2 \frac{1}{2} \succ'_2 0,$$

$$\frac{1}{4} \succ'_3 0, \qquad \frac{2}{3} \succ'_4 \frac{1}{2} \succ'_4 1, \qquad \frac{2}{3} \succ'_5 1.$$

The algorithm sets $B^* = \{1, 2\}$, $R^* = \{4\}$ and $C_0 = \{1, 4\}$. Then, in the second phase agent 5 has an IS-deviation to C_0: we have $\{1, 4, 5\} \succ_5 \{5\}$, $\{1, 4, 5\} \succ_4 \{1, 4\}$ and $\{1, 4, 5\} \sim_1 \{1, 4\}$. Since agents 2 and 3 do not have an IS-deviation to $\{1, 4, 5\}$, the algorithm stops and outputs $(\{1, 4, 5\}, \{2\}, \{3\})$.

In Example 4.2, our algorithm outputs a Pareto optimal outcome; however, our next example shows that this is not always the case.

Example 4.3. Consider an HDG with $B = \{1, 2, 3, 4\}$, $R = \{5\}$, where $\frac{1}{5} \succ'_i 0 \succ'_i \frac{1}{2}$ for each $i \in B$, and the red agent's most preferred ratio is $\frac{1}{5}$. Then, for every agent the formation of the grand coalition is the most preferred outcome. However, as $B^* = \varnothing$, the algorithm sets $C_0 = \varnothing$ and hence outputs the partition consisting of five singletons.

Interestingly, the algorithm of Bredereck et al. would output the grand coalition on the instance from Example 4.3. However, it is not the case that on any single-peaked instance the output of Bredereck et al.'s algorithm Pareto-dominates the output of Algorithm 1; there exists an example where the converse is true. This means, in particular, that neither of these algorithms is guaranteed to output a Pareto-optimal outcome on single-peaked instances; this is in contrast with the algorithm of Bogomolnaia and Jackson (2002), which, given a single-peaked anonymous game, always outputs an IS outcome that is also Pareto optimal.

Better-Response Dynamics We note that one can also attempt to reach an IS outcome by a sequence of IS deviations: starting from an arbitrary partition, we check if the current partition is individually stable, and if not, we pick an agent

who has an IS-deviation and allow her to perform it. We will refer to this procedure as the *better-response dynamics (IS-BRD)*. While IS-BRD is a very general algorithm that can be used for arbitrary hedonic games, it may fail to converge even if an IS outcome exists, and it may need a superpolynomial number of iterations to converge (Gairing and Savani 2010); however, no results concerning its convergence are known for hedonic diversity games or for anonymous games.

The following proposition, which applies to arbitrary dichotomous hedonic games, makes partial progress towards the understanding of the performance of IS-BRD: it shows that for such games IS-BRD always converges as long as in the initial partition all agents belong to the grand coalition. We note that existence of IS outcomes in dichotomous hedonic games has been established by Peters (2016); his proof provides a polynomial-time algorithm for finding an IS outcome, but this algorithm differs from IS-BRD.

Proposition 4.4. *For every dichotomous hedonic game with a set of agents N, any sequence of IS deviations starting from the grand coalition converges to an IS outcome after at most $|N|$ iterations.*

We will revisit IS-BRD in Section 6, where we compare different approaches to finding IS outcomes in HDGs.

5 Diversity Games With k Classes

Bredereck, Elkind, and Igarashi (2019) define hedonic diversity games for two agent classes. However, diversity-related considerations remain relevant in the presence of three or more classes: for instance, in the example discussed in the beginning of the paper, the visiting students may come from several different countries. To capture such settings, we need to reason about games with k agent classes for $k > 2$: e.g., for $k = 3$ we may have red, blue and green agents. A direct generalization of the model of Bredereck et al. is to allow a red agent to base her preferences on the ratio of red, blue and green agents; a more restrictive approach is to assume that a red agent only cares about the fraction of red agents in her coalition. We will now explore both of these approaches; as we feel that the latter approach is closer in spirit to the original HDG model, we reserve the term k-*HDG* to refer to the more restrictive model and refer to games where agents can have arbitrary preferences over ratios as k-*tuple HDGs*.

Definition 5.1. A k-*tuple hedonic diversity game* is a hedonic game $(N, (\succeq_i)_{i \in N})$ where N can be partitioned into k pairwise disjoint sets $R_1, \ldots, R_k$ so that for each agent $i \in N$ and every pair of coalitions $C, D \in \mathcal{N}(i)$ with $\frac{|C \cap R_s|}{|C|} = \frac{|D \cap R_s|}{|D|}$ for each $s \in [k]$ we have $C \sim_i D$.

Given an n-agent k-tuple HDG and an agent $i \in N$, we can map a coalition $C \in \mathcal{N}(i)$ to a k-tuple of fractions $\left(\frac{|C \cap R_\ell|}{|C|} \right)_{\ell \in [k]}$: i is indifferent between two coalitions that map to the same tuple. Hence, i's preferences can be described by a partial order over such tuples. As the number of tuples of this form is bounded by n^{2k}, if k is bounded by a constant, the size of this representation is polynomial in n. On the other hand, every hedonic game with n agents

is an n-tuple HDG, as we can simply place each agent in a separate class.

Our XP result for Nash stability extends to this more general setting: if the total size of the smallest $k - 1$ classes can be bounded by a constant, a Nash stable outcome can be found in time polynomial in the input representation size, which, in turn, can be bounded as $\mathcal{O}(n^3)$ in this case. The proof (omitted) is a simple generalization of the proof of Theorem 3.1.

Theorem 5.2. *Given a k-tuple HDG G with set of agents N, $|N| = n$, and classes $R_1, \ldots, R_k$ such that $\min_{j \in [k]} |N \setminus R_j| = p$, we can decide whether G has a Nash stable outcome in time $(np)^p \cdot \text{poly}(n)$.*

In contrast, the following example shows that Theorem 4.1 does not extend to $k > 2$, i.e., k-tuple HDGs may fail to have an individually stable outcome if $k > 2$.

Example 5.3. Consider a 3-tuple HDG G with $N = R_1 \cup R_2 \cup R_3$, $R_1 = \{r\}$, $R_2 = \{b\}$, $R_3 = \{g\}$, where the agents have the following preferences over 3-tuples:

$$\left(\frac{1}{2}, \frac{1}{2}, 0 \right) \succ'_r \left(\frac{1}{2}, 0, \frac{1}{2} \right) \succ'_r (1, 0, 0) \succ'_r \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right);$$

$$\left(0, \frac{1}{2}, \frac{1}{2} \right) \succ'_b \left(\frac{1}{2}, \frac{1}{2}, 0 \right) \succ'_b (0, 1, 0) \succ'_b \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right);$$

$$\left(\frac{1}{2}, 0, \frac{1}{2} \right) \succ'_g \left(0, \frac{1}{2}, \frac{1}{2} \right) \succ'_g (0, 0, 1) \succ'_g \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right).$$

That is, the top choice of agent r is to be in coalition $\{r, b\}$, followed by $\{r, g\}$, followed by the singleton coalition, followed by the grand coalition. It is immediate that this game has no IS outcome. Indeed, every agent has an IS-deviation from the grand coalition, and if each agent is in a singleton coalition, r has an IS-deviation to $\{b\}$. Further, if an outcome consists of a coalition of size 2 and a singleton coalition, for one of the agents in the coalition of size 2 joining the singleton agent is an IS-deviation. To extend this example to $k > 3$ classes, we can simply add agents who prefer to be in homogeneous coalitions.

Moreover, for every fixed $k \geq 5$, deciding the existence of an IS outcome in a k-tuple HDG is NP-complete.

Theorem 5.4. *Given a k-tuple HDG G with $k \geq 5$, it is NP-complete to decide whether G has an individually stable outcome.*

In k-tuple HDGs, an agent may have very complex preferences over ratios of agents from different classes. In our second model, an agent does not distinguish among classes other than her own and hence only cares about the fraction of members of her class in her coalition; we will refer to these games as k-HDGs. We note that a similar approach has been used recently to provide a game-theoretic model of Schelling segregation with $k > 2$ agent classes (Elkind et al. 2019; Echzell et al. 2019); this line of work, while similar in spirit to hedonic diversity games, is, however, very different from a technical perspective.

Definition 5.5. A *hedonic diversity game with k-classes (k-HDG)* is a hedonic game $(N, (\succeq_i)_{i \in N})$ where N can be partitioned into k pairwise disjoint sets $R_1, \ldots, R_k$ so that for

each $j \in [k]$, each agent $i \in R_j$, and every pair of coalitions $C, D \in \mathcal{N}(i)$ we have $C \sim_i D$ as long as $\frac{|C \cap R_j|}{|C|} = \frac{|D \cap R_j|}{|D|}$.

By definition, every k-HDG is a k-tuple HDG, but the converse is not true, e.g., the game in Example 5.3 is not a k-HDG for any value of k. Further, a 2-HDG is simply an HDG as defined by Bredereck et al. Unlike k-tuple HDGs, k-HDGs admit a succinct representation even if k is not bounded by a constant: for each agent $i \in R_j$, her preferences over coalitions in $\mathcal{N}(i)$ can be described by her preferences over fractions of the form $\frac{\ell}{m}$, where $m \in [n]$, $\ell \in [\min\{|R_j|, m\}]$.

Remarkably, this formalism captures anonymous games.

Proposition 5.6. *Every anonymous game can be represented as a k-HDG.*

Proof. Recall that an anonymous game $G = (N, (\succeq_i)_{i \in N})$ with $|N| = n$ can be equivalently represented by a collection $(\succeq_i^*)_{i \in N}$ of weak orders over $[n]$: $C \succeq_i D$ if and only if $|C| \succeq_i^* |D|$. It follows that G can be viewed as an n-HDG with partition $N = R_1 \cup \ldots \cup R_n$, so that $R_i = \{i\}$ for each $i \in N$. Indeed, for each $i \in N$ and each coalition $C \in \mathcal{N}(i)$, the fraction of agents from class R_i in C is exactly $\frac{1}{|C|}$, so two coalitions $C, D \in \mathcal{N}(i)$ have the same fraction of agents from R_i if and only if they have the same size. $\square$

It follows that k-HDGs inherit negative results for anonymous games, such as non-existence of IS outcomes (Bogomolnaia and Jackson 2002) and hardness of deciding whether a given game has an IS outcome (Ballester 2004).

Corollary 5.7. *There exists a k-HDG that has no individually stable outcome. Moreover, deciding if a given k-HDG has an individually stable outcome is NP-complete.*

Now, Bogomolnaia and Jackson (2002) show that every single-peaked anonymous game has an IS outcome. The definition of single-peaked preferences extends naturally to k-HDGs, i.e., we can use essentially the same definition as for HDGs. However, it is not clear if the algorithm of Bogomolnaia and Jackson (2002) for finding an IS outcome in single-peaked anonymous games can be extended to single-peaked k-HDGs; this is an interesting question for future work. Note also that the hardness result of Corollary 5.7 only holds for $k = n$; it is not clear if the problem of finding an IS outcome in k-HDGs remains hard for small values of k (e.g., $k = 3$). Further, the anonymous game with no IS outcomes constructed by Bogomolnaia and Jackson (2002) has 63 agents and therefore translates into a 63-HDG; it remains an open problem whether k-HDGs with $k < 63$ are guaranteed to have an IS outcome.

6 Empirical Analysis

The algorithm for computing individually stable outcomes described in Section 4 has two substantial advantages over the algorithm of Bredereck et al.: first, it works for general preferences, and second, it is much simpler. However, as illustrated by Example 4.3, Bredereck et al.'s algorithm may result in higher agents' satisfaction. The IS-BRD algorithm is even simpler, but we do not know if it always converges

to an IS outcome. To better understand the performance of these three algorithms, in this section, we empirically compare them with respect to three measures: the average social welfare, the average coalition size and the average diversity.

Preference Models As Bredereck et al.'s algorithm is only defined for single-peaked HDGs, we only use single-peaked instances in our analysis. We consider three intuitively appealing ways of sampling strict preferences over ratios that are single-peaked on Θ.

Uniform single-peaked preferences (uSP). For each agent i, we sample $\succ_i'$ uniformly at random among all single-peaked strict orders on Θ, using the algorithm of Walsh (2015) (see also Lackner and Lackner (2017)).

Uniform-peak single-peaked preferences (upSP). To generate $\succ_i'$, we first select i's most preferred ratio by choosing a point in Θ uniformly at random. We then continue to place elements of Θ in positions $2, \ldots, |\Theta|$ of $\succ_i'$ one by one; when $k < |\Theta|$ elements have been ranked, there are at most two elements of Θ that can be placed in position $k + 1$ so that the resulting ranking is single-peaked on Θ, and we choose between them with equal probability. This approach to sampling single-peaked preferences was popularized by Conitzer (2009).

Symmetric single-peaked preferences (symSP). For each agent i, we choose her preferred point θ_i from the uniform distribution on $[0, 1]$ and define the relation $\succeq_i'$ so that $\theta \succeq_i' \theta'$ if and only if $|\theta - \theta_i| \leq |\theta' - \theta_i|$. While theoretically the resulting relation may have ties, in our experiments this approach always generated strict orders.

The first two distributions are quite different from each other, e.g., in the uSP model a ranking where 0 appears first is exponentially less likely than a ranking where $\frac{1}{2}$ appears first, while in the upSP model we are equally likely to see 0 and $\frac{1}{2}$ ranked first. On the other hand, upSP and symSP appear to be fairly similar, but our experiments show that the three algorithms behave differently on them.

Performance measures The primary measure we are interested in is social welfare, i.e., the sum of agents' utilities. However, in general, this measure is difficult to define, since in HDGs agents' preferences over coalitions are given by weak orders rather than numerical values. For symmetric single-peaked preferences, we can circumvent this difficulty by defining an agent's disutility as the difference between the fraction of red agents in her coalition and her ideal ratio, so, given a partition π, for each $i \in N$ we set

$$\omega(i) = 1 - \left| \frac{|\pi_i \cap R|}{|\pi_i|} - \theta_i \right|.$$

For uniform and uniform-peak single-peaked preferences, we identify the utility of agent i in partition π with the Borda score of $\theta(\pi_i) = \frac{|\pi_i \cap R|}{|\pi_i|}$ in $\succ_i'$: for each $\theta \in \Theta$, we set

$$\beta(i, \theta) = |\{\theta' \in \Theta : \theta \succ_i' \theta'\}| \text{ and } \omega(i) = \frac{\beta(i, \theta(\pi_i))}{|\Theta| - 1}.$$

In both cases, we define the average welfare as

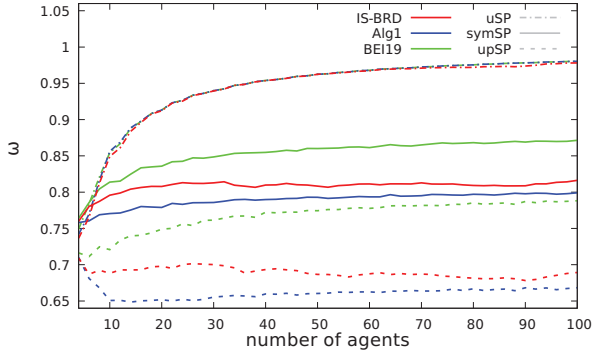$$\omega(\pi) = \frac{1}{n} \sum_{i \in N} \omega(i).$$
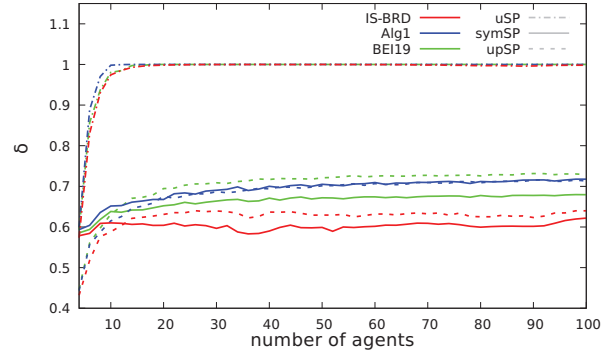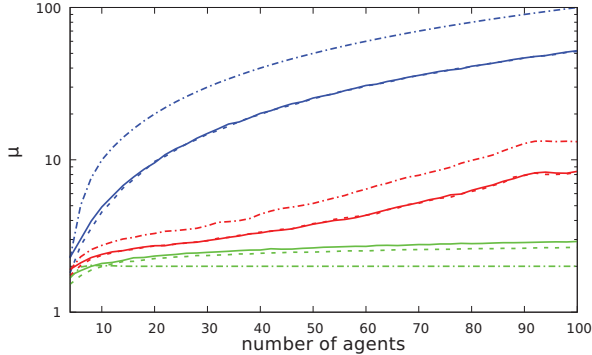
Figure 1: Average social welfare

Figure 2: Average coalition size

To gain additional insight into the behavior of our algorithms, we also consider two other measures, namely, the average coalition size and the average diversity. We measure the diversity of a coalition C as $\delta(C) = 1 - 2\left|\frac{1}{2} - \frac{|C \cap R|}{|C|}\right|$; note that $\delta(C) = 0$ if C is homogeneous and $\delta(C) = 1$ if C is balanced. Both for the average size and for the average diversity, we average over all agents in the partition rather than all coalitions, as we focus on the experience of an individual agent. That is, the average coalition size and the average diversity are defined, respectively, as

$$\mu(\pi) = \frac{1}{n}\sum_{i \in N}|\pi_i| \quad \text{and} \quad \delta(\pi) = \frac{1}{n}\sum_{i \in N}\delta(\pi_i).$$

Results The results of our experiments are shown in Figures 1–4. In all graphs, BEI19 refers to the algorithm of Bredereck et al., Alg1 refers to our Algorithm 1 and IS-BRD refers to the IS-BRD algorithm that starts with a uniformly random partition of agents and at each iteration chooses the deviation uniformly at random from among all available deviations. Remarkably, on every instance we have generated, IS-BRD converges in $\mathcal{O}(n^2)$ iterations.

In Figures 1–3, we first consider HDGs with the same number of red and blue agents. For each $s = 2, \ldots, 50$, we generate 1000 HDGs with s red and s blue agents; thus, $n = 2s$ takes even values from 4 to 100.

Figure 3: Average diversity

Our results for social welfare are shown in Figure 1. For uSP, all algorithms perform similarly and provide fairly high social welfare. Intuitively, this is because under this preference model the agents are likely to rank the ratio $\frac{1}{2}$ highly, and therefore they are quite happy to be in a coalition with an approximately equal number of red and blue agents. Fortunately, all of our algorithms are good at partitioning the agents into such coalitions. In particular, under this distribution it is likely that each agent weakly prefers a balanced coalition to a homogeneous coalition, so Algorithm 1 will be able to stop after setting $C_0 = N$, and the grand coalition will be well-liked by most agents. However, the other two distributions tell a different story: under these distributions, the algorithm of Bredereck et al. substantially outperforms the other two algorithms, with IS-BRD being consistently better than Algorithm 1. Thus, if the distribution of agents' top choices is close to uniform and the numbers of red and blue agents are equal, the more complex algorithm of Bredereck et al. has an advantage over our approach.

Figure 2 shows that the three algorithms are very different in terms of the sizes of coalitions they produce: Algorithm 1 tends to produce the largest coalitions, and the algorithm of Bredereck et al. tends to produce the smallest coalitions; this holds irrespective of how we sample the agents' preferences.

Figure 3 shows that all three algorithms achieve almost perfect diversity under the uSP distribution; we have suggested possible reasons for this when discussing the social welfare. For the other two distributions, Algorithm 1 and the algorithm of Bredereck et al. perform similarly, while IS-BRD results in somewhat less diverse partitions.

We also investigate what happens if the two agent classes have different sizes. In Figure 4, we graph the average social welfare of individually stable outcomes obtained by the three algorithms as the number of agents is fixed to $n = 50$ and the fraction of red agents in the game (denoted by $\theta(N)$) changes from 0 to 0.5. In this setting, too, the algorithm of Bredereck et al. outperforms the other two algorithms with respect to social welfare, for almost all values of $\theta(N) = \frac{|R|}{|R|+|B|}$; the difference is the most pronounced for the upSP distribution. However, for the uSP distribution, there is a range of values of $\theta(N)$ where Algorithm 1 has better performance than the algorithm of Bredereck et al.;
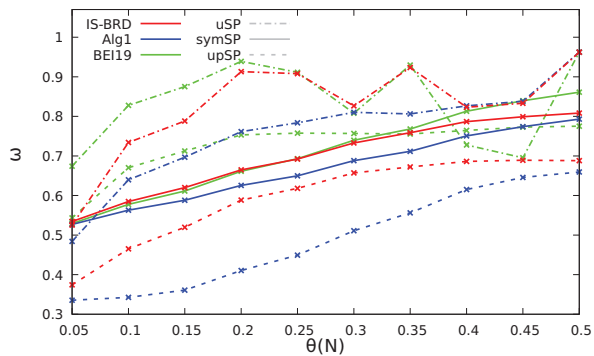
Figure 4: Average social welfare

this can be attributed to the fact that the latter algorithm only produces coalitions that are very unbalanced or perfectly balanced, while Algorithm 1 can output coalitions with more complex ratios.

Finally, we note that all of our algorithms are very fast; we have not explicitly measured their performance, but on every input size we have considered, each of them converges in a few milliseconds. Comparing the three algorithms, Algorithm 1 converges approximately three times faster than the algorithm of Bredereck et al. and 50 times faster than IS-BRD.

7 Conclusions

Our work contributes to the study of hedonic diversity games—an interesting class of coalition formation games introduced by Bredereck et al. We have focused on stability concepts that deal with deviations by individual agents, namely, Nash stability and individual stability. Remarkably, while our algorithm for finding IS outcomes is theoretically more appealing than the algorithm of Bredereck et al., in that it is simpler and more general, empirically the latter algorithm has better performance for the domain on which it is defined. Perhaps the most interesting open question concerning individual stability is whether IS-BRD is guaranteed to converge to an IS outcome, and if yes, whether convergence always happens after polynomially many iterations; we note that, in our experiments, this is always the case. The same question can be asked in the context of anonymous games. We also feel that the k-HDG model deserves further attention: in particular, we would like to understand the complexity of finding IS outcomes for small values of k and/or for single-peaked preferences.

References

Aziz, H., and Savani, R. 2016. Hedonic games. In Brandt, F.; Conitzer, V.; Endriss, U.; Lang, J.; and Procaccia, A., eds., *Handbook of Computational Social Choice*. Cambridge University Press. chapter 15.

Aziz, H.; Brandl, F.; Brandt, F.; Harrenstein, P.; Olsen, M.; and Peters, D. 2019. Fractional hedonic games. *ACM Transactions on Economics and Computation* 7(2):6:1–6:29.

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

Boehmer, N., and Elkind, E. 2019. Individual-based stability in hedonic diversity games. *arXiv preprint arXiv:1911.08669*.

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

Bredereck, R.; Elkind, E.; and Igarashi, A. 2019. Hedonic diversity games. In *Proceedings of the 18th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 565–573.

Conitzer, V. 2009. Eliciting single-peaked preferences using comparison queries. *Journal of Artificial Intelligence Research* 35:161–191.

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

Echzell, H.; Friedrich, T.; Lenzner, P.; Molitor, L.; Pappik, M.; Schöne, F.; Sommer, F.; and Stangl, D. 2019. Convergence and hardness of strategic Schelling segregation. In *Proceedings of the 16th Conference on Web and Internet Economics (WINE)*.

Elkind, E.; Gan, J.; Igarashi, A.; Suksompong, W.; and Voudouris, A. A. 2019. Schelling games on graphs. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, 266–272.

Gairing, M., and Savani, R. 2010. Computing stable outcomes in hedonic games. In *Proceedings of the 3rd International Symposium on Algorithmic Game Theory (SAGT)*, 174–185.

Lackner, M.-L., and Lackner, M. 2017. On the likelihood of single-peaked preferences. *Social Choice and Welfare* 48(4):717–745.

Peters, D. 2016. Complexity of hedonic games with dichotomous preferences. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, 579–585.

Walsh, T. 2015. Generating single peaked votes. *CoRR* abs/1503.02766.