Mechanism Design for Ordinal Classes of Hedonic Games

Martin Bullinger
University of Oxford
Oxford, United Kingdom
martin.bullinger@cs.ox.ac.uk

Edith Elkind
Northwestern University
Evanston, United States
edith.elkind@northwestern.edu

ABSTRACT

We study coalition formation in the framework of hedonic games from a mechanism design perspective. We consider three classes of games in which preferences are derived from ordinal information: anonymous hedonic games, hedonic diversity games, and $\mathcal{W}$-hedonic games. Our goal is to find strategyproof mechanisms that output individually rational outcomes and offer nontrivial welfare guarantees. While all classes of games we consider admit individually rational outcomes whose welfare is within a constant factor from optimal, our basic family of mechanisms only achieves a linear-factor welfare approximation. However, these basic mechanisms can be refined to be strategyproof, individually rational, and to provide a constant-factor welfare approximation when considering single-peaked domains or when allowing randomization.

KEYWORDS

Algorithmic game theory, coalition formation, hedonic games, mechanism design, strategyproofness

1 INTRODUCTION

The formation of a cohesive community is essential for the welfare or social good of a society. It aids towards improving health [16], the well-being of older people [17], and is a crucial in the fight against climate change [35]. Social cohesion can be defined as the "quality of social cooperation" [21]. We aim at a better understanding of cooperation by studying coalition formation through the lens of algorithmic game theory and mechanism design.

Coalition formation captures settings where a set of agents needs to be partitioned into disjoint coalitions, and agents are endowed with preferences over the resulting partitions. Typical examples in the cooperative game theory literature involve dividing a set of workers into groups performing independent tasks, or assigning a set of students to group projects. However, coalition formation games also offer a representation formalism for social networks, where coalitions may describe communities of agents [4] or a framework for clustering, a major task in machine learning with far-ranging applications such as image analysis [1, 2].

We study coalition formation in the framework of hedonic games. This model assumes that the only information relevant for agents' preferences is the coalition they belong to, i.e., they do not care how agents outside of their coalition partition themselves into groups [22]. The criteria for evaluating the resulting partition include *stability*, i.e., absence of incentives for agents to move to a different coalition, and *optimality*, i.e., global guarantees, which can be measured, for instance, by utilitarian welfare.

Hedonic games literature typically assumes availability of full and correct information about all agents' preferences. However, in practice, agents cannot always be expected to report their preferences truthfully. Therefore, in this work, we take a *mechanism design* perspective: our goal is to design algorithms that can find good partitions even if agents are *strategic*, i.e., they may misreport their preferences if this will lead to them being placed in a more desirable coalition. This perspective was already present in the early literature on hedonic games, where the goal was to identify strategyproof mechanisms that produce stable outputs [3, 10, 15]. More recent work has focused on classes of hedonic games with a cardinal utility representation of preferences, such as additively separable and fractional hedonic games [25, 26, 40]. In particular, Flammini et al. [25] and Flammini et al. [26] seek to find strategyproof mechanisms whose outputs yield good approximations of utilitarian welfare.

We complement this work by providing an in-depth investigation of classes of hedonic games in which preferences are derived from ordinal rankings. Specifically, we study anonymous hedonic games, hedonic diversity games, and $\mathcal{W}$-hedonic games [10, 12, 15]. In these classes of games, agents submit rankings over their preferred coalition sizes, proportion of two agent types, and worst-case coalition partners, respectively.

We want our mechanisms to satisfy three key properties: (i) strategyproofness, i.e., an agent cannot misrepresent their preferences to improve their outcome, (ii) individual rationality, i.e., an agent likes their assigned coalition at least as much as being on their own, and (iii) high welfare.

Since preferences in the games we consider are not directly associated with a utility function, they lack an endogenous quantitative measure of welfare. Hence, following a common approach in the literature on hedonic games [see, e.g., 18, 31], we exogenously assign *Borda utilities* to the agents' submitted ordinal rankings and define the Borda welfare of an outcome of the sum of Borda utilities.[1] Beyond their intuitive appeal, using Borda utilities have significant support as a natural and reasonable approach. Indeed, it has recently been argued that reinforcement learning from human feedback aggregates over hidden contexts using the Borda rule [38], i.e., the Borda rule emerges naturally in aggregation contexts even without being explicitly introduced. It also admits several attractive axiomatic characterizations [24, 32, 41].

We first show that individual rationality is compatible with a constant-factor approximation of the Borda welfare. We then study the compatibility of welfare guarantees with strategyproofness. For each class of games we consider, we present a family of mechanisms that are strategyproof, individually rational, and achieve a linear-factor approximation of the maximum Borda welfare. Moreover, we carve out two possibilities to improve these mechanisms to yield a constant-factor approximation of the maximum Borda welfare by (1) restricting the domain to single-peaked preferences

[1] Borda utilities in the presence of ordinal rankings have also been used in other areas of social choice, such as fair division [19, 36].

and (2) enhancing the capabilities of mechanisms by allowing for randomization.

2 RELATED WORK

Coalition formation is a central concern in game theory, and can be traced back as far as to the seminal work by von Neumann and Morgenstern [39] eight decades ago. The framework of hedonic games is significantly more recent, and was introduced by Drèze and Greenberg [22]. After key publications by Bogomolnaia and Jackson [10], Banerjee et al. [8], and Cechlárová and Romero-Medina [15], hedonic games have received a steady stream of attention; see the book chapters by Aziz and Savani [6] and Bullinger et al. [13] for an overview.

A major challenge in the anaysis of hedonic games is to define useful and computationally tractable preference representation formalisms. This is important because agents have to express preferences over all possible coalitions they can be part of, i.e., over a set of exponential size with respect to the number of agents. Over the years, a rich variety of succinct classes of hedonic games have been introduced, capturing various priorities in a coalition formation process [see, e.g., 4, 10, 12, 14, 20, 23]. The most common approach is to ask each agent to supply a ranking over a small set of objects, and then deduce the agent's preferences over all possible coalitions from that ranking. Sometimes this ranking is given by specifying cardinal utilities for individual agents, as in additively separable hedonic games [10] and fractional hedonic games [4]; other classes of games rely on an ordinal ranking that is not associated explicitly with a utility interpretation.

As described in the introduction, in our work we consider three classes of games of the latter type, namely, anonymous hedonic games [10], hedonic diversity games [12], and $\mathcal{W}$-hedonic games [15]. Previous research on these classes of games has focused on identifying outcomes that satisfy stability and optimality criteria, such as individual stability or Pareto optimality [see, e.g., 5, 7, 9, 11]. However, strategyproofness was also already considered in the early work on anonymous hedonic games: In their concluding remarks, Bogomolnaia and Jackson [10] warn that strategyproofness can be a demanding desideratum as it is incompatible with individual stability for single-peaked anonymous hedonic games. Moreover, Cechlárová and Romero-Medina [15] also propose a class of games based on best players, and present an algorithm similar to Gale's top-trading algorithm [37] that is strategyproof and yields outcomes in the strict core.

Subsequently, strategyproofness was used for axiomatic characterizations. Alcalde and Revilla [3] single out the top covering algorithm as the only strategyproof algorithm returning stable outcomes for a class of hedonic games based on so-called top-responsive preferences. Further, Rodríguez-Álvarez [30] characterizes strategyproof mechanisms on domains generalizing additively separable hedonic games. Notably, he uses axioms similar to our desiderata: Like us, he demands individual rationality, but, in addition, he requires a form of Pareto efficiency instead of aiming at an approximation of welfare.

A more recent stream of work studies strategyproofness in hedonic games with a cardinal utility representation, which, in contrast to our classes of games, admit an endogenous notion of welfare. Specifically, Wright and Vorobeychik [40] consider additively separable hedonic games with positive utilities and bounded coalition sizes. Flammini et al. [25, 26] are interested in strategyproof mechanisms with good approximations of utilitarian welfare in additively separable and fractional hedonic games. Since the cardinal utility representation of these games has a direct encoding as weighted graphs, their mechanisms make use of the combinatorial structure: They compute partitions based on connected components or matchings. While matching-based algorithms turn out to perform well for $\mathcal{W}$-hedonic games, we employ novel ideas for anonymous hedonic games and hedonic diversity games.

3 PRELIMINARIES

For a positive integer $k \in \mathbb{N}$, we write $[k] := \{1, \dots, k\}$. Moreover, given a strict order $\succ$, we write $a \succeq b$ as a shorthand for '$a \succ b$ or $a = b$'.

3.1 Hedonic Games

Hedonic games aim at partitioning a set of agents into pairwise disjoint coalitions of agents according to the agents' preferences. We consider a finite set $N = \{a_i : i \in [n]\}$ of n agents; later, we will use the agents' indices for the design of our mechanisms. A nonempty subset of N is called a *coalition*. A *partition* of N is a subset $\mathfrak{C} \subseteq 2^N$ such that $\bigcup_{C \in \mathfrak{C}} C = N$, and for every pair of coalitions $C, D \in \mathfrak{C}$ with $C \neq D$ it holds that $C \cap D = \emptyset$. We denote the set of all partitions of N by $\mathfrak{P}(N)$. Given a partition $\mathfrak{C} \in \mathfrak{P}(N)$, we denote by $\mathfrak{C}(i)$ the coalition containing agent i. A coalition of size 1 is called a *singleton coalition*, and the partition that consists of singleton coalitions is called the *singleton partition*. The partition $\{N\}$ consisting of a single coalition is called the *grand coalition*.

Let $\mathcal{N}_i$ denote the set of all possible coalitions containing agent i, i.e., $\mathcal{N}_i := \{C \subseteq N : i \in C\}$. A *hedonic game* is defined by a pair $(N, \succsim)$, where N is a set of agents and $\succsim = (\succsim_i)_{i \in N}$ is a profile of weak orders: $\succsim_i$ is a weak order over $\mathcal{N}_i$ that represents the preferences of agent i. These weak orders induce weak orders over partitions of N: we set $\mathfrak{C}' \succsim_i \mathfrak{C}$ if and only if $\mathfrak{C}'(i) \succsim_i \mathfrak{C}(i)$. A coalition $C \in \mathcal{N}_i$ is said to be *individually rational* for agent i if $C \succsim_i \{i\}$; a partition $\mathfrak{C}$ is said to be *individually rational* if $\mathfrak{C}(i)$ is individually rational for each $i \in N$. Given a partition $\mathfrak{C}$, we denote by $\mathcal{I}(\mathfrak{C})$ the set of individually rational coalitions contained in $\mathfrak{C}$.

As $|\mathcal{N}_i| = 2^{n-1}$ for all $i \in N$, representing $\succsim_i$ explicitly is impractical for large n. Therefore, from an algorithmic perspective, it is natural to consider classes of hedonic games that admit a succinct encoding. In this work, we will consider three such classes, defined below.

In *anonymous hedonic games* (AHGs), introduced by Bogomolnaia and Jackson [10], the agents only care about the size of the coalition they belong to. Formally, each agent $i \in N$ is equipped with a strict[2] order $\succ_i^S$ over the integers in $[n]$ (superscript S for sizes) such that

$$C \succsim_i C' \text{ if and only if } |C| \succsim_i^S |C'|.$$

Note that a strict order over coalition sizes induces a weak order over coalitions as an agent is indifferent between any two coalitions

[2] One can also define AHGs and the subsequently defined classes of hedonic games based on weak orders. For our treatment of welfare, it is, however, preferable to assume strict orders.

of the same size. We represent an AHG by the pair $(N, \succ^S)$ where $\succ^S = (\succ_i^S)_{i \in N}$ is the profile of the agents' orders over coalition sizes.

In *hedonic diversity games* (HDGs), introduced by Bredereck et al. [12], the agents are divided into two different *types* (or *colors*). Formally, we partition N as $N = B \cup R$, where $R \cap B = \emptyset$; the agents in B and R are called *blue* and *red* agents, respectively. We denote the cardinalities of these sets by $r := |R|$ and $b := |B|$. For a nonempty coalition $C \subseteq N$, let $f_C := \frac{|R \cap C|}{|C|}$ be the fraction of red agents in C. In an HDG, each agent only cares about the proportion of red agents present in their own coalition, i.e., for each agent $i \in N$ there exists a strict order $\succ_i^F$ over the fractions in $F := \left\{ \frac{p}{p+q} : p \in [r] \cup \{0\}, q \in [b] \cup \{0\}, p + q \geq 1 \right\}$ such that

$$C \succsim_i C' \text{ if and only if } f_C \succeq_i^F f_{C'}.$$

An HDG is said to be *balanced* if $r = b$, i.e., if the number of agents of different types is equal.[3] We represent an HDG by the pair $(N, \succ^F)$ where $\succ^F = (\succ_i^F)_{i \in N}$ is the profile of the agents' orders over coalition ratios. Note that the fractions of 0 and 1 correspond to coalitions consisting of only blue and only red agents, respectively. Hence, a fraction of 0 (or 1) can never be attained by the coalition of a red (or blue) agent and we omit them from their respective rankings.

In *$\mathcal{W}$-hedonic games* (WHGs), introduced by Cechlárová and Romero-Medina [15], agents care about the worst agent in their coalition. Formally, each agent $i \in N$ is equipped with a strict order $\succ_i^A$ over the agent set N (including themselves). For a coalition $C \in \mathcal{N}_i$ with $C \neq \{i\}$, we define $\min_{\succ_i^A}(C)$ as the worst agent in $C \setminus \{i\}$ according to $\succ_i^A$. Moreover, let $\min_{\succ_i^A}(\{i\}) = i$. Then, agent i's preference over coalitions in $\mathcal{N}_i$ is defined so that

$$C \succsim_i C' \text{ if and only if } \min_{\succ_i^A}(C) \succeq_i^A \min_{\succ_i^A}(C').$$

In other words, agents rank coalitions by the worst agent in the coalition excluding themselves. Thus, a nonsingleton coalition $C \in \mathcal{N}_i$ is individually rational for i if and only if i ranks the worst agent in $C \setminus \{i\}$ above themselves. We represent a $\mathcal{W}$-hedonic game by the pair $(N, \succ^A)$ where $\succ^A = (\succ_i^A)_{i \in N}$ is the profile of the agents' orders over N.

3.2 Mechanism Design and Objectives

The key idea of the mechanism design perspective on hedonic games is that a game is not given explicitly, by listing all agents' preferences. Instead, the agents' preferences need to be elicited from the agents themselves, and the agents may lie about their preferences if they can benefit from doing so.

We now consider mechanisms for AHGs, HDGs, or WHGs; we denote a generic game of one of these classes by $(N, \succ^X)$, where $X \in \{S, F, A\}$, which induces the generic hedonic game $(N, \succsim)$. A (deterministic) *mechanism* $\mathcal{M}$ for a class of hedonic games (e.g., AHGs) elicits the agents' preferences over the relevant set of objects (e.g., in case of AHGs, over coalition sizes), transforms them into a game in that class, and maps this game to an output partition; thus, we can view a mechanism as a mapping from games to partitions. We denote by $\mathcal{M}(N, \succ^X) \in \mathfrak{P}(N)$ the output of the mechanism $\mathcal{M}$ for game $(N, \succ^X)$.

The first property that we want a mechanism to satisfy is that agents cannot benefit from declaring preferences different from their true orders. Given a preference profile $\succ = (\succ_i)_{i \in N}$ and a preference order $\hat{\succ}_i$ for agent i, we denote by $(\succ_{-i}, \hat{\succ}_i)$ the preference profile where agent i has preference order $\hat{\succ}_i$ and, for $j \in N \setminus \{i\}$, agent j has preference order $\succ_j$. A mechanism $\mathcal{M}$ is said to be *strategyproof* if for every game $(N, \succ^X)$, agent $i \in N$, and preference order $\hat{\succ}_i^X$, it holds that $\mathcal{M}(N, \succ^X) \succsim_i \mathcal{M}(N, (\succ_{-i}^X, \hat{\succ}_i^X))$. Hence, by reporting truthfully, the agent obtains a weakly best outcome among the ones they can achieve, holding other agents' reports fixed.

Our second key property is individual rationality, which we view as a nonnegotiable requirement. A mechanism $\mathcal{M}$ is said to be *individually rational* if for every game $(N, \succ^X)$ it holds that $\mathcal{M}(N, \succ^X)$ is individually rational.

Finally, we want mechanisms to provide welfare guarantees. Following previous work on hedonic games, we obtain a quantification of preferences by associating to the objects in the ordinal rankings of the agents a Borda score measuring its position from the bottom [see, e.g., 18, 31]. More formally, we define the *Borda utility* of an agent i for a coalition C as

$$u_i(C) := \begin{cases} |\{s \in [n] \colon |C| \succ_i^S s\}| & \text{for AHGs,} \\ |\{f \in F \colon f_C \succ_i^F f\}| & \text{for HDGs,} \\ |\{j \in N \colon \min_{\succ_i^A}(C) \succ_i^A j\}| & \text{for WHGs.} \end{cases}$$

Hence, the utility of a coalition measures how many other coalition sizes, coalition fractions, or agents are beaten by its own size, fraction, or worst agent. Moreover, we define the *Borda welfare* of a partition $\mathfrak{C}$ as

$$\mathcal{SW}(\mathfrak{C}) := \sum_{i \in N} u_i(\mathfrak{C}(i)).$$

Since this is the only type of utility and welfare that we consider in this paper, we refer to Borda utilities and Borda welfare simply as utilities and welfare, respectively.

Our welfare objective is to be as close as possible to the maximum possible welfare. We distinguish two guarantees of how close we are to this goal. We say that a mechanism achieves a *constant-factor approximation* of the maximum welfare if there exist constants $\gamma \geq 1$ and $n_0 \in \mathbb{N}$ such that for every game $(N, \succ^X)$ with $n \geq n_0$ it holds that

$$\gamma \cdot \mathcal{SW}(\mathcal{M}(N, \succ^X)) \geq \max_{\mathfrak{C} \in \mathfrak{P}(N)} \mathcal{SW}(\mathfrak{C}). \tag{1}$$

If Equation (1) is replaced by $\gamma m \cdot \mathcal{SW}(\mathcal{M}(N, \succ^X)) \geq \max_{\mathfrak{C} \in \mathfrak{P}(N)} \mathcal{SW}(\mathfrak{C})$, we speak of a *linear-factor approximation*. There, m is the number of alternatives ranked by the agents, i.e., $m = n$ for AHGs and WHGs and $m = |F| - 1 = \Theta(n^2)$ for HDGs.

In addition to deterministic mechanisms, we also consider *randomized mechanisms*. A randomized mechanism can sample from a distribution $\mathcal{U}$ and use the resulting random sample $u \sim \mathcal{U}$ in its decisions; its output on a game $(N, \succ^X)$ is then a probability distribution $\mathcal{M}(N, \succ^X, u)$ over $\mathfrak{P}(N)$. Note that, by fixing u, we

[3] Note that for $p, q \in \mathbb{N}$, p and q are coprime if and only if p and $p + q$ are coprime. Moreover, the proportion of coprime numbers, i.e., the fraction of coprime pairs in $[k] \times [k]$ converges to $\frac{6}{\pi^2}$ as k tends to infinity [29]. Hence, in balanced HDGs, agents provide rankings over a set of $\Theta(n^2)$ alternatives.

transform a randomized mechanism $\mathcal{M}$ into a deterministic mechanism $\mathcal{M}_u(N, \succ^X) := \mathcal{M}(N, \succ^X, u)$. We say that $\mathcal{M}$ is *universally strategyproof* (or *universally individually rational*) if $\mathcal{M}_u$ is strategyproof (or individually rational) for each $u \in \mathcal{U}$. Finally, we say that a randomized mechanism achieves a *constant-factor approximation* of the maximum welfare if Equation (1) holds in expectation over $\mathcal{U}$.

4 WELFARE OF INDIVIDUALLY RATIONAL OUTCOMES

We start by investigating welfare guarantees and show that there always exists an individually rational coalition structure that approximates the maximum welfare within a constant factor. The proof relies on a counting argument that provides a threshold such that some alternative within a sufficiently large set of alternative is ranked above the threshold by a constant fraction of voters. This is made rigorous in Lemma 4.1. All missing proofs here and in the following can be found in the appendix.

Lemma 4.1. *Let M be a set of m alternatives. Let $\alpha, \beta \in (0,1)$ and $H \subseteq M$ be a subset of alternatives with $|H| \geq \lceil \alpha m \rceil$. Set $\gamma = (1-\alpha) + \alpha\beta$. Then, in every game where n agents provide rankings over M, some alternative in H is ranked in position at most $\lceil \gamma m \rceil$ by at least βn agents.*

We use Lemma 4.1 to show the existence of individually rational coalition structures with a constant-factor approximation of the maximum welfare. To apply Lemma 4.1, we consider a set H consisting of coalition sizes or coalition fractions that lead to small coalitions. Hence, one such size or fraction has to be ranked high by a constant fraction of agents. We can use it to extract a partition of high welfare. We illustrate the idea with the proof for AHGs and defer the case of HDGs to the appendix.

Theorem 4.2. *For AHGs (or balanced HDGs), there exists an individually rational partition that achieves a constant-factor approximation of the maximum welfare.*

Proof for AHGs. In these games, agents provide rankings of length n. Since the maximum welfare is bounded by $n(n-1)$, it suffices to prove the existence of an individually rational partition with a welfare of $\Theta(n^2)$.

Consider the set $H = \left[\left\lceil \frac{n}{4} \right\rceil\right]$. By Lemma 4.1 for $\alpha = \frac{1}{4}$ and $\beta = \frac{1}{2}$, some coalition size $s \in H$ is ranked at position at most $\left\lceil \frac{7}{8}n \right\rceil$ by at least $\frac{n}{2}$ agents. Assume first that at least half of these $\frac{n}{2}$ agents, i.e., at least $\frac{n}{4}$ agents prefer s to a coalition size of 1. Since $s \leq \left\lceil \frac{n}{4} \right\rceil$, we can place at least half of these $\frac{n}{4}$ agents, i.e., a total of $\frac{n}{8}$ agents, to individually rational coalitions of size s. This achieves a welfare of at least $\frac{n}{8}\left\lfloor \frac{1}{8}n \right\rfloor = \Theta(n^2)$. Otherwise, there are at least $\frac{n}{4}$ agents that achieve a utility of $\left\lfloor \frac{1}{8}n \right\rfloor$ in a singleton coalition, and then the singleton partition achieves a welfare of at least $\frac{n}{4}\left\lfloor \frac{1}{8}n \right\rfloor = \Theta(n^2)$. □

We remark that the assumption that HDGs are balanced is crucial for achieving high welfare in HDGs. Otherwise, imposing the requirement of individual rationality may lower the welfare by a factor that is linear in n.

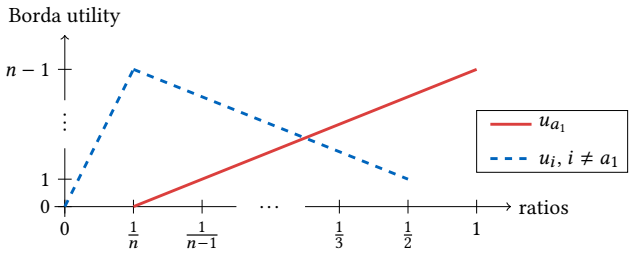

Figure 1: Visualization of Borda utilities in Example 4.3.

Example 4.3. Consider an HDG with n agents: one red agent a_1 and $n-1$ blue agents. Their Borda utilities are depicted in Figure 1. The red agent a_1 prefers larger over smaller fractions, i.e., their top-ranked fraction is 1. Then the only individually rational coalition for a_1 is the singleton coalition. If this coalition forms, then all other agents are in coalitions of ratio 0 since they are all blue. Suppose that all blue agents prefer smaller ratios to larger ratios, except for a ratio of 0, which is the least preferred. That is, we have $\frac{1}{n} \succ_i^F \frac{1}{n-1} \succ_i^F \cdots \succ_i^F \frac{1}{3} \succ_i^F \frac{1}{2} \succ_i^F 0$ for $i \in N \setminus \{a_1\}$. Then, the welfare in every individually rational coalition structure is $n-1$, whereas the welfare of the grand coalition is $(n-1)^2$. ◁

For WHGs, the technique used for deriving Theorem 4.2 no longer works: Knowing that one agent (say, a) is ranked highly by many agents does not imply that we can create a large individually rational coalition in which each agent achieves the utility associated with a: When forming large coalitions including a, the coalition members can negatively influence each other's utility. In fact, determining large individually rational coalitions (even when containing a specific agent) is computationally hard, even when only an approximation is required, see Appendix B. Still, we obtain a result analogous to Theorem 4.2 with a different proof technique: our algorithm finds a matching of the agents that achieves high welfare.

More precisely, our algorithm considers agents one by one and forms singleton coalitions of high utility or individually rational pairs that yield high utility for one of its members. This part is called the main stage of the algorithm. After the main stage, we form singleton coalitions with the agents not assigned to coalitions, yet. Clearly, the resulting partition is individually rational. Moreover, it can have high welfare for two reasons. First, if we create many coalitions in the main stage, we immediately have a high welfare. Second, we can have a large number of agents not assigned to coalitions in the main stage. These agents do not form individually rational coalitions with a large fraction of other agents. Hence, accumulating these large fractions achieves a high welfare because of individual rationality.

Theorem 4.4. *For WHGs, there exists an individually rational partition that achieves a constant-factor approximation of the maximum welfare.*

5 STRATEGYPROOF MECHANISMS

For the remainder of the paper, our goal is to come up with strategyproof mechanisms. In addition, we demand individual rationality. One easy way to achieve this is to always output the singleton

partition. However, this does not offer any worst-case welfare guarantees. Hence, our third goal is approximation guarantees regarding welfare.

5.1 General Deterministic Mechanisms

We first define a simple class of mechanisms for AHGs, which we will repeatedly consider throughout the paper.

Mechanism $\mathcal{M}_x^{AHG}$. Let $x \in [n] \setminus \{1\}$. The mechanism $\mathcal{M}_x^{AHG}$ proceeds as follows. First, it identifies the set $S_x := \{i \in N \colon x \succ_i^S 1\}$. Let $s_x := |S_x|$. Then, $\mathcal{M}_x^{AHG}$ picks the $x \cdot \left\lfloor \frac{s_x}{x} \right\rfloor$ agents in S_x with the lowest indices, and places them into $\left\lfloor \frac{s_x}{x} \right\rfloor$ coalitions of size x. The remaining agents are placed in singleton coalitions.

We will now consider the properties of mechanisms of this form.

Proposition 5.1. *For $x \in [n] \setminus \{1\}$, the mechanism $\mathcal{M}_x^{AHG}$ for AHGs is strategyproof, individually rational, and achieves a welfare of at least $n - x + 1$.*

Proof. Individual rationality is straightforward because every agent can only be in a coalition of size x or of size 1, and the former is only possible for agents i with $x \succ_i^S 1$.

For strategyproofness, note that every agent ends up in a coalition of size x or of size 1. Also, the mechanism only relies on the preferences concerning these coalition sizes. Consider an agent i. If i reports that they prefer 1 to x, they are guaranteed to end up in a singleton coalition. Hence, if i's truthful preference is $1 \succ_i x$ then i ends up in a coalition of their preferred size, out of the two possible options. Further, if $x \succ_i 1$, then i can only change the outcome (relative to truthful reporting) if they report that they prefer 1 to x. But then they will end up in a singleton coalition, which is not an improvement. Hence, the mechanism is strategyproof.

Finally, we consider welfare. Note that the utility can only be 0 for an agent that prefers x to 1 but is placed in a singleton coalition. As there can be at most $x - 1$ such agents and the utility of all other agents is at least 1, the claim follows. □

Notably, our bound on the welfare is tight. In an instance where all agents rank 1 and x in the last two positions, and $n-x+1$ agents rank 1 before x, $\mathcal{M}_x^{AHG}$ creates the singleton partition and achieves a welfare of $n - x + 1$.

Moreover, $\mathcal{M}_2^{AHG}$ achieves a welfare of $n - 1$, which is only a linear factor worse than the maximum welfare, which is bounded by $n(n-1)$ in any instance.

Corollary 5.2. *There exists a strategyproof and individually rational mechanism for AHGs that achieves a linear-factor approximation of the maximum welfare.*

A similar class of mechanisms can be defined for HDGs. There, we form one coalition that is as large as possible subject to satisfying a given ratio of red and blue agents.

Mechanism $\mathcal{M}_f^{HDG}$. Given an HDG, let $f \in F \setminus \{0, 1\}$ be a feasible fraction not corresponding to a singleton coalition. Suppose that $f = \frac{r'}{r'+b'}$ where $r' \in [r]$, $b' \in [b]$, and r' and b' are coprime. The mechanism $\mathcal{M}_f^{HDG}$ proceeds as follows.

First, it identifies $R_f := \{r \in R \colon f \succ_r 1\}$ and $B_f := \{b \in B \colon f \succ_b 0\}$. Let $\alpha_f := \max\{\alpha \in \mathbb{N} \colon \alpha r' \leq |R_f|, \alpha b' \leq |B_f|\}$. Hence, we have sufficiently many agents in R_f and B_f to form a coalition of proportion f with $\alpha r'$ red and $\alpha b'$ blue agents, but no larger individually rational coalition of this proportion can be formed.

Let $\hat{R}_f$ be the $\alpha r'$ agents in R_f with smallest indices. Similarly, let $\hat{B}_f$ be the $\alpha b'$ agents in B_f with smallest indices. The mechanism $\mathcal{M}_f^{HDG}$ then forms the coalition $\hat{R}_f \cup \hat{B}_f$, and places all other agents into singleton coalitions.

Proposition 5.3 shows that $\mathcal{M}_f^{HDG}$ achieves a combination of desirable properties that is similar to that of $\mathcal{M}_x^{AHG}$.

Proposition 5.3. *Let $r' \in [r]$ and $b' \in [b]$ be coprime, and set $f = \frac{r'}{r'+b'}$. Then the mechanism $\mathcal{M}_f^{HDG}$ for HDGs is strategyproof, individually rational, and achieves a welfare of at least $n - \max\{|R_f| + b' - 1, |B_f| + r' - 1\}$.*

Proof. The proof of individual rationality and strategyproofness is analogous to the respective proofs in Proposition 5.1.

For the bound on the welfare, first observe that every agent in $N \setminus (R_f \cup B_f)$ achieves a utility of 1. Moreover, by construction of the coalition $\hat{R}_f \cup \hat{B}_f$, it holds that $|B_f \setminus \hat{B}_f| \leq b' - 1$ or $|R_f \setminus \hat{R}_f| \leq r' - 1$. Hence, at most $\max\{|R_f| + b' - 1, |B_f| + r' - 1\}$ agents in $R_f \cup B_f$ do not achieve a utility of at least 1. Note that this proof in particular includes the case where $R_f = \emptyset$ or $B_f = \emptyset$. For instance, if $R_f = \emptyset$, then $|R_f \setminus \hat{R}_f| = 0 \leq r' - 1$, and at most $|B_f| \leq |B_f| + r' - 1$ agents do not achieve a welfare of 1. □

Again, the bound on the welfare is tight. More precisely, there are instances where $\mathcal{M}_f^{HDG}$ only achieves a welfare of $n - \max\{r + b' - 1, b + r' - 1\}$. For instance, if $r + b' - 1 \geq b + r' - 1$, consider an instance where $R_f = R$ and B_f contains exactly $b' - 1$ agents.

It may appear that the bound on the welfare established in Proposition 5.3 is quite weak. Nevertheless, $\mathcal{M}_{1/2}^{HDG}$ achieves reasonably good welfare that is, in particular, a linear-factor approximation of the maximum welfare for balanced HDGs.

Corollary 5.4. *There exists a strategyproof and individually rational mechanism for HDGs that achieves a welfare of $n - \max\{r, b\}$.*

For WHGs, there does not seem to be a natural analogue of the mechanisms $\mathcal{M}_x^{AHG}$ and $\mathcal{M}_f^{HDG}$. First, this leads to similar computational boundaries as discussed before Theorem 4.4 and made rigorous in Appendix B: We cannot simply fix some agent $\hat{a}$ and efficiently determine a large individually rational coalition containing this agent. Moreover, even if we were able to perform this computational task, creating such a coalition raises strategyproofness concerns because the utility of $\hat{a}$ depends on the composition of their coalition. Instead, we obtain strategyproofness by starting with a fixed coalition structure (chosen independently of agents' preferences), checking which of the coalitions in that coalition structure are individually rational (based on the reported preferences), and then keeping these coalitions and splitting all other coalitions into singletons.

Mechanism $\mathcal{M}_{\mathfrak{C}^*}^{WHG}$. Let $\mathfrak{C}^*$ be any coalition structure. Recall that $\mathcal{I}(\mathfrak{C}^*)$ is the set of individually rational coalitions in $\mathfrak{C}^*$. Then, the mechanism $\mathcal{M}_{\mathfrak{C}^*}^{WHG}$ creates the coalition structure $\mathcal{I}(\mathfrak{C}^*) \cup \{\{i\} \colon C \in \mathfrak{C}^* \setminus \mathcal{I}(\mathfrak{C}^*), i \in C\}$.

Proposition 5.5. *Let $\mathfrak{C}^*$ be a partition. The mechanism $\mathcal{M}^{WHG}_{\mathfrak{C}^*}$ is strategyproof, individually rational, and achieves a welfare of at least $|\{C \in \mathfrak{C}^* : |C| > 1\}|$.*

Proof. Individual rationality of $\mathcal{M}^{WHG}_{\mathfrak{C}^*}$ follows from its definition. For strategyproofness, we observe that any report of an agent can only lead to two outcomes: being placed in their coalition in $\mathfrak{C}^*$ or being placed in a singleton coalition. If their coalition in $\mathfrak{C}^*$ is not individually rational for another agent, the agent's report has no impact and reporting strategically cannot improve their outcome. Otherwise, the agent accomplishes the more preferred outcome among the two possible ones by reporting truthfully. Hence, the mechanism is strategyproof.

For the welfare bound, consider any coalition $C \in \mathfrak{C}^*$ with $|C| > 1$. If $C \in \mathcal{I}(\mathfrak{C}^*)$, then all agents in C have a utility of at least 1. Otherwise, there exists at least one agent in C that receives a utility of at least 1 in a singleton coalition. □

Once again, we obtain a linear-factor welfare approximation for a suitable parameter of this mechanism. Since the welfare of the produced partitions depends on the number of proposed nonsingleton coalitions, we can simply propose a large matching. Given a hedonic game, we define the matching partition

$$\mathfrak{M} := \begin{cases} \left\{\{a_{2i-1}, a_{2i}\} : 1 \leq i \leq \frac{n}{2}\right\} & \text{if } n \text{ is even,} \\ \left\{\{a_{2i-1}, a_{2i}\} : 1 \leq i \leq \frac{n-1}{2}\right\} \cup \{\{a_n\}\} & \text{if } n \text{ is odd.} \end{cases}$$

By considering $\mathcal{M}^{WHG}_{\mathfrak{M}}$, we obtain the following corollary.

Corollary 5.6. *There exists a strategyproof and individually rational mechanism for WHGs that achieves a linear-factor approximation of the maximum welfare.*

5.2 Constant-Factor Welfare Approximation

In Section 5.1, we have seen simple mechanisms that combine strategyproofness and individual rationality while achieving a linear-factor approximation of the maximum welfare. However, in Section 4, we have shown that individual rationality is compatible with a constant-factor welfare approximation. Therefore, a natural follow-up question is whether there exist strategyproof and individually rational mechanisms with a constant-factor approximation of the maximum welfare. While we leave the ultimate answer to this question open, we show two weaker possibility results: one for the case of single-peaked domains and one for randomized mechanisms.

5.2.1 Single-Peaked Preferences. A frequently considered preference restriction is single-peakedness. Intuitively, in a single-peaked AHG, an agent has a most preferred coalition size, i.e., their peak. Moreover, when comparing two coalition sizes on the same side of the peak, they prefer the one closer to the peak. For HDGs and WHGs, the definition is similar, but formulated in terms of coalition ratios and agent indices, respectively. It is well known from the social choice literature that single-peakedness can lead to strategyproofness [28]. Moreover, this preference restriction seems natural and has been studied for both AHGs and HDGs [10–12]. For WHGs it could be interpreted as saying that agents are ordered by some qualitative intrinsic feature, e.g., how competitive they are.

Formally, an AHG (resp., HDG) is said to be *single-peaked* if for every agent $i \in N$ there exists a coalition size $p_i \in [n]$ (resp., a coalition ratio $p_i \in F$) such that for all $x, y \in [n]$ (resp., $x, y \in F$) with $x < y \leq p_i$ or $p_i \geq y > x$, it holds that $y \succ_i^S x$ (or $y \succ_i^F x$). Moreover, a WHG is said to be *single-peaked* if for every agent $i \in N$ there exists an agent index $p_i \in [n]$ such that for all $x, y \in [n]$ with $x < y \leq p_i$ or $p_i \geq y > x$, it holds that $a_y \succ_i^A a_x$.

We will now investigate the performance of the mechanisms in Section 5.1 on single-peaked instances. For AHGs, we observe that all agents achieve a high utility for the coalition size $\lfloor \frac{n}{2} \rfloor$. Similarly, all agents in balanced HDGs obtain a high utility for the fraction $\frac{1}{2}$. Hence, we obtain a constant-factor approximation by considering $\mathcal{M}^{AHG}_{\lfloor \frac{n}{2} \rfloor}$ and $\mathcal{M}^{HDG}_{\frac{1}{2}}$. For WHGs, we show that $\mathcal{M}^{WHG}_{\mathfrak{M}}$ has the desired welfare guarantee when applied to single-peaked instances.

Theorem 5.7. *There is a strategyproof and individually rational mechanism for single-peaked AHGs (or single-peaked, balanced HDGs, or single-peaked WHGs) that achieves a constant-factor approximation of the maximum welfare.*

Proof. We start with the consideration of AHGs. Let $\mathcal{M} = \mathcal{M}^{AHG}_{\lfloor \frac{n}{2} \rfloor}$. By Proposition 5.1, $\mathcal{M}$ is strategyproof and individually rational. Hence, we only have to prove our claim about its welfare approximation guarantee.

The key insight is that, for single-peaked preferences, every agent achieves a utility of at least $\lfloor \frac{n}{2} \rfloor - 1$ for the coalition size $\lfloor \frac{n}{2} \rfloor$. This is true because there are $\lfloor \frac{n}{2} \rfloor - 1$ smaller and $\lceil \frac{n}{2} \rceil \geq \lfloor \frac{n}{2} \rfloor$ larger coalition sizes. Moreover, for single-peaked preferences, $\lfloor \frac{n}{2} \rfloor$ is more preferred than all smaller or all larger sizes.

If an agent prefers the coalition size 1 over $\lfloor \frac{n}{2} \rfloor$, then they achieve a utility of at least $\lfloor \frac{n}{2} \rfloor$ in the coalition structure produced by $\mathcal{M}$. Moreover, at most $\lfloor \frac{n}{2} \rfloor - 1$ of the agents preferring $\lfloor \frac{n}{2} \rfloor$ over 1 are assigned to a singleton coalition. All other agents with these preferences are assigned to a coalition of size $\lfloor \frac{n}{2} \rfloor$ and therefore achieve a utility of at least $\lfloor \frac{n}{2} \rfloor - 1$. Hence, together, the partition produced by $\mathcal{M}$ has a welfare of at least

$$\left(n - \left(\left\lfloor \frac{n}{2} \right\rfloor - 1\right)\right)\left(\left\lfloor \frac{n}{2} \right\rfloor - 1\right) \geq \frac{n^2}{4} - \frac{9}{4}.$$

Since the maximum welfare is $n(n-1)$, this yields a constant-factor approximation to the maximum welfare.

Next, we consider HDGs, for which we analyze $\mathcal{M}^{HDG}_{\frac{1}{2}}$. By Proposition 5.3, we only have to prove the welfare guarantee of $\mathcal{M}^{HDG}_{\frac{1}{2}}$. For balanced HDGs, exactly half of the fractions in $F \setminus \{0, \frac{1}{2}, 1\}$ are smaller and larger than $\frac{1}{2}$. Hence, for single-peaked preferences, every agent achieves at least a utility of $\frac{|F|-3}{2}$ when assigned to a coalition of proportion $\frac{1}{2}$ or when preferring a singleton coalition over the fraction $\frac{1}{2}$.

Since this happens to all agents of at least one type, the partition produced by $\mathcal{M}^{HDG}_{\frac{1}{2}}$ achieves a welfare of at least $\frac{n}{2}\frac{|F|-3}{2}$. Since the maximum welfare of any partition is bounded by $n(|F|-1)$, it follows that $\mathcal{M}^{HDG}_{\frac{1}{2}}$ is a constant-factor approximation of the maximum welfare.

Finally, we consider WHGs. We show that $\mathcal{M}^{WHG}_{\mathfrak{M}}$ fulfills the desired properties. By Proposition 5.5, we only have to prove the welfare guarantee. Let $i \in \left[\left\lfloor \frac{n}{4} \right\rfloor\right]$. Then, there are at least $2i-2$ agents with lower index than a_{2i-1}'s potential partner a_{2i} in $\mathfrak{M}$. Moreover, there are at least $\frac{n}{2}$ agents with higher index than a_{2i}. By single-peakedness, a_{2i-1} and a_{2i} each achieve a utility of at least $i-2$ in $\mathfrak{M}$. If, however, $\{a_{2i-1}, a_{2i}\} \notin \mathcal{I}(\mathfrak{M})$, then at least one of a_{2i-1} and a_{2i} achieves a utility of at least $2i-1$ in the partition produced by $\mathcal{M}^{WHG}_{\mathfrak{M}}$. Hence, the welfare of this partition is at least

$$\sum_{i=1}^{\lfloor \frac{n}{4} \rfloor} 2i-1 \geq \sum_{i=1}^{\frac{n}{4}-1} 2i-1 = \frac{1}{16}(n-4)^2.$$

Since the maximum welfare of any partition is bounded by $n(n-1)$, it follows that $\mathcal{M}^{HDG}_{\mathfrak{M}}$ is a constant-factor approximation of the maximum welfare. □

We can prove an analogue of Theorem 5.7 for AHGs with *single-crossing preferences*—another prominent domain restriction in social choice [33]. Even though this preference restriction is less well-studied than single-peakedness, the study of mechanism design for single-crossing instances offers additional insights into the power and limitations of strategyproof mechanisms. We defer the formal treatment of this setting to Appendix C.

5.2.2 Randomized Mechanisms. Instead of restricting the domain on which our mechanisms have to exhibit desirable properties, we can increase the capabilities of the mechanisms we consider. One popular way of doing so is to investigate mechanisms that use randomization. It turns out that this approach, too, enables us to simultaneously achieve strategyproofness, individual rationality, and a constant-factor approximation of the maximum welfare.

The idea is to run our previously proposed mechanisms $\mathcal{M}^{AHG}_{x}$, $\mathcal{M}^{HDG}_{f}$, and $\mathcal{M}^{WHG}_{\mathfrak{C}^*}$ with parameters x, f, and $\mathfrak{C}^*$ drawn from a carefully designed distribution. We can then apply techniques similar to the ones in the proof of Theorem 4.2 to derive the welfare bound. For AHGs and HDGs, we achieve high expected welfare when running the respective mechanisms for a sufficiently large proportion of the possible parameters. For WHGs, we use a random matching as our starting point.

Theorem 5.8. *There exists a randomized mechanism for AHGs (or balanced HDGs or WHGs) that is universally strategyproof, universally individually rational, and achieves a constant-factor approximation of the maximum welfare.*

Proof. We first consider AHGs. Set $s := \left\lfloor \frac{n}{4} \right\rfloor$. Our randomized mechanism first selects $x \in [s]$ uniformly at random and then executes $\mathcal{M}^{AHG}_{x}$ for the selected x, where $\mathcal{M}^{AHG}_{1}$ is defined as the mechanism that simply selects the singleton partition. By Proposition 5.1, the mechanism is strategyproof and individually rational for every outcome of the randomization and therefore universally strategyproof and universally individually rational.

It remains to prove the bound on the welfare. By Lemma 4.1, for every subset H of $\left\lceil \frac{n}{8} \right\rceil$ alternatives, there exists $\gamma \in (0,1)$ such that at least one alternative in H is ranked in position at most $\lceil \gamma n \rceil$ by at least $\frac{n}{2}$ agents. Let $T \subseteq [s]$ be the subset of $[s]$ of alternatives that are ranked in position at most $\lceil \gamma n \rceil$ by at least $\frac{n}{2}$ agents. By the choice of γ, we have that $|[s] \setminus T| \leq \left\lceil \frac{n}{8} \right\rceil - 1 \leq \left\lfloor \frac{n}{8} \right\rfloor$. Hence, $|T| \geq \left\lfloor \frac{n}{4} \right\rfloor - \left\lfloor \frac{n}{8} \right\rfloor$.

Now, consider a fixed alternative $x \in T$. Then, from the (at least) $\frac{n}{2}$ agents ranking x in position at most $\lceil \gamma n \rceil$, at least $\frac{n}{2} - x + 1 \geq \frac{n}{2} - s + 1 > \frac{n}{4}$ receive a utility of at least $n - \lceil \gamma n \rceil$. Hence, the expected welfare of the partition produced by our randomized mechanism is at least

$$\frac{|T|}{s}\frac{n}{4}(n - \lceil \gamma n \rceil) = \frac{\left\lfloor \frac{n}{4} \right\rfloor - \left\lfloor \frac{n}{8} \right\rfloor}{\left\lfloor \frac{n}{4} \right\rfloor}\frac{n}{4}(n - \lceil \gamma n \rceil) = \Theta(n^2).$$

Since the maximum welfare that can be achieved by any partition is $n(n-1)$, this proves the assertion.

We now consider balanced HDGs. Consider

$$H := \left\{ \frac{p}{p+q} : 1 \leq p, q \leq \frac{n}{8} \right\}.$$

Our randomized mechanism first selects $f \in H$ uniformly at random and then executes $\mathcal{M}^{HDG}_{f}$. By Proposition 5.3, the mechanism is strategyproof and individually rational for every outcome of the randomization and therefore universally strategyproof and universally individually rational.

It remains to consider the welfare bound. Since the proportion of coprime numbers converges to $\frac{6}{\pi^2}$ [29, see also Footnote 3], we know that there exists $n_0 \in \mathbb{N}$ and $\alpha \in (0,1)$ such that $\frac{1}{2}|H| \geq \lceil \alpha n^2 \rceil$ for all $n \geq n_0$. Hence, by Lemma 4.1, there exists $\gamma \in (0,1)$, such that, whenever there are $n \geq n_0$ agents, for every subset $U \subseteq H$ with $|U| \geq \frac{1}{2}|H|$ some fraction $f \in U$ is ranked in position at most $\lceil \gamma |F| \rceil$ by at least $\frac{7n}{8}$ agents.

Now, consider a given HDG $(N, \succ^F)$ and let $T \subseteq H$ be the subset of fractions that are ranked in position at most $\lceil \gamma |F| \rceil$ by at least $\frac{7n}{8}$ agents. By our choice of γ, it holds that $|T| \geq \frac{1}{2}|H|$.

We now show that $\mathcal{M}^{HDG}_{f}(N, \succ^F)$ achieves a high welfare for every fraction $f \in T$. This implies the desired bound because we run $\mathcal{M}^{HDG}_{f}$ for some $f \in T$ with probability at least $\frac{1}{2}$.

Let $f \in T$ and let $N_f \subseteq N$ be the set of agents that rank f in position at most $\lceil \gamma |F| \rceil$. Recall that $|N_f| \geq \frac{7n}{8}$. Hence, because the game is balanced, N_f contains at least $\frac{3n}{8}$ agents of each type.

Assume first that at least half of the blue agents and half of the red agents in N_f prefer a proportion of f to being in a singleton coalition. Hence, there are at least $\frac{3n}{16}$ such agents of each type. Recall that $\mathcal{M}^{HDG}_{f}$ forms an individually rational coalition C of ratio f that is as large as possible. By design of the set H, we can enlarge C whenever we have $\frac{n}{8}$ agents of each type that are in N_f but not added to C. Hence, there exists a type of which at least $\frac{n}{16}$ agents are in C and in N_f. Since each of these agents achieves a utility of $\lfloor (1-\gamma)|F| \rfloor$, it holds that

$$\mathcal{SW}(\mathcal{M}^{HDG}_{f}(N, \succ^F)) \geq \frac{n}{16}\lfloor (1-\gamma)|F| \rfloor. \tag{2}$$

Otherwise, it is the case that at least half of the blue or half of the red agents in M, i.e., a total number of at least $\frac{3n}{16}$ agents rank a singleton coalition above a coalition of ratio f. By our choice of f, each of these agents achieves a utility of at least $\lfloor (1-\gamma)|F| \rfloor$ in an individually rational partition. Hence, since $\mathcal{M}^{HDG}_{f}(N, \succ^F)$ is

individually rational, we conclude that

$$\mathcal{SW}(\mathcal{M}_f^{HDG}(N, \succ^F)) \geq \frac{3n}{16}\lfloor(1-\gamma)|F|\rfloor. \tag{3}$$

Combining Equations (2) and (3), we obtain

$$\mathbb{E}[\mathcal{SW}] \geq \frac{|T|}{|H|}\frac{n}{16}\lfloor(1-\gamma)|F|\rfloor \geq \frac{n}{32}\lfloor(1-\gamma)|F|\rfloor.$$

Since the maximum welfare is bounded by $n(|F|-1)$, we conclude that our randomized mechanism achieves a constant-factor approximation of the maximum welfare.

Finally, we consider WHGs. We define a randomized algorithm $\mathcal{M}_{rand}^{WHG}$ for $\mathcal{W}$-hedonic games. First, $\mathcal{M}_{rand}^{WHG}$ selects a permutation $\sigma\colon [n] \to [n]$ uniformly at random. Define

$$\mathfrak{M}_\sigma := \begin{cases} \left\{\{a_{\sigma(2i-1)}, a_{\sigma(2i)}\} : 1 \leq i \leq \frac{n}{2}\right\} & \text{if } n \text{ even,} \\ \left\{\{a_{\sigma(2i-1)}, a_{\sigma(2i)}\} : 1 \leq i \leq \frac{n-1}{2}\right\} \cup \{\{a_{\sigma(n)}\}\} & \text{if } n \text{ odd.} \end{cases}$$

In other words, $\mathfrak{M}_\sigma$ essentially is a uniformly selected matching among all maximal matchings. Then, $\mathcal{M}_{rand}^{WHG}$ runs $\mathcal{M}_{\mathfrak{M}_\sigma}^{WHG}$. Note that σ can be sampled efficiently by uniformly selecting agents one by one from the nonselected agents.

By Proposition 5.5, the mechanism is strategyproof and individually rational for every outcome of the randomization and therefore universally strategyproof and universally individually rational. It remains to prove the desired guarantee for the welfare.

Consider a WHG $(N, \succ^A)$ and let $\mathfrak{C} := \mathcal{M}_{rand}^{WHG}(N, \succ^A)$ be the random partition produced by our mechanism.

In the random matching (before transitioning to individually rational coalitions), each agent is matched with any other agent with equal probability, and, if the number of agents is odd, this also equals the probability of being the unique agent in a singleton coalition. Hence, if n is odd, then $\mathbb{E}_\sigma[\mathcal{SW}(\mathfrak{M}_\sigma)] = n\frac{1}{n}\sum_{i=1}^{n} i - 1 = \frac{n(n-1)}{2}$. Moreover, if n is even, then, with equal probability an agent has a utility equal to every number in $\{0, 1, \dots, n-1\}$ except for the number corresponding to a singleton coalition. Bounding with the case where a singleton coalition gives a utility of $n-1$, we estimate $\mathbb{E}_\sigma[\mathcal{SW}(\mathfrak{M}_\sigma)] \geq n\frac{1}{n-1}\sum_{i=1}^{n-1} i - 1 = \frac{n(n-2)}{2}$.

Together, we have shown that

$$\mathbb{E}_\sigma[\mathcal{SW}(\mathfrak{M}_\sigma)] \geq \frac{n(n-2)}{2}. \tag{4}$$

We want to compute the welfare achieved by our mechanism by subtracting the loss in welfare due to not individually rational coalitions.

Assume first that $\frac{n}{10}$ agents receive a utility of at least $\frac{n}{10}$ when assigned to a singleton coalition. Then, since the mechanism is universally individually rational, it immediately follows that $\mathbb{E}_\sigma[\mathcal{SW}(\mathfrak{C})] \geq \frac{n^2}{100}$.

Hence, we may assume that at least $\frac{9n}{10}$ agents receive a utility of less than $\frac{n}{10}$ from being in a singleton coalition. Our strategy is to consider the welfare achieved by $\mathfrak{M}_\sigma$ as computed above and to subtract the loss caused by dissolving some of the coalitions because of them not being individually rational for some of their members.

There are a total of at most $\frac{9n}{10}\frac{n}{10} + \frac{n}{10}n \leq \frac{1}{5}n^2$ coalitions that can be dissolved. The first part of this sum accounts for the $\frac{9n}{10}$ agents for which at most $\frac{n}{10}$ other agents are less preferred than being in a singleton coalition. For the remaining $\frac{n}{10}$ agents, it is possible that all coalitions containing them get dissolved. Each dissolved coalition can cause a loss in welfare of at most $2(n-1)$. Moreover, if n is even, then each coalition (of size 2) occurs in $\mathfrak{M}_\sigma$ with probability $\frac{1}{n-1}$ and if n is odd, then each pair coalition occurs in $\mathfrak{M}_\sigma$ with probability $\frac{1}{n}$.

We conclude that

$$\begin{aligned}\mathbb{E}_\sigma[\mathcal{SW}(\mathfrak{C})] &\geq \mathbb{E}_\sigma[\mathcal{SW}(\mathfrak{M}_\sigma)] - \frac{1}{n-1}\frac{n^2}{5}2(n-1) \\ &\geq \frac{n(n-2)}{2} - \frac{2}{5}n^2 = \Theta(n^2).\end{aligned}$$

The second inequality follows from Equation (4). Since the welfare of any outcome is bounded by n^2, we conclude that $\mathcal{M}_{rand}^{WHG}$ achieves the desired welfare approximation. □

6 CONCLUSION

We have studied mechanisms for anonymous hedonic games, hedonic diversity games, and $\mathcal{W}$-hedonic games. Our goal was to achieve strategyproofness, individual rationality, and good welfare guarantees. In a general voting setting, the famous theorem by Gibbard [27] and Satterthwaite [34] states that strategyproof voting rules are dictatorial or are duple, i.e., limit the choice to two alternatives only. Our mechanisms are similar in spirit to duple rules: for each voter, only the preference over two possible (sets of) outcomes matters. One of these outcomes always corresponds to being in a singleton coalition.

In contrast, dictatorships do not seem to lead to strategyproof and individually rational rules in our setting. This is because following the dictator's wishes may result in outcomes that are not individually rational for other agents. If the top choice of the dictator is only implemented if it leads to an individually rational outcome, the dictator has an incentive to manipulate their top choice. Finally, implementing the best choice of the dictator leading to an individually rational outcome causes incentives for other agents to manipulate the set of individually rational outcomes.

Our duple mechanisms are strategyproof and individually rational while providing a linear-factor approximation of the maximum welfare. If we restrict attention to single-peaked domains or if we allow randomized mechanisms, we even obtain a constant-factor approximation of the maximum welfare. It is an intriguing open question whether deterministic strategyproof mechanisms can achieve a constant-factor approximation on the full preference domains. Another direction for future research is to study the compatibility of strategyproofness with objectives beyond welfare, such as fairness.

ACKNOWLEDGMENTS

Martin Bullinger was supported by the AI Programme of The Alan Turing Institute and Edith Elkind was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/X038548/1.

REFERENCES

[1] Anders Aamand, Justin Y. Chen, Allen Liu, Sandeep Silwal, Pattara Sukprasert, Ali Vakilian, and Fred Zhang. 2023. Constant Approximation for Individual Preference Stable Clustering. In *Proceedings of the 37th Conference on Neural Information Processing Systems (NeurIPS)*. Forthcoming.

[2] Saba Ahmadi, Pranjal Awasthi, Samir Khuller, Matthäus Kleindessner, Jamie Morgenstern, Pattara Sukprasert, and Ali Vakilian. 2022. Individual preference stability for clustering. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*. 197–246.

[3] José Alcalde and Pablo Revilla. 2004. Researching with whom? Stability and manipulation. *Journal of Mathematical Economics* 40, 8 (2004), 869–887.

[4] Haris Aziz, Florian Brandl, Felix Brandt, Paul Harrenstein, Martin Olsen, and Dominik Peters. 2019. Fractional Hedonic Games. *ACM Transactions on Economics and Computation* 7, 2 (2019), 1–29.

[5] Haris Aziz, Felix Brandt, and Paul Harrenstein. 2013. Pareto Optimality in Coalition Formation. *Games and Economic Behavior* 82 (2013), 562–581.

[6] Haris Aziz and Rahul Savani. 2016. Hedonic Games. In *Handbook of Computational Social Choice*, Felix Brandt, Vincent Conitzer, U. Endriss, J. Lang, and Ariel D. Procaccia (Eds.). Cambridge University Press, Chapter 15.

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

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

[9] Niclas Boehmer and Edith Elkind. 2020. Individual-Based Stability in Hedonic Diversity Games. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*. 1822–1829.

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

[11] Felix Brandt, Martin Bullinger, and Anaëlle Wilczynski. 2023. Reaching Individually Stable Coalition Structures. *ACM Transactions on Economics and Computation* 11, 1–2 (2023), 4:1–65.

[12] Robert Bredereck, Edith Elkind, and Ayumi Igarashi. 2019. Hedonic Diversity Games. In *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. 565–573.

[13] Martin Bullinger, Edith Elkind, and Jörg Rothe. 2024. Cooperative Game Theory. In *Economics and Computation: An Introduction to Algorithmic Game Theory, Computational Social Choice, and Fair Division*, Jörg Rothe (Ed.). Springer, Chapter 3, 139–229.

[14] Katarína Cechlárová and Jana Hajduková. 2004. Stable partitions with $\mathcal{W}$-preferences. *Discrete Applied Mathematics* 138, 3 (2004), 333–347.

[15] Katarína Cechlárová and Antonio Romero-Medina. 2001. Stability in Coalition Formation games. *International Journal of Game Theory* 29 (2001), 487–494.

[16] Ying-Chih Chuang, Kun-Yang Chuang, and Tzu-Hsuan Yang. 2013. Social cohesion matters in health. *International Journal for Equity in Health* 12, 87 (2013).

[17] Jane M. Cramm and Anna P. Nieboer. 2015. Social cohesion and belonging predict the well-being of community-dwelling older people. *BMC Geriatrics* 15, 30 (2015).

[18] Andreas Darmann. 2023. Stability and Welfare in (Dichotomous) Hedonic Diversity Games. *Theory of Computing Systems* 67, 6 (2023), 1133–1155.

[19] A. Darmann and C. Klamler. 2016. Proportional Borda Allocations. *Social Choice and Welfare* 47 (2016), 543–558.

[20] Dinko Dimitrov, Peter Borm, Ruud Hendrickx, and Shao C. Sung. 2006. Simple Priorities and Core Stability in Hedonic Games. *Social Choice and Welfare* 26, 2 (2006), 421–433.

[21] Georgi Dragolov, Zsófia S. Ignácz, Jan Lorenz, Jan Delhey, Klaus Boehnke, and Kai Unzicker. 2016. *Social cohesion in the western world: What holds societies together: Insights from the social cohesion radar.* Springer.

[22] Jacques H. Drèze and Joseph Greenberg. 1980. Hedonic Coalitions: Optimality and Stability. *Econometrica* 48, 4 (1980), 987–1003.

[23] Edith Elkind and Michael Wooldridge. 2009. Hedonic coalition nets. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. 417–424.

[24] Daniel Farkas and Shmuel Nitzan. 1979. The Borda rule and Pareto stability: a comment. *Econometrica* 47, 5 (1979).

[25] Michele Flammini, Bojana Kodric, Gianpiero Monaco, and Qiang Zhang. 2021. Strategyproof Mechanisms for Additively Separable and Fractional Hedonic Games. *Journal of Artificial Intelligence Research* 70 (2021), 1253–1279.

[26] Michele Flammini, Bojana Kodric, and Giovanna Varricchio. 2022. Strategyproof mechanisms for friends and enemies games. *Artificial Intelligence* 302 (2022), 103610.

[27] Allan Gibbard. 1973. Manipulation of Voting Schemes: A General Result. *Econometrica* 41, 4 (1973), 587–601.

[28] Hervé Moulin. 1980. On strategy-proofness and single peakedness. *Public Choice* 35, 4 (1980), 437–455.

[29] James E. Nymann. 1972. On the probability that k positive integers are relatively prime. *Journal of number theory* 4, 5 (1972), 469–473.

[30] Carmelo Rodríguez-Álvarez. 2009. Strategy-Proof Coalition Formation. *International Journal of Game Theory* 38 (2009), 431–452.

[31] Jörg Rothe, Hilmar Schadrack, and Lena Schend. 2018. Borda-induced hedonic games with friends, enemies, and neutral players. *Mathematical Social Sciences* 96 (2018), 21–36.

[32] D. G. Saari. 1990. The Borda dictionary. *Social Choice and Welfare* 7, 4 (1990), 279–317.

[33] Alejandro Saporiti and Fernando Tohmé. 2006. Single-crossing, strategic voting and the median choice rule. *Social Choice and Welfare* 26 (2006), 363–383.

[34] Mark A. Satterthwaite. 1975. Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions. *Journal of Economic Theory* 10, 2 (1975), 187–217.

[35] Willemijn Schreuder and Lummina G. Horlings. 2022. Transforming places together: transformative community strategies responding to climate change and sustainability challenges. *Climate Action* 1, 24 (2022).

[36] E. Segal-Halevi, A. Hassidim, and H. Aziz. 2020. Fair Allocation with Diminishing Differences. *Journal of Artificial Intelligence Research* 67 (2020), 471–507.

[37] Lloyd Shapley and Herbert Scarf. 1974. On cores and indivisibility. *Journal of mathematical economics* 1, 1 (1974), 23–37.

[38] Anand Siththaranjan, Cassidy Laidlaw, and Dylan Hadfield-Menell. 2024. Distributional preference learning: Understanding and accounting for hidden context in RLHF. In *Proceedings of the 12th International Conference on Learning Representations (ICLR)*.

[39] John von Neumann and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton University Press.

[40] Mason Wright and Yevgeniy Vorobeychik. 2015. Mechanism design for team formation. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI)*.

[41] H. Peyton Young. 1974. An axiomatization of Borda's rule. *Journal of Economic Theory* 9, 1 (1974), 43–52.

[42] David Zuckerman. 2006. Linear degree extractors and the inapproximability of max clique and chromatic number. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*. 681–690.

APPENDIX

In the appendix, we provide proofs missing in the main paper as well as additional results.

A MISSING PROOFS IN SECTION 4

We start with the counting lemma that is useful for deriving welfare possibilities.

Lemma 4.1. *Let M be a set of m alternatives. Let $\alpha, \beta \in (0,1)$ and $H \subseteq M$ be a subset of alternatives with $|H| \geq \lceil \alpha m \rceil$. Set $\gamma = (1-\alpha) + \alpha\beta$. Then, in every game where n agents provide rankings over M, some alternative in H is ranked in position at most $\lceil \gamma m \rceil$ by at least βn agents.*

Proof. Assume that M, H, α, and β are given like in the assumptions of the lemma.

Assume for contradiction that every alternative in H is ranked in position at most $\lceil \gamma m \rceil$ by less than βn agents. We will derive a contradiction by counting how the first $\lceil \gamma m \rceil$ spots in the rankings of all n agents are filled by showing that a total of less than $\lceil \gamma m \rceil n$ of these spots are filled.

First, by assumption, the alternatives in H fill less than $|H|\beta n = \lceil \alpha m \rceil \beta n$ of these spots. Second, consider alternatives in $M \setminus H$. The total number of spots in any position filled by alternatives in $M \setminus H$ is at most $|M \setminus H| n = \lfloor (1-\alpha)m \rfloor n$, which also bounds the number of spots in position at most $\lceil \gamma m \rceil$. Together, the total number of spots filled in the ranking of any agent in position at most $\lceil \gamma m \rceil$ is less than

$$\lfloor (1-\alpha)m \rfloor n + \lceil \alpha m \rceil \beta n.$$

Next, we claim that

$$(1-\alpha)m - \lfloor (1-\alpha)m \rfloor = \lceil \alpha m \rceil - \alpha m.$$

Clearly, this is true if αm is an integer. Otherwise, there exist $k \in \mathbb{N}$ and $q \in (0,1)$ such that $(1-\alpha)m = k + q$ and $\alpha m = (m-k-1) + (1-q)$. Note that $(1-q) \in (0,1)$ as well. Hence, $(1-\alpha)m - \lfloor (1-\alpha)m \rfloor = q = \lceil \alpha m \rceil - \alpha m$.

Consequently, as $\beta \in (0,1)$, it follows that $\left((1-\alpha)m - \lfloor (1-\alpha)m \rfloor\right) n \geq \left(\lceil \alpha m \rceil - \alpha m\right) \beta n$.

We conclude that

$$\begin{aligned} &\lfloor (1-\alpha)m \rfloor n + \lceil \alpha m \rceil \beta n \\ &\leq (1-\alpha)mn + \alpha m \beta n \\ &= (1 - \alpha + \alpha\beta) mn \\ &= \gamma m n \leq \lceil \gamma m \rceil n. \end{aligned}$$

This is a contradiction because all $\lceil \gamma m \rceil n$ positions have to be filled by some alternative. □

We apply the lemma to prove welfare possibilities of individually rational outcomes in AHGs and balanced HDGs.

Theorem 4.2. *For AHGs (or balanced HDGs), there exists an individually rational partition that achieves a constant-factor approximation of the maximum welfare.*

Proof for HDGs. We complete the proof by considering balanced HDGs. Recall that red and blue agents provide rankings of the set $F \setminus \{0\}$ and $F \setminus \{1\}$ of their possible fractions, respectively. As the game is balanced, there are $\frac{n}{2}$ agents of each color. Hence,

$$|F| \leq (|R|+1)(|B|+1) = \left(\frac{n}{2}+1\right)^2 \leq n^2.$$

Consider

$$H := \left\{ \frac{p}{p+q} : 1 \leq p, q \leq \frac{n}{8} \right\}.$$

Since the proportion of coprime numbers converges to $\frac{6}{\pi^2}$ [29, see also Footnote 3], we know that there exists $n_0 \in \mathbb{N}$ and $\alpha \in (0,1)$ such that $|H| \geq \lceil \alpha n^2 \rceil$ for all $n \geq n_0$. Hence, by Lemma 4.1 for this α and $\beta = \frac{7}{8}$, there exists $\gamma \in (0,1)$, such that, whenever there are $n \geq n_0$ agents, some fraction in $f \in H$ is ranked in position at most $\lceil \gamma |F| \rceil$ by at least $\frac{7n}{8}$ agents. Let N' be the set of these agents. Note that, because the game is balanced, N' contains at least $\frac{3n}{8}$ agents of each type. We make a case distinction similar to AHGs.

Assume first that at least half of the blue agents and half of the red agents in N' prefer a proportion of f to being in a singleton coalition. Hence, there are at least $\frac{3n}{16}$ such agents of each type. Now, assume that we form an individually rational coalition C of ratio f with agents in N' that is as large as possible. Then, by design of the set H, there exists a type of which there are at most $\frac{n}{8}$ agents in N' that are not in C but find f individually rational. Hence, since $\frac{3n}{16}$ agents of this type are in N', it holds that at least $\frac{n}{16}$ agents of this type are in C. Each of these agents achieves a utility of at least $\lceil (1-\gamma)|F| \rceil - 1$. Hence, the coalition C can be extended to an individually rational partition (e.g., by forming singleton coalitions with the remaining agents) that achieves a constant fraction of the maximum welfare.

Otherwise, it is the case that at least half of the blue and red agents in N', i.e., a total number of at least $\frac{3n}{8}$ agents achieve a utility of $\lceil (1-\gamma)|F| \rceil - 1$ in a singleton coalition. Then, the singleton partition achieves a constant fraction of the welfare. □

Now, we provide the proof that individually rational outcomes with a constant-factor welfare approximation exist in WHGs.

Theorem 4.4. *For WHGs, there exists an individually rational partition that achieves a constant-factor approximation of the maximum welfare.*

Proof. Consider a WHG. We assume that $n \geq 2$ as otherwise the statement is trivial.

We find a partition with the following algorithm. For each $i \in [n]$, if a_i already is in a coalition, we move to the next agent. If not, then a_i checks two possible cases. First, if forming a singleton coalition achieves a utility of $\frac{n}{2}$, then a_i forms a singleton coalition. Second, consider the other agents that are not in a coalition, yet, and for which forming a coalition of size 2 with a_i is individually rational for both them and a_i. If a_i achieves a utility of $\frac{n}{2}$ in a coalition of size 2 with any such agent, then such a coalition is formed. If none of these two cases applies, a_i is not assigned to a coalition and we continue with the consideration of the next agent. We refer to the part of the algorithm until every agent has been considered as its *main stage*. Once every agent has been considered, we form singleton coalitions with all agents not assigned to coalitions during the main stage.

By construction, this algorithm is individually rational. We claim that it achieves a welfare of at least $\frac{n^2}{16}$ for $n \geq 3$.

For proving this, we consider two cases. First, if, during the main stage of the algorithm, at least $\frac{n}{8}$ agents form a coalition in which they achieve a utility of $\frac{n}{2}$, the claim is true.

Hence, assume now that less than $\frac{n}{8}$ agents achieve this. Recall that, in the main stage, the algorithm only forms coalitions of size 1 or 2 which yield a utility of $\frac{1}{2}$ for one contained agent. Consequently, as less than $\frac{n}{8}$ agents achieve a utility of $\frac{n}{2}$, less than $\frac{n}{4}$ agents are assigned to coalitions in the main stage. Thus, there are at least $\frac{3n}{4}$ agents that are not assigned a coalition in the main stage. These agents achieve a utility of less than $\frac{n}{2}$ for forming a singleton coalition and could not find a coalition partner that yields a utility of $\frac{n}{2}$ and for which forming a coalition of size 2 with them is an individually rational coalition.

Let a be some such agent. Since a singleton coalition yields a utility of less than $\frac{n}{2}$, there are at least $\left\lfloor \frac{n}{2} \right\rfloor$ agents, which yield a utility of $\frac{n}{2}$ for a (which is then also individually rational for a). Since a total number of less than $\frac{n}{4}$ agents form coalitions in the main stage, there must exist at least $\left\lfloor \frac{n}{2} \right\rfloor - \frac{n}{4} \geq \frac{n}{4} - \frac{1}{2}$ agents that were considered by a to form a coalition of size 2. All of these coalitions have not been formed which means that all of these agents prefer being in a singleton coalition over forming a coalition with a.

We can bound the welfare by accumulating the utility gained through preferring the output to being with agents not assigned in the main stage. Hence, since the algorithm outputs an individually rational partition, it must have a welfare of at least $\frac{3n}{4}\left(\frac{n}{4} - \frac{1}{2}\right) = \frac{n^2}{16} + \left(\frac{n^2}{8} - \frac{3n}{8}\right) \geq \frac{n^2}{16}$ for $n \geq 3$. This is a $\frac{1}{16}$-approximation of the maximum welfare because the maximum welfare is bounded by $n(n-1)$. □

B COMPUTATION OF LARGE INDIVIDUALLY RATIONAL COALITIONS IN $\mathcal{W}$-HEDONIC GAMES

The individually rational partitions achieving a constant-factor approximation of maximum welfare in AHGs and HDGs (cf. Theorem 4.2) and the algorithms $\mathcal{M}_x^{AHG}$ and $\mathcal{M}_f^{HDG}$ defined in Section 5.1 rely on choosing a coalition size or coalition fraction and creating coalitions of this size or fraction with as many agents as possible. In a similar vein, one could try to fix some agent and create a large individually rational coalition containing this agent. However, this approach faces computational difficulties. In fact, the maximum size of an individually rational coalition is even hard to approximate within a factor of $n^{1-\epsilon}$.

Theorem B.1. *Let $\epsilon \in (0,1)$. The following problem is* NP*-complete: given a WHG $(N, \succ^A)$ and a positive integer $q \in \mathbb{N}$, decide whether $(N, \succ^A)$ admits an individually rational coalition of size at least $\frac{q}{n^{1-\epsilon}}$.*

Proof. Let $\epsilon > 0$. Clearly, our problem is contained in NP because a coalition of size $\frac{q}{n^{1-\epsilon}}$ can be verified to be individually rational in polynomial time by checking individual rationality for each contained agent.

For hardness, we reduce from the approximate MaxClique problem. The input is a graph G and a positive integer $t \in \mathbb{N}$. An instance (G, t) is a Yes-instance if there exists a clique of size at least $\frac{t}{n^{1-\epsilon}}$. This problem is known to be NP-complete [42].

We now describe the reduction. Assume we are given an instance (G, t) of MaxClique, where $G = (V, E)$ is an unweighted graph. We construct a WHG $G' = (N, \succ^A)$ as follows. The set of agents is $N = V$. Moreover, the ranking of agent $i \in N$ is chosen in any way such that j is ranked above i if $\{i, j\} \in E$ and j is ranked below i if $\{i, j\} \notin E$.

For the threshold $q = t$, we ask whether the reduced game admits an individually rational coalition of size at least $\frac{q}{n^{1-\epsilon}}$.

Consider any coalition C of N. If C is individually rational, then for every pair of agents $i, j \in C$ it holds that $\{i, j\} \in E$, i.e., C forms a clique in G. Conversely, every clique in G induces an individually rational coalition in G'. Hence, there exists a clique of size at least $\frac{t}{n^{1-\epsilon}}$ in G if and only if there exists an individually rational coalition of size at least $\frac{q}{n^{1-\epsilon}}$ in G'. □

One can enhance this construction with an additional special agent that is ranked at the top by all other agents and for which every coalition is individually rational (i.e., they are ranked bottom for themselves). Then, for the threshold $q = r + 1$, there exists a clique of size at least $\frac{t}{n^{1-\epsilon}}$ in G if and only if there exists an individually rational coalition of size at least $\frac{q}{n^{1-\epsilon}}$ in G'. Hence, we also obtain the following variant of the previous theorem.

Theorem B.2. *Let $\epsilon \in (0,1)$. The following problem is* NP*-complete: given a WHG $(N, \succ^A)$, a special agent $i^* \in N$ and a positive integer $q \in \mathbb{N}$, decide whether $(N, \succ^A)$ admits an individually rational coalition containing i^* of size at least $\frac{q}{n^{1-\epsilon}}$.*

C CONSTANT-FACTOR WELFARE APPROXIMATION FOR SINGLE-CROSSING AHGS

In Section 5.2, we have seen that there are strategyproof and individually rational mechanisms achieving a constant-factor approximation of the maximum welfare if we consider single-peaked games. We now show that the same can be achieved for AHGs in single-crossing domains. Afterwards, we show that the same approach does not work for single-crossing, balanced HDGs.

Single-crossing preferences capture the idea that agents are ordered such that for any pair of alternatives $\{a, b\}$ all agents who prefer a to b precede the agents who prefer b to a or vice versa. In our definition, we additionally assume that the preferences are single-crossing with respect to the given order of the agents by indices. This is without loss of generality as long as we assume that the single-crossing axis of the voters is part of the input.

Formally, an AHG (or HDG) is said to be *single-crossing* if for every pair of coalition sizes $a, b \in [n]$ (or pair of ratios $a, b \in F$) with $a \succ_{a_1}^X b$ (where $X = S$ or $X = F$) there exists $j \in [n]$ such that $\{i \in N \colon a \succ_i b\} = \{a_i \colon i \in [j]\}$.

The key insight why we can obtain mechanisms with a constant-factor welfare for single-crossing domains is that the middle agent $a_{\lceil \frac{n}{2} \rceil}$'s pairwise preferences are similar to large proportions of agents. Hence, by choosing a coalition size or ratio that is good for

the middle agent, we also achieve a significant welfare gain from other agents.

Theorem C.1. *There is a strategyproof and individually rational mechanism for single-crossing AHGs that achieves a constant-factor approximation of the maximum welfare.*

Proof. As in the proof of Theorem 5.8, we define $\mathcal{M}_1^{AHG}$ as the mechanism that selects the singleton partition.

Now, let $s := \lfloor \frac{n}{4} \rfloor$ and $k := \lceil \frac{n}{2} \rceil$. Our mechanism first selects the most preferred alternative $x \in [s]$ according to a_k and then executes $\mathcal{M}_x^{AHG}$, however, with the modification that a_k has the highest priority to be assigned to a coalition of size x.

By Proposition 5.1, this mechanism is individually rational. Note, however, that strategyproofness does not follow from Proposition 5.1, because the mechanism first chooses the parameter x. We will argue next that the mechanism is still strategyproof.

First, for $i \in [n] \setminus \{k\}$, agent a_i does not have an influence on the selected parameter. Hence, with the analogous proof as for Proposition 5.1, they cannot benefit by misrepresenting their preferences.

Now, consider agent a_k. Assume that x is the best alternative in $[s]$ according to a_k. If $x = 1$, then a_k is guaranteed to be in a singleton coalition. If $x > 1$, then, since the preferences are single-crossing, at least $\frac{n}{2}$ agents prefer x over 1. As $x \leq \frac{n}{4}$, at least one coalition of size x is formed, which has to contain a_k according to our modification to $\mathcal{M}_x^{AHG}$. Hence, a_k ends up in a coalition of size x if they state x as their most preferred alternative among $[s]$. This implies strategyproofness because only the preferences of a_k over the set $[s]$ matter.

It remains to establish the welfare approximation. Without loss of generality, we may restrict attention to AHGs where the preferences over $[n] \setminus [s]$ occupy the first $n-s$ positions of every agent. Clearly, the welfare of the partition produced by the mechanism in any other instance can be bounded by the welfare of such a restricted instance that maintains the preferences over $[s]$.

Now, let $z \in [s] \setminus \{x\}$. Since preferences are single-crossing, the agents preferring z to x either all have indices smaller than k or larger than k. Hence, the welfare of a partition where every agent achieves at least the utility of a coalition size of x is at least $\frac{n}{2}(s-1)$. In the partition produced by $\mathcal{M}_x^{AHG}$, at most $x-1 \leq s-1$ agents do not receive a utility of at least their utility for x. Since we assume that alternatives in $[s]$ are ranked last, the utility of such an agent for x can be at most $s-1$. Together, the welfare achieved by the partition produced by the mechanism is at least

$$\begin{aligned}
&\frac{n}{2}(s-1) - (s-1)(s-1) \\
&\geq 2s(s-1) - (s-1)(s-1) \\
&= s^2 - 1 \geq \left(\frac{n}{4} - 1\right)^2 - 1.
\end{aligned}$$

Since the maximum welfare is bounded by $n(n-1)$, this yields a constant-factor approximation of the maximum welfare. □

We conclude with an example showing that the same approach does not work for balanced HDGs.

Example C.2. We define a mechanism $\mathcal{M}$ for HDGs. Given an HDG, let $H \subseteq F \setminus \{0,1\}$ be a nonempty subset of feasible ratios and $k := \lceil \frac{n}{2} \rceil$. Consider the mechanism that first determines the most-preferred ratio f of a_k in the set S and then runs $\mathcal{M}_f^{HDG}$.

We define a single-crossing and balanced family of instances, where this algorithm does only achieve a welfare of $\frac{n}{2}$, which is a factor of $\Theta(|F|) = \Theta(n^2)$ worse than the maximum welfare.

Let $\ell \in \mathbb{N}$ with $\ell \geq 2$ and consider the balanced instance with $n = 2\ell$ agents, ℓ of which are red and blue. We assume that $R = \{a_i \colon 1 \leq i \leq \ell\}$ and $B = \{a_i \colon \ell + 1 \leq i \leq 2\ell\}$. Let $f^* \in H$ be any ratio and let $f' \in F \setminus \{0, 1, f^*\}$ (F contains another ratio because $\ell \geq 2$). The preferences of all red agents have f' ranked top, f^* ranked bottom, and 1 at the second last position. The preferences of all blue agents have f^* ranked first, then f', and a ratio of 0 ranked last. Clearly, given these constraints, the preferences can be extended to single-crossing preferences.

In this instance, since any coalition of ratio f^* would require at least one red agent, $\mathcal{M}$ produces the singleton coalition, which achieves a welfare of $\frac{n}{2}$. On the other hand, there exists a coalition of ratio f' that encompasses at least half of the red or half of the blue agents. This coalition can be extended by singleton coalitions to a partition $\mathfrak{C}$ with $\mathcal{SW}(\mathfrak{C}) = \frac{n}{4}(|F| - 3)$. ◁

We remark that the mechanism considered in Example C.2 is not even strategyproof for most sets S, even those containing small ratios. Like in Example C.2, there might not exist any individually rational coalition of some ratio $f_1 \in S$, while such a coalition exists for other ratios $f_2 \in S$. In this case, if f_1 is a_k's top-ranked alternative among S, they have an incentive to misreport f_2 above f_1.