Journal of Artificial Intelligence Research 75 (2022) 1139-1176 Submitted 02/2022; published 11/2022

Learning to Design Fair and Private Voting Rules

Farhad Mohsin MOHSIF@RPI.EDU
Ao Liu LIUA6@RPI.EDU
Rensselaer Polytechnic Institute

Pin-Yu Chen PIN-YU.CHEN@IBM.COM
Francesca Rossi FRANCESCA.ROSSI2@IBM.COM
IBM Research

Lirong Xia XIALIRONG@GMAIL.COM
Rensselaer Polytechnic Institute

Abstract

Voting is used widely to identify a collective decision for a group of agents, based on their preferences. In this paper, we focus on evaluating and designing voting rules that support both the privacy of the voting agents and a notion of fairness over such agents. To do this, we introduce a novel notion of group fairness and adopt the existing notion of local differential privacy. We then evaluate the level of group fairness in several existing voting rules, as well as the trade-offs between fairness and privacy, showing that it is not possible to always obtain maximal economic efficiency with high fairness or high privacy levels. Then, we present both a machine learning and a constrained optimization approach to design new voting rules that are fair while maintaining a high level of economic efficiency. Finally, we empirically examine the effect of adding noise to create local differentially private voting rules and discuss the three-way trade-off between economic efficiency, fairness, and privacy.

1. Introduction

Voting is one of the most used and well-studied methods to make a collective decision (Brandt et al., 2016). In the relevant literature, voting rules have been designed, evaluated, and compared w.r.t. various desirable normative properties. Some desiderata concern economic efficiency; for example, Condorcet efficiency requires that a decision (also usually termed "alternative") that beats other alternatives in pairwise comparisons be chosen. Others concern fairness; for example, anonymity requires that the selection of the collective decision should only be based on the agents' reported preferences and must be insensitive to their identities or features.

While anonymity preserves the "one person, one vote" principle (often used in political elections), it may not be suitable for many other applications of voting. In fact, the use of anonymous voting rules may lead to the well-known "tyranny of the majority" (Mill, 1859) issue, where the majority of voting agents dictates the decision, which may not be favorable to the minority. This motivated us to consider other properties that ensure fairness over groups of agents.

Recently, various notions of *group fairness* (Chouldechova & Roth, 2020) have been proposed to address fairness in algorithmic decision-making, mainly focusing on classification problems in machine learning (ML). Due to bias in data or training methodologies, an algorithmic decision-making system can be biased towards one group of people in terms of

important metrics such as accuracy, positive predictive value, etc. This can lead to unfair decisions, creating discrimination. To avoid this, fairness is defined over protected features (e.g., gender, race, etc.) that indicate group membership. For example, we may require that the prediction accuracy is equal for different groups, such as men and women.

While voting does not have metrics like prediction accuracy, we can consider an analogous scenario where the average utility received by different groups of agents is equal. We, therefore, define a novel way to measure group fairness in voting (Definitions 1–2), focusing on how voting outcomes (that is, the collective decisions) affect different groups of agents. We then investigate existing voting rules in terms of the trade-off between fairness and economic efficiency and find worst-case results for fairness (Theorems 1–6) that show that well-studied voting rules can be very unfair in the worst case.

This way of defining and achieving fairness over groups of agents needs to expose features of agents, since such features define the group that each agent belongs to. This means that the voting process is not anonymous, which leads to privacy concerns. To circumvent this, we employ the well-studied notion of *differential privacy (DP)* (Dwork et al., 2006) to ensure that an adversary cannot learn too much about the voting behaviors of the agents from the voting outcome. This is usually done by adding noise to the outcome of the collective decision process, to ensure that it does not compromise privacy regarding individual voting behavior. However, since aggregating votes is a centralized process where an aggregator collects all votes and applies a voting rule to determine the collecting decision, adversarial attacks on the aggregator can be privacy-compromising as well. This compels us to consider an extension of *local differential privacy (local DP)* (Evfimievski et al., 2003) where noise is rather added to individual votes, so even an attack on the aggregator is not privacy-compromising for the agents.

We study theoretically the effect of adding local differential privacy to fair voting rules and analyze the three-way trade-off between fairness, economic efficiency, and privacy (Theorem 7). We show that, while a high privacy requirement results in higher efficiency loss, we can have moderate privacy with a small decrease in either efficiency or fairness.

To add to our theoretical work, on the practical front we present two frameworks to design voting rules with varying levels of fairness, privacy, and economic efficiency. For the first framework, we define a family of voting rules that maximize fairness under efficiency constraints and can be thought of as a natural extension of positional scoring rules such as Plurality, Borda, etc. The extension comes from looking at alternative scores as indicators for group utilities. The second framework employs a machine learning-based approach that allows us to design fair and efficient voting rules that go beyond just positional scoring rules and work with fairness definitions under more general notions of utility. Similarly, notions of economic efficiency other than utilities, e.g., Condorcet efficiency, are considered. We use a mixture of synthetic voting data that is fair and efficient to learn voting rules that achieve different levels of efficiency and group fairness.

Experimentally, we show that the learned family of voting rules succeeds in achieving high fairness and efficiency satisfaction levels, based on simulations on synthetic data. In particular, our newly designed voting rules are never dominated by a voting rule that focuses on purely economic efficiency or group fairness. Additionally, we analyze the effect of adding local noise to design local differentially private voting rules. Finally, we experimentally verify our theoretical results for the fairness-efficiency-privacy trade-off, showing that for

moderate privacy requirements (when the noise level is not very high), the loss in efficiency and fairness is small.

To summarize, the goal of this paper is to study and design collective decision-making processes that support group fairness, economic efficiency, and privacy. Often these properties will conflict with each other. So, we first discuss and define notions of group fairness and privacy in voting frameworks and theoretically analyze how they affect voting rules. We then empirically analyze the trade-off between efficiency, fairness, and privacy, and propose methods to design fair and private voting rules.

Economic efficiency, other notions of fairness, and privacy have all been studied before. However, we are not aware of previous work that designs and studies voting rules with a focus on group fairness in a single-winner setting, or that discusses the three-way trade-off between efficiency, fairness, and privacy.

2. Related Work

Fairness in Voting and Beyond. The term fairness lends itself to various concepts, some related to our work, in voting theory. Diversity constraints and fairness constraints including statistical parity have been discussed by Bredereck et al. (2018), Celis et al. (2018) in a multi-winner setting. Other ways of looking at fairness in voting have been introduced: justified representation in multi-winner scenarios (Chamberlin & Courant, 1983; Monroe, 1995; Skowron et al., 2015; Sánchez-Fernández et al., 2017; Aziz et al., 2017, 2018), sortition (Stone, 2016), proportional apportionment (Balinski & Young, 2010; Cembrano et al., 2021). Many of these works focus on multi-winner elections, and some assume implicit group membership among the agents. Fair social choice in a dynamic setting has also been studied, where in a scenario of repeated voting, the goal is to get fair outcomes in the long term (Parkes & Procaccia, 2013; Freeman et al., 2017; Lackner, 2020). In comparison to this, we focus on a single-winner in a static setting and explicit group membership in terms of protected features. Still, the motivation behind these works is relevant to our work.

Our approach to fairness in voting can also be considered a form of egalitarian voting (Myerson, 1981; Brams et al., 2007; Contucci et al., 2016). However, we present a setting with an explicit specification of group memberships for agents, motivated by recent work in the machine learning literature about group fairness. When individuals are divided into groups according to certain *protected features* (e.g., gender, race), group fairness requires (approximate) parity between the groups for some statistical measure. We refer to Chouldechova (2017), Corbett-Davies et al. (2017), Kleinberg (2018), Verma and Rubin (2018), Chouldechova and Roth (2020) for more exposition to the topic. The notion of treating different groups of agents similarly leads to a discussion about group equitability and group envy-freeness (Hossain et al., 2020; Aleksandrov & Walsh, 2018) and also fairness gerrymandering (Kearns et al., 2018). We study similar notions in a voting based decision-making scenario.

Differential Privacy (DP). A vast number of mechanisms have been proposed to achieve DP (Dwork et al., 2006) by perturbing the output of voting rules (Shang et al., 2014; Hay et al., 2017; Lee, 2015; Birrell & Pass, 2011). However, these mechanisms usually cannot guarantee privacy in distributed settings like collective decisions, where the central aggregator may become compromised. Local DP (Evfimievski et al., 2003) is a similar notion as

DP, designed to avoid the potential privacy risks in distributed systems, and widely used to protect the sensitive personal data in the procedure of data collection (Ye et al., 2019; Cormode et al., 2018; Gursoy et al., 2019). Joseph et al. (2018) first introduced the concept of local DP to voting systems. However, they only studied voting settings with two possible decisions (agree and disagree). Recently, local DP has become a common privacy notion for voting systems, especially in the application scenarios where the rank-aggregators are not trustworthy. Yan et al. (2020) proposed a local DP voting rule able to aggregate pairwise comparisons. Wang et al. (2019) demonstrated the usefulness, soundness, truthfulness and indistinguishability properties of local DP voting rules. Besides this, Liu et al. (2020) theoretically connected privacy with the intrinsic randomness of votes. In this paper, we study the local DP properties of general voting rules and focus on exploring the trade-off among fairness, economic efficiency, and local DP.

Automated Mechanism Design. Voting rules fall under the umbrella of social choice mechanisms that are used for making decisions with desired properties from agent inputs. The notion of automating the mechanism design (Conitzer & Sandholm, 2002, 2003) process has a growing body of literature, but most of the work has focused on mechanisms involving money. For example, in the auction setting, incentive-compatible revenue maximizing techniques have been developed under various constraints using deep neural networks (Dütting et al., 2019; Shen et al., 2019; Curry et al., 2020). In social choice, (Feng et al., 2018; Golowich et al., 2018) used deep learning to design mechanisms without money for problems like multi-facility location, double-sided matching, social choice for single-peaked preferences, etc. Machine learning frameworks have also been explored to learn specific families of voting rules (scoring rules, tree-based rules) (Procaccia et al., 2009; Jha & Zick, 2020). The idea of incorporating social choice axioms into an ML framework through a synthetic data generation process has been considered for designing voting rules as well (Xia, 2013; Armstrong & Larson, 2019). Anil and Bao (2021) use neural network architectures such as deep sets, set transformers, and graph neural networks to learn utility maximizing voting rules. As opposed to this final work, we consider feature representations of preference profiles so that the learning mechanism is scalable to large elections. We expand the idea of designing learned voting rules from synthetic data by adding our novel fairness criterion as additional requirements to the system.

3. Preliminaries

In the preliminaries section, we present some background literature on voting rules and the concept of local differential privacy.

3.1 Voting Rules

Let $\mathcal{A} = \{a_1, \ldots, a_m\}$ be the set of m alternatives and $\mathcal{L}$ be the set of all rankings or linear orders (linear orders are anti-symmetric, transitive, and total binary relations) over $\mathcal{A}$. There are n agents (voters), each provides a full ranking, $R \in \mathcal{L}$ over $\mathcal{A}$ as her vote. In a ranking R, if alternative a is preferred to another alternative b, we write $a \succ_R b$. A collection of n votes, $P \in \mathcal{L}^n$ is called a preference profile. A voting rule is a mapping $r : \mathcal{L}^n \mapsto \mathcal{A}$ that chooses a winner from the preference profile. To indicate protected features, we assume

each agent is a member of one of two disjoint groups with group sizes n_1, n_2 ($n_1 + n_2 = n$). Each group has preference profile $P_k \in \mathcal{L}^{n_k}$ for $k = 1, 2$. To consider group memberships, we redefine *voting rule* as a mapping from a collection of preference profiles to a winner, $r : \mathcal{L}^{n_1} \times \mathcal{L}^{n_2} \mapsto \mathcal{A}$. We refer to P as the preference profile with all agents. For voting rules that do not use group membership, $r(P)$ and $r(P_1, P_2)$ are the same. Most of the theoretical and experimental work in the paper is presented for two groups, so we focus the preliminaries on two-group scenarios as well. We also briefly discuss how to extend this notion to more groups in Section 4.2.

A common family of voting rules is **positional scoring rules**, which have score vector $\vec{s} = \langle s_1, \ldots, s_m \rangle$ such that $s_1 \geq \cdots \geq s_m$ and $s_1 > s_m$. For each ranking R, the j-th ranked alternative gets a score of s_j. Given a preference profile, the alternative with maximum total score will be the winner. Some popular scoring rules are: **Plurality**, with scoring vector $\langle 1, 0, \ldots, 0 \rangle$; **Borda**, with $\langle m-1, m-2, \ldots, 1, 0 \rangle$; **Veto**, with $\langle 1, 1, \ldots, 1, 0 \rangle$.

Condorcet rules is a family of voting rules that are defined by a different measure of efficiency called the Condorcet criterion. For a preference profile, if an alternative beats all other alternatives in pairwise comparison, it is called the *Condorcet winner*. An alternative a_i beats another alternative a_ℓ if $N_{(a_j \succ a_\ell)} > N_{(a_\ell \succ a_j)}$, where $N_{(a_i \succ a_\ell)}$ is the number of agents who prefer a_i to a_ℓ in preference profile P. A voting rule satisfies the Condorcet criterion if it always selects the Condorcet winner whenever it exists. For example, the **Copeland** rule chooses the alternative that maximizes the number of alternatives that it beats in pairwise comparisons. Formally, define the pairwise comparison variable as follows:

$$w_C(a_j, a_\ell, P) = \begin{cases} 1 & \text{if } N_{(a_j \succ a_\ell)} > N_{(a_\ell \succ a_j)}, \\ 0 & \text{otherwise.} \end{cases}$$

The Copeland winner is the alternative that maximizes the Copeland score, $score_C(a_j, P) = \sum_{\ell \neq j} w_C(a_j, a_\ell, P)$.

3.2 Economic Efficiency in Voting

In this paper, we consider two types of economic efficiency that are popular in the social choice literature, and both are related to the two families of voting rules that we mentioned before: Condorcet rules and positional scoring rules.

The first is *Condorcet efficiency*, which is dependent on the Condorcet criterion. For a preference profile, P, the Condorcet winner exists only if there is an alternative that beats all other alternatives in pairwise comparison. In cases where the Condorcet winner exists, the winner according to a voting rule r, $r(P)$, may be the same or different from the Condorcet winner. As mentioned before, Copeland satisfies the Condorcet criterion and will always choose the Condorcet winner when it exists, but the same is not true for Borda, Plurality, etc. So, we measure Condorcet efficiency as the fraction of preference profiles where a voting rule winner is identical to the Condorcet winner. For n agents and m alternatives, Condorcet efficiency (CE) is defined as

$$CE(r) = \frac{\sum_{P \in \mathcal{L}^n} r(P) \text{ is the Condorcet winner}}{\sum_{P \in \mathcal{L}^n} P \text{ where the Condorcet winner exists}}.$$

This is an efficiency measure as efficient decisions (output of voting rules) should be preferred to all other alternatives. Condorcet rules like Copeland have a CE value of 1, whereas positional scoring rules have CE values less than one. It is usually not analytically or computationally possible to compute Condorcet efficiency exactly. In this paper, we empirically estimate Condorcet efficiency for voting rules using sample preference profiles instead of counting for all preference profiles in $\mathcal{L}^n$.

The other efficiency notion takes a utilitarian view (Boutilier et al., 2015), where each agent can receive different cardinal utilities from the alternatives. For an agent, a utility function $u : \mathcal{A} \times \mathcal{L} \mapsto \mathbb{R}$ defines alternative a's utility to agent i. A utility function, u, is consistent with a ranking R if for all $a, b \in \mathcal{A}$, $a \succ_R b \implies u(a, R) > u(b, R)$. For this paper, we limit ourselves to every agents having the same utility function, which is dependent only on the rank. Thus, we assume that a utility function u is defined by a vector $\vec{u} = \langle u_1, \ldots, u_m \rangle$ such that $u_1 \geq \cdots \geq u_m$ and $u_1 > u_m$. If an alternative a is ranked j-th in R, then $u(a, R) = u_j$.

We denote the family of all such utility functions as $\mathcal{U}$, hereon. All $u \in \mathcal{U}$ use a vector similar in definition to score vectors of positional scoring rules. To reduce confusion, we will use $\vec{s}$ for the score vectors and $\vec{u}$ for utility functions.[1] Given a preference profile $P \in \mathcal{L}^n$ and utility function u, the *social welfare (SW)* is $\sum_{R \in P} u(a, R)$, the sum of utilities for all agents. In a utility maximization problem, the social welfare maximizing alternative will be the winner. However, we are concerned with the amount of utilities received by different groups. The range of social welfare for different groups will vary greatly based on the size of the groups. To circumvent this, we define average utility as follows. The *average utility* for alternative a is $W(a, P) = \frac{1}{n} \sum_{R \in P} u(a, R)$. For preference profile P_k of group k, average utility $W(a, P_k)$ is a measure of satisfaction for group k when a is the winner of the election.

3.3 Local Differential Privacy (DP)

We adapt the formal definition of local differential privacy (Evfimievski et al., 2003) to the domain of voting and state its difference from standard DP. Recall that a voting rule for n agents is a mapping $r : \mathcal{L}^n \to \mathcal{A}$.

The notion of differential privacy usually requires a private mechanism to be random. Thus, we require a randomized version of a voting rule, where the output will be a probability distribution over the set of alternatives. The randomization usually comes from adding noise to the original non-private method (Figure 1). A randomized voting rule r is said to be ε -local DP, if for any agent j, any alternative $a \in \mathcal{A}$, and any pair of rankings $R, R' \in \mathcal{L}$, the following holds: $\Pr[r(R, P_{-j}) = a] \leq \exp(\varepsilon) \cdot \Pr[r(R', P_{-j}) = a]$, where P_{-j} is the preference profile with all agents other than agent j. In particular, this indicates that the vote of any single agent will be hard to infer from the outcome. Hence, it gives a privacy guarantee to the agents. Smaller ε means stronger privacy guarantee. Local DP is stronger than standard DP (Dwork, 2006), because it requires the individual votes to be private, and therefore implies that the aggregation of agents' data is private due to the *post processing* property (Dwork et al., 2014). On the other hand, standard DP does not

1. We want to emphasize that we use different vectors $\vec{s}$ and $\vec{u}$ for the score vectors and utilities intentionally. While the positional scoring rules has the notion of maximizing some sort of utility measure, the scoring vector used may be different from a true utility function.

guarantee privacy at the level of an individual agent. Rather, a voting rule is ε -DP if for any two preference profiles P, P' that differ on only one vote, for all alternatives $a \in \mathcal{A}$, $\Pr[r(P) = a] \leq \exp(\varepsilon) \cdot \Pr[r(P') = a]$. Figure 1 illustrates the comparison between standard and local DP. We note that voting rules with local DP (or standard DP) must be randomized. So, our local differentially private voting rules can only be used where some randomization would not be very problematic.

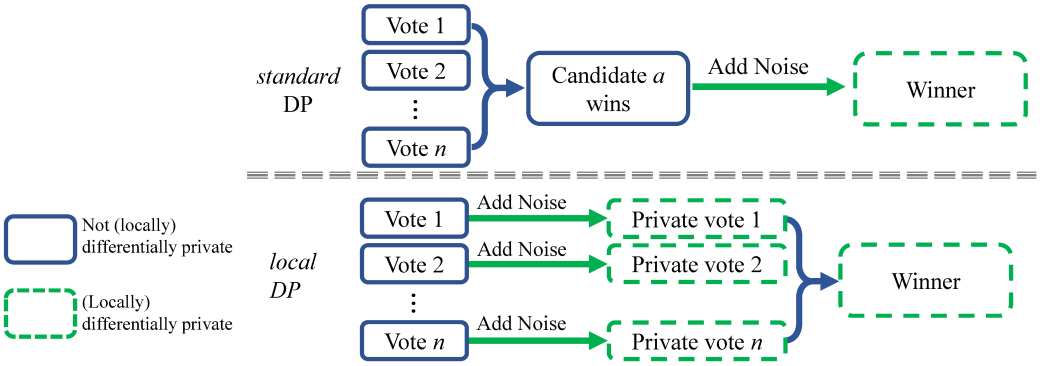


Figure 1: Comparison between local DP and standard DP.

4. Group Fairness in Voting

In this section, we formally define the notion of group fairness in voting. To do this, we consider the difference in utilities that different groups of agents receive from a voting outcome. Initially, we define group fairness in terms of two groups and then extend the definition to work for more groups.

4.1 Fairness for Two Groups

In presence of pre-defined groups among agents, traditional voting rules that are anonymous and do not differentiate between different agents may be unfair to some groups. Particularly, if one group is large, their preferences may always get higher preference and lead to tyranny of the majority, as shown in Example 1.

Example 1. *Consider the scenario presented in Table 1 with two groups of agents and three alternatives. The utility function in consideration here is based on the utility vector* $\langle 2, 1, 0\rangle$ *for* $m = 3$ *alternatives. We see that alternative A receives high utility from the larger group,* G_1*, and in turn has the highest average utility for the whole population. However, A has zero utility for the smaller group ,*G_2*, and this decision would be unfair to* G_2*. We notice that alternative B is not ranked lowest in terms of average utility to either group and can be viewed as a more fair decision than A.*

This idea of treating different groups similarly is captured in the notion of *statistical parity*. Statistical parity is a desirable property on the high level because it indicates there is no broad statistical discrimination between the utilities received by different groups.

	#agents	Ranking	Utility A	B	C
Group G_1	30	A $\succ$ B $\succ$ C	60	30	0
	30	B $\succ$ A $\succ$ C	30	60	0
	20	A $\succ$ C $\succ$ B	40	0	20
Group G_2	10	B $\succ$ C $\succ$ A	0	20	10
	10	C $\succ$ B $\succ$ A	0	10	20
G_1 average utility			1.625	1.125	0.25
G_2 average utility			0	1.5	1.5
Overall average utility			1.3	1.2	0.5

Table 1: Sample preference profile and group utilities

As mentioned in the introduction, group fairness has become an important notion in the domain of algorithmic decision-making. In this work, we consider this notion in the domain of collective decision-making, with consideration of the group identity of the agents. And this is one of the properties that motivate our definition of group-imbalance and imbalance-induced fairness. Statistical parity is formalized in the following way: for two groups, an election winner a satisfies statistical parity in terms of group utility if there exists small ξ_k such that $|W(a, P_k) - W(a, P)| < \xi_k$ for both groups, $k = 1, 2$. That is, statistical parity requires that both group's average utilities be close to the overall average utility.

Definition 1 (Group imbalance)**.** Given a utility function u, an alternative $a \in \mathcal{A}$, and preference profiles P_1, P_2 for two groups of agents, imbalance between the two groups in terms of u for a is

$$\text{Imb}(u, a, P_1, P_2) = \begin{cases} \frac{|W(a,P_1) - W(a,P_2)|}{W(a,P)} & \text{if } W(a, P) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

where P is the combined preference profile for all agents.

Remark 1. For any $\delta \to 0$, if we have $\text{Imb}(u, a, P_1, P_2) < \delta$, then there exist $\xi_1, \xi_2 \to 0$ such that statistical parity is satisfied for the group utilities, i.e., $|W(a, P_k) - W(a, P) < \xi_k|$ for $k = 1, 2$.

Remark 1 indicates how group imbalance can be useful for defining group fairness in a similar note to statistical parity. Imbalance indicates the unfairness of an alternative in terms of two groups. The imbalance values will differ based on how different pairs of groups value an alternative. An alternative can be considered fair if it has low imbalance for the two groups. In Lemma 1 and Example 2, we try to give an idea about what kind of values imbalance can have.

Lemma 1. *Given any utility function, u, for any preference profiles $P_1 \in \mathcal{L}^{n_1}, P_2 \in \mathcal{L}^{n_2}$, for all $a \in \mathcal{A}$, $0 \leq Imb(u, a, P_1, P_2) \leq \frac{n_1+n_2}{\min(n_1,n_2)}$.*

Proof. W.l.o.g., assume that $n_1 \geq n_2$ and that the score vector associated with u is $\langle u_1, \ldots, u_m \rangle$. Now, the average utility for the full preference profile, $W(a, P) = \frac{n_1}{n_1+n_2} W(a, P_1) + \frac{n_2}{n_1+n_2} W(a, P_2)$. We have two cases.

1. $W(a, P_1) \geq W(a, P_2)$,
2. $W(a, P_1) < w(a, P_2)$.

For case 1,

$$\text{Imb}(u, a, P_1, P_2) = \frac{(n_1 + n_2)(W(a, P_1) - W(a, P_2))}{n_1 W(a, P_1) + n_2 W(a, P_2)}.$$

If n_1, n_2, and $W(a, P_2)$ are all fixed, then this is a monotonically increasing function for $W(a, P_1)$ when $W(a, P_1) > W(a, P_2)$. Similarly for fixed n_1, n_2 and $W(a, P_1)$, this is a monotonically decreasing function for $W(a, P_2)$ when $0 < W(a, P_2) < W(a, P_1)$. Both these facts combined, we observe that the worst-case (largest) imbalance occurs when when $W(a, P_1) = \max(u_1, \ldots, u_m) = u_1$ and $W(a, P_2) = \min(u_1, \ldots, u_m) = u_m$. Plugging in these values gives us maximum imbalance of $\frac{(n_1+n_2)(u_1-u_m)}{n_1 u_1}$. Using the fact that $u_m \geq 0$ for any utility function, we get the following bound: $\text{Imb}(u, a, P_1, P_2) \leq \frac{n_1+n_2}{n_1}$.

For case 2, we have

$$\text{Imb}(u, a, P_1, P_2) = \frac{(n_1 + n_2)(W(a, P_2) - W(a, P_1))}{n_1 W(a, P_1) + n_2 W(a, P_2)}.$$

A similar argument as in case 1 gives maximum imbalance of $\frac{n_1+n_2}{n_2}$. Since $\frac{1}{n_2} \geq \frac{1}{n_1}$, this gives the final upper bound in the Lemma. □

Example 2. *For Example 1, we have group sizes of* 80 *and* 20. *We notice that from Lemma 1, this means that the worst-case imbalance can be 5. This will occur for some alternative that has highest utility for everyone in the minority group and zero utility for everyone in the majority group. This leads to a high imbalance and is an undesired output. On the other hand, for the opposite case of an alternative with high utility for majority group and zero utility for the minority group will have an imbalance value of* $\frac{5}{4}$. *This is still high imbalance (as a fair alternative ideally has close to zero imbalance), but numerically, this alternative receives a lower imbalance value than the other one. And if these two were the only available alternatives, choosing the alternative preferred by the minority group would actually cause an ironic tyranny of the minority and hence has the higher imbalance value.*

We define imbalance-based group fairness of a voting rule for a utility function in terms of the worst-case imbalance achieved by the voting rule winner for any preference profile. While our notion of fairness can be defined and studied for the general q group setting, we start with cases with only two group. Since imbalance is always between pairs of groups, we can work on extending our work naturally to a higher number of groups, which we discuss in Section 4.2.

Example 2 and Lemma 1 indicate that the ratio of group sizes $z = \frac{\max(n_1,n_2)}{\min(n_1,n_2)}$ is important in fairness consideration. We have $\max(n_1, n_2) = \frac{z}{1+z}n$ and $\min(n_1, n_2) = \frac{1}{1+z}n$. Thus, total number of agents, n, and group ratio, z, determine the group sizes.

We now define a normalized notion of fairness for a voting rule, given group size parameters, n and z, and a utility function, u. This is done by taking considering the worst-case imbalance achieved by the voting rule winner out of all possible preference profiles.

Definition 2 (Group imbalance-based fairness)**.** Given a voting rule, r, utility function, u, n total agents, and group ratio, z, the group imbalance-based fairness is

$$F(r, u, n, z) = 1 - \tfrac{1}{1+z} \cdot \max_{\substack{P_1 \in \mathcal{L}^{n_1} \\ P_2 \in \mathcal{L}^{n_2}}} \text{Imb}(u, r(P_1, P_2), P_1, P_2).$$

Definition 2 gives a notion of fairness that is not specific to a preference profile, but rather looks at the worst-case imbalance. Not all values of n, z will lead to valid group sizes, but we still consider all z values, because with $n \to \infty$, any $z > 1$ will lead to valid preference profiles and considered for worst-case performance. Because of Lemma 1, it is guaranteed that $F(r, u, n, z)$ will have a value between 0 and 1 for any r, u, n, and z. A low fairness value indicates that in the worst case the voting rule winner will have high group imbalance. On the other hand, a high fairness value indicates that, even in the worst case, group imbalance for the voting rule winner will not be very high. We will present theoretical results for fairness for different voting rules in Section 5.

4.2 Fairness for More than Two Groups

While the initial discussion concerned two groups, fairness in cases where there are more than two groups is also of interest. Our definition can be extended to more than two groups in two ways, both of which have their own merit. Let there be q groups with preference profiles $P_1, \ldots, P_q$. For a voting rule r, $r(P_1, \ldots, P_q)$ is the winner given all q groups. For each group preference profile P_k, we define its complement P_k^C as the union of all other preference profiles.

Group-complement imbalance: For more than two groups, we can define imbalance from each group's perspective with respect to its complement; for group k it becomes $\text{Imb}(u,a, P_k, P_k^C)$ and thus we define the generalized Fairness of Voting Rule in presence of q groups of size $n_1, \ldots, n_q$.

For each group, we can think whether the collective decision is fair for that group specifically by looking at the imbalance between a group and its complement. Thus a fair collective decision is the one that minimizes imbalance for all group-complement pairs. Based on this idea, we can have the following definition of generalized group fairness of voting rules.

Given utility function u, group sizes $n_1, \ldots, n_q$ with $n = \sum_k n_k$, we define group fairness for voting rule r as

$$F(r, u, n_1, \ldots, n_q) = 1 - \max_{P_1, \ldots, P_q} \tfrac{\min(n_1, \ldots, n_q)}{n} \cdot \max_{k \le q} \text{Imb}(u, r(P_1, \ldots, P_q), P_k, P_k^C).$$

As can be expected, the worst-case fairness value comes from the most imbalanced group. For group-complement imbalance, we define the group size ratio z to be $\frac{n - \min(n_1, \ldots, n_q)}{\min(n_1, \ldots, n_q)}$. As will be apparent in the following section, when we discuss fairness results based on group imbalance, the group size ratio z will be an important parameter.

Group-pairwise imbalance: On the other hand, we may be interested in pairwise group interaction. For example, a collective decision, while not unfair to a group overall, may be unfair with when compared to some other group. In this scenario, given utility

function u, group sizes $n_1, \ldots, n_q$ with $n = \sum_k n_k$, we define group fairness for voting rule r as

$$F(r, u, n_1, \ldots, n_q) = 1 - \max_{P_1,\ldots,P_q} \min_{k,\ell \leq q} \frac{n_\ell}{n_k + n_\ell} \cdot \mathrm{Imb}(u, r(P_1, \ldots, P_q), P_k, P_l).$$

We define group size ratio, z, for this scenario as $\max_{k,l} \frac{n_k}{n_l}$, the worst-case ratio between any two group sizes.

For most of our theoretical results, the arguments will be based on a two-group scenario. However, since the notion of imbalance-based fairness is extended to more groups as above, we will present the results for multi-group scenarios.

5. Fairness Results for Voting Rules

In this section, we consider fairness results for various voting rules, with the aim to analyze how unfair some voting rule may be in the worst case. First, we present fairness results as function of a voting rule, r, specific utility function, u and preference profile parameters n, z. As mentioned in the previous section, n is the size of the number of total agents and z is the group size ratio parameter.

Before we delve into the theorems and proofs, we briefly discuss the proof technique here. Most of the proofs of our fairness theorems follow a similar strategy. We will show the detailed proof for two groups and later describe how the result generalizes to our two fairness definitions for more groups. We always consider the two cases as mentioned in the proof for Lemma 1, which corresponds to one group receiving a higher average utility compared to the other. And then, we find the maximum imbalance for each case which gives us the fairness result of a voting rule. In particular, we try to see if the worst-case imbalance value $\frac{n}{n_2}$ can be reached, which gives 0 as fairness value. We will also repeatedly make use of the fact that the imbalance function is monotonic in terms of $W(a, P_1)$ and $W(a, P_2)$ when the other terms are fixed. In the proof for Theorem 3, we explicitly address the extension of the results to group imbalance-based fairness for more than two groups. The argument is similar for all of the other theorems about fairness results. So for them, the proofs are given only for two groups.

5.1 Positional Scoring Rules

Our first results are for positional scoring rules. We first present the fairness results for popular positional scoring rules like Plurality and Borda. We consider the utility functions defined by $\vec{u}_{top} = \langle 1, 0, \ldots, 0 \rangle$ and $\vec{u}_{rank} = \langle m-1, m-2, \ldots, 0 \rangle$.

With the top-1 utility function, the agents only receive utility if their top ranked alternative wins. A fairness value of 0 for Borda means that, for some preference profile, the Borda winner will have the worst-case imbalance value, being highly unfair. But if we choose the Plurality winner, at least in some cases we are guaranteed a positive fairness value. On the other hand, with the rank utility function, where the utility linearly decreases along with rank, we see that the Borda winner has the same fairness value as the most fair voting rule, whereas Plurality's fairness value is 0. In both cases, the voting rule that maximizes the utility function has better fairness results. See Table 2 for a summary of these results.

	u_{top} (Theorem 1)	u_{rank} (Theorem 2)
Plurality	0 if $z < m-1$ $1-1/z$, otherwise	0
Borda	0 if $m \geq 3$ $1-1/z$ otherwise	$1-1/z$

Table 2: $F(r, u, n, z)$ for Plurality and Borda under u_{top} and u_{rank}.

The fairness results for Plurality (r_{Plu}) and Borda (r_{Bor}) under these two utility functions are formally stated and proven in the following two theorems.

Theorem 1. *Given n total agents and group size ratio, z, for m alternatives and u_{top} utility function,*

$$F(r_{Plu}, u_{top}, n, z) = \begin{cases} 1 - \frac{1}{z} & \text{if } z \leq m-1, \\ 1 - \frac{(m-1)}{z} & \text{otherwise.} \end{cases}$$

$$F(r_{Bor}, u_{top}, n, z) = \begin{cases} 0 & \text{if } m \geq 3. =, \\ 1 - \frac{1}{z} & \text{otherwise,} \end{cases}$$

Proof. Assume that we have two groups with preference profiles $P_1 \in \mathcal{L}_1^n$ and $P_2 \in \mathcal{L}_2^n$. W.l.o.g, assume that $n_1 \geq n_2$ and thus $z = \frac{n_1}{n_2}$. [2]

Plurality. When $n_2 \geq \frac{n}{m}$ and for some $a \in \mathcal{A}$, $u(a, P_2) = 1$, such an alternative can be Plurality winner with 0 average top-1 utility for agents in P_1. A preference profile where this happens has $\text{Imb}(u, a, P_1, P_2) = \frac{n}{n_2}$ which gives the first part of the result.

For $n_2 < \frac{n}{m}$, this is not possible. For alternatives with $u(a, P_1) > u(a, P_2)$ for any preference profile, we know that the worst case imbalance is $\frac{n}{n_1}$. We need to check what the worst case can be when $u(a, P_2) > u(a, P_1)$. For the worst case, assume $u(a, P_2) = 1$ and $u(a, P_1) = \frac{n_1'}{n_1}$, which we get when n_1' agents in P_1 has a on top of their ranking. For a to be a Plurality winner, we need

$$\begin{aligned} n_2 + n_1' &\geq \frac{n}{m} \\ \implies \quad n_1' &\geq \frac{n}{m} - n_2 \\ \implies u_{top}(a, P_1) &\geq \frac{1}{n_1}(\frac{n}{m} - n_2). \end{aligned}$$

Taking the smallest possible value for $u_{top}(a, P_1)$ gives–

$$\begin{aligned} \text{Imb}(u_{top}, a, P_1, P_2) &= \frac{1 - \frac{1}{n_1}(\frac{n}{m} - n_2)}{\frac{1}{m}} \\ &= \frac{(m-1)n}{n_1} \leq \frac{n}{n_1}. \end{aligned}$$

2. How to extend this proof for the general case of more groups are discussed in detail at the end of the proof of Theorem 3.

So, this is the worst case imbalance, which gives the fairness result.

Borda. For the Borda winner, we show that group imbalance can be $\frac{n}{n_2}$, leading to worst case fairness of 0 for $m \geq 3$. The minimum rank utility (or Borda count) the Borda winner needs is $\frac{n(m-1)}{2}$. Consider the preference profile where for alternative a, $u_{top}(a, P_2) = 1, u_{top}(a, P_1) = 0$ but all agents in P_1 ranks a second. So, although a would receive no top-1 utility from agents in P_1, it would contribute to Borda count. For a to be the Borda winner, we need $n_2(m-1) + n_1(m-2) \geq \frac{n(m-1)}{2}$. This is true for all $m \geq 3$. So, for $m \geq 3$, such P_1, P_2 exist that group imbalance is the maximum and thus worst case fairness is 0 for Borda.

For $m = 2$, Borda reduces to Plurality and the result immediately follows. □

Theorem 2. *Given n total agents and group size ratio, z, for m alternatives and u_{rank} utility function,*

$$F(r_{Plu}, u_{rank}, n, z) = 0,$$
$$F(r_{Bor}, u_{rank}, n, z) = 1 - \tfrac{1}{z}.$$

Proof. As in Theorem 1, we present the argument for a two group scenario, with group sizes n_1 and n_2, where $n_1 \geq n_2$.

Plurality It is trivial from the definition of plurality that as long as $n_2 \geq \frac{n}{m}$, an alternative ranked at top for every ranking in P_2 and ranked at the bottom for every ranking in P_1 can be the Plurality winner. This observation indicates that the Plurality winner achieves the worst-case imbalance for some preference profile. This results in the fairness value for Plurality being 0.

Borda We again break down into the two cases and find which one leads to a worst possible imbalance. Assume that the Borda winner is a. Case-1: $u_{rank}(a, P_1) > u_{rank}(a, P_2)$, which gives the worst case imbalance of $\frac{n}{n_1}$.

Case-2: $u_{rank}(a, P_2) > u_{rank}(a, P_1)$. For the Borda winner, minimum Borda score (and thus, minimum sum of rank utility) needs to be $\frac{n(m-1)}{2}$. Even if all rankings in P_2 has a on top, $u_{rank}(a, P_2) = n_2(m-1) \leq \frac{n(m-1)}{2}$. Thus, $u_{rank}(a, P_1) \geq \frac{1}{n_1}(\frac{n(m-1)}{2} - n_2(m-1)) = (m-1)\frac{n_1 - n_2}{2n_1}$. This gives, in the worst case,

$$\begin{aligned}
\text{Imb}(u_{rank}, a, P_1, P_2) &= \frac{(m-1) - (m-1)\frac{n_1-n_2}{2n_1}}{\frac{m-1}{2}} \\
&= \frac{1 - \frac{n_1-n_2}{2n_1}}{\frac{1}{2}} \\
&= \frac{n_1 + n_2}{n_1} = \frac{n}{n_1},
\end{aligned}$$

which gives the same bound as in case-1. This leads to the fairness result for Borda.

□

In our next theorem, we present general bounds for the fairness value of any positional scoring rule under any utility function. Let $s_{[k:m]} = \frac{s_k + \cdots + s_m}{m-k+1}$ and $u_{[k:m]} = \frac{u_k + \cdots + u_m}{m-k+1}$. Here, $s_{[k:m]}$ is the mean score for the lowest $m-k+1$ alternatives for score vector $\vec{s}$. Similarly, $u_{[k:m]}$ is the mean utility for the lowest $m-k+1$ alternatives for utility vector u. In

Theorem 3, we see that for any positional scoring rule, based on which interval the group size ratio, z, falls in, we get different upper bounds for the group fairness level. There are $m-1$ intervals that we consider:

$$\left[1, \frac{s_1 - s_m}{s_1 - s_{[m-1:m]}}\right], \left[\frac{s_1 - s_m}{s_1 - s_{[m-1:m]}}, \frac{s_1 - s_m}{s_1 - s_{[m-2:m]}}\right], \ldots, \left[\frac{s_1 - s_m}{s_1 - s_{[2:m]}}, \infty\right)$$

For any utility function, the fairness upper bound for different positional scoring rules can be significantly different, even for the same number of total agents, n, and group size ratio, z.

Theorem 3 (**Fairness for Positional Scoring Rules**). *Given utility function u, for n total agents and group size ratio, z, for any positional scoring rule, r_s with score vector $\langle s_1, \ldots, s_m \rangle$, we have*

$$F(r_s, u, n, z) \leq \begin{cases} \min\left(1 - \frac{u_1 - u_m}{z u_1 + u_m}, 1 - \frac{u_1 - u_{[k:m]}}{z u_{[k:m]} + u_1}\right) & \text{if } z \in \left[\frac{s_1 - s_m}{s_1 - s_{[k+1:m]}}, \frac{s_1 - s_m}{s_1 - s_{[k:m]}}\right] \\ & \text{for } k = m-1, \ldots, 2, \\ 1 - \frac{u_1 - u_m}{z u_1 + u_m} & \text{otherwise.} \end{cases}$$

Proof. We assume two groups with n_1 and n_2 agents with $n_1 \geq n_2$. We start the proof for positional scoring rules with an intermediate result. For a preference profile, given a positional scoring rule with score vector $\vec{s}$, let S_a be the average score for alternative $a \in \mathcal{A}$, i.e., $S_a = \frac{1}{n} \sum_{j=1}^{n} S_j(a)$, where $S_j(a) = s_k$ if a was ranked k-th by agent j. Now, call $S = [S_{a_1} \ldots, S_{a_m}]$ the average score vector. Then we have the following result.

Claim 1. *For a positional scoring rule with score vector $\vec{s}$, as $n \to \infty$, we will always have a preference profile $P \in \mathcal{L}^n$ with average score vector arbitrarily close to any vector $\langle s'_1, \ldots, s'_m \rangle$ as long as $\sum_{i=1}^{m} s'_i = \sum_{i=1}^{m} s_i$ and $\forall i$, $s_1 \geq s'_i \geq s_m$.*

Proof. For the set of alternatives $\mathcal{A}$, there are $m!$ rankings $R \in \mathcal{L}$. Assume that there are n_k agents whose ranking is R_k. So, $\sum_{k=1}^{m!} n_k = n$. For each ranking R_k, the score that each alternative gets is a permutation of $\langle s_1, \ldots, s_m \rangle$. We denote the permuted score vector associated with ranking R_k as $\vec{\pi}_k$. E.g., for $R_1 = a_1 \succ \cdots \succ a_m$, $\vec{\pi}_1 = \langle s_1, \ldots, s_m \rangle$, whereas for $R_{m!} = a_m \succ \cdots \succ a_1$, $\vec{\pi}_{m!} = \langle s_m, \ldots, s_1 \rangle$. Thus, the $m!$ rankings can be mapped to $m!$ such ordering of scores. If we call the ordered score vector associated with order R_k as $\vec{\pi}_k$, then we can say that

$$\begin{bmatrix} | & \cdots & | \\ \pi_i & \ddots & \pi_{m!} \\ | & \cdots & | \end{bmatrix} \begin{bmatrix} n_1/n \\ \vdots \\ n_{m!}/n \end{bmatrix} = [S_{a_1} \ldots S_{a_m}]$$

We assume

$$\Pi = \begin{bmatrix} | & \cdots & | \\ \pi_i & \ddots & \pi_{m!} \\ | & \cdots & | \end{bmatrix}$$

Each column of the matrix Π is a permutation of the score vector, $\vec{s}$. This matrix is trivially always of rank m. Since S_{a_i} for any i is the average score it gets from all alternatives, we

know that $s_m \leq S_{a_i} \leq s_1$ for any a_i. Since the rank of matrix Π is m, for any valid average score vector S (i.e., if $\sum_{i=1}^{m} S_{a_i} = \sum_{i=1}^{m} s_i$, and $s_1 \geq S_i \geq s_m$ for all i), we will find real solutions for $n_1/n, \dots n_{m!}/n$. When $n \to \infty$, we can ensure that they lead to integer values for each n_k, which means that for any score vector S, there exist preference profiles such that the score values get arbitrarily close to those in S. This concludes the proof. □

This property for existence of an appropriate preference profile will be used to complete the proof of the theorem. First, we give a generic bound, which is the second part of the theorem. In the proof of Remark 1, we showed how extreme values of group utilities lead to worst imbalance, and thus worst fairness values. Now, for positional scoring rules, we consider what extremities may occur. The winning alternative has to be more preferred by either group. So, we have two cases.[3]

The first case is where the winner is most preferred in preference profile P_1. In that case, for the winner a, $W(a, P_1) > W(a, P_2)$. Thus the imbalance becomes $n \cdot \frac{W(a,P_1)-W(a,P_2)}{n_1 W(a,P_1)+n_2 W(a,P_2)}$. From similar arguments to what we made in the proof for Remark 1, we can say that the worse case occurs when $W(a, P_1) = u_1$, $W(a, P_2) = u_m$. The preference profile that has a at top of all rankings in P_1 and at the bottom of all rankings in P_2 will achieve that if every other alternatives are placed randomly. This will occur because $s_1 \geq \frac{s_2+\dots+s_{m-1}}{m-2}$ for all positional scoring rules. Thus, this gives a bound for all scoring rules. The imbalance in this case becomes $\frac{n(u_1-u_m)}{n_1 u_1+n_2 u_m}$. Thus, the fairness value is $1 - \frac{u_1-u_m}{\frac{n_1}{n_2}u_1+u_m}$. This gives the bound for the $z > \dfrac{s_1 - s_m}{s_1 - s_{[2:m]}}$ case (the otherwise case) of the theorem.

The second case is where $W(a, P_2) > W(a, P_1)$. For this part, w.l.o.g., assume that $W(a_1, P_1) \geq W(a_2, P_1) \geq \dots \geq W(a_m, P_1)$. The winner a is such that $W(a, P_2) > W(a, P_1)$. High imbalance would occur in the case where a is highly scored in the rankings in P_2 and lowly scored in P_1. First, similar to our definition of average score of an alternative S_a, let $S(a_a, P_1)$ be the average score for alternative a in preference profile P_1, and similarly denote $S(a_a, P_2)$.

The lowest scored alternative in preference profile, P_1, is a_m. However, it trivially follows that when $n_1 > n_2$, a_m can never be the winner for any scoring rule if $W(a_m, P_1) = u_m$. So, we try to find what is the lowest ranked a_k that can be the winner. Note that a_{m-1} will be a winner if $n_1 S(a_1, P_1) + n_2 S(a_1, P_2) \leq n_2 S(a_{m-1}, P_1) + n_2 S(a_{m-1}, P_2)$. Additionally, the average score values of $S(a_1, P_1) = s_1, S(a_1, P_2) = s_m, S(a_{m-1}, P_2) = s_1$ lead to maximum extremities. So, we get the following condition:

$$\frac{n_1}{n_2} \leq \frac{s_1 - s_m}{s_1 - S(a_{m-1}, P_1)}.$$

Now, $S(a_{m-1}, P_1)$ is the score for the second lowest scored alternative in a preference profile. Dependent on $\vec{s}$, we may get different values of how low this can be. But one value that we can always get is if $S(a_{m-1}, P_1) = S(a_m, P_1) = \frac{s_{m-1}+s_m}{2}$. This will happen when the last two alternatives are tied, and because of previously proven property of existence of preference

3. We may have an alternative most preferred by both groups, but that does not cause worst-case imbalance, so we ignore that case.

profiles with arbitrarily close score vectors, we can ensure such a preference profile exists. This analysis led to the case for $k = m - 1$ in the first part of the theorem.[4]

So, when $z \in \left[1, \frac{s_1 - s_m}{s_1 - \frac{s_{m-1}+s_m}{2}}\right]$, then

$$F(r_s, u, n, z) \leq \min\left(1 - \frac{u_1 - u_m}{zu_1 + u_m}, 1 - \frac{u_1 - \frac{u_{m-1}+u_m}{2}}{z \cdot \frac{u_{m-1}+u_m}{2} + u_1}\right).$$

When a_{m-1} cannot be a possible winner (when $\frac{n_1}{n_2}$ does not fall into this interval) an earlier alternative may be the lowest ranked winner. For any alternative a_k, we can generalize this analysis. we will again assume that $S(a_1, P_1) = s_1, S(a_1, P_2) = s_m, S(a_k, P_2) = s_1$. Again, a_k is guaranteed to have a worst-case $S(a_k, P_1)$ of $\frac{s_k+\cdots+s_m}{m-k+1}$ in the case that the last $m-k+1$ alternatives are tied. Now, if $S(a_k, P_1) = \frac{s_k+\cdots+s_m}{m-k+1}$, then obviously $W(a_k, P_1) = \frac{u_k+\cdots+u_m}{m-k+1}$ for the utility. Thus, we would have $W(a_1, P_1) = u_1, W(a_1, P_2) = u_m, W(a_k, P_2) = u_1$. Using these values to calculate group imbalance fairness results in the reported fairness bound. A minimum is taken over this value and the generic bound for positional scoring rules that we got in the earlier part of the proof to complete the proof of the theorem.

For the extension to general forms of fairness for more than two groups, we consider both definitions of fairness for more groups. For each definition, we get a different group ratio parameter z, dependent on the sizes of each group. The imbalance for a full preference profile is the worst case over all of the pairwise imbalance values calculated for each group. So, the worst-case result would come for the most imbalanced group size. Since that is determined by group ratio z, we see that the result holds directly for both our extensions for group fairness for higher number of groups. This extension will be the same for all our theorems for fairness results, so for the following theorems, we will only discuss the two-group scenarios in the proof.

□

Theorem 3 provides a quantitative formula for an upper bound to the fairness level under a fixed utility function. The bounds in Theorem 3 are not tight, and the actual fairness value can be lower. We give an example of this in Table 3. This theorem is however still important, because it provides a general result, that works for any positional scoring rule, under any utility function.

	u_{top}	u_{rank}
Plurality	$1 - 1/z$	$1 - 1/z$
Borda (m = 4)	0 for $z \in [1, 3/2]$ $1 - 1/z$, otherwise	$1 - 1/z$

Table 3: Upper bounds for $F(r, u, n, z)$ for Plurality and Borda under u_{top} and u_{rank}, according to Theorem 3. Bounds for Borda is dependent on m, and the table shows the result for $m = 4$.

Table 3 shows the fairness upper bounds for Plurality and Borda for the top-1 and rank utility functions, as provided by Theorem 3. While for Plurality we get the same bound

4. When $n \to \infty$ is not true, the exact low value given by this bound may not be attainable because of only discrete values of scores being possible. So, in those cases the upper bound for fairness may be slightly higher, and dependent on the exact values of n_1 and n_2.

for any number of alternatives m, for Borda, the bound depends on m. As an example, we show the bounds for $m = 4$ for Borda. We can compare these bounds to the tight fairness bounds provided by Theorems 1 and 2 (See Table 2). For Plurality, we see that for top-1 utility, the bound is equal to the tight one $(1 - 1/z)$ when $z > m - 1$, while is higher than the tight bound for u_{rank}. On the other hand, for Borda, the bound is tight only when $z \in [1, 3/2]$ for top-1 utility. However, we get the tight bound for rank utility.

5.2 Condorcet Rules

Next, we present a fairness result that is true for all Condorcet rules.

Theorem 4 (**Fairness for Condorcet Rules**). *Given utility function u, n total agents and group size ratio, z, for any Condorcet rule*

$$F(r, u, n, z) \leq \min\left(1 - \frac{u_1 - u_m}{zu_1 + u_m}, 1 - \frac{u_1 - u_{m-k^*}}{zu_{m-k^*} + u_1}\right)$$

where $k^* = \lceil \frac{m-1}{2}(1 - \frac{1}{z}) \rceil$.

Proof. We assume two groups with n_1 and n_2 agents with $n_1 \geq n_2$.

Now, the first element in the min function is the same generic bound that we found for positional scoring rules. Whenever everyone from the majority group will have an alternative at top, that alternative must be the Condorcet winner (beats all other alternatives with at least $n_1 \geq n/2$ votes). The result immediately follows from this case.

The second part is from when we have $W(a, P_2) > W(a, P_1)$ for the winner a. It makes sense in that case that $W(a, P_2) = u_1$, because that will lead to high imbalance. So, a gets n_2 pairwise preferences over all other alternatives from P_2. To be Condorcet winner, a requires at least $\frac{n}{2} - n_2$ pairwise preferences over each of the other $m - 1$ alternatives in P_1. Since there are n_1 agents in P_1, this implies that a should have at least $\frac{(\frac{n}{2} - n_2)(m-1)}{n_1} = \frac{m-1}{2}(1 - \frac{1}{z})$ pairwise preferences over each other alternatives for every agent in $P - 1$. Let $k^* = \lceil \frac{m-1}{2}(1 - \frac{1}{z}) \rceil$. If a is ranked at position $m - k^*$ for all rankings in P_1, and all other alternatives are ranked uniformly randomly, a gets the necessary amount of pairwise preferences over each alternative. Considering all other alternatives to be ranked uniformly randomly in both P_1, P_2 makes sure that a beats every other alternative in pairwise comparison as is the unique Condorcet winner. This is also the least k^* for such preference profiles, where we have this guarantee. Computing the imbalance for this scenario gives the second term inside the min function in the theorem. □

Theorem 4 indicates how low the fairness values may be for any Condorcet rule. For specific Condorcet voting rules, e.g., Copeland, the fairness value can be even worse than the general result. For example, if the utility function is u_{top} with only the top ranked alternative giving utility, then the fairness value for Copeland is 0.

Consider for example two groups with preference profiles P_1 and P_2, with group 1 being the majority group, and assume that every agent in P_2 ranks alternative a at the top, while every agent in P_1 ranks a second. For the rest of the $m - 1$ alternatives, assume that the rankings are evenly distributed for both P_1 and P_2. That is, all of the $(m - 1)!$ possible rankings for the rest of the alternatives are found equally when constrained to just those

alternatives. This preference profile will lead to a being the winner, with $W(a, P_1) = 0$ and $W(a, P_2) = 1$. This leads to the fairness value being 0 for Copeland under u_{top}.

5.3 Fair Voting Rules

Both scoring rules and Condorcet rules maximize different efficiency measures. Based on our definition of imbalance, we can define a family of fairness maximizing voting rules. The imbalance values are calculated using specific utility functions u. To create a high-fairness voting rule, we define a family of voting rules as below.

Definition 3 (u-fair voting rules)**.** For utility function u, the u-fair voting rule, $r^{\mathrm{u}}_{\mathrm{fair}}$, is a voting rule that chooses the alternative with minimum imbalance with respect to utility function u for any preference profile.

Remark 2. For a fixed utility function u, for n agents, group size ratio z, the u-fair voting rule ($r^{\mathrm{u}}_{\mathrm{fair}}$) has maximum fairness value $F(r, u, n, z)$ out of all voting rules.

Thus, knowing the utility function allows us to calculate fairness bounds for specific voting rules. Assume the utility function defined by $\vec{u}_{top} = \langle 1, 0 \ldots, 0\rangle$ where only the top alternative gives utility when chosen. For this, the u_{top}-fair voting rule would be of interest. Similarly, for $\vec{u}_{rank} = \langle m-1, m-2, \ldots, 0\rangle$, u_{rank}-fair voting rule would be of interest.

Theorem 5 (**Fairness for u-fair Voting Rules**)**.** *Given n total agents, and group size ratio z,*

$$F(r^{top}_{fair}, u_{top}, n, z) = 1 - \frac{1}{z},$$
$$F(r^{rank}_{fair}, u_{rank}, n, z) = 1 - \frac{1}{z}.$$

Proof. We will give a constructive proof for the fairness value for the u_{top}-fair voting rule, assuming two groups with preference profiles P_1 and P_2 and group size n_1 and n_2 respectively. W.l.o.g., assume that $n_1 \geq n_2$ and that all agents in P_1 ranks a_1 at top, while all agents in P_2 ranks a_2 at top. This gives $\mathrm{Imb}(u_{top}, a_1, P_1, P_2) = \frac{n}{n_1}$ and $\mathrm{Imb}(u_{top}, a_1, P_1, P_2) = \frac{n}{n_2}$. Since $n_1 \geq n_2$, if a_1 and a_2 are the only alternatives, the winner $r^{\mathrm{top}}_{\mathrm{fair}}(P_1, P_2) = a_1$. For $m > 2$ alternatives, for any preference profiles in $\mathcal{L}^n_1, \mathcal{L}^n_2$, there exists at least one alternative a, such that $W(a, P_1) > W(a, P_2)$. For preference profiles where no a exists such that $W(a, P_1) = W(a, P_2)$, one of the alternatives favored in P_1 will be the u_{top}-fair winner. Any such alternative has a maximum imbalance bound of $\frac{n}{n_1}$ as indicated by Lemma 1, which is the same as the bound above for $m = 2$. This proves the fairness result for u_{top}-fair voting rule.

For the u_{rank}-fair voting rule, the utility function is u_{rank}. By definition, Borda's score vector is the same as the utility vector for u_{rank}, thus Borda will be the social welfare maximizer. With that in mind, we first calculate the fairness value for Borda under u_{rank} Assume that the Borda winner is a. case 1: $W(a, P_1) > W(a, P_2)$, which gives the worst-case imbalance of $\frac{n}{n_1}$, which is what we would get from Theorem 3. case 2: $W(a, P_2) > W(a, P_1)$. Since it maximizes the Borda score, the least average score of the Borda winner needs to be $\frac{1}{m}((m-1)+\cdots+1) = \frac{m-1}{2}$. Even if all rankings in P_2 has a on top, $W(a, P_2) = n_2(m-1) \leq$

$\frac{n(m-1)}{2}$. Thus, $W_{rank}(a, P_1) \geq \frac{1}{n_1}(\frac{n(m-1)}{2} - n_2(m-1)) = (m-1)\frac{n_1-n_2}{2n_1}$. This gives, in the worst case,

$$\text{Imb}(u_{rank}, a, P_1, P_2) = \frac{(m-1) - (m-1)\frac{n_1-n_2}{2n_1}}{\frac{m-1}{2}} = \frac{1 - \frac{n_1-n_2}{2n_1}}{\frac{1}{2}} = \frac{n_1+n_2}{n_1} = \frac{n}{n_1}.$$

This is the same bound as in case 1.

Now, we show an instance where the u_{rank}-fair rule winner that has imbalance of $\frac{n}{n_1}$. Consider the case where $m = 2$, all rankings in P_1 has $a_1 \succ a_2$ and all rankings in P_2 has $a_2 \succ a_1$. Here, $r^{\text{rank}}_{\text{fair}}(P_1, P_2) = a_1$ and imbalance is $\frac{n}{n_1}$. This is the same value as the worst-case imbalance for Borda under this utility function. The worst-case for u_{rank}-fair cannot be worse than that of Borda according to Remark 2. Thus we have found the worst-case imbalance for the u_{rank}-fair voting rule. □

Comparing the fairness results for the u-fair voting rules to those for positional scoring rules, we notice the following. From Theorem 1, we see that the fairness value for Plurality is $1-1/z$ when $z > m-1$ and we use the top-1 utility. This is the same as the fairness value for u_{top}-fair for top-1 utility, and hence optimal. Similarly, from Theorem 2, we see that the fairness value for Borda is the same as that of u_{rank}-fair for rank utility. Plurality and Borda are respectively the social welfare maximizer for the top-1 and rank utility functions, since they share the same vectors as utility vector and score vector. So, for many z values, the social welfare maximizer is also optimal in terms of fairness. Whether this is a property of these two specific utility functions or it generalizes is an interesting question for the future. However, this also means that just looking at worst-case fairness does not provide a complete picture, as this indicates that even the fairest voting rule is sometimes only as fair as the most efficient voting rule. This motivates us to also consider average-case fairness, and to study whether we can design voting rules that are fairer in average. We will discuss average-case fairness and the trade-off between fairness and efficiency in depth in Sections 6 and 8.

5.4 General Result

These worst-case bound for u-fair rules give us a reference point as to how unfair other rules can be compared to the most fair rule. Again, referring to Theorems 1–4, we see that the traditional voting rules can be very unfair, leading to 0 fairness values in some cases. This negative result can be summarized in the following theorem, which indicates that for any positional scoring rule or Condorcet voting rule, there exists utility functions such that the fairness value is 0.

Theorem 6. *If r is any positional scoring rule or Condorcet voting rule, there exist some utility function $u \in \mathcal{U}$, and some group size parameters, n and z, such that*

$$\min_{\substack{u \in \mathcal{U} \\ n \in \mathbb{N}, z \geq 1}} F(r, u, n, z) = 0.$$

Proof. For each case, we give an example of u, n, z for which $F(r, u, n, z)$ becomes 0.

For any positional scoring rule, assume $z \in [1, \frac{s_1 - s_m}{s_1 - \frac{s_{m-1}+s_m}{2}}]$, now, if $u_{m-1} = u_m = 0$, $F(r, u, n, z) \leq \min(1 - \frac{1}{z}, 0) = 0$.

For any Condorcet rule, for all $k \geq m - k^*$, take $u_k = 0$, where k^* is defined as in Theorem 4. For any preference profile where the Condorcet winner (assume, a) exists, such a utility function will result in $\text{Imb}(u, a, P_1, P_2) = 1 + z$. Which implies that $F(r, u, n, z) = 0$ for any Condorcet voting rule. □

As we note from the theorems and the examples presented in Table 2 for two groups, none of the considered voting rules guarantee non-zero group imbalance-based fairness for arbitrary utility functions. Theorem 6 gives a somewhat negative result, in that all traditional efficiency-maximizing voting rules may turn out to be unfair under some circumstance. Hence, we also want to learn about general imbalance-based fairness of voting rules instead of worst-case only. The empirical analysis regarding both traditional and u-fair voting rules is in Section 8.

6. Designing Fair and Efficient Voting Rules

The theoretical results in the previous section are about how different voting rules perform in the worst case in terms of fairness. While traditional voting rule focus on maximizing some efficiency measure, our proposed u-fair rules maximize group imbalance-based fairness for different utility function assumptions. We showed via an example that traditional voting rules can be more unfair. u-fair rules, however, can lead to poor efficiency (in terms of either social welfare or Condorcet efficiency), because they do not consider maximization of efficiency as an objective. As we will see in our empirical experiments in Section 8, in the average case there is a large difference between the most fair and most efficient voting rules in terms of fairness and efficiency.

In this section, we want to find a compromise between the two; for that we propose two frameworks to design new voting rules with various efficiency and fairness values.

6.1 Framework 1: Utility-constrained Fair Voting Rules

Based on an assumed utility function, u, we can define constrained fair voting rules as a compromise between positional scoring rules and u-fair rules. Note that we are not assuming access to ground truth utility function of agents, but rather assuming a utility function here, similar to how we defined the u-fair voting rules in Definition 3.

Definition 4 (α-efficient fair Borda (α-FB))**.** Given preference profiles P_1, P_2 for two groups of agents, $\alpha \in [0, 1]$, the α-efficient Fair Borda winner, $r_{\alpha-FB}(P_1, P_2)$ is given by

$$\begin{aligned} &\underset{a \in \mathcal{A}}{\text{minimize}} \quad \text{Imb}(u_{\text{rank}}, a, P_1, P_2) \\ &\text{subject to} \quad W_{\text{rank}}(a, P) \geq \alpha \cdot \max_{a' \in \mathcal{A}} W_{\text{rank}}(a', P). \end{aligned} \tag{1}$$

Thus, for different utility functions, we can design various new voting rules with different α parameters. Here α is a measure of how economically efficient we require the voting rule to be, and varying α should give different levels of fairness and efficiency. When $\alpha = 1$, it is the same as Borda, without any fairness requirements. When $\alpha = 0$, it is the u_{rank}-max-fair

rule. We also define the constrained fair rule in terms of top-1 utility, which is a fairness constrained version of Plurality, as below.

Definition 5 (α-efficient fair Plurality)**.** Given preference profiles P_1, P_2, $\alpha \in [0,1]$ the α-efficient Fair Plurality winner, $f^{top}_{\alpha-FP}(P_1, P_2)$ is given by

$$\begin{aligned} &\underset{a\in\mathcal{A}}{\text{minimize}} \quad \text{Imb}(u_{top}, a, P_1, P_2) \\ &\text{subject to} \quad W_{top}(a, P) \geq \alpha \cdot \max_{a_k\in\mathcal{A}} W_{top}(a_k, P). \end{aligned} \tag{2}$$

6.2 Framework 2: ML-based Framework for Fair-efficient Voting Rules

We also explore the other direction of designing a fair-efficient voting rule using machine learning. Framework 1 is particularly helpful when we have a utility function that we can make use of and the economic efficiency of interest is social welfare. However, if the economic efficiency is measured in terms of Condorcet efficiency, and we want to design a voting rule that explores the trade-off between Condorcet efficiency and fairness, we cannot use such a constrained method easily. Here, the learning framework is much more useful.

We note that a voting rule r can be viewed as a multi-class classifier: the input is a preference profile P and the classes are the alternatives in $\mathcal{A}$. From this viewpoint, we propose a learning framework that generates synthetic data with random preference profiles. As feature vectors, we can use summary features for preference profiles like weighted majority graph or positional score matrix.

Also, to learn such classifiers, we need proper features to represent voting rules. In particular, since a preference profile with a large number of agents will have a very high dimension, we choose to represent preference profiles using some summary features, that explain the profiles to a large extent. And for many voting rules, the winner can be determined directly using these features without access to the preference profile themselves. We define two such features here, that we will use for our learning methods.

Definition 6 (Positional score matrix)**.** Define positional score matrix, $S_{m\times m}$ such that $S_{j\ell}$ is the number of agents who has ranked alternative a_i at the ℓ-th position

Definition 7 (Weighted majority graph (WMG))**.** Given any preference profile P, the WMG is a directed graph where the nodes are all the alternatives, there exists edges in both directions for each pair of alternatives and the weight for edge (a_i, a_ℓ) is $D(a_i, a_\ell) = N_{(a_i \succ a_\ell)} - N_{(a_\ell \succ a_i)}$, where $N_{(a_i \succ a_\ell)}$ is the number of agents who prefer a_i over a_ℓ in their ranking.

To create the training data, the high level idea is that we want to create a mixed synthetic dataset, where part of the labels come from fair winners and the rest of the labels come from efficient winners. We can choose our notion of efficiency for training this. In particular, this method is more useful for efficiency measures like Condorcet efficiency. Because in the previous framework, we directly depended on our ability to compute utilities and thus, social welfare from a preference profile. Since Condorcet efficiency is not defined on a single preference profile, we cannot take a similar approach for that. Hence, the ML-based approach would be more useful.

We present two ML-based methods. The first one is β-Mix , where $\beta \in [0, 1]$ is a mixing parameter that determines how economically efficient the learned voting rule should be. The learning method is defined in Algorithm 1 for the version where the efficiency measure is Condorcet efficiency. The sampling method (see line 4 of Algorithm 1) is separately presented in Algorithm 2. Note that while Algorithms 1 and 2 are presented for the version with Condorcet efficiency, it can be modified slightly to work for any economic efficiency measure. For example, if the economic efficiency measure is ranked utility, checking if a Condorcet winner exists will be unnecessary, and instead of picking the Condorcet winner, we will choose the Borda winner.

The second method is β-Soft, which uses a soft-labeling method where $\beta \in [0, 1]$ is now a weight parameter. For every preference profile, we add a sample of weight β for the efficient alternative, and add a sample of weight $(1 - \beta)$ for the fair alternative. This method is summarized in Algorithm 3.

Algorithm 1 Learning framework with sample mixing: β-mix.

1: **Inputs:** Mixing parameter β and learning algorithm F.
2: Generate feature set of random preference profiles $\mathcal{P}$
3: For each $P \in \mathcal{P}$, compute fair winner and Condorcet winner (if exists)
4: β-sampling labels from fair and Condorcet winners to get label set Y, with mixture of Condorcet and fair data (Algorithm 2)
5: Train multi-class classifier $\mathcal{P}, Y$ using learning algorithm F to learn model H
6: **Output:** Learned voting rule, H

Algorithm 2 Data Set Generation with β-sampling

1: **Inputs:** Mixture parameter β, fair voting rule
2: **for** $i \leftarrow 1$ to T **do**
3: Sample P uniformly
4: Compute winner F for given fair voting rule
5: **if** Condorcet winner exists **then**
6: Compute Condorcet winner, C
7: **if** $C = F$ **then**
8: $w \leftarrow C$
9: **else**
10: with probability β, $w \leftarrow C$
11: with probability $1 - \beta$, $w \leftarrow F$
12: **end if**
13: **else**
14: $w \leftarrow F$
15: **end if**
16: Append (P, w) to data set
17: **end for**
18: **Output:** Training data $\{(P, w)\}$ for designing new rule

Algorithm 3 Learning framework with soft labeling: β-soft.

1: **Inputs:** Mixing parameter β and learning algorithm F.
2: Generate feature set of random preference profiles $\mathcal{P}$
3: For each $P \in \mathcal{P}$, compute fair winner and Condorcet winner (if exists) (or efficient winner)
4: Add two samples, P with efficient winner, P with fair winner with weights β and $1-\beta$ respectively.
5: Train multi-class classifier $\mathcal{P}, Y$ using learning algorithm F to learn model H
6: **Output:** Learned voting rule, H

Instantiations of both frameworks have been developed and extensively experimented on using synthetic datasets. We share our results in Section 8.

7. Adding Privacy to the Collective Decision-Making Framework

In this section, we consider how to add privacy to voting rules. As mentioned before, we adopt local differential privacy (DP) as our privacy measure as it protects even against attacks on an aggregator. Local DP requires the messages sent from every agent to be differentially private, so we add noise locally and send the noisy vote to the aggregator. Then, the aggregator applies a voting rule. For example, Figure 2 shows how our framework for utility-constrained voting rules can be modified to also be local diferentially private.

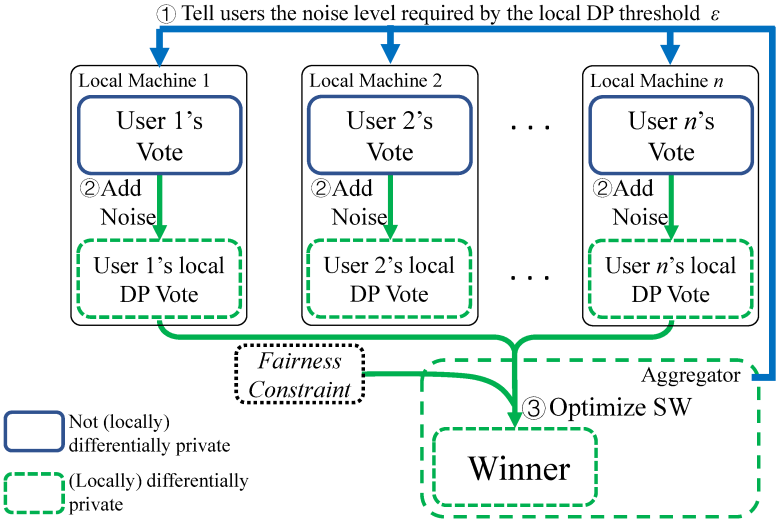


Figure 2: The structure of our collective decision framework with the property of local DP. Label ① - ③ shows the order between steps.

Next, we introduce flipping-coin algorithm, which adds noise to votes. In general, flipping-coin algorithm $f_p(x)$ is a two-step approach:

1. Output the input x with probability p.
2. Otherwise, output uniformly at random from $\mathcal{X}$, which is the support of f_p.

The support set $\mathcal{X}$ contains all the possible votes a single agent can give as input. When collective decision take agents' full ranking as inputs, we have $|\mathcal{X}| = m!$.

We first prove some initial results and definitions for local DP in terms of voting.

Lemma 2. *Flipping coin algorithm with* $p = \frac{\exp(\varepsilon)-1}{|\mathcal{X}|+\exp(\varepsilon)-1}$ *can provide* ε*-local DP.*

Proof. By $f_p(\cdot)$'s definition, for any $x' \neq x$ and $x' \in \mathcal{X}$,

$$\Pr[f_p(x) = x] = \frac{\exp(\varepsilon)}{|\mathcal{X}| + \exp(\varepsilon) - 1} \quad \text{and} \quad \Pr[f_p(x) = x'] = \frac{1}{|\mathcal{X}| + \exp(\varepsilon) - 1}.$$

Then, Lemma 2 follows by the definition of local DP. □

As shown above, better privacy can be achieved when $|\mathcal{X}|$ is smaller. Actually, the aggregator does not always require all information from full rankings. For example, top-1 utility only requires the top alternative to calculate average utility and imbalance. Similarly, Plurality only requires the top alternative to calculate the winner. Thus, users need to only upload the top alternative if using the social welfare maximizing rule for top-1 utility, which can make $|\mathcal{X}| = m$. Following this idea, we revise our protocol by only uploading the needed information to calculate winner, average utilities and imbalance.

Next, we introduce our method to recover group (and average) utility and fairness from noisy votes. To simplify notation, we let $f_p(P)$ to denote a profile where flipping-coin algorithm is applied to all votes in P.

We use $\hat{}$ on top of one value to denote our estimate from the noisy profile. The values without $\hat{}$ denote the ground truth value computed from P. Our estimators are formally defined as below:

- $\hat{W}(a, P) = \frac{W(a, f_p(P))}{p} - \frac{1-p}{p} \cdot \frac{||\vec{u}||_1}{m}$,
- $\hat{\text{Imb}}(u, a, P_1, P_2) = \frac{|\hat{W}(a,P_1) - \hat{W}(a,P_2)|}{\hat{W}(a,P)}$,

where $\vec{u}$ defines the utility function. The estimators are unbiased estimators for the ground truth, i.e., $\mathbb{E}\left[\hat{W}(a, P)\right] = W(a, P)$, as stated in Lemma 3.

Lemma 3 (Expectation for Estimators)**.** *Using the notations above, we have,*

$$\mathbb{E}\left[\hat{W}(a, P)\right] = W(a, P).$$

Proof. Given that group utility is the mean of all agents' utilities, we only need to check the expected utility for one vote. In this proof, we use R_j to denote the j-th vote in profile P, and we have,

$$\mathbb{E}[W(a, f_p(R_j))] = \sum_{R \in \mathcal{L}} \Pr[f_p(R_j) = R] \cdot W(a, R).$$

By the definition of $f_p(\cdot)$, we know there is probability p to report P_k truthfully and probability $1-p$ for reporting uniformly at random. Thus, we have, $\mathbb{E}[W(a, f_p(R_j))] = p \cdot W(a, R_j) +$

$\sum_{R\in\mathcal{L}} \frac{1-p}{m!} \cdot W(a,R)$. By the definition of $W(a,R)$, we have, $\sum_{R\in\mathcal{L}} W(a,R) = \frac{||\vec{u}||_1 \cdot m!}{m}$. Combining these two relations, we have,

$$\mathbb{E}[W(a, f_p(R_j))] = p \cdot W(a,R_j) + (1-p) \cdot \frac{||\vec{u}||_1}{m}. \tag{3}$$

Using the definition of $\hat{W}(a,P)$, we get

$$\mathbb{E}[\hat{W}(a,P)] = \frac{\mathbb{E}[W(a,f_p(P))]}{p} - \frac{1-p}{p} \cdot \frac{||\vec{u}||_1}{m}$$

Calculating $\mathbb{E}[[W(a,f_p(P))]$ using Equation 3, the result in Lemma 3 directly follows from here. □

Our framework solves a similar optimization problem as in Problem 1, but instead of using Imb and W, we would use $\hat{\text{Imb}}$ and $\hat{W}$. Thus, the complete input to our framework are: group preference profiles P_1, P_2, privacy requirement ε, the inferred noisy profile $f_p(P)$ and a threshold α for utility.

In the next theorem, we show how privacy threshold ε influences our maximized efficiency and maximized fairness. This result also shows the role of privacy in fairness-privacy-efficiency trade-off. For this analysis, the efficiency measure that we consider is average utility or social welfare, as opposed to Condorcet efficiency. Since the results are about average utility and imbalance rather than fairness, we present all results in term of two groups of size n_1, n_2.

Theorem 7 (Fairness-Privacy-utility Trade-off)**.** *For any ε-local DP requirement on making collective decisions with two groups, we have the following:*

$$(1)\ \Pr\left[\hat{W}(a,P) \geq W(a,P) - t\right] \geq 1 - \exp\left[-\frac{2t^2p^2n}{(\Delta u_{\max})^2}\right],$$

$$(2)\ \Pr\left[\hat{Imb}(u,a,P_1,P_2) \leq Imb(u,a,P_1,P_2) + \frac{\Delta u_{\max}}{p} \cdot \frac{(Imb(u,a,P_1,P_2)+1)(n_1^{-0.3}+n_2^{-0.3})}{W(a,P) - \frac{\Delta u_{\max}}{p}\left(n_1^{-0.3}+n_2^{-0.3}\right)}\right] \geq 1 - 2\exp\left(-2n_1^{0.4}\right) - 2\exp\left(-2n_2^{0.4}\right),$$

where $p = \frac{\exp(\varepsilon)-1}{|\mathcal{X}|+\exp(\varepsilon)-1}$ *and* $\Delta u_{\max} \triangleq \max_{i,j} |u_i - u_j| = u_1 - u_m$.

Proof. To simplify notations, we let $W_k = W(a,P_k)$ and $W = W(a,P)$. Similarly, let $\hat{W}_k = \hat{W}(a,P_k)$ and $\hat{W} = \hat{W}(a,P)$. First, we apply the Hoeffding bound (Hoeffding, 1963) to get an upper bound on the probability of how large $W(a,f_p(P))$ can become. Given independent random variables $X_i, \cdots, X_n$ such that $c_i \leq X_i \leq d_i$ with $S_n = \sum_{i=1}^n X_n$. Then the Hoeffding bound theorem states that for all $t \geq 0$,

$$\Pr[S_n \geq \mathbb{E}[S_n] + t] \leq \exp\left[-\frac{2t^2}{\sum_{i=1}^n (d_i - c_i)^2}\right], \text{and}$$
$$\Pr[S_n \leq \mathbb{E}[S_n] - t] \leq \exp\left[-\frac{2t^2}{\sum_{i=1}^n (d_i - c_i)^2}\right].$$

The noise from different agents are added independently. We also note that $W(a, P_1), W(a, P_2)$ and $W(a, P)$ are all average utilities. So, $u_1 \geq W(a, \cdot) \geq u_m$, since u_1 and u_m are respectively the highest and lowest possible utility values. For a preference profile P and alternative a, let $X_i = \frac{W(a,R_i)}{|P|}$ for agent i's vote. Here $|P|$ is the size of the preference profile, P. Thus, applying the Hoeffding bound theorem, we have,

$$\Pr\left[W(a, f_p(P)) \geq \mathbb{E}[W(a, f_p(P))] + t\right] \leq \exp\left[-\frac{2t^2|P|}{(\Delta u_{\max})^2}\right].$$

Here, for the preference profile of all agents, $|P| = n$, and then $|P_1| = n_1$ and $|P_2| = n_2$ respectively for two groups. Then, using the definition of $\hat{W}(a, P)$ and the relation between $\mathbb{E}[\hat{W}(a, P)]$ and $W(a, f_p(P))$ from Lemma 3, we have,

$$\Pr\left[\hat{W}(a, P) \geq W(a, P) + t\right] \leq \exp\left[-\frac{2t^2p^2|P|}{(\Delta u_{\max})^2}\right]. \tag{4}$$

The p^2 factor in $\exp[\cdot]$ comes from the $\frac{1}{p}$ factor in the definition of $\hat{W}(a, p)$. Part (1) of the theorem directly follows from here.

Next, we prove the concentration bound to our imbalance estimator. Inequality (4) gives individual bounds for $\hat{W}_1$ and $\hat{W}_2$. By letting $t = \frac{\Delta u_{\max}}{p} n_i^{-0.3}$ for group i in Inequality 4, we get,

$$\Pr\left[\left|\hat{W}_1 - W_1\right| \geq \frac{(\Delta u_{\max})}{p} n_1^{-0.3}\right] \leq 2\exp(-2n_1^{0.4}) \quad \text{and}$$
$$\Pr\left[\left|\hat{W}_2 - W_2\right| \geq \frac{(\Delta u_{\max})}{p} n_1^{-0.3}\right] \leq 2\exp(-2n_2^{0.4}).$$

Both inequalities can be proved in a similar way as (4) using both branches of the Hoeffding bound. From the union bound (Boole, 1847), for any two events A and B, $\Pr[A \cup B] \leq \Pr[A] + \Pr[B]$. Then, applying the union bound on the bounds for each group's utility, we have,

$$\begin{aligned}\Pr\left[\left(\left|\hat{W}_1 - W_1\right| \geq \frac{(\Delta u_{\max})}{p} n_1^{-0.3}\right) \bigcup \left(\left|\hat{W}_2 - W_2\right| \geq \frac{(\Delta u_{\max})}{p} n_2^{-0.3}\right)\right] \\ \leq\ 2\exp(-2n_1^{0.4}) + 2\exp(-2n_2^{0.4}).\end{aligned} \tag{5}$$

In the next (technical) Lemma, we connect $\hat{\text{Imb}}(u, a, P_1, P_2)$ with the above probability.

Lemma 4. *When*

$$\hat{Imb}(u, a, P_1, P_2) \geq \frac{|W_1 - W_2| + \frac{\Delta u_{\max}}{p}(n_1^{-0.3} + n_2^{-0.3})}{W - \frac{\Delta u_{\max}}{p} \cdot \frac{|n_1^{0.7} - n_2^{0.7}|}{n}}, \tag{6}$$

we must have either $\left|\hat{W}_1 - W_1\right| \geq \frac{(\Delta u_{\max})}{p} n_1^{-0.3}$ *or* $\left|\hat{W}_2 - W_2\right| \geq \frac{(\Delta u_{\max})}{p} n_2^{-0.3}$ *to be true.*

Proof. According to the definition of $\hat{Imb}$ and $\hat{W}$, we know that

$$\hat{\mathrm{Imb}}(u,a,P_1,P_2)=\frac{\left|\hat{W}_1-\hat{W}_2\right|}{\hat{W}}=\frac{\left|\hat{W}_1-\hat{W}_2\right|}{\frac{n_1}{n}\hat{W}_1+\frac{n_2}{n}\hat{W}_2}.$$

If $W_1\geq W_2$, the "worst-case" for $\hat{\mathrm{Imb}}(u,a,P_1,P_2)$ is that $\hat{W}_1>W_1$ while $\hat{W}_2<W_2$. Then, we know that if

$$\begin{aligned}\hat{\mathrm{Imb}}(u,a,P_1,P_2)&\geq\frac{\left(W_1+\frac{(\Delta u_{\max})}{p}n_1^{-0.3}\right)-\left(W_2-\frac{(\Delta u_{\max})}{p}n_2^{-0.3}\right)}{\frac{n_1}{n}\left(W_1+\frac{(\Delta u_{\max})}{p}n_1^{-0.3}\right)+\frac{n_2}{n}\left(W_2-\frac{(\Delta u_{\max})}{p}n_2^{-0.3}\right)}\\&=\frac{W_1-W_2+\frac{\Delta u_{\max}}{p}(n_1^{-0.3}+n_2^{-0.3})}{W+\frac{\Delta u_{\max}}{p}\cdot\frac{n_1^{0.7}-n_2^{0.7}}{n}},\end{aligned}\tag{7}$$

we must have either $\left(\hat{W}_1-W_1\geq\frac{(\Delta u_{\max})}{p}n_1^{-0.3}\right)$ or $\left(W_2-\hat{W}_2\geq\frac{(\Delta u_{\max})}{p}n_2^{-0.3}\right)$. Since the condition in Lemma 4, (6), is stronger than (7), Lemma 4 directly follows by repeating the above analysis for the $W_2\geq W_1$ case. □

Combining Lemma 4 with our union bound (5), we have,

$$\Pr\left[\hat{\mathrm{Imb}}(u,a,P_1,P_2)\geq\frac{|W_1-W_2|+\frac{\Delta u_{\max}}{p}(n_1^{-0.3}+n_2^{-0.3})}{W-\frac{\Delta u_{\max}}{p}\cdot\frac{\left|n_1^{0.7}-n_2^{0.7}\right|}{n}}\right]\leq 2\exp(-2n_1^{0.4})+2\exp(-2n_2^{0.4}).$$

We also have

$$\begin{aligned}&\frac{|W_1-W_2|+\frac{\Delta u_{\max}}{p}(n_1^{-0.3}+n_2^{-0.3})}{W-\frac{\Delta u_{\max}}{p}\cdot\frac{\left|n_1^{0.7}-n_2^{0.7}\right|}{n}}\\&=\frac{|W_1-W_2|-\frac{\Delta u_{\max}}{p}\cdot\frac{\left|n_1^{0.7}-n_2^{0.7}\right|}{n}\cdot\frac{|W_1-W_2|}{W}}{W-\frac{\Delta u_{\max}}{p}\cdot\frac{\left|n_1^{0.7}-n_2^{0.7}\right|}{n}}\\&\qquad+\frac{\frac{\Delta u_{\max}}{p}\cdot\frac{\left|n_1^{0.7}-n_2^{0.7}\right|}{n}\cdot\frac{|W_1-W_2|}{W}+\frac{\Delta u_{\max}}{p}(n_1^{-0.3}+n_2^{-0.3})}{W-\frac{\Delta u_{\max}}{p}\cdot\frac{\left|n_1^{0.7}-n_2^{0.7}\right|}{n}}\\&=\frac{|W_1-W_2|}{W}+\frac{\Delta u_{\max}}{p}\cdot\frac{n_1^{-0.3}+n_2^{-0.3}+\frac{\left|n_1^{0.7}-n_2^{0.7}\right|}{n}\cdot\frac{|W_1-W_2|}{W}}{W-\frac{\Delta u_{\max}}{p}\cdot\frac{\left|n_1^{0.7}-n_2^{0.7}\right|}{n}}\\&=\mathrm{Imb}(u,a,P_1,P_2)+\frac{\Delta u_{\max}}{p}\cdot\frac{n_1^{-0.3}+n_2^{-0.3}+\frac{\left|n_1^{0.7}-n_2^{0.7}\right|}{n}\cdot\mathrm{Imb}(u,a,P_1,P_2)}{W-\frac{\Delta u_{\max}}{p}\cdot\frac{\left|n_1^{0.7}-n_2^{0.7}\right|}{n}}\\&\leq\mathrm{Imb}(u,a,P_1,P_2)+\frac{\Delta u_{\max}}{p}\cdot\frac{n_1^{-0.3}+n_2^{-0.3}+\left(n_1^{-0.3}+n_2^{-0.3}\right)\cdot\mathrm{Imb}(u,a,P_1,P_2)}{W-\frac{\Delta u_{\max}}{p}\cdot\left(n_1^{-0.3}+n_2^{-0.3}\right)}\\&=\mathrm{Imb}(u,a,P_1,P_2)+\frac{\Delta u_{\max}}{p}\cdot\frac{\mathrm{Imb}(u,a,P_1,P_2)+1}{W-\frac{\Delta u_{\max}}{p}\cdot\left(n_1^{-0.3}+n_2^{-0.3}\right)}\cdot\left(n_1^{-0.3}+n_2^{-0.3}\right).\end{aligned}$$

Then, Theorem 7 directly follows by the observation that

$$\hat{\text{Imb}}(u,a,P_1,P_2) \geq \text{Imb}(u,a,P_1,P_2) + \frac{\Delta u_{\max}}{p} \cdot \frac{\text{Imb}(u,a,P_1,P_2)+1}{W - \frac{\Delta u_{\max}}{p} \cdot \left(n_1^{-0.3} + n_2^{-0.3}\right)} \cdot \left(n_1^{-0.3} + n_2^{-0.3}\right)$$

is a stronger condition than $\hat{\text{Imb}}(u,a,P_1,P_2) \geq \frac{|W_1 - W_2| + \frac{\Delta u_{\max}}{p}(n_1^{-0.3}+n_2^{-0.3})}{W - \frac{\Delta u_{\max}}{p} \cdot \frac{\left|n_1^{0.7} - n_2^{0.7}\right|}{n}}$. □

From part (1) of Theorem 7, we see that with high probability, the utility estimator is close to the actual utility. And the probability decreases exponentially with the number of agents n. It also implies that the smaller ε is (stronger privacy, also corresponds to the smaller value of p), the higher is the probability for our utility estimator to be inaccurate. From part (2) of the theorem, the additive part in the bound for the imbalance estimate is:

$$\begin{aligned} &\frac{\Delta u_{\max}}{p} \cdot \frac{(\text{Imb}(u,a,P_1,P_2)+1)(n_1^{-0.3} + n_2^{-0.3})}{W - \frac{\Delta u_{\max}}{p} \cdot \left(n_1^{-0.3} + n_2^{-0.3}\right)} \\ =&\frac{\Delta u_{\max}}{p} \cdot \frac{(W + |W_1 - W_2|)(n_1^{-0.3} + n_2^{-0.3})}{W\left(W - \frac{\Delta u_{\max}}{p}(n_1^{-0.3} + n_2^{-0.3})\right)}. \end{aligned}$$

For smaller values of p (better privacy), the additive term gets larger, and the probability for the imbalance estimator to be inaccurate gets higher. However, for high n_1 and n_2 values, this term is low compared to the range of imbalance values. Additionally, we note that this bound is tighter when the actual imbalance value is small and the average utility is large. This means that for alternatives with low imbalance (fair alternatives) and high utility (efficient alternatives), the imbalance estimate will tend to be less noisy. So, for these alternatives, the loss in fairness due to privacy is expected to be low.

The privacy framework uses an assumed utility function, similar to what we have done for designing constrained fair voting rules. The inaccuracy for the private voting rule will be in terms of this assumed utility function, which adds to the loss in efficiency and fairness due to arbitrary ground truth utility functions as discussed in Section 5.

We run experiments with synthetic data with α-FP fair voting rules (Definition 5), analyzing the three-way trade-off between privacy, group fairness and efficiency. These experimental results are presented in Section 8. Now, even though our theoretical and experimental results regarding privacy discuss only the constrained fair voting rule framework, we suggest that it can be effectively applied to voting rules designed with the ML-based framework as well. Since the learned voting rule works by first converting votes to the features weighted majority graph and positional score matrix, both of which depend on the preference of each alternative in each group (i.e., they can be calculated from the $W(a,P)$ values for each group), thus the effect of adding local noise would be similar and the features would also have unbiased estimators. Thus, we can also design local differentially private versions of voting rules designed using the ML framework with the flipping-coin algorithm.

8. Experimental Results

While our theoretical results deal with worst-case analysis, we contend that average-case analysis is important as well. If we assume that the agents come from some underlying

distributions, the expected fairness and expected efficiency are interesting metrics to check the trade-off. However, theoretically analyzing expected fairness and efficiency is difficult. So, we do empirical analysis on synthetic data to get an idea about the average fairness-efficiency trade-off. All our experimental results focus on two-group scenarios. We consider two type of ranking models to generate all the synthetic election data, as described below.

- **Uniform**: For any set of all alternatives $\mathcal{A}$ with $|\mathcal{A}| = m$, the set of all linear orders $\mathcal{L}$ define all possible rankings over the set of alternatives. In uniform sampling, a ranking is sampled uniformly at random from $\mathcal{L}$. This sampling technique is also known as impartial culture.

- **Plackett-Luce (PL) (Plackett, 1975; Luce, 1959)**: The PL model is a widely used model for human preferences. For the PL model, the parameter space is $\Theta = \{\vec{\theta} = \{\theta^j | 1 \leq j \leq m\}\}$ and the sample space is $\mathcal{L}$. Given the parameter $\vec{\theta} \in \Theta$, the probability of any full ranking $\sigma = a_{j_1} \succ a_{j_2} \succ \cdots \succ a_{j_m}$ is $\Pr_{\text{PL}}(\sigma|\vec{\theta}) = \prod_{p=1}^{m-1} \frac{\exp(\theta^{j_p})}{\sum_{q=p}^{m} \exp(\theta^{j_q})}$.

  The PL model can be intuitively thought of as giving a score to each alternative. Then the alternatives with higher scores have a higher probability of being ranked towards the top.

For the experiments, we first do uniform sampling for all agents from both groups. This is representative of scenarios where both groups have similar ranking behavior and behave uniformly at random. On the other hand, we use the PL model for simulating group behavior. We assume that agents in a group inherently have similar preferences, while the two groups in general behave differently. So, first we first get two separate Plackett-Luce models (with randomly sampled parameters $\vec{\theta}$) for the two groups. Then, each agent's vote is sampled from using their group's PL model.

While creating training and test data, for each data point, first new PL parameters are sampled randomly. Then, a single preference profile is sampled using these group PL parameters. This ensures that the training and test data come from different distributions.

The average fairness values that we present for a voting rule are computed from the winners of the sample preference profiles. For a voting rule r, for particular group size parameters, n and z, if we have K sample preference profiles, $\{(P_1^k, P_2^k)\}_{k=1}^K$, then the average fairness is:

$$F^{Average}(r, u, n, z) = 1 - \frac{1}{K} \cdot \frac{1}{1+z} \cdot \sum_{k=1}^{K} \text{Imb}(u, r(P_1^k, P_2^k), P_1^k, P_2^k).$$

8.1 Trade-off between Fairness and Efficiency

The experiments above consider fairness results for voting rules under the same utility function. As a contrast to that, we also consider many different utility functions. For this experiment, for each utility function, we calculate a sample worst-case imbalance value through simulation, and take the mean over all the sample worst-case imbalance values. We present the results in Table 4. For u-fair rules, on top of u_{top}, u_{rank}, we consider u_{veto},

Voting Rule	u_{top}-fair	u_{rank}-fair	u_{veto}-fair	Plurality	Borda	Copeland
Fairness	0.77	**0.85**	0.81	0.68	0.78	0.78

Table 4: Average fairness value over sampled utility functions

which is defined by $\vec{u} = \langle 1, 1, 1, 0 \rangle$ (for $m = 4$). For traditional voting rules, we consider Plurality, Borda and Copeland.

We see that, as expected, even for arbitrary utility functions, the average fairness value is higher for the u-fair voting rules compared to efficiency-based rules. Among the u-fair rules, we get the highest average fairness for rank utility. This indicates that rank utility can work as a good utility function assumption for our voting rule designing frameworks.

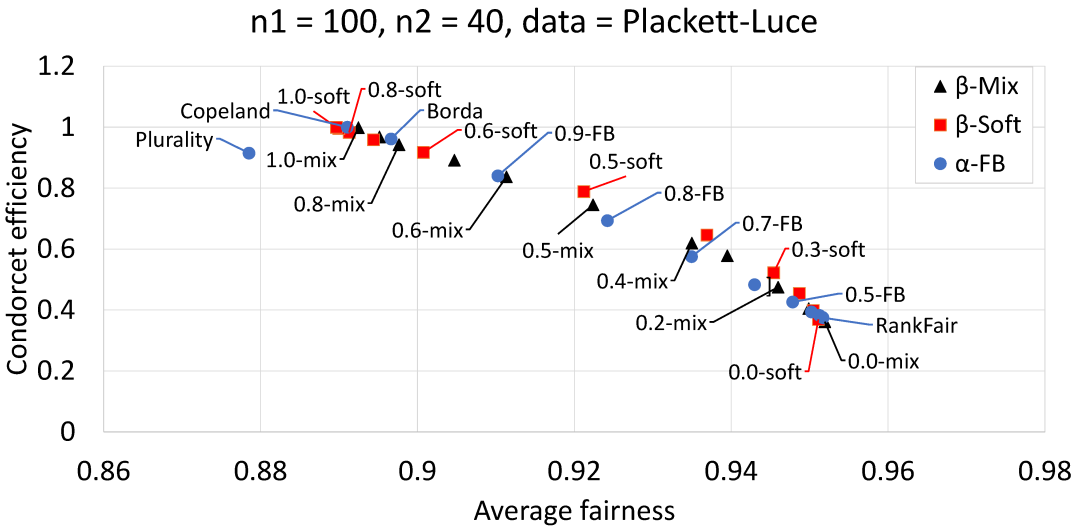
For experiment for various group sizes, varying the n_1/n_2 ratio from 100/20 to 100/100, and we consider both uniform and Plackett-Luce distributions for randomized preference profiles to understand the behavior of the voting rules under different circumstances. We compute average fairness and mean social welfare (presented as average utility). We compute Condorcet Efficiency for a voting rule as the ratio of the number of preference profiles where a voting rule chooses the Condorcet winner and the number of profiles where the Condorcet winner exists. First, we run simulations for different α-FB rules.

For our experiments on β-ML rules, we chose boosted gradient trees for learning in Algorithm 1, making use of the XGBoost (Chen & Guestrin, 2016) library. For each setting, we generated 2.4 million data points to learn from. Based on the learned voting rule, we compute Condorcet efficiency and average rank utility for the preference profiles in the test set. Results for various n_1/n_2 ratios have similar characteristics, and we present the results for $n_1 = 100$ and $n_2 = 40$ in Figure 3.

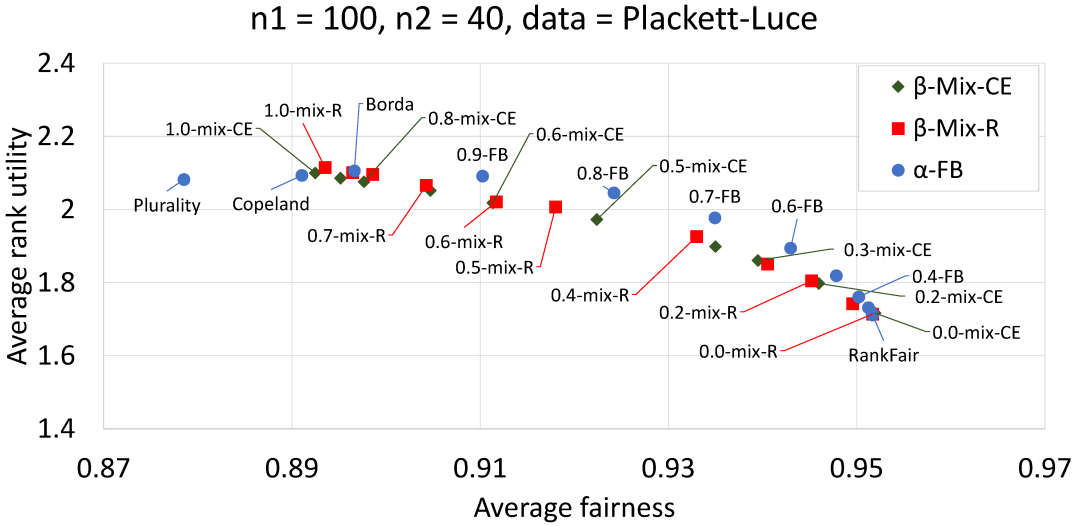
Figure 3a shows that the learned voting rules from both learning methods, β-Mix and β-Soft , mostly dominate α-FB methods, and provide a good improvement in terms of fairness compared to Copeland (a Condorcet consistent rule) while achieving almost similar levels of Condorcet efficiency.

Figure 3b on the other hand shows that the constrained rule (α-FB) achieves the goal of improving average utility while keeping high fairness values. The learned voting rules here are trained with the β-Mix method and we show the voting rules for two cases: trained with Condorcet efficiency as the efficiency measure, and trained with rank utility as the efficiency measure. We note that even with rank utility as the efficiency measure, the learned algorithm is outperformed by the α-FB voting rules. We have also repeated this experiment with the β-Soft method with very similar results.

In both Figures 3a and 3b, we see that none of the voting rules focusing purely on either economic efficiency or fairness (Plurality, Borda, Copeland and RankFair) dominates the newly designed voting rules in both economic efficiency and fairness. This is also seen true for simple mixture type randomized voting rules (e.g., a voting rule that outputs the Borda winner half of the time and the RankFair winner the other half) from experiments. So, we find that the newly designed can achieve different levels of economic efficiency and fairness that is not possible with purely efficient or fair voting rules or simple mixtures of them.



(a) Condorcet efficiency vs average fairness.



(b) Average rank utility vs average fairness.

Figure 3: Trade-off between economic efficiency and fairness for various voting rules under Plackett-Luce synthetic data. Voting rules considered are Plurality, Borda, Copeland, α-efficient FB rules and the learned rules. RankFair indicates the u_{rank}-fair voting rule. (a) The learned rules use the β-Mix and β-Soft method, both using Condorcet efficiency as the efficiency measure while training. (b) The learned rules use the β-Mix method for training. β-Mix-CE uses Condorcet efficiency as the efficiency measure, while β-Mix-R uses rank utility as the efficiency measure.

Finally, we show the trade-off comparison for different agent distributions in Figure 4, for synthetic data generated using uniform and PL-sampling. We note that for uniform distribution, the difference in average fairness between Borda and the fair rule is not high. That indicates that if there is no high dissent among the preferences between the two

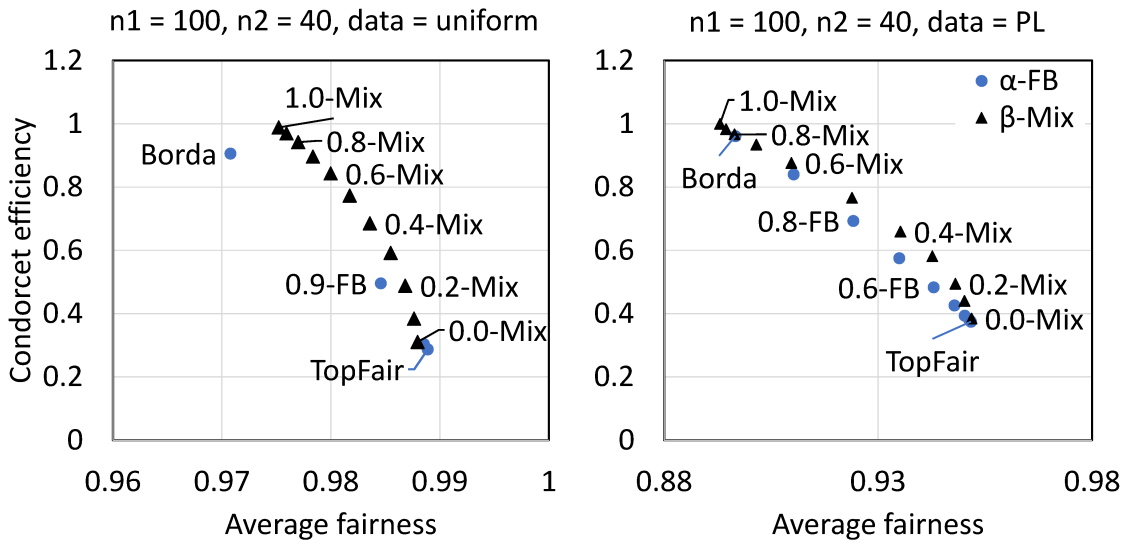


Figure 4: Trade-off between Condorcet efficiency and average fairness for various newly designed voting rules under uniformly sampled and PL-sampled distribution of voting agents. TopFair indicates the u_{top}-fair voting rule. The β-Mix rules are all trained with Condorcet efficiency as the efficiency measure.

groups, efficient voting rules are not very unfair. However, when the voting behavior of the two groups is different, as we see in the PL distribution experiments, the difference between fair voting rules and efficient voting rules becomes clearer. Nonetheless, in both scenarios, we see that our designed voting rules provide a good trade-off between fairness and efficiency. That is, we can design voting rules that have different levels of average fairness and efficiency, as compared to Borda or u_{top}-fair voting rules, which focus purely on efficiency or fairness. We have also repeated all our experiments for different group size ratios and have seen similar results as reported here.

8.2 Trade-off between Fairness, Efficiency and Privacy

From our discussions about the flipping-coin algorithm (Lemma 2), we know that better privacy can be achieved for Plurality since the support set for its randomization is smaller. Thus, we run the privacy experiments for α-efficient Fair Plurality. We sample profiles randomly and run this experiment repeatedly for values of $\varepsilon = 5$ (moderate privacy) to $\varepsilon = 2$ (strong privacy) and compare the results with non-private voting rules ($\varepsilon = \infty$). From the results in Figure 5, we see that for low privacy requirement ($\varepsilon = 5$), the loss in terms of average fairness and average utility is minimal. For higher privacy requirements, both fairness and efficiency suffer from the noise-adding mechanism. However, for $\varepsilon = 3$, which is a moderately standard setting, we see relatively low loss ($5 \sim 15\%$) for both average utility and fairness. The loss is particularly more pronounced for more private voting rules (low α values) rather than more efficient voting rules (1.0-FP voting rule is the same as Plurality, which sees the smallest decrease in both fairness and utility).

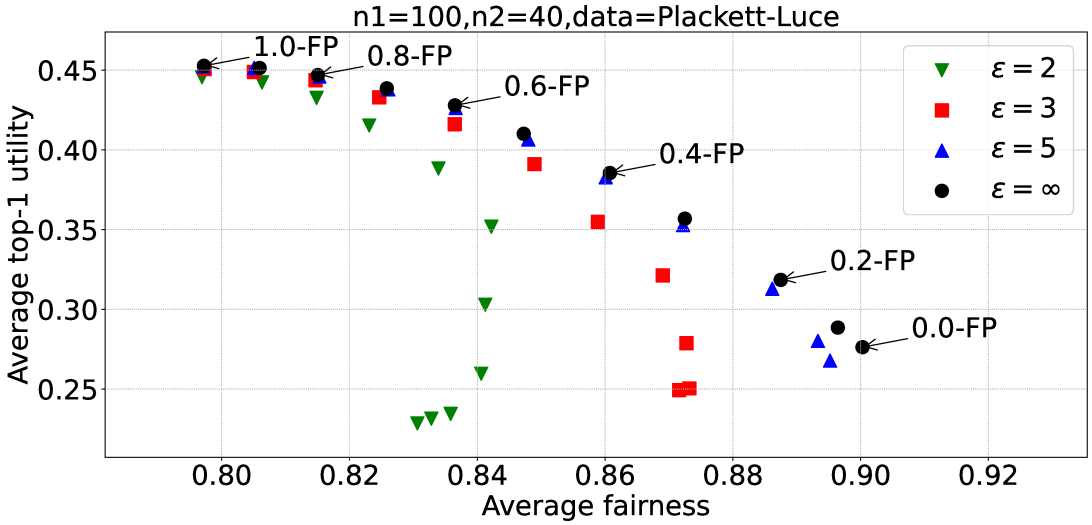


Figure 5: Efficiency (Average top-1 utility) and average fairness for various levels of privacy. High ε values indicate lower privacy requirements with $\varepsilon = \infty$ indicating non-private voting rule.

This same trade-off is empirically observed for different group size ratios. Also, we note that while local differential privacy can be achieved for Plurality and fair versions of Plurality, we cannot expect this from Borda. We also ran similar experiments on Borda and α-efficient Fair Borda that we do not report here. As expected, since the support set for Borda is much larger (the size of the support set is $m!$ for m alternatives), the flipping-coin algorithm requires a high level of noise. Thus, the decrease in utility and fairness is high for even moderate privacy requirements. However, we can use the privacy measures effectively in other voting rules with low support sizes for decisions, such as approval voting.

9. Conclusions and Future Work

In this paper, we study how to design and learn voting rules that embed desirable properties related to fairness, privacy, and economic efficiency. We present a new notion of group imbalance-based fairness in the social choice domain. We see that most traditional voting rules, due to being anonymous in nature, are not fair in terms of groups in worst-case scenarios. However, we show that it is possible to work on the trade-off between economic efficiency (indicated by social welfare or Condorcet efficiency) and group fairness and design new voting rules with good average fairness and efficiency values. One of our methods for designing new voting rules is a data-driven procedure for learning new voting rules from synthetically generated voting data. Finally, we consider local differentially private versions of the voting rules and examine the three-way trade-off between economic efficiency, group fairness, and privacy.

Going forward, we think there are interesting questions in all of the directions we considered. While we say the notion of group imbalance-based fairness satisfies some desirable properties, we keep axiomatic consideration of the group fairness property for future work. Additionally, we discuss homogeneous utility functions for the agents, that is, all agents

have the same utility function. While we still consider the worst-case analysis over all such utility functions, it would be interesting to see how the fairness bounds are for more general utility definitions, such as metric preferences (Anshelevich et al., 2018). More general analysis similar to social choice distortion (Procaccia & Rosenschein, 2006) is also an interesting direction for future work. We can also consider the idea of fairness under composition when there are multiple properties across which the agents may be grouped, e.g., *male vs female* and *old vs young*. We can compute different imbalance values for an alternative for both properties. Thus, an alternative that is fair across gender may be unfair across race. Whether designing fair voting rules under composition leads to more complicated scenarios is an interesting question.

Regarding the data-driven method for learning new voting rules, our work in using learning to design voting rules with new properties can be extended. Both of our current methods focus on getting good fairness and efficiency values in average and does not focus on individual preference profiles. In the future, we can consider a more general loss-function-based approach with loss for both group-fairness and efficiency properties such that even for individual preference profiles, the learned voting rules manage to find alternatives that are both somewhat fair and efficient.

Finally, when considering local differential privacy, we note that getting a more private voting rule has a trade-off with having high efficiency and high fairness. However, work in differential privacy literature has suggested that the randomization required for privacy may be applied in ways to also preserve like fairness (e.g., in (Zhu et al., 2022)). So, it will be interesting to see whether local differential privacy can be obtained at the same time with group fairness properties in randomized voting rules.

Acknowledgments

The authors are grateful to Michael Hind for discussions and the reviewers for their insightful comments and suggestions. This work was supported by the Rensselaer-IBM AI Research Collaboration, part of the IBM AI Horizons Network. Ao Liu is supported by the RPI-IBM AI Horizons scholarship. Lirong Xia acknowledges NSF 1453542, 1716333, 2007994, 2106983, and a Google gift fund for support.

References

Aleksandrov, M., & Walsh, T. (2018). Group envy freeness and group pareto efficiency in fair division with indivisible items. In *Proceedings of the 41st Joint German/Austrian Conference on Artificial Intelligence (Künstliche Intelligenz)*, pp. 57–72. Springer.

Anil, C., & Bao, X. (2021). Learning to elect. In *Proceedings of the 35th Annual Conference on Neural Information Processing Systems*, pp. 8006–8017. NeurIPS.

Anshelevich, E., Bhardwaj, O., Elkind, E., Postl, J., & Skowron, P. (2018). Approximating optimal social choice under metric preferences. *Artificial Intelligence*, *264*, 27–51.

Armstrong, B., & Larson, K. (2019). Machine learning to strengthen democracy. In *AI for Social Good workshop at NeurIPS-2019*.

Aziz, H., Brill, M., Conitzer, V., Elkind, E., Freeman, R., & Walsh, T. (2017). Justified representation in approval-based committee voting. *Social Choice and Welfare*, *48*(2), 461–485.

Aziz, H., Elkind, E., Huang, S., Lackner, M., Fernández, L. S., & Skowron, P. (2018). On the complexity of extended and proportional justified representation. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, pp. 902–909. AAAI.

Balinski, M. L., & Young, H. P. (2010). *Fair representation: meeting the ideal of one man, one vote*. Brookings Institution Press.

Birrell, E., & Pass, R. (2011). Approximately strategy-proof voting. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, pp. 67–72. AAAI/IJCAI.

Boole, G. (1847). *The mathematical analysis of logic*. Philosophical Library.

Boutilier, C., Caragiannis, I., Haber, S., Lu, T., Procaccia, A. D., & Sheffet, O. (2015). Optimal social choice functions: A utilitarian view. *Artificial Intelligence*, *227*, 190–213.

Brams, S. J., Kilgour, D. M., & Sanver, M. R. (2007). A minimax procedure for electing committees. *Public Choice*, *132*(3), 401–420.

Brandt, F., Conitzer, V., Endriss, U., Lang, J., & Procaccia, A. D. (2016). *Handbook of computational social choice*. Cambridge University Press.

Bredereck, R., Faliszewski, P., Igarashi, A., Lackner, M., & Skowron, P. (2018). Multiwinner elections with diversity constraints. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, pp. 933–940. AAAI.

Celis, L. E., Huang, L., & Vishnoi, N. K. (2018). Multiwinner voting with fairness constraints. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pp. 144–151. IJCAI.

Cembrano, J., Correa, J., Díaz, G., & Verdugo, V. (2021). Proportional apportionment: A case study from the chilean constitutional convention. In *Proceedings of the 1st Equity and Access in Algorithms, Mechanisms, and Optimization*, pp. 1–9. ACM.

Chamberlin, J. R., & Courant, P. N. (1983). Representative deliberations and representative decisions: Proportional representation and the borda rule. *American Political Science Review*, *77*(3), 718—-733.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 785–794. ACM.

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, *5*(2), 153–163.

Chouldechova, A., & Roth, A. (2020). A snapshot of the frontiers of fairness in machine learning. *Communications of the ACM*, *63*(5).

Conitzer, V., & Sandholm, T. (2002). Complexity of mechanism design. In *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, pp. 103—-110. Morgan Kaufmann.

Conitzer, V., & Sandholm, T. (2003). Applications of automated mechanism design. In *Workshop on Bayesian Modeling Applications at UAI-2003*.

Contucci, P., Panizzi, E., Ricci-Tersenghi, F., & Sîrbu, A. (2016). Egalitarianism in the rank aggregation problem: a new dimension for democracy. *Quality & Quantity*, *50*, 1185–1200.

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017). Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 797–806. ACM.

Cormode, G., Jha, S., Kulkarni, T., Li, N., Srivastava, D., & Wang, T. (2018). Privacy at scale: Local differential privacy in practice. In *Proceedings of the 44th International Conference on Management of Data*, pp. 1655–1658. ACM.

Curry, M., Chiang, P.-Y., Goldstein, T., & Dickerson, J. (2020). Certifying strategyproof auction networks. In *Proceedings of the 34th Annual Conference on Neural Information Processing Systems*, pp. 4987–4998. NeurIPS.

Dütting, P., Feng, Z., Narasimhan, H., Parkes, D. C., & Ravindranath, S. S. (2019). Optimal auctions through deep learning. In *Proceedings of the 36th International Conference on Machine Learning*, pp. 1706–1715. PMLR.

Dwork, C. (2006). Differential Privacy. In *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming*, pp. 1–22. Springer.

Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Proceedings of the 3rd Theory of Cryptography Conference*, pp. 265–284. Springer.

Dwork, C., Roth, A., et al. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, *9*(3-4), 211–407.

Evfimievski, A., Gehrke, J., & Srikant, R. (2003). Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the 22nd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pp. 211—-222. ACM.

Feng, Z., Narasimhan, H., & Parkes, D. C. (2018). Deep learning for revenue-optimal auctions with budgets. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems*, pp. 354–362. IFAAMAS.

Freeman, R., Zahedi, S. M., & Conitzer, V. (2017). Fair social choice in dynamic settings. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pp. 4580–4587. IJCAI.

Golowich, N., Narasimhan, H., & Parkes, D. C. (2018). Deep learning for multi-facility location mechanism design. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pp. 261–267. IJCAI.

Gursoy, M. E., Tamersoy, A., Truex, S., Wei, W., & Liu, L. (2019). Secure and utility-aware data collection with condensed local differential privacy. *IEEE Transactions on Dependable and Secure Computing*, *18*(5), 2365–2378.

Hay, M., Elagina, L., & Miklau, G. (2017). Differentially private rank aggregation. In *Proceedings of the 17th SIAM International Conference on Data Mining*, pp. 669–677. SIAM.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, *58*(301), 13–30.

Hossain, S., Mladenovic, A., & Shah, N. (2020). Designing fairly fair classifiers via economic fairness notions. In *Proceedings of The Web Conference 2020*, p. 1559–1569. ACM.

Jha, T., & Zick, Y. (2020). A learning framework for distribution-based game-theoretic solution concepts. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pp. 355–377. ACM.

Joseph, M., Roth, A., Ullman, J., & Waggoner, B. (2018). Local differential privacy for evolving data. In *Proceedings of the 32nd Annual Conference Neural Information Processing Systems*, pp. 2381—-2390. NeurIPS.

Kearns, M., Neel, S., Roth, A., & Wu, Z. S. (2018). Preventing fairness gerrymandering: Auditing and learning for subgroup fairness. In *Proceedings of the 35th International Conference on Machine Learning*, pp. 2564–2572. PMLR.

Kleinberg, J. (2018). Inherent trade-offs in algorithmic fairness. In *Abstracts of the 2018 ACM International Conference on Measurement and Modeling of Computer Systems*, pp. 40–40. ACM.

Lackner, M. (2020). Perpetual voting: Fairness in long-term decision making. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, pp. 2103–2110. AAAI.

Lee, D. T. (2015). Efficient, private, and e-strategy proof elicitation of tournament voting rules. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, pp. 2026–2032. AAAI/IJCAI.

Liu, A., Lu, Y., Xia, L., & Zikas, V. (2020). How private are commonly-used voting rules?. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence*, pp. 629–638. PMLR.

Luce, R. D. (1959). *Individual choice behavior - A Theoretical Analysis*. Wiley.

Mill, J. S. (1859). *On liberty*. Longman, Roberts, & Green Co.

Monroe, B. L. (1995). Fully proportional representation. *American Political Science Review*, *89*(4), 925—940.

Myerson, R. B. (1981). Utilitarianism, egalitarianism, and the timing effect in social choice problems. *Econometrica*, *49*(4), 883–897.

Parkes, D. C., & Procaccia, A. D. (2013). Dynamic social choice with evolving preferences. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, pp. 767–773. AAAI.

Plackett, R. L. (1975). The analysis of permutations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *24*.

Procaccia, A. D., & Rosenschein, J. S. (2006). The distortion of cardinal preferences in voting. In *Proceedings of the 10th International Workshop on Cooperative Information Agents*, pp. 317—-331. Springer.

Procaccia, A. D., Zohar, A., Peleg, Y., & Rosenschein, J. S. (2009). The learnability of voting rules. *Artificial Intelligence*, *173*(12-13), 1133–1149.

Sánchez-Fernández, L., Elkind, E., Lackner, M., Fernández, N., Fisteus, J. A., Val, P. B., & Skowron, P. (2017). Proportional justified representation. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, pp. 670–676. AAAI.

Shang, S., Wang, T., Cuff, P., & Kulkarni, S. (2014). The application of differential privacy for rank aggregation: Privacy and accuracy. In *Proceedings of the 17th International Conference on Information Fusion*, pp. 1–7. IEEE.

Shen, W., Tang, P., & Zuo, S. (2019). Automated mechanism design via neural networks. In *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems*, pp. 215–223. IFAAMAS.

Skowron, P., Faliszewski, P., & Slinko, A. (2015). Achieving fully proportional representation: Approximability results. *Artificial Intelligence*, *222*, 67–103.

Stone, P. (2016). Sortition, voting, and democratic equality. *Critical review of international social and political philosophy*, *19*(3), 339–356.

Verma, S., & Rubin, J. (2018). Fairness definitions explained. In *Proceedings of the 2018 IEEE/ACM International Workshop on Software Fairness*. ACM.

Wang, S., Du, J., Yang, W., Diao, X., Liu, Z., Nie, Y., Huang, L., & Xu, H. (2019). Aggregating votes with local differential privacy: Usefulness, soundness vs. indistinguishability. arXiv preprint arXiv:1908.04920.

Xia, L. (2013). Designing social choice mechanisms using machine learning. In *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems*, pp. 471–474. IFAAMAS.

Yan, Z., Li, G., & Liu, J. (2020). Private rank aggregation under local differential privacy. *International Journal of Intelligent Systems*, *35*(10), 1492–1519.

Ye, Q., Hu, H., Meng, X., & Zheng, H. (2019). Privkv: Key-value data collection with local differential privacy. In *Proceedings of the 40th IEEE Symposium on Security and Privacy*, pp. 317–331. IEEE.

Zhu, T., Ye, D., Wang, W., Zhou, W., & Yu, P. S. (2022). More than privacy: Applying differential privacy in key areas of artificial intelligence. *IEEE Transactions on Knowledge and Data Engineering*, *34*(6), 2824–2843.