Subset Selection Based On Multiple Rankings in the Presence of Bias: Effectiveness of Fairness Constraints for Multiwinner Voting Score Functions

Niclas Boehmer [1] L. Elisa Celis [2] Lingxiao Huang [3] Anay Mehrotra [2] Nisheeth K. Vishnoi [2]

Abstract

We consider the problem of subset selection where one is given multiple rankings of items and the goal is to select the highest "quality" subset. Score functions from the multiwinner voting literature have been used to aggregate rankings into quality scores for subsets. We study this setting of subset selection problems when, in addition, rankings may contain systemic or unconscious biases toward a group of items. For a general model of input rankings and biases, we show that requiring the selected subset to satisfy group fairness constraints can improve the quality of the selection with respect to unbiased rankings. Importantly, we show that for fairness constraints to be effective, different multiwinner score functions may require a drastically different number of rankings: While for some functions, fairness constraints need an exponential number of rankings to recover a close-to-optimal solution, for others, this dependency is only polynomial. This result relies on a novel notion of "smoothness" of submodular functions in this setting that quantifies how well a function can "correctly" assess the quality of items in the presence of bias. The results in this paper can be used to guide the choice of multiwinner score functions for the subset selection setting considered here; we additionally provide a tool to empirically enable this.

1. Introduction

The task of selecting a size-k subset from a set of m items is a basic problem in machine learning and computer science with applications arising in recommender systems (McSherry, 2002; El-Arini et al., 2009), feature selection (Guyon & Elisseeff, 2003), search engines (Agrawal et al., 2009), data summarization (Elhamifar & Kaluza, 2017; Angelidakis et al., 2022), and algorithmic hiring (Schumann et al., 2019; Raghavan et al., 2020; Deczynski, 2021). The simplest formulation of this problem assumes that every item has some reported utility, which leads to the selection of k items with the highest utility. In several applications, however, items are evaluated from multiple points of view, e.g., different users evaluating items in a group recommendation or shortlisting setting (Jameson & Smyth, 2007; Raghavan et al., 2020; Streviniotis & Chalkiadakis, 2022a), or a single user evaluating items from different perspectives (e.g., in personalized recommendations, a user might be interested in being recommended items from different categories, and separate orderings for categories are produced (El-Arini et al., 2009; Streviniotis & Chalkiadakis, 2022b; Gawron & Faliszewski, 2022)). Moreover, these different evaluations oftentimes come in the form of *rankings* of the m items instead of numerical scores (Chakraborty et al., 2019; Streviniotis & Chalkiadakis, 2022b;a). Asking for rankings instead of numerical scores for items is a popular approach (e.g., as witnessed by its usage to train large language models, like ChatGPT with human feedback (Ramponi, 2022), and in reinforcement learning (Christiano et al., 2017)) because numerical scores have a higher elicitation cost and can lead to aggregation and calibration issues (Griffin & Brenner, 2004; Mitliagkas et al., 2011; Wang & Shah, 2019).

A principled method to aggregate n rankings of m items into a single quality score function for subsets of $\{1, \ldots, m\}$ is the usage of "score functions" studied in the multiwinner voting literature within social choice theory (Faliszewski et al., 2017; Chakraborty et al., 2019; Mondal et al., 2021; Streviniotis & Chalkiadakis, 2022b; Gawron & Faliszewski, 2022; Streviniotis & Chalkiadakis, 2022a; Lackner & Skowron, 2023). Here, many score functions have been proposed and their properties and merits have been extensively discussed (Elkind et al., 2017b;a; Lackner & Skowron, 2019). These functions, for instance, allow for specifying which part of each ranking to take into account, and how many of the selected items are relevant from the perspective of each ranking. For instance, a score function

[1]TU Berlin [2]Yale University [3]State Key Laboratory of Novel Software Technology, Nanjing University, Nanjing, China. Correspondence to: L. Elisa Celis <ecelis21@gmail.com>, Lingxiao Huang <huanglingxiao1990@126.com>, Nisheeth K. Vishnoi <nisheeth.vishnoi@gmail.com>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

may only take into account the top choice of each ranking, and the quality of a subset is then the number of rankings whose top choice it contains (this score function is referred to as single non-transferable vote, or SNTV). More generally, a score function may take into account the entire ranking. For instance, each ranking may add $m-1$ points to the quality of a subset if its top choice is present in it, another $m-2$ points if its second most preferred choice is present, and so on (the resulting score function is called the Borda count). While SNTV and Borda lead to modular set functions, more general score functions from the multiwinner voting literature give rise to general submodular functions (Definition 2.2), and a multitude of techniques have been developed to optimize the quality of the selection with respect to the given rankings (Procaccia et al., 2008; Skowron et al., 2015; Aziz et al., 2015).

However, there is growing evidence that rankings of items – whether generated by humans or ML algorithms – may contain implicit and/or other forms of systemic biases against socially-salient groups (e.g., those defined by race, gender, opinions) (Rooth, 2010; Kite & Whitley, 2016; Régner et al., 2019). For instance, in the context of subset selection, Capers IV et al. (2017) detected that, despite identical qualifications, members of admission committees rank White medical students above African Americans, and Moss-Racusin et al. (2012) found that faculty members evaluate male applicants for a job as more qualified than female ones. The latter bias has also been observed in Amazon's former resume screening tool (Dastin, 2018), as human biases can find their way into the output of machine learning algorithms in case they are trained on biased or systematically skewed data (Dastin, 2018; Bogen & Rieke, 2018; Dressel & Farid, 2018; Raghavan et al., 2020; Jiang & Nachum, 2020). A question arises: in which situations can bias be mitigated in subset selection based on aggregating multiple (biased) rankings using multiwinner score functions, and in these situations, how can the selection of a high-quality committee with respect to the unbiased rankings be guaranteed?

A model to incorporate such biases for the simplest (numerical) formulation of subset selection was considered by Kleinberg & Raghavan (2018): Items are partitioned into an advantaged and a disadvantaged group and item i has latent utility W_i. For members of the advantaged group, the observed utility is the same as the latent utility, whereas, for members of the disadvantaged group, the observed utility is $\theta \cdot W_i$ for some global bias parameter $\theta \in [0, 1]$. Kleinberg & Raghavan (2018) show that when optimizing for the summed observed utility, the latent utility of the selected subset drops sharply in the presence of implicit bias. These suboptimal selections as well as strong negative consequences for the affected groups have also been observed in practice (Rooth, 2010; Kang et al., 2011; Chapman et al., 2013; Kite & Whitley, 2016; Greenwald & Lai, 2020).

To mitigate suboptimal selections in the presence of bias, one popular intervention is the use of representational constraints, which are a type of group fairness constraint requiring that a certain fraction of the selected items need to be part of the disadvantaged group (used, e.g., in the form of the Rooney Rule in shortlisting job applicants (Collins, 2007; Waldstein, 2015), or in primary elections (Evéquoz et al., 2022)). Kleinberg & Raghavan (2018) studied the effectiveness of such representational constraints in their bias model and demonstrated that, while for worst-case W_i's these constraints may not improve the latent utility of the selection, under certain conditions, e.g., assuming that the W_i's are all drawn from the same distribution, representational constraints can increase the expected latent quality of the returned selection. Subsequent works have confirmed the power of representational constraints to reduce the effect of bias in more involved settings such as online subset selection (Salem & Gupta, 2020), with intersectional groups (Mehrotra et al., 2022), producing a ranking of items (Celis et al., 2020), and subset selection if biases come in the form of evaluation uncertainty (Emelianov et al., 2022a;b).

Representational constraints have also already been considered in the context of subset selection based on score functions from multiwinner voting, albeit with a focus on the computational complexity of selecting subsets maximizing some goal function while respecting such constraints (Bredereck et al., 2018; Celis et al., 2018). In contrast to these works, following the work of Kleinberg & Raghavan (2018), we analyze whether and when representational constraints can debias subset selection based on multiple input rankings and help to select high-quality subsets. Notably, previous results on selection in the presence of bias do not apply to our setting, as we assume that the input comes from multiple sources, in the form of rankings (and not numerical values of items), which are aggregated using multiwinner score functions.

Our contributions. We study the power of representational constraints for the following problem: Given n potentially biased rankings over m items, an integer k, and a score function F, select a size-k subset of the items with a high score according to F with respect to the latent rankings. As in the work of Kleinberg & Raghavan (2018), without distributional assumptions on the rankings and an explicit model of bias (that quantifies the relation between latent and observed rankings), no algorithm can guarantee that the returned subset has even a small fraction of optimum latent quality even with representational constraints (see, e.g., Section 3.1). We consider a model where each latent ranking in the input is generated i.i.d. from a given distribution and then bias is introduced into each of these rankings in a "controlled" manner, inspired by the work of Kleinberg & Raghavan (2018); our model is more general and is presented in Section 2. We then give an algorithm and prove that there is a representational constraint such that, if the number of input rankings is large enough, the algorithm outputs an "approximately" optimal solution with respect to the latent rankings; see Theorem 3.2. We note that while the algorithm takes as input biased rankings and aims to find the subset with the maximum (biased) score subject to the representational constraint, its guarantee is with respect to the (unconstrained) optimal solution when the input rankings are unbiased. Importantly, the *number of input rankings* the above result requires to hold depends on the chosen multiwinner score function. To mathematically capture these differences between (submodular) multiwinner score functions, we introduce a new notion of "smoothness" (Definition 3.1), which quantifies how well the score function can "correctly distinguish" the strength of candidates among the same group under the latent and biased preferences. A particular challenge here is that by considering submodular instead of modular functions, assessing the strength of a candidate becomes more difficult as a candidate's "quality" depends on the other candidates present in the subset, and that which of two candidates has a higher marginal contribution to some subset might change by biasing the rankings. We complement our algorithmic result with an impossibility result that establishes the need for a large number of input rankings for certain multiwinner score functions; see Theorem D.3.

As a simple corollary of our results, we show a stark contrast between the two popular multiwinner score functions SNTV and Borda: Under a utility-based model generalizing the model of Kleinberg & Raghavan (2018) falling into our general class of generative models, for SNTV an exponential number of input rankings (in the number of items) is needed for representational constraints to recover a close-to-optimal solution, whereas for Borda the dependency is only polynomial (Theorems 3.3 and D.3). This difference is also intuitive, as under SNTV only a local snapshot of each ranking, i.e., who is ranked in the first position, is taken into account making fine-grained distinctions of candidates' quality harder, whereas for Borda the full ranking matters.

In summary, we contribute to the growing literature on designing algorithms in the presence of bias by showing the effectiveness of representational constraints in ranking-based subset selection. Distinguishing the effectiveness of representational constraints for different multiwinner score functions based on a new notion of smoothness, we contribute to a better understanding of multiwinner score functions, thereby providing further guidance for the selection of such a rule in the desired context. Allowing for a convenient extension of our theoretical results, we give code that, given a bias model and scoring function as input, can be used as a tool to study the smoothness of the score function and the effectiveness of representational constraints with respect to the given bias model via simulations (Section G).

Other related work. We present additional related works on the study of representational constraints beyond subset selection, submodular maximization, learning user preferences in the presence of noise, and other forms of subset selection in Section A.

2. Models of Score Functions, Bias, and Representational Constraints

In this section, we formally introduce our setting and the studied problem. For the sake of consistency with the literature, we use terminologies from the multiwinner voting literature. We start by introducing the family of score functions in Section 2.1, then our bias model in Section 2.2, and lastly representational constraints in Section 2.3. We present an overview of all used notation in Table 1 in the appendix.

2.1. A Family of Score Functions

Given a set C of candidates (also called items), let $\mathcal{L}(C)$ be the set of all strict and complete orders over C. We refer to elements of $\mathcal{L}(C)$ as preference lists (or rankings) and to subsets of C as committees. Given a preference list $\succ \in \mathcal{L}(C)$, we use $\mathrm{pos}_{\succ}(c)$ to denote the position of $c \in C$ in $\succ$ and use $\mathrm{pos}_{\succ}(S) \coloneqq (\mathrm{pos}_{\succ}(c))_{c\in S}$ to denote the vector of positions of each $c \in S$.

In the classic multiwinner voting problem, there is (i) a set C of m candidates, (ii) a set V of n voters together with a preference profile $R = \{\succ_v \in \mathcal{L}(C) : v \in V\}$ where $\succ_v$ is the preference list of voter v, (iii) a function $F : 2^C \to \mathbb{R}_{\geq 0}$ with an evaluation oracle that depends on R, and (iv) a desired committee size $k \in [m]$. The goal is then to select a committee S that maximizes $F(S)$ subject to having size-k. Our results apply to a general class of submodular functions F. For its definition, we use the following notion to compare the quality of committees.

Definition 2.1 (**Position-wise dominance).** Given $\succ \in \mathcal{L}(C)$ and two committees $S = \{c_1, \dots, c_k\}$ and $S' = \{c'_1, \dots, c'_k\}$ of equal size, satisfying $\mathrm{pos}_{\succ}(c_1) > \cdots > \mathrm{pos}_{\succ}(c_k)$ and $\mathrm{pos}_{\succ}(c'_1) > \cdots > \mathrm{pos}_{\succ}(c'_k)$, we say that S *position-wise dominates* S' *with respect to* $\succ$ if for every $\ell \in [k]$, $\mathrm{pos}_{\succ}(c_\ell) \leq \mathrm{pos}_{\succ}(c'_\ell)$.

We consider the following general family of score functions.

Definition 2.2 (**Multiwinner score function).** Given a set C of candidates and the voters' preference profile $R = \{\succ_v \in \mathcal{L}(C) : v \in V\}$, we say a function $F : 2^C \to \mathbb{R}_{\geq 0}$ is a *multiwinner score function* if F satisfies the following:

- (Separability) there is a function $f : 2^C \times \mathcal{L}(C) \to \mathbb{R}_{\geq 0}$ mapping committees and preference lists to scores such that for every $S \subseteq C$, $F(S) = \sum_{v\in V} f(S, \succ_v)$. Here, $f_v(S) \coloneqq f(S, \succ_v) = f(\mathrm{pos}_{\succ_v}(S))$[1] is the score voter v

[1]We only consider functions f of the positions $\mathrm{pos}_{\succ_v}(S)$ of

awards to S that depends on their preference list $\succ_v$;

- (Monotone submodular) for every voter $v \in V$, f_v is a monotone submodular function;[2]
- (Domination sensitive) for every $v \in V$ and committees $S, S' \subseteq C$ with $|S| = |S'|$ and set $T \subseteq C \setminus S \cup S'$, if S position-wise dominates S' with respect to $\succ_v$, we have $f_v(S) - f_v(S') \geq f_v(S \cup T) - f_v(S' \cup T) \geq 0$.[3]

Among others, Definition 2.2 captures the following score functions $F(S) = \sum_{v \in V} f(S, \succ_v)$ relevant to our work:[4]

- If $f(S, \succ) = \sum_{c \in S} \mathbb{1}_{\mathrm{pos}_\succ(c)=1}$, F is the SNTV rule.
- If $f(S, \succ) = \sum_{c \in S} (m - \mathrm{pos}_\succ(c))$, F is the Borda rule.
- If $f(S, \succ) = \max_{c \in S} \{m - \mathrm{pos}_\succ(c)\}$, F is the ℓ_1-Chamberlin-Courant rule (or ℓ_1-CC rule) introduced by Chamberlin & Courant (1983).

We conclude by defining the marginal contribution of a candidate: Given a function $f : 2^C \times \mathcal{L}(C) \to \mathbb{R}_{\geq 0}$, preference list $\succ \in \mathcal{L}(C)$, committee $S \subsetneq C$ and candidate $c \in C \setminus S$, we define $f_S(c, \succ) \coloneqq f(S \cup \{c\}, \succ) - f(S, \succ)$ to be the marginal contribution of c to S with respect to f and $\succ$.

2.2. Multiwinner Voting in the Presence of Bias

In this section, we define the general problem of subset selection based on multiple input rankings in the presence of implicit bias. Our general approach is in line with previous works on subset selection (Kleinberg & Raghavan, 2018; Celis et al., 2020), and assumes that the candidates are partitioned into an advantaged group G_1 and a disadvantaged group G_2, where disadvantaged candidates $c \in G_2$ face systematic biases. Henceforth, we denote by $\succ_v$ the "true" or *latent* preference of voter v, and by $\not\succ_v$ their *biased* or *observed* preference. We consider the following problem.

Problem 1 (**Multiwinner voting in the presence of bias**). Let $R = \{\succ_v \in \mathcal{L}(C) \colon v \in V\}$ be the voter's latent preference lists and $\widehat{R} = \{\not\succ_v \in \mathcal{L}(C) \colon v \in V\}$ be the voter's biased or observed preference lists. Further, let $F \colon 2^C \to \mathbb{R}_{\geq 0}$ be a multiwinner score function, and $k \geq 1$ be the desired committee size. The goal of the *multiwinner voting problem in the presence of bias* is to select a size-k committee S that (approximately) maximizes $F(S)$, while only taking $\widehat{R}$ as input (and not knowing R).

2.2.1. Models for latent and biased preferences

We consider a general setting of how the latent and biased preferences of voters are generated. Our main assumption is that there are no differences between voters, implying that all voters sample their latent and biased preferences from the same distribution. This is a common assumption both in the line of work on selection in the presence of bias (Kleinberg & Raghavan, 2018; Celis et al., 2020; Emelianov et al., 2020) and in many mechanisms to sample preferences in social choice theory (Szufa et al., 2020). We first define our generative model for latent preferences:

Definition 2.3 (**Generative model for latent preferences).** A generative model μ for latent preferences is defined by a distribution π over $\mathcal{L}(C)$: under μ, each voter $v \in V$ draws its latent preference list $\succ_v$ from π independently.

Similarly, in our generative model for biased preferences, we only require that all voters with the same latent preferences sample their biased preferences from the same distribution. Formally, our generative model for biased preferences takes a latent preference list $\succ$ as input and draws a biased preference list $\not\succ$ from a certain conditional distribution on $\succ$. For instance, this would allow for swapping down a candidate from the disadvantaged group uniformly at random in the given, latent preferences.

Definition 2.4 (**Generative model for biased preferences).** A generative model $\widehat{\mu}$ for biased preferences is defined by a distribution $\widehat{\pi}$ over $\mathcal{L}(C) \times \mathcal{L}(C)$: under $\widehat{\mu}$, each voter $v \in V$ with latent preference $\succ_v$ draws its biased preference list $\not\succ_v$ from $\widehat{\pi} \mid \succ_v$ independently.[5]

Note that our generative model for biased preferences also allows that some voters are biased and others unbiased (or even that different voters have opposing biases). For instance, $\widehat{\mu}$ could be defined by a mixture of distributions, where each voter v draws $\not\succ_v$ according to $\widehat{\pi}_1 \mid \succ_v$ with probability $\frac{1}{2}$ and according to $\widehat{\pi}_2 \mid \succ_v$ with probability $\frac{1}{2}$. In this case, roughly half of the voters would generate their biased preferences according to $\widehat{\pi}_1$, and the remaining voters would generate them using $\widehat{\pi}_2$ and, hence, display no bias (or a potentially opposite) bias than the other voters.

2.2.2. Utility-Based Model

In this section, we give a specific example of what a generative model for latent and biased preferences could look like. For this, we present a natural adaption of the bias model considered by Kleinberg & Raghavan (2018) to our setting. In the resulting utility-based model, we assume that each voter $v \in V$ has a non-zero latent utility $w_{v,c}$ (respectively biased utility $\widehat{w}_{v,c}$) for each candidate $c \in C$, and

S in $\succ_v$, imposing that f is symmetric with respect to candidate indices. Such functions are known as neutral functions in social choice.

[2]Here, "monotone" means that for any $S \subsetneq C$ and candidate $c \in C \setminus S$, $f_v(S \cup \{c\}) \geq f_v(S)$; "submodular" means that for any $S \subsetneq T \subsetneq C$ and candidate $c \in C \setminus T$, $f_v(T \cup \{c\}) - f_v(T) \leq f_v(S \cup \{c\}) - f_v(S)$.

[3]Voters award a higher score to the better committee S than to the worse committee S'. Adding the same candidates to S and S' decreases the difference between their awarded scores (as the two become "more similar").

[4]As discussed in Section B.1, Definition 2.2 also covers the larger classes of committee-scoring rules and approval-based rules.

[5]Definition 2.4 can be extended to condition on additional information beyond $\succ_v$. For instance, Definition 2.6, considers the conditional distribution with respect to a latent utility vector w_v (which specifies $\succ_v$) instead of $\succ_v$.

that in their latent (respectively biased) preferences voters rank candidates sorted in the non-increasing order of $w_{v,c}$ (respectively $\widehat{w}_{v,c}$) with breaking ties arbitrarily. We start by explaining how the latent utilities are generated:

Definition 2.5 (**Utility-based latent generative model).** Every candidate $c \in C$ has an intrinsic utility $\omega_c \geq 0$. For each voter $v \in V$ and candidate $c \in C$, the utility $w_{v,c}$ is generated independently by $w_{v,c} = \eta \cdot \omega_c$, where η is a random variable drawn uniformly from $[0, 1]$.[6]

As in the work of Kleinberg & Raghavan (2018), applying bias then boils down to scaling down the utilities of disadvantaged candidates.

Definition 2.6 (**Utility-based biased generative model).** Given a bias parameter $\theta \in [0, 1]$, for every $v \in V$, $\widehat{w}_{v,c} = w_{v,c}$ for all $c \in G_1$, and $\widehat{w}_{v,c} = \theta \cdot w_{v,c}$ for all $c \in G_2$.

The definition of the utility-based bias model implies that the intrinsic utilities of all disadvantaged candidates reduce by a factor of θ. However, note that this seemingly uniform reduction can have different effects on different candidates: As an extreme example suppose that $G_1 = \{c_3, c_4, \dots, c_m\}$ and $G_2 = \{c_1, c_2\}$ with intrinsic utility values $\omega_1 = 10$ and $\omega_2 = \omega_3 = \cdots = \omega_m = 1$. For any θ, the probability that c_1's position in $\not\succ$ is worse than their position in $\succ$ is $1 - \Theta(\theta)$. Whereas the probability that the other disadvantaged candidate c_2 is placed at a worse position in $\not\succ$ compared to $\succ$ is with $1 - \Theta(\theta^m)$ much higher.

2.2.3. Order Preservation

To analyze the capabilities of representational constraints to mitigate bias, different ways in which the used multiwinner score function interacts with the used generative distributions for latent and biased preferences are relevant. Capturing this interplay, in this section we introduce two order-preservation properties of multiwinner score functions. Both of these properties hold for the utility-based model, but they may not hold for all generative models (Definitions 2.3 and 2.4). In the remainder of this section and in Sections 3.1 and 3.2, we present our results for general generative models. In Section 3.3 we describe the implications of our general results for the utility-based model and in Section F, we describe simplified variants of our general arguments and properties tailored to the utility-based model.

In order to be able to compute a committee with close-to-optimal utility in our algorithmic analysis, it will be crucial that if the number n of voters is large enough, then the optimal committee consists of the k candidates with the highest expected individual scores $\mathbb{E}_\mu [f(c, \succ)]$ for $c \in C$, as otherwise even if we would have access to the latent generative model computing an optimal committee might be computationally intractable. To ensure this, it will turn out that it is sufficient to assume that each candidate has an intrinsic quality (i.e., $\mathbb{E}_\mu [f(c, \succ)]$) and that this quality determines the ordering of candidates by relative marginal contributions to committees. More formally, we require that if c has a higher intrinsic quality than c', then c's marginal contribution to a committee S (for any S) is higher than that of c' (Item 1 in Definition 2.7). However, this still does not restrict how the difference between the intrinsic quality of candidates influences the difference between their marginal contributions to some committee. To address this, we also require that the difference in the marginal contributions of c and c' to committees is a non-increasing function with respect to adding candidates to the committee (Item 2 in Definition 2.7).

Definition 2.7 (**Order-preservation with respect to latent preferences).** Given a generative model μ for latent preference and a multiwinner score function $F = \sum_{v \in V} f(\cdot, \succ_v)$, we say F *is order-preserving with respect to* μ if for any two candidates $c \neq c' \in C$ belonging to the same group (G_1 or G_2) **if** $\mathbb{E}_\mu [f(c, \succ)] \leq \mathbb{E}_\mu [f(c', \succ)]$, **then** for any subsets $T \subseteq S \subseteq C \setminus \{c, c'\}$

1. $\mathbb{E}_\mu [f_S(c, \succ)] \leq \mathbb{E}_\mu [f_S(c', \succ)]$;
2. $\mathbb{E}_\mu [f_S(c', \succ)] - \mathbb{E}_\mu [f_S(c, \succ)]$
$$\leq \mathbb{E}_\mu [f_T(c', \succ)] - \mathbb{E}_\mu [f_T(c, \succ)] .$$

We prove in Lemma B.2 that any multiwinner score function F is order-preserving with respect to the utility-based latent generative model (Definition 2.5). The reason for this is that in the utility-based model the intrinsic quality of a candidate c is an increasing function of ω_c, as $\omega_c \leq \omega_{c'}$ implies that $\mathrm{pos}(c', \succ) < \mathrm{pos}(c, \succ)$ with probability at least $\frac{1}{2}$. The order preservation can then be shown using the domination sensitivity of F (Definition 2.2).

As an additional ingredient, we need a second (order-preservation) property controlling the relation between the latent μ and biased $\widehat{\mu}$ generative models: Otherwise, if we would allow $\widehat{\mu}$ to be arbitrarily different from μ, then one cannot hope to identify a committee with high latent utility by just observing biased preferences (generated according to $\widehat{\mu}$). In particular, our second order-preservation property restricts by how much the ratio between the marginal contribution of two candidates c and c' to some set S is allowed to change when bias is applied.

Definition 2.8 (**Order-preservation between latent and biased preferences).** Given a generative model μ for latent preference lists and a generative model $\widehat{\mu}$ for biased preference lists, a multiwinner score function $F = \sum_{v \in V} f(\cdot, \succ_v)$, and numbers $\beta, \gamma \in [0, 1]$, we say F is (β, γ) *order preserving between* μ *and* $\widehat{\mu}$ if for any two candidates $c \neq c' \in C$ belonging to the same group (G_1 or G_2) and any $S \subseteq C \setminus \{c, c'\}$, **if** $\beta \cdot \mathbb{E}_\mu [f_S(c', \succ)] \geq \mathbb{E}_\mu [f_S(c, \succ)] > 0$, **then** $\gamma \cdot \mathbb{E}_{\widehat{\mu}} [f_S(c', \not\succ)] \geq \mathbb{E}_{\widehat{\mu}} [f_S(c, \not\succ)]$.

[6]More generally, one can consider other distributions and, we do so, in Definitions B.3, B.8 and C.7.

In this definition, β specifies the range of candidates affected by the property and $1-\gamma$ is the allowed gap between their relative marginal contributions that can emerge by applying bias: When β is close to 1 then we consider candidate pairs c, c' and sets S with a large range of $\frac{\mathbb{E}_\mu[f_S(c,\succ)]}{\mathbb{E}_\mu[f_S(c',\succ)]}$. In contrast, as γ goes to one, expected marginal contributions are allowed to become more and more similar even in case there is a substantial gap under μ (and, hence, candidates get harder to distinguish).

In case no bias is applied to the preferences, for each candidate c and set S the expected marginal contribution of c to S is the same in both μ and $\hat{\mu}$ (notably, this can also happen in other cases, for instance, if both the latent and biased preferences of a voter are independently drawn from the uniform distribution on $\mathcal{L}(C)$). If all expected marginal contributions remain unchanged, every multiwinner score function is (β, β) order preserving between μ and $\widehat{\mu}$ for any $\beta \in (0, 1]$. As shown in Lemma B.2, many multiwinner score functions (including the examples in Section 2.1) are (β, γ) order preserving between the utility-based latent and biased generative model for any $0 \leq \beta \leq 1-\lambda$ and $1 - m^{-\Theta(1)} \cdot \lambda \leq \gamma \leq 1$ where $\lambda = m^{-\Theta(1)}$. Intuitively, this is because we apply the same multiplicative bias to all members of the disadvantaged group.

Recall that we have observed in Section 2.2.1 that Definition 2.4 also allows the biased preference lists to be drawn from a mixture of two or more distributions. In the case of mixtures of utility-based models, certain order-preservation properties translate from the individual components to the mixture. Concretely, assume that $(\mu, \widehat{\mu}_1)$ and $(\mu, \widehat{\mu}_2)$ are two utility-based generative models with the same intrinsic values ω but heterogeneous bias parameters θ_1 and θ_2 respectively. Let us consider mixtures of $(\mu, \widehat{\mu}_1)$ and $(\mu, \widehat{\mu}_2)$. Using linearity of expectation, it can be shown that if F is (β_i, γ_i) order-preserving between μ and $\widehat{\mu}_i$ for each $i \in \{1, 2\}$, then F is $(\min\{\beta_1, \beta_2\}, \max\{\gamma_1, \gamma_2\})$ order-preserving between μ and $\delta\widehat{\mu}_1 + (1-\delta)\widehat{\mu}_2$ for any $\delta \in (0, 1)$.

Swapping-based bias model. In Section B.3, we present a swapping-based bias model inspired by the popular Mallows model (Mallows, 1957), which is also captured by the general bias model. Given a latent preference list $\succ \in \mathcal{L}(C)$, for $t \geq 1$ iterations, this model selects two candidates $c \in G_1$ and $c' \in G_2$ with c' being ranked above c and swaps their position, where the average difference in the position of swapped candidates can be controlled by a parameter $\phi \in [0, 1]$. We prove in Lemma B.13 that for $\lambda = \Theta(t\phi)$, all multiwinner score functions are $(1-\lambda, 1-\frac{\lambda}{2})$ order preserving between the utility-based latent generative model μ and swapping-based biased generative model $\widehat{\mu}$. Intuitively, β is bounded away from 1 because, unlike in the utility-based bias model, in the swapping-based model bias acts non-uniformly across members of the disadvantaged groups (depending on their positions relative to advantaged candidates in voters' preference lists).

2.3. Representational Constraints

Let $S^\star := \arg\max_{S \subseteq C: |S|=k} F(S)$ denote the optimal solution of the underlying multiwinner voting problem, and let $\mathrm{OPT} := F(S^\star)$. Furthermore, let $M \subseteq C$ denote the collection of k candidates c with the highest value $\mathbb{E}_\mu[f(c, \succ)]$ (as discussed above if F satisfies order-preservation with respect to the latent generative model μ and n is high enough, $S^\star$ is likely to be equal to M).

A natural approach to solve the multiwinner voting problem in the presence of bias is to directly solve $\widehat{S} := \arg\max_{S \subseteq C:\, |S|=C} \widehat{F}(S)$. However, due to biases, this may lead to committees with poor latent quality. In fact, we argue in Remark D.2 that one can lose significant value by not selecting candidates from G_2, i.e., $F(\widehat{S}) \leq (1 - \frac{|G_2 \cap S^\star|}{2k}) \cdot \mathrm{OPT}$. To mitigate such adverse effects of bias, a popular intervention are *representational constraints* that require the output subset to include at least a specified number of candidates from the disadvantaged group.[7]

Definition 2.9 (**Representational constraints).** Given integer $0 \leq \ell \leq k$, the ℓ-representational constraint requires $|S \cap G_2| \geq \ell$ for the selected committee $S \subseteq C$. Let $\mathcal{K}(\ell)$ denote the collection of subsets that satisfy the ℓ-representational constraint.

In the presence of the ℓ-representational constraint, the straightforward output subset, say $\widehat{S}_\ell$, is a solution to the following optimization problem: $\max_{S \in \mathcal{K}(\ell): |S|=k} \widehat{F}(S)$. This paper centers around the following problem:

Problem 2 (**Effectiveness of representational constraints**). Is there an integer $0 \leq \ell \leq k$ such that $F(\widehat{S}_\ell) \approx \mathrm{OPT}$ (with high probability), where the inputs are as specified in Problem 1? Moreover, is there a polynomial-time algorithm that outputs a set $S \in K(\ell)$ with $F(S) \approx \mathrm{OPT}$?

3. Algorithmic Results for Problem 2

In this section, we show that representational constraints can help mitigate the adverse effects of bias in ranking-based subset selection with multiwinner score functions. Specifically, in Theorem 3.2, we provide a sufficient condition and an efficient algorithm for Problem 2. Our algorithmic result is based on a new notion of smoothness (Definition 3.1), which distinguishes the capabilities of multiwinner score

[7]Note that enforcing representational constraints requires knowledge of the sets of advantaged and disadvantaged candidates. While they may not be available or costly to obtain in some contexts, e.g., web-search (see (Mehrotra & Vishnoi, 2022) and the references therein), they are available in relevant contexts such as elections when groups are based on (combinations of) socially-salient attributes (such as race, gender, age, and disability) (Evequoz et al., 2022).

functions to be debiased by representational constraints. Afterward, in Section 3.3, we discuss specific implications of Theorem 3.2, among others, highlighting that under the utility-based model the sufficient number of voters to recover a close-to-optimal utility for SNTV or Chamberlin-Courant is exponential in m, whereas this dependence is only polynomial for Borda.

3.1. Smoothness

In a preliminary theoretical and empirical analysis, we observed that SNTV requires a significantly higher number of voters to recover the same solution quality as Borda when using representational constraints. We provide a theoretical demonstration of this contrast in Theorem 3.3. To quantitatively measure such differences between functions, we introduce the following notion of smoothness of a multiwinner score function F, which quantifies the ability of representational constraints to recover a latent utility that is close to optimal under F.

To build some intuition, let us focus on a case where F is order-preserving with respect to μ and $(1, 0.01)$ order-preserving between $(\mu, \hat{\mu})$. Then, the set M of candidates with the highest expected value $\mathbb{E}_{\mu}\left[f(c, \succ)\right]$ is likely to be an optimal solution and those candidates are also likely to have the highest expected value $\mathbb{E}_{\widehat{\mu}}\left[f(c, \succ)\right]$. In this case, our algorithm only needs to identify all and, in particular, the weakest candidate c from M using samples in $\widehat{R} = \{\not\succ_v \in \mathcal{L}(C) \colon v \in V\}$. The marginal contribution of this candidate can be as low as $\mathbb{E}_{\widehat{\mu}}\left[f_{C\setminus\{c\}}(c, \not\succ)\right]$. Let $\tau_1(f)$ be the maximum possible expected score $\mathbb{E}_{\mu}\left[f(c, \succ)\right]$ of a candidate. To capture how many samples are needed to identify the weakest candidate c from M, we introduce a new parameter α which needs to be at least $\alpha \geq \frac{1}{\tau_1(f)} \min_{c\in M} \mathbb{E}_{\widehat{\mu}}\left[f_{C\setminus\{c\}}(c, \not\succ)\right]$ (note that $\tau_1(f)$ acts as a normalization). When α is close to 0, it is "more difficult" to distinguish candidates in M from candidates in $C \setminus M$ since margin values of some candidates from M are more likely to be close to 0, and hence, a larger number of voters is required to distinguish them.

This observation can be related to our initial finding regarding SNTV and Borda. SNTV requires more voters to identify the strength of a candidate, as only the top choice of a voter is taken into account. For instance, in the presence of a strong candidate $c^\star$, who is ranked first with a high probability, it may take many samples to observe the first vote where $c^\star$ is not ranked first. In contrast, for Borda, candidate strength can be distinguished much more easily, as every sampled vote provides new information on all candidates.

Note that in a different direction, we also need a sufficient number of voters to ensure that candidates from M are likely to constitute an optimal solution. For this, we again need that the sampled votes are enough so that candidates from M have a higher contribution than candidates from $C \setminus M$ with respect to the latent votes; accordingly, similar to the above we also require that $\alpha \geq \frac{1}{\tau_1(f)} \min_{c\in M} \mathbb{E}_{\mu}\left[f_{C\setminus\{c\}}(c, \succ)\right]$.

Combining these observations with the order-preservation properties, we propose the following definition of smoothness. On an intuitive level, our smoothness definition boils down to how well the multiwinner score function F can "correctly distinguish" the strength of candidates among the same group under the latent and biased preferences.

Definition 3.1 (Smoothness). Let $F = \sum_{v\in V} f(\cdot, \succ_v)$ be a multiwinner score function. Let $(\mu, \widehat{\mu})$ be a generative model defined in Definitions 2.7 and 2.8. Given $\alpha, \beta, \gamma \in [0, 1]$, we say F is (α, β, γ)-smooth with respect to $(\mu, \widehat{\mu})$ if the following holds

- $\alpha \geq \frac{1}{\tau_1(f)} \min_{c\in M} \mathbb{E}_{\widehat{\mu}}\left[f_{C\setminus\{c\}}(c, \not\succ)\right]$ and $\alpha \geq \frac{1}{\tau_1(f)} \min_{c\in M} \mathbb{E}_{\mu}\left[f_{C\setminus\{c\}}(c, \succ)\right]$,
- F is order-preserving with respect to μ; and
- F is (β, γ) order preserving between μ and $\widehat{\mu}$.

We have already argued before that α influences the number of voters required to identify a close-to-optimal committee. The parameter γ also influences the number of voters needed but in this case, the larger γ gets the more votes are needed: If γ is close to 1, then two candidates c, c' with $\mathbb{E}_{\mu}\left[f(c, \succ)\right] \ll \mathbb{E}_{\mu}\left[f(c', \succ)\right]$, may have a similar quality in the presence of bias, $\mathbb{E}_{\widehat{\mu}}\left[f(c, \not\succ)\right] \approx \mathbb{E}_{\widehat{\mu}}\left[f(c', \not\succ)\right]$, making it harder to identify the stronger candidate needed for a close-to-optimal solution. In contrast to α and γ, factor β bounds how close to the optimum one can get when observing biased voters: Intuitively, a value of β close to 1 implies that for two candidates the order of their marginal contributions remains unchanged when applying the bias. In contrast to this, for $\beta < 1$, for two candidates c, c' with $\mathbb{E}_{\mu}\left[f_S(c', \succ)\right] > \mathbb{E}_{\mu}\left[f_S(c, \succ)\right]$ and $\frac{\mathbb{E}_{\mu}[f_S(c,\succ)]}{\mathbb{E}_{\mu}[f_S(c',\succ)]} \geq \beta$ for some set S, we allow the ratio of their marginal contributions to S to change arbitrarily in the presence of bias. Consequently, for such pairs of candidates c and c', even if the number of observed biased voters goes to infinity, we will not be able to distinguish which of the two is the stronger candidate under the latent distribution, leading to a potential multiplicative loss of β in the latent quality of the output solution due to wrongfully including c instead of c' in the returned solution.

3.2. Main Theorem

As discussed above, for an (α, β, γ)-smooth function, the values of α and γ determine the number of samples required for representational constraints to return an approximately optimal solution, while the value of β bounds the achievable multiplicative approximation factor. Our main algorithmic result matches these intuitions, and we provide a sufficient condition on n under which representational constraints recover a solution S that is close to optimal. The proof of this result appears in Section C.1.

Theorem 3.2 (Algorithmic result for Problem 2). *Let $F\colon 2^C \to \mathbb{R}_{\geq 0}$ be a multiwinner score function. Let μ and $\widehat{\mu}$ be generative models of latent and biased preference lists respectively. Suppose F is (α, β, γ)-smooth with respect to $(\mu, \widehat{\mu})$ for some $\alpha, \beta, \gamma \in [0,1]$. For any $0 < \varepsilon, \delta < 1$, if*

$$n \geq \frac{16k}{(\alpha \min\{\varepsilon, 1-\gamma\})^2} \cdot \log \frac{m}{\delta},$$

there is an algorithm that given groups G_1, G_2, numbers k, $\ell = |M \cap G_2|$, and a value oracle $\mathcal{O}$ to $\widehat{F}(\cdot)$ as input, outputs a subset $S \in \mathcal{K}(\ell)$ of size k such that

$$\Pr_{\mu,\widehat{\mu}}\left[F(S) \geq (\beta - \varepsilon) \cdot \mathrm{OPT}\right] \geq 1 - \delta.$$

The algorithm makes $O(mk)$ calls to oracle $\mathcal{O}$ and performs $O(m \log m)$ arithmetic operations.

The algorithm underlying Theorem 3.2 is a standard greedy algorithm (Algorithm 1 in Section C.1) that maximizes $\widehat{F}$ subjected to representational constraint $\ell = |M \cap G_2|$. The key idea used in the proof of Theorem 3.2 is that, due to the smoothness of F, when Algorithm 1 adds the ith candidate to the committee, the incurred marginal contribution with respect to the latent preferences is at least a β fraction compared to when building M and adding the ith highest scoring candidate to a set of the $(i-1)$ highest scoring candidates (Lemma C.4). Note that due to the greedy nature of our algorithm, the output solution S may not be identical to $\widehat{S}_\ell$ for some ℓ (recall $\widehat{S}_\ell = \arg\max_{S \in \mathcal{K}(\ell):|S|=k} \widehat{F}(S)$). However, for modular score functions such as SNTV and Borda, the algorithm always outputs $S = \widehat{S}_{|M \cap G_2|}$. Hence, for some multiwinner score functions, Theorem 3.2 also implies $F(\widehat{S}_\ell) \geq (\beta - \varepsilon) \cdot \mathrm{OPT}$ for $\ell = |M \cap G_2|$, which partially addresses the first question of Problem 2.

Note that the value $|M \cap G_2|$ is unknown in advance. While in general, this value depends on F and the generative models $(\mu, \widehat{\mu})$, there are also natural special cases where it is independent. For instance, if we assume that preference lists drawn from μ are not systematically skewed toward candidates in either group, as may be the case in the real world (Evequoz et al., 2022), then $|M \cap G_2| \approx k \cdot \frac{|G_2|}{m}$ with probability $1 - o_k(1)$. Moreover, in applications such as recommendation systems that use multiwinner scoring functions (Streviniotis & Chalkiadakis, 2022a), ℓ may be tuned via A/B testing by trying different ℓ, estimating latent quality from user engagement, and selecting the value of ℓ that maximizes the latent quality.

It is worth noting that although α has a similar form as the curvature of submodular functions, there are some differences between them that render the curvature ineffective in measuring the effectiveness of representational constraints. For instance, the curvature is unable to distinguish modular functions. We discuss this and the relevance of Definition 3.1 to research on the effect of noise on multiwinner voting in Section C.2.

Proof overview of Theorem 3.2. The smoothness condition is defined with respect to expectations over the generative models but in Theorem 3.2 we have only access to n samples. The proof's first component is a concentration inequality showing that expectations over samples are close to the true expectations: for any candidate $c \in C$ and committee $T \subseteq C$ "of interest," $\mathbb{E}_\mu[f_T(c, \succ)] \approx \frac{1}{n} F_T(c)$ and $\mathbb{E}_{\widehat{\mu}}[f_T(c, \not\succ)] \approx \frac{1}{n} \widehat{F}_T(c)$ (Lemma C.3), where $F_T(c) := F(T \cup \{c\}) - F(T)$ and $\widehat{F}_T(x) := \widehat{F}(T \cup \{c\}) - \widehat{F}(T)$. Let $S \in \mathcal{K}(\ell)$ be the subset output by the algorithm. Recall that $M \subseteq C$ is the set of k candidates c with the highest value $\mathbb{E}_\mu\left[f(c, \succ)\right]$ and $\mathrm{OPT} := \max_{S \subseteq C:|S|=k} F(S)$. The proof strategy is to show that $F(M) \approx \mathrm{OPT}$ (Lemma C.2) and to compare the utility of "prefixes" of S to "prefixes" of M (Lemma C.4). For this, for large enough n, we show that our algorithm has the following property (Equation (41)): there is an ordering $m_1, \ldots, m_k$ of elements in M such that

$$\forall_{i \in [\ell]}, \quad \widehat{F}_{\{s_1, \ldots, s_{i-1}\}}(s_i) \geq \gamma \cdot \widehat{F}_{\{s_1, \ldots, s_{i-1}\}}(m_i), \qquad (1)$$

where s_j is the j-th item added to S for any j. If F is modular, the remainder of the proof is straightforward: Due to order preservation between μ and $\widehat{\mu}$, Equation (1) implies that $F_{\{s_1,\ldots,s_{i-1}\}}(s_i) \geq \beta \cdot F_{\{s_1,\ldots,s_{i-1}\}}(m_i)$ for each i, and, since F is modular, $F(S) = \sum_i F(s_i) \geq \beta \cdot \sum_i F(m_i) = \beta \cdot F(M)$. When F is not modular, we need to show that Equation (1) implies that $F_{\{s_1,\ldots,s_{i-1}\}}(s_i) \geq \beta \cdot F_{\{m_1,\ldots,m_{i-1}\}}(m_i)$ (note the change in the base). We do so in Lemma C.4 and Equations (52) and (53) using order preservation with respect to μ.

3.3. Applications of Theorem 3.2

In this section, we focus on the utility-based generative models $(\mu, \widehat{\mu})$ (Definition 2.6) and derive bounds for some specific multiwinner score functions; see Tables 2 and 3 in the appendix for a summary of all results. Lemmas B.2 and B.7 give bounds on β and γ for this case. We provide the missing values of α for some multiwinner score functions and the resulting sufficient numbers of voters using Theorem 3.2 in the following result, whose proof appears in Section C.3.

Theorem 3.3 (Algorithmic result for the utility-based generative model; Informal). *Let $(\mu, \widehat{\mu})$ be the utility-based generative models from Definitions 2.5 and 2.6 with bias parameter $\theta \in (0,1]$. For ℓ_1-CC and SNTV it holds that $\alpha \geq \Theta\left(\theta^{-2(m-1)}\right)$, and for Borda that $\alpha \geq \Theta\left(\theta^{-2}\right)$.*

Using this and Lemma C.8, Theorem 3.2 applies for

1. *SNTV and ℓ_1-CC with $n \geq \theta^{-2(m-1)} \cdot m^{\Theta(1)} \varepsilon^{-2}$ and $\beta = 1 - m^{-\Theta(1)}$; and*
2. *Borda with $n \geq \theta^{-2} \cdot m^{\Theta(1)} \varepsilon^{-2}$ and $\beta = 1 - m^{-\Theta(1)}$.*

Note that it is quite intuitive that the above computed α values depend on θ as in the utility-based model θ controls the multiplicative gap between the scores awarded to candidates from G_2 compared to G_1 and, thus, the value of $\frac{1}{\tau_1(f)} \min_{c\in M} \mathbb{E}_{\widehat{\mu}} \left[f_{C\setminus\{c\}}(c, \not\succ)\right]$.

The sufficient number of voters in the above result varies significantly depending on the multiwinner score function: on the one hand, for ℓ_1-CC rule and SNTV the dependence is $\theta^{-O(m)}$, on the other hand, the dependence is only θ^{-2} for the Borda rule. We can also prove that these dependencies are not only sufficient but also necessary by providing an impossibility result (Theorem D.3) that shows that representational constraints cannot recover an (approximately) optimal solution if n is "substantially" smaller than these bounds; see Section D for more discussions. Combined with our impossibility result, Theorem 3.3 shows a stark contrast between different score functions, e.g., SNTV and Borda, under the utility-based model, implying that the latter function is advantageous in the presence of implicit bias.

The above results extend to certain mixtures of generative models. For instance, if F is $(\alpha_1, \beta_1, \gamma_1)$-smooth with respect to $(\mu, \widehat{\mu}_1)$ and $(\alpha_2, \beta_2, \gamma_2)$-smooth with respect to $(\mu, \widehat{\mu}_2)$, then for any $\delta \in (0, 1)$, it can be shown that F is $(\delta\alpha_1 + (1-\delta)\alpha_1, \min\{\beta_1, \beta_2\}, \max\{\gamma_1, \gamma_2\})$-smooth with respect to the mixture $(\mu, \delta\widehat{\mu}_1 + (1-\delta)\widehat{\mu}_2)$; this follows from Definition 3.1 and linearity of expectation.

A tool for analyzing multiwinner score functions (Section G). In addition to the above computations for *existing* bias models and multiwinner score functions, we also provide code to estimate the smoothness of *new* multiwinner score functions with respect to *new* generative models (Section G). The code takes as input oracles that (1) evaluate the multiwinner score function F and (2) sample from generative models $(\mu, \widehat{\mu})$. First, for specified m and k, it outputs estimates $(\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\gamma})$ of the smoothness of F with respect to $(\mu, \widehat{\mu})$, along with corresponding confidence intervals (implied by a concentration inequality; Lemma C.3). This allows for theoretical estimates of the capabilities of representational constraints using our main result (Theorem 3.2).

Second, given values of n, m, and k, it estimates the fraction of the optimal score recovered by representational constraints for F with respect to the given $(\mu, \widehat{\mu})$. In Section G, we illustrate the code using a set of latent generative models provided by Szufa et al. (2020) in combination with the swapping-based bias model (Definition B.12). In line with Theorem 3.3, our observations in these simulations show a stark contrast between SNTV and Borda: for all values of n and generative models we consider, representational constraints recover a significantly larger fraction of the maximum achievable latent score for Borda than for SNTV.

4. Conclusion and Future Work

In this paper, we have investigated the effectiveness of representational constraints in the presence of bias, in a variant of the subset selection problem with rankings of items as input. To convert multiple rankings into scores for subsets, we leverage ideas from multiwinner voting. Extending a line of research on (re-)designing algorithms when the inputs might suffer from biases (Kleinberg & Raghavan, 2018), we demonstrate that representational constraints continue to have the power to improve the latent quality of the solution in the setting of multiple inputs and submodular (instead of modular) score functions. Our work brings out differences in the effectiveness of representational constraints depending on the (sub)modular score function used and can be used to guide the choice of multiwinner score functions for subset selection; we further provide a tool to enable the latter. To carry out this analysis, we develop a notion of smoothness of a submodular function that might be of further interest.

In addition to the already covered submodular multiwinner score functions, it would be interesting to extend our work to sequential rules such as STV, greedy CC, and greedy Monroe (Faliszewski et al., 2017; Lackner & Skowron, 2023). Designing interventions for debiasing outcomes of these rules seems challenging as it is not clear how representational constraints can be implemented here (Celis et al., 2018). As already discussed in the introduction, instead of assuming that voters rank the items, each voter could also provide a numerical (utility) score for each candidate (Kleinberg et al., 2004; Skowron et al., 2016). The class of submodular functions (Definition 2.2) we study contains some functions relevant in this setting: if all voters assign the same utility to their ith most preferred candidate for every i. On a more general note, studying whether representational constraints continue to be effective for further classes of (non-separable) submodular functions is an interesting direction but is likely to require a new approach, as our smoothness definition and bias model assume that the function is separable.

Finally, we remark that we study only one aspect of real-world selection problems and other aspects must be carefully considered to avoid possible negative impacts of intervention constraints. Indeed, intervention constraints can help if they are used appropriately but may potentially also have negative consequences, e.g., imposing intervention constraints for one disadvantaged group may harm a different not-considered disadvantaged group.

Acknowledgements

This project is supported in part by NSF Awards (CCF-2112665 and IIS-2045951). NB is supported by the DFG project ComSoc-MPMS (NI 369/22).

References

Agrawal, R., Gollapudi, S., Halverson, A., and Ieong, S. Diversifying search results. In *Proceedings of the Second International Conference on Web Search and Web Data Mining, WSDM 2009, Barcelona, Spain, February 9-11, 2009*, pp. 5–14. ACM, 2009.

Allouche, T., Lang, J., and Yger, F. Multi-winner approval voting goes epistemic. In *Uncertainty in Artificial Intelligence, Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence, UAI 2022, 1-5 August 2022, Eindhoven, The Netherlands*, volume 180 of *Proceedings of Machine Learning Research*, pp. 75–84. PMLR, 2022.

Angelidakis, H., Kurpisz, A., Sering, L., and Zenklusen, R. Fair and fast k-center clustering for data summarization. In *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pp. 669–702. PMLR, 2022.

Awasthi, P., Blum, A., Sheffet, O., and Vijayaraghavan, A. Learning mixtures of ranking models. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pp. 2609–2617, 2014.

Aziz, H., Gaspers, S., Gudmundsson, J., Mackenzie, S., Mattei, N., and Walsh, T. Computational aspects of multi-winner approval voting. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2015, Istanbul, Turkey, May 4-8, 2015*, pp. 107–115. ACM, 2015.

Bhattacharyya, A. and Dey, P. Predicting winner and estimating margin of victory in elections using sampling. *Artif. Intell.*, 296:103476, 2021.

Blum, A. and Stangl, K. Recovering from biased data: Can fairness constraints improve accuracy? In *1st Symposium on Foundations of Responsible Computing, FORC 2020, June 1-3, 2020, Harvard University, Cambridge, MA, USA (virtual conference)*, volume 156 of *LIPIcs*, pp. 3:1–3:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.

Bogen, M. and Rieke, A. Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias, December 2018. https://www.upturn.org/reports/2018/hiring-algorithms/.

Bredereck, R., Faliszewski, P., Igarashi, A., Lackner, M., and Skowron, P. Multiwinner elections with diversity constraints. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pp. 933–940. AAAI Press, 2018.

Burges, C. J. From RankNet to LambdaRank to LambdaMART: An Overview. *Learning*, 2010.

Busa-Fekete, R., Hüllermeier, E., and Szörényi, B. Preference-based rank elicitation using statistical models: The case of mallows. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, volume 32 of *JMLR Workshop and Conference Proceedings*, pp. 1071–1079. JMLR.org, 2014.

Capers IV, Q., Clinchot, D., McDougle, L., and Greenwald, A. G. Implicit racial bias in medical school admissions. *Academic Medicine*, 92(3):365–369, 2017.

Caragiannis, I. and Micha, E. Learning a ground truth ranking using noisy approval votes. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, IJCAI'17, pp. 149–155. AAAI Press, 2017a. ISBN 9780999241103.

Caragiannis, I. and Micha, E. Learning a ground truth ranking using noisy approval votes. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pp. 149–155. ijcai.org, 2017b.

Caragiannis, I., Procaccia, A. D., and Shah, N. When do noisy votes reveal the truth? *ACM Trans. Econ. Comput.*, 4(3), mar 2016. ISSN 2167-8375. doi: 10.1145/2892565. URL https://doi.org/10.1145/2892565.

Caragiannis, I., Kaklamanis, C., Karanikolas, N., and Krimpas, G. A. Evaluating approval-based multiwinner voting in terms of robustness to noise. *Auton. Agents Multi Agent Syst.*, 36(1):1, 2022.

Celis, L. E., Huang, L., and Vishnoi, N. K. Multiwinner voting with fairness constraints. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, pp. 144–151. ijcai.org, 2018.

Celis, L. E., Mehrotra, A., and Vishnoi, N. K. Interventions for ranking in the presence of implicit bias. In *FAT* '20: Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, January 27-30, 2020*, pp. 369–380. ACM, 2020.

Celis, L. E., Hays, C., Mehrotra, A., and Vishnoi, N. K. The effect of the rooney rule on implicit bias in the long term. In *FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event / Toronto, Canada, March 3-10, 2021*, pp. 678–689. ACM, 2021.

Chakraborty, A., Patro, G. K., Ganguly, N., Gummadi, K. P., and Loiseau, P. Equality of voice: Towards fair representation in crowdsourced top-k recommendations. In *Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* 2019, Atlanta, GA, USA, January 29-31, 2019*, pp. 129–138. ACM, 2019.

Chamberlin, J. R. and Courant, P. N. Representative deliberations and representative decisions: Proportional representation and the borda rule. *American Political Science Review*, 77(3):718–733, 1983. doi: 10.2307/1957270.

Chapman, E. N., Kaatz, A., and Carnes, M. Physicians and implicit bias: how doctors may unwittingly perpetuate health care disparities. *Journal of general internal medicine*, 28(11):1504–1510, 2013.

Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., and Amodei, D. Deep reinforcement learning from human preferences. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Collas, F. and Irurozki, E. Concentric mixtures of mallows models for top-k rankings: sampling and identifiability. In *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pp. 2079–2088. PMLR, 2021.

Collins, B. W. Tackling unconscious bias in hiring practices: The plight of the rooney rule. *NYUL Rev.*, 82:870, 2007.

Conforti, M. and Cornuéjols, G. Submodular set functions, matroids and the greedy algorithm: Tight worst-case bounds and some generalizations of the rado-edmonds theorem. *Discrete Applied Mathematics*, 7(3):251–274, 1984.

Dastin, J. Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*, 2018.

Deczynski, R. Hiring Is a Pain–These 5 Companies Say A.I. Can Make It Better. *Inc.*, 2021.

Dey, P., Kar, D., and Sanyal, S. Sampling-based winner prediction in district-based elections. *CoRR*, abs/2203.00083, 2022.

Dressel, J. and Farid, H. The accuracy, fairness, and limits of predicting recidivism. *Science advances*, 4(1):eaao5580, 2018.

El-Arini, K., Veda, G., Shahaf, D., and Guestrin, C. Turning down the noise in the blogosphere. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, June 28 - July 1, 2009*, pp. 289–298. ACM, 2009.

Elhamifar, E. and Kaluza, M. C. D. P. Subset selection and summarization in sequential data. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pp. 1035–1045, 2017.

Elkind, E., Faliszewski, P., Laslier, J., Skowron, P., Slinko, A., and Talmon, N. What do multiwinner voting rules do? an experiment over the two-dimensional Euclidean domain. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*, pp. 494–501. AAAI Press, 2017a.

Elkind, E., Faliszewski, P., Skowron, P., and Slinko, A. Properties of multiwinner voting rules. *Soc. Choice Welf.*, 48(3):599–632, 2017b.

Emelianov, V., Gast, N., Gummadi, K. P., and Loiseau, P. On fair selection in the presence of implicit variance. In *EC '20: The 21st ACM Conference on Economics and Computation, Virtual Event, Hungary, July 13-17, 2020*, pp. 649–675. ACM, 2020.

Emelianov, V., Gast, N., Gummadi, K. P., and Loiseau, P. On fair selection in the presence of implicit and differential variance. *Artif. Intell.*, 302:103609, 2022a.

Emelianov, V., Gast, N., and Loiseau, P. Fairness in selection problems with strategic candidates. In *EC '22: The 23rd ACM Conference on Economics and Computation, Boulder, CO, USA, July 11 - 15, 2022*, pp. 375–403. ACM, 2022b.

Evéquoz, F., Rochel, J., Keswani, V., and Celis, L. E. Diverse representation via computational participatory elections - lessons from a case study. In *Equity and Access in Algorithms, Mechanisms, and Optimization, EAAMO 2022, Arlington, VA, USA, October 6-9, 2022*, pp. 12:1–12:11. ACM, 2022.

Evequoz, F., Rochel, J., Keswani, V., and Celis, L. E. Diverse representation via computational participatory elections - lessons from a case study. In *Equity and Access in Algorithms, Mechanisms, and Optimization*, EAAMO '22, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450394772.

Faenza, Y., Gupta, S., and Zhang, X. Reducing the feeder effect in public school admissions: A bias-aware analysis for targeted interventions. *CoRR*, abs/2004.10846, 2020.

Faliszewski, P., Skowron, P., Slinko, A., and Talmon, N. Multiwinner voting: A new challenge for social choice theory. *Trends in computational social choice*, 74(2017): 27–47, 2017.

Fujishige, S. *Submodular functions and optimization*. Elsevier, 2005.

Garg, N., Li, H., and Monachou, F. Standardized tests and affirmative action: The role of bias and variance. In *FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event / Toronto, Canada, March 3-10, 2021*, pp. 261. ACM, 2021.

Gawron, G. and Faliszewski, P. Using multiwinner voting to search for movies. In *Multi-Agent Systems - 19th European Conference, EUMAS 2022, Düsseldorf, Germany, September 14-16, 2022, Proceedings*, volume 13442 of *Lecture Notes in Computer Science*, pp. 134–151. Springer, 2022.

Greenwald, A. G. and Lai, C. K. Implicit social cognition. *Annual Review of Psychology*, 71(1):419–445, 2020.

Griffin, D. and Brenner, L. Perspectives on probability judgment calibration. *Blackwell handbook of judgment and decision making*, 199:158–177, 2004.

Guiver, J. and Snelson, E. L. Bayesian inference for plackett-luce ranking models. In *Proceedings of the 26th Annual International Conference on Machine Learning, ICML 2009, Montreal, Quebec, Canada, June 14-18, 2009*, volume 382 of *ACM International Conference Proceeding Series*, pp. 377–384. ACM, 2009.

Guyon, I. and Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.*, 3:1157–1182, 2003.

Heidari, H. and Kleinberg, J. Allocating opportunities in a dynamic model of intergenerational mobility. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, pp. 15–25, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450383097. doi: 10.1145/3442188.3445867. URL https://doi.org/10.1145/3442188.3445867.

Jameson, A. and Smyth, B. Recommendation to groups. In *The adaptive web*, pp. 596–627. Springer, 2007.

Jiang, H. and Nachum, O. Identifying and correcting label bias in machine learning. In *The 23rd International Conference on Artificial Intelligence and Statistics, AISTATS 2020, 26-28 August 2020, Online [Palermo, Sicily, Italy]*, volume 108 of *Proceedings of Machine Learning Research*, pp. 702–712. PMLR, 2020.

Kang, J., Bennett, M., Carbado, D., Casey, P., and Levinson, J. Implicit bias in the courtroom. *UCLa L. rev.*, 59:1124, 2011.

Kay, M., Matuszek, C., and Munson, S. A. Unequal representation and gender stereotypes in image search results for occupations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI 2015, Seoul, Republic of Korea, April 18-23, 2015*, pp. 3819–3828, Seoul, Republic of Korea, 2015. ACM.

Kite, M. E. and Whitley, B. E. *Psychology of prejudice and discrimination*. Routledge, 2016.

Kleinberg, J. M. and Raghavan, M. Selection problems in the presence of implicit bias. In Karlin, A. R. (ed.), *9th Innovations in Theoretical Computer Science Conference, ITCS 2018, January 11-14, 2018, Cambridge, MA, USA*, volume 94 of *LIPIcs*, pp. 33:1–33:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.

Kleinberg, J. M., Papadimitriou, C. H., and Raghavan, P. Segmentation problems. *J. ACM*, 51(2):263–280, 2004.

Krause, A. and Golovin, D. Submodular function maximization. *Tractability*, 3:71–104, 2014.

Lackner, M. and Skowron, P. A quantitative analysis of multi-winner rules. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pp. 407–413. ijcai.org, 2019.

Lackner, M. and Skowron, P. *Multi-Winner Voting with Approval Preferences*. Springer Briefs in Intelligent Systems. Springer, 2023.

Liu, A. and Moitra, A. Efficiently learning mixtures of mallows models. In *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pp. 627–638. IEEE Computer Society, 2018.

Liu, T.-Y. Learning to rank for information retrieval. *Foundations and Trends® in Information Retrieval*, 3(3):225–331, 2009. ISSN 1554-0669. doi: 10.1561/1500000016.

Lu, T. and Boutilier, C. Effective sampling and learning for mallows models with pairwise-preference data. *J. Mach. Learn. Res.*, 15(1):3783–3829, 2014.

Magrino, T. R., Rivest, R. L., and Shen, E. Computing the margin of victory in IRV elections. In *2011 Electronic Voting Technology Workshop / Workshop on Trustworthy Elections, EVT/WOTE '11, San Francisco, CA, USA, August 8-9, 2011*. USENIX Association, 2011.

Mallows, C. L. Non-null ranking models. I. *Biometrika*, 44 (1/2):114–130, 1957.

McSherry, D. Diversity-conscious retrieval. In *European Conference on Case-Based Reasoning*, pp. 219–233. Springer, 2002.

Mehrotra, A. and Vishnoi, N. K. Fair ranking with noisy protected attributes. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A. (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 31711–31725. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/cdd0640218a27e9e2c0e52e324e25db0-Paper-Conference.pdf.

Mehrotra, A. and Vishnoi, N. K. Maximizing submodular functions for recommendation in the presence of biases. In *Proceedings of the ACM Web Conference 2023*, WWW '23, pp. 3625–3636, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450394161. doi: 10.1145/3543507.3583195. URL https://doi.org/10.1145/3543507.3583195.

Mehrotra, A., Pradelski, B. S. R., and Vishnoi, N. K. Selection in the presence of implicit bias: The advantage of intersectional constraints. In *FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul, Republic of Korea, June 21 - 24, 2022*, pp. 599–609. ACM, 2022.

Mitliagkas, I., Gopalan, A., Caramanis, C., and Vishwanath, S. User rankings from comparisons: Learning permutations in high dimensions. In *49th Annual Allerton Conference on Communication, Control, and Computing, Allerton 2011, Allerton Park & Retreat Center, Monticello, IL, USA, 28-30 September, 2011*, pp. 1143–1150. IEEE, 2011.

Mondal, A. S., Bal, R., Sinha, S., and Patro, G. K. Two-sided fairness in non-personalised recommendations (student abstract). In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pp. 15851–15852. AAAI Press, 2021.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., and Handelsman, J. Science faculty's subtle gender biases favor male students. *Proceedings of the national academy of sciences*, 109(41):16474–16479, 2012.

Motwani, R. and Raghavan, P. *Randomized algorithms*. Cambridge university press, 1995.

Nemhauser, G. L., Wolsey, L. A., and Fisher, M. L. An Analysis of Approximations for Maximizing Submodular Set Functions - I. *Mathematical Programming*, 14(1): 265–294, 1978.

Patro, G. K., Porcaro, L., Mitchell, L., Zhang, Q., Zehlike, M., and Garg, N. Fair Ranking: A Critical Review, Challenges, and Future Directions. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, pp. 1929–1942, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450393522.

Pitoura, E., Stefanidis, K., and Koutrika, G. Fairness in Rankings and Recommendations: An Overview. *The VLDB Journal*, 2021.

Procaccia, A. D., Rosenschein, J. S., and Zohar, A. On the complexity of achieving proportional representation. *Soc. Choice Welf.*, 30(3):353–362, 2008.

Procaccia, A. D., Reddi, S. J., and Shah, N. A maximum likelihood approach for selecting sets of alternatives. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence, Catalina Island, CA, USA, August 14-18, 2012*, pp. 695–704. AUAI Press, 2012.

Raghavan, M., Barocas, S., Kleinberg, J., and Levy, K. Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, pp. 469–481, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450369367.

Ramponi, M. How chatgpt actually works. *Assembly Ai*, 2022.

Régner, I., Thinus-Blanc, C., Netter, A., Schmader, T., and Huguet, P. Committees with implicit biases promote fewer women when they do not believe gender bias exists. *Nature human behaviour*, 3(11):1171–1179, 2019.

Rooth, D.-O. Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics*, 17(3): 523–534, 2010.

Salem, J. and Gupta, S. Closing the gap: Mitigating bias in online résumé-filtering. In *Web and Internet Economics - 16th International Conference, WINE 2020, Beijing, China, December 7-11, 2020, Proceedings*, volume 12495 of *Lecture Notes in Computer Science*, pp. 471. Springer, 2020.

Schumann, C., Counts, S. N., Foster, J. S., and Dickerson, J. P. The diverse cohort selection problem. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '19, Montreal, QC, Canada, May 13-17, 2019*, pp. 601–609. International Foundation for Autonomous Agents and Multiagent Systems, 2019.

Skowron, P., Faliszewski, P., and Slinko, A. M. Achieving fully proportional representation: Approximability results. *Artif. Intell.*, 222:67–103, 2015.

Skowron, P., Faliszewski, P., and Lang, J. Finding a collective set of items: From proportional multirepresentation to group recommendation. *Artif. Intell.*, 241:191–216, 2016.

Steck, H. Calibrated recommendations. In *Proceedings of the 12th ACM Conference on Recommender Systems*, RecSys '18, pp. 154–162, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450359016. doi: 10.1145/3240323.3240372. URL https://doi.org/10.1145/3240323.3240372.

Streviniotis, E. and Chalkiadakis, G. Preference aggregation mechanisms for a tourism-oriented bayesian recommender. In *PRIMA 2022: Principles and Practice of Multi-Agent Systems - 24th International Conference, Valencia, Spain, November 16-18, 2022, Proceedings*, volume 13753 of *Lecture Notes in Computer Science*, pp. 331–346. Springer, 2022a.

Streviniotis, E. and Chalkiadakis, G. Multiwinner election mechanisms for diverse personalized bayesian recommendations for the tourism domain. In *Proceedings of the Workshop on Recommenders in Tourism (RecTour 2022) co-located with the 16th ACM Conference on Recommender Systems (RecSys 2022), Seattle, WA, USA and Online, September 22, 2022*, volume 3219 of *CEUR Workshop Proceedings*, pp. 65–82. CEUR-WS.org, 2022b.

Szufa, S., Faliszewski, P., Skowron, P., Slinko, A., and Talmon, N. Drawing a map of elections in the space of statistical cultures. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems, AAMAS '20, Auckland, New Zealand, May 9-13, 2020*, pp. 1341–1349. International Foundation for Autonomous Agents and Multiagent Systems, 2020.

Vitelli, V., Sørensen, Ø., Crispino, M., Frigessi, A., and Arjas, E. Probabilistic preference learning with the mallows rank model. *J. Mach. Learn. Res.*, 18:158:1–158:49, 2017.

Vondrák, J. Submodularity and curvature: The optimal algorithm (combinatorial optimization and discrete algorithms). *RIMS Kokyuroku Bessatsu*, 23:253–266, 2010.

Waldstein, D. Success and shortfalls in effort to diversify n.f.l. coaching. *The New York Times*, 2015.

Wang, J. and Shah, N. B. Your 2 is my 1, your 3 is my 9: Handling arbitrary miscalibrations in ratings. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '19, Montreal, QC, Canada, May 13-17, 2019*, pp. 864–872. International Foundation for Autonomous Agents and Multiagent Systems, 2019.

Xia, L. Computing the margin of victory for various voting rules. In *Proceedings of the 13th ACM Conference on Electronic Commerce, EC 2012, Valencia, Spain, June 4-8, 2012*, pp. 982–999. ACM, 2012.

Zehlike, M., Yang, K., and Stoyanovich, J. Fairness in Ranking, Part II: Learning-to-Rank and Recommender Systems. *ACM Comput. Surv.*, apr 2022. Just Accepted.

Contents

A. Other Related Work

Comparison to (Mehrotra & Vishnoi, 2023). A recent work (Mehrotra & Vishnoi, 2023) studies a variant of subset selection where the goal is to select a size-k subset that maximizes the value of a submodular function. Like this work Mehrotra & Vishnoi (2023) also study the effectiveness of representational constraints for mitigating the adverse effects of bias. They, however, consider a family of submodular functions that is relevant to recommendation and web search, which is different from score functions considered in this work (Definition 2.2).

Specifically, in their setting, each item (such as websites, products, movies, or candidates) has m attributes (such as topics, product-category, genres, or skills). Items and attributes in their setting map to candidates and voters respectively in the context of this work. Accordingly, we refer to items and attributes as candidates and voters respectively in the following. Mehrotra & Vishnoi (2023) specify a submodular function $F\colon 2^C \to \mathbb{R}_{\geq 0}$ by n non-decreasing functions $g_1, g_2, \ldots, g_n \colon \mathbb{R} \to \mathbb{R}_{\geq 0}$ (which measure the utility for each voter) and an $n \times m$ utility matrix w (capturing the utility of candidates to each voter) as follows: $F(S) = g_1\left(\sum_{c \in S} W_{1,c}\right) + g_2\left(\sum_{c \in S} W_{2,c}\right) + \cdots + g_n\left(\sum_{c \in S} W_{n,c}\right)$.

The primary difference between the two families is that in their family each voter provides us with a cardinal utility for each candidate (scores $W_{1,c}, W_{2,c}, \ldots, W_{n,c}$ for each candidate c) whereas we consider ordinal utilities specified by n rankings of the m candidates. Numerical scores are more accurate but can lead to serious aggregation and calibration issues (Griffin & Brenner, 2004; Mitliagkas et al., 2011; Steck, 2018; Wang & Shah, 2019). Moreover, while numerical scores are available in recommendation and web-search contexts, in contexts most relevant to this work, such as elections, numerical scores have a high elicitation cost (Griffin & Brenner, 2004; Mitliagkas et al., 2011; Wang & Shah, 2019). Due to this difference, the family of submodular functions considered by Mehrotra & Vishnoi (2023) is incomparable to the multiwinner scoring functions considered in this work (Definition 2.2).

- On the one hand, in Mehrotra & Vishnoi (2023)'s model, voters can have the same utility for two or more candidates and can also have an arbitrarily large difference between the utilities of two candidates. However, because the functions we consider are defined by strict rankings over candidates, voters cannot be indifferent between two candidates. Moreover, as preferences are captured by rankings, we do not capture the magnitude of the difference in the utilities of two candidates in, say, adjacent positions in a ranking.
- On the other hand, since in Mehrotra & Vishnoi (2023)'s model, the utility of a committee S to a voter v is a function of the sum $\sum_{i \in S} W_{v,c}$, it cannot capture other utility functions such as the $\max$ utility a voter derives for a selected candidate or the sum of utilities for the best t candidates included for some $1 < t < k$ which are required, e.g., to capture the ℓ_1-CC rule and its extensions.

That said, the SNTV rule, the Borda rule, and other committee scoring rules (Section B.1), can be captured by both the family of submodular functions in (Mehrotra & Vishnoi, 2023) and Definition 2.2. Here, Mehrotra & Vishnoi (2023)'s and our results complement each other. The bounds established by (Mehrotra & Vishnoi, 2023) on the number of voters that are needed for representational constraints to recover a close-to-optimal utility degrades with $\mathrm{poly}(n/k)$[8] and, since the number of voters, n, is much higher than the number of selected candidates, k, in real-world election contexts, (Mehrotra & Vishnoi, 2023)'s results lead to vacuous bounds in these contexts. In contrast, our main algorithmic result (Theorem 3.2) degrades with $\frac{1}{\mathrm{poly}(n/k)}$ and, hence, it gives a meaningful bound in contexts where $n \gg k$ but is vacuous if $k \gg n$.

The reason why we are able to obtain results that degrade as $\mathrm{poly}\left(\frac{1}{n}\right)$ in our setting is because we require preference lists to be i.i.d., because of which a *fixed* representational constraint is sufficient to recover high latent utility in our setting. In contrast, Mehrotra & Vishnoi (2023) allow the utility matrix (that captures preferences in their model) to be non-i.i.d. and show that, because of this, no fixed representational constraint is sufficient to recover any constant fraction of the optimal latent utility. Instead, they give an algorithm that, given functions $g_1, \ldots, g_n$ and biased or observed utilities, determines the relevant representational constraint. Such algorithms may be reasonable in certain contexts (such as recommendation and web search) where the selection does not require direct human feedback but is not suitable in contexts that require direct human feedback (such as elections) and where representational constraints must be fixed before human feedback (e.g. votes) are collected.

Further related work on the study of representational constraints, learning preferences in the presence of noise, and other works of subset selection. Beyond the study of the ability of representational constraints to increase the utility

[8]Concretely, their positive result degrades with $O\left(n^2 k^{-1/4}\right)$

of selection, their effect on the decisionmaker's bias over the long term has also been studied (Celis et al., 2021; Heidari & Kleinberg, 2021). Moreover, apart from representational constraints, the power of various other interventions to debias decisions based on biased inputs has also been studied, e.g., by Faenza et al. (2020); Garg et al. (2021) in the context of school and college admission and by Blum & Stangl (2020) in the context of classification.

The problem of selecting a subset maximizing a multiwinner score function is a special case of submodular maximization. There is a rich literature on submodular maximization. In this literature, optimization in the presence of cardinality constraints has been extensively studied (Fujishige, 2005; Krause & Golovin, 2014) and there is a standard $(1 - \frac{1}{e})$-approximation algorithm for finding a size-k subset maximizing a submodular function (Nemhauser et al., 1978).

Closer to our work from a methodological perspective, there are many works that want to learn user preferences, by e.g., fitting some parameterized generative model to the data (Guiver & Snelson, 2009; Lu & Boutilier, 2014; Vitelli et al., 2017; Allouche et al., 2022). Particularly closely related to the present paper are works on the sample complexity of such learning algorithms (Awasthi et al., 2014; Busa-Fekete et al., 2014; Caragiannis & Micha, 2017b; Liu & Moitra, 2018; Collas & Irurozki, 2021), and the capabilities of different rules to recover a ground truth (Procaccia et al., 2012; Caragiannis et al., 2022). Notably, the problem that we study also captures some of these learning problems by interpreting latent rankings as the underlying ground truth and biased rankings as voters' noisy estimates of the ground truth. Thereby, we recover settings studied, e.g., by Procaccia et al. (2012) and Caragiannis et al. (2022).

Finally, there is also a large body of work studying the generalization of the simplest (numerical) formulation of subset selection from outputting a subset to outputting a ranking (Liu, 2009; Burges, 2010). Within this body of works, the problem of biases in the output rankings as well as different types of interventions, including representational constraints, to mitigate these biases have been studied (Kay et al., 2015; Pitoura et al., 2021; Zehlike et al., 2022; Patro et al., 2022). This problem is different from the variant of subset selection we study in two ways: (1) the input is a single numerical score for each item, and (2) the output is a ranking instead of a subset.

B. Additional Materials for Section 2

In this section, we present additional examples and missing proofs from Section 2. For ease of reference, we also include the following table of notations.

B.1. Examples of Multiwinner Score Functions

We list some well-known multiwinner score functions that are specific cases of Definition 2.2.

Committee scoring functions. A committee scoring function awards a score to each committee as

$$\mathrm{CS}(S) = \sum_{v \in V} g(\mathrm{pos}_{\succ_v}(S))$$

for some function $g : [m]^k \to \mathbb{R}_{\geq 0}$. The followings are some examples of committee scoring rules.

Example B.1 (**Examples of committee scoring functions**). Let $s \in \mathbb{R}^m_{\geq 0}$ be a vector. We have the following examples of committee scoring functions.

- If $s_1 = 1$ and $s_i = 0$ for $i \geq 2$, and $g(i_1, \ldots, i_k) = \sum_{l \in [k]} s_{i_l}$, we call F the SNTV rule.
- If $s_i = m - i$ for $i \in [m]$, and $g(i_1, \ldots, i_k) = \sum_{l \in [k]} s_{i_l}$, we call F the Borda rule.
- If $s_i = m - i$ for $i \in [m]$, and $g(i_1, \ldots, i_k) = \max_{l \in [k]} s_{i_l}$, we call F the ℓ_1-CC rule.

Approval-based functions. Suppose each voter $v \in V$ approves a subset A_v of m' candidates (that are the first $m' \in [m]$ candidates in $\succ_v$; note that m' is the same for all voters).

An approval-based function awards a score to each committee as

$$\mathrm{App}(S) = \sum_{v \in V} g(|S \cap A_v|),$$

(a) Basic notation

Symbol	Meaning
n	Number of voters
m	Number of candidates
k	Number of selected candidates
V	Set of voters
C	Set of candidates
$\succ$	A "latent" ranking of all candidates in C
$\not\succ$	A "biased" ranking of all candidates in C
μ	Generative model of latent preferences (Definition 2.3)
$\widehat{\mu}$	Generative model of biased preferences (Definition 2.4)
$\mathcal{L}(C)$	Set of all strict and complete orders over C
G_1, G_2	Advantaged and disadvantaged groups, respectively. Disjoint subsets of C.
F	Multiwinner score function $F(S) = \sum_{v\in V} f(S, \succ_v)$ (Definition 2.2)

(b) Notation specific to multiwinner scoring functions

Symbol	Meaning
f	A function $f\colon 2^C \times \mathcal{L}(C) \to \mathbb{R}_{\geq 0}$ such that $F(S) = \sum_{v\in V} f(S, \succ_v)$; see Definition 2.2
$f_S(c, \succ)$	The marginal contribution of c to S with respect to f and $\succ$: $f_S(c, \succ) := f(S \cup \{c\}, \succ) - f(S, \succ)$.
$\mathrm{pos}_\succ(c)$	Position of c in the preference list $\succ$
$\tau_1(f)$	The maximum possible expected score $\mathbb{E}_\mu[f(c, \succ)]$ of a candidate (in the case of multiwinner score functions this is the score which function $f(\cdot, \succ)$ awards to the set consisting of the candidate ranked in the first position of $\succ$)

(c) Notation specific to the utility-based model

Symbol	Meaning
θ	Bias parameter (Definition 2.6)
ω_c	Intrinsic utility of $c \in C$ (Definition 2.5)
$w_{v,c}$	Latent utility of $c \in C$ observed by voter v (Definition 2.5)
$\hat{w}_{v,c}$	Biased utility of $c \in C$ observed by voter v (Definition 2.6)

(d) Notation specific to the swapping-based model

Symbol	Meaning
ϕ	Bias parameter (Definition B.12)
t	Number of swaps (Definition B.12)
$A(\succ)$	The collection of all pairs (i, j) such that there exist $c \in G_1$ and $c' \in G_2$ with $\mathrm{pos}_\succ(c) = i > j = \mathrm{pos}_\succ(c')$ (Definition B.12)
$Z(\succ)$	Normalization factor: $\sum_{(i',j')\in A(\succ)} \phi^{i'-j'}$ (Definition B.12)

(e) Voting rules and smoothness parameters

Symbol	Meaning
SNTV	Single non-transferable vote; defined by $f(S, \succ) = \sum_{c\in S} \mathbb{1}_{\mathrm{pos}_\succ(c)=1}$
Borda	A multiwinner score functions defined by $f(S, \succ) = \sum_{c\in S} (m - \mathrm{pos}_\succ(c))$
ℓ_1-CC	ℓ_1 Chamberlin-Courant rule; defined by $f(S, \succ) = \max\limits_{c\in S} \{m - \mathrm{pos}_\succ(c)\}$
α	Definition 3.1
β	Definition 2.8
γ	Definition 2.8

(f) Special subsets

Symbol	Meaning
R	Voters' latent preference profile $\{\succ_v \in \mathcal{L}(C) : v \in V\}$
$\widehat{R}$	Voters' biased preference profile $\{\not\succ_v \in \mathcal{L}(C) : v \in V\}$
$S^\star$	$\arg\max_{S\subseteq C : \lvert S\rvert = k} F(S)$
$\widehat{S}$	$\arg\max_{S\in\mathcal{K}(\ell) : \lvert S\rvert = k} \widehat{F}(S)$
M	The collection of k candidates c with the highest value $\mathbb{E}_\mu[f(c, \succ)]$
$\mathcal{K}(\ell)$	The collection of subsets of C that have at least ℓ candidates from G_2

Table 1. Table of notations.

where $g : [k] \to \mathbb{R}_{\geq 0}$ is a non-decreasing concave function. The above definition captures the class of OWA-rules, which are parameterized by an OWA-vector $\lambda = (\lambda_1, \ldots, \lambda_k)$ with $\lambda_1 \geq \cdots \geq \lambda_k \geq 0$ and correspond to $g(i) = \sum_{j=1}^{i} \lambda_j$. Among others, the class of OWA-functions contains the popular PAV ($\lambda = (1, \frac{1}{2}, \ldots, \frac{1}{k})$) and $\ell_{\min}$-CC ($\lambda = (1, 0, \ldots, 0)$) rules.

We note that our main result Theorem 3.2 can also be extended to approval-based rules where the size of the approval set can be different for different voters.

B.2. Order-Preserving Properties of a Utility-Based Model

In this section, we show that the utility-based generative models (Definitions 2.5 and 2.6) are order-preserving with respect to μ and between μ and $\widehat{\mu}$. We divide the section into two parts corresponding to each proof.

B.2.1. Order Preservation With Respect to μ.

In this section, we prove the following lemma.

Lemma B.2 (**Order-preserving properties of the latent utility-based model**). *Let $F = \sum_{v\in V} f(\cdot, \succ_v)$ be a multiwinner score function. Let μ be a utility-based generative model defined in Definition 2.5. F is order-preserving with respect to μ.*

Recall that in the utility-based model (Definition 2.5), the variable η is drawn from the uniform distribution on $[0, 1]$. We will, in fact, prove a more general version of the above lemma that holds for any distribution of η that satisfies certain properties (Definition B.3). With some abuse of notation, we use η to denote both the distribution and a value drawn from distribution η (independent of all other randomness).

Definition B.3 (**A family of distributions** η). Let η be the distribution on $\mathbb{R}_{\geq 0}$ from Definition 2.6 that parameterizes the generative model μ. Let $\mathrm{cdf}_\eta \colon \mathbb{R} \to [0, 1]$ be the cumulative distribution function of η. We define the following properties of η.

- (Order preserving A) η is order preserving if for all $\varepsilon \in [0, 1]$ and $a_2 \geq a_1 \geq 0$,

$$\Pr_{X,Y\sim\eta}\left[X > Y(1-\varepsilon) \mid X, Y \in [a_1, a_2]\right] \geq \frac{1}{2}.$$

- (Order preserving B) We say η is order preserving if for all $0 \leq a_1 < a_2 \leq a_3 < a_4$ and all $\varepsilon \in [0, 1]$, the following holds:
  - Let $\mathscr{E}$ be the event that either $X \in [a_3, a_4]$ and $Y(1-\varepsilon) \in [a_1, a_2]$ or $X \in [a_1, a_2]$ and $Y(1-\varepsilon) \in [a_3, a_4]$
  - $\Pr_{X,Y\sim\eta}\left[X \in [a_3, a_4] \text{ and } Y \in [a_1, a_2] \mid \mathscr{E}\right] \;\geq\; \Pr_{X,Y\sim\eta}\left[X \in [a_1, a_2] \text{ and } Y(1-\varepsilon) \in [a_3, a_4] \mid \mathscr{E}\right].$

Several distributions on $\mathbb{R}_{\geq 0}$ including the uniform distribution on $[0, 1]$ and the exponential distributions satisfy the properties in Definition B.3. Order preservation properties A and B are used to ensure that if $\omega_c > \omega_{c'}$ (for any $c, c' \in C$ in the same group), then conditioned on the event, say $\mathscr{F}$, that $\{c, c'\}$ appear in positions $\{\ell_1, \ell_2\}$ (for any $1 \leq \ell_1 < \ell_2 \leq m$), then c' is more likely to appear in position ℓ_1 than in ℓ_2. (Note that conditioned on $\mathscr{F}$, c and c' may appear in positions ℓ_1 and ℓ_2 respectively or in positions ℓ_2 and ℓ_1 respectively)

Fix the following parameters.

1. Any multiwinner score function F;
2. Any distribution η satisfying the properties in Definition B.3;
3. Any pair of candidates $c, c' \in C$ in the same group (G_1 or G_2); and
4. Any subset $S \subseteq C \setminus \{c, c'\}$.

To prove that F is order-preserving with respect to μ, we need to show that if $\mathbb{E}_\mu\left[f(c, \succ)\right] < \mathbb{E}_\mu\left[f(c', \succ)\right]$, then the following two conditions hold

1. (Property A) for any subsets $R \subseteq S \subseteq C \setminus \{c, c'\}$, $\mathbb{E}_\mu\left[f_S(c, \succ)\right] \leq \mathbb{E}_\mu\left[f_S(c', \succ)\right];$
2. (Property B) for any subsets $R \subseteq S \subseteq C \setminus \{c, c'\}$, $\mathbb{E}_\mu\left[f_S(c', \succ)\right] - \mathbb{E}_\mu\left[f_S(c, \succ)\right] \leq \mathbb{E}_\mu\left[f_R(c', \succ)\right] - \mathbb{E}_\mu\left[f_R(c, \succ)\right].$

Proof of Property A. The following is the main lemma used to prove Property A.

Lemma B.4. *For any $d, d' \in C$ in the same group (G_1 or G_2) and any set $T \subseteq C \setminus \{c, c'\}$, if $\omega_{d'} \geq \omega_d$, then the following holds*

$$\mathbb{E}_\mu \left[f_T(c', \succ)\right] \geq \mathbb{E}_\mu \left[f_T(c, \succ)\right] \quad \text{and} \quad \mathbb{E}_{\widehat{\mu}} \left[f_T(c', \not\succ)\right] \geq \mathbb{E}_{\widehat{\mu}} \left[f_T(c, \not\succ)\right].$$

Property A straightforwardly follows from the above lemma.

Proof of Property A. We claim that $\mathbb{E}_\mu \left[f(c, \succ)\right] < \mathbb{E}_\mu \left[f(c', \succ)\right]$, implies that $\omega_c \leq \omega_{c'}$. To see this, suppose $\omega_c > \omega_{c'}$, then from Lemma B.4 (invoked with $T = \emptyset$) it follows that $\mathbb{E}_\mu \left[f(c, \succ)\right] \geq \mathbb{E}_\mu \left[f(c', \succ)\right]$, which is a contradiction. Hence, $\omega_c \leq \omega_{c'}$. Now, using Lemma B.4 with the set $T = S$ and the fact that $\omega_c \leq \omega_{c'}$, Property A follows. □

Proof of Property B. We first derive an equivalent version of Property B that is easier to prove.

Step 1 (An alternate version of Property B): By rearranging the terms in the inequality in Property B and using the definition of the marginal score, we get the following equivalent version of Property B

$$\forall_{R \subseteq S \subseteq C \setminus \{c, c'\}}, \quad \mathbb{E}_\mu \left[f(S \cup c', \succ)\right] - \mathbb{E}_\mu \left[f(R \cup c', \succ)\right] \leq \mathbb{E}_\mu \left[f(S \cup c, \succ)\right] - \mathbb{E}_\mu \left[f(R \cup c, \succ)\right].$$

Using the definition of marginal score again, we get following equivalent version of Property B

$$\forall_{R \subseteq S \subseteq C \setminus \{c, c'\}}, \quad \mathbb{E}_\mu \left[f_{R \cup c'}(S \setminus R, \succ)\right] \leq \mathbb{E}_\mu \left[f_{R \cup c}(S \setminus R, \succ)\right].$$

Observe that the set $S \setminus R$ is disjoint from $R \cup c'$ and $R \cup c$. Defining $A \coloneqq R$ and $B \coloneqq S \setminus R$, we get the following equivalent version

$$\forall_{A, B \subseteq C \setminus \{c, c'\} \colon A \cap B = \emptyset}, \quad \mathbb{E}_\mu \left[f_{A \cup c'}(B, \succ)\right] \leq \mathbb{E}_\mu \left[f_{A \cup c}(B, \succ)\right]. \tag{2}$$

We can prove the above inequality, however, the fact that B can have multiple candidates makes the proof unnecessarily technical. Instead, we will simplify the above version of Property B using the following lemma.

Lemma B.5. *For any $d, d' \in C$, the following holds*

$$\begin{aligned} &\forall_{X, Y \subseteq C \setminus \{c, c'\} \colon X \cap Y = \emptyset}, \quad \mathbb{E}_\mu \left[f_{X \cup c'}(Y, \succ)\right] \leq \mathbb{E}_\mu \left[f_{X \cup c}(Y, \succ)\right], \\ \iff \quad &\forall_{Y \subseteq C \setminus \{c, c'\}} \ \forall_{y \in C \setminus X \setminus \{c, c'\}}, \quad \mathbb{E}_\mu \left[f_{X \cup c'}(y, \succ)\right] \leq \mathbb{E}_\mu \left[f_{X \cup c}(y, \succ)\right]. \end{aligned}$$

Proof. We are required to prove (1) and equivalent (2), where (1) and (2) are the following conditions respectively

$$\forall_{X, Y \subseteq C \setminus \{c, c'\} \colon X \cap Y = \emptyset}, \quad \mathbb{E}_\mu \left[f_{X \cup c'}(Y, \succ)\right] \leq \mathbb{E}_\mu \left[f_{X \cup c}(Y, \succ)\right], \tag{3}$$

$$\forall_{Y \subseteq C \setminus \{c, c'\}} \ \forall_{y \in C \setminus X \setminus \{c, c'\}}, \quad \mathbb{E}_\mu \left[f_{X \cup c'}(y, \succ)\right] \leq \mathbb{E}_\mu \left[f_{X \cup c}(y, \succ)\right]. \tag{4}$$

Step A ($2 \implies 1$)**:** Fix any disjoint $X, Y \subseteq C \setminus \{c, c'\}$. Let $Y \coloneqq \{y_1, y_2, \ldots, y_t\}$. It holds that

$$\begin{aligned} \mathbb{E}_\mu \left[f_{X \cup c'}(Y, \succ)\right] &= \sum_{i \in [t]} \mathbb{E}_\mu \left[f_{X \cup c' \cup y_1 \cup y_2 \cup \cdots \cup y_{i-1}}(y_i, \succ)\right] \\ &\leq \sum_{i \in [t]} \mathbb{E}_\mu \left[f_{X \cup c \cup y_1 \cup y_2 \cup \cdots \cup y_{i-1}}(y_i, \succ)\right] && \text{(Using Equation (3) to switch } c' \text{ with } c) \\ &\leq \mathbb{E}_\mu \left[f_{X \cup c}(Y, \succ)\right] \end{aligned}$$

(Using that $F_S(T \cup R) = F_S(T) + F_{S \cup T}(R)$ for all sets S, R, T and any submodular function F)

Step B ($1 \implies 2$)**:** Equation (4) follows by setting $Y = \{y\}$ in Equation (3). □

Step 2 (Proving alternate version of Property B): Thus, from Equation (2) and Lemma B.5, it follows that the following condition is equivalent to Property B.

$$\forall_{A \subseteq C \setminus \{c, c'\}} \ \forall_{b \in C \setminus X \setminus \{c, c'\}}, \quad \mathbb{E}_\mu \left[f_{A \cup c'}(b, \succ)\right] \leq \mathbb{E}_\mu \left[f_{A \cup c}(b, \succ)\right]. \tag{5}$$

Notation. We first define some notation that is used to express $\mathbb{E}_\mu\left[f_{A\cup c'}(b,\succ)\right]$ and $\mathbb{E}_\mu\left[f_{A\cup c}(b,\succ)\right]$. Fix any voter v and consider $\succ_v :=\succ$ where $\succ\sim\mu$. Define $w_{(i)}$ as the i-th largest order statistic among the utilities of candidates in $C\setminus\{c,c'\}$ for each $i\in[m-1]$. In other words, $w_{(i)}$ is the i-th largest value in the set

$$\mathcal{W}\coloneqq\left\{w_{v,d}\mid d\in C\setminus\{c,c'\}\right\}.$$

By this definition, we have the inequalities

$$w_{(1)}\geq w_{(2)}\geq\cdots\geq w_{(m)}.$$

Define the interval $I_\ell\coloneqq\left[w_{(\ell)},w_{(\ell+1)}\right]$ for each $\ell\in[m-2]$. Define $\mathscr{E}_{\ell,k}$, for each $\ell,k\in[m-2]$ as the event that

$$w_{v,c'}\in I_\ell\quad\text{and}\quad w_{v,c}\in I_k. \tag{Event $\mathscr{E}_{\ell,k}$}$$

Define $\mathscr{F}$ as the event that

$$w_{v,c'}>w_{v,c}. \tag{Event $\mathscr{F}$}$$

Note that $\mathscr{F}$ is equivalent to the event $\mathrm{pos}_\succ(c')<\mathrm{pos}_\succ(c)$. Finally, for each $\ell\in[m]$, define the random variable

$$\tau_{\mathrm{pos}_\succ(b)}(\mathrm{pos}_\succ(A),\ell)\coloneqq f_{A\cup i(\ell)}(b,\succ).$$

Where $i(\ell)$ is the candidate at the ℓ-th position in $\succ$. The randomness in $\tau_S(\ell)$ due to the randomness in $\mathrm{pos}_\succ(A)$ and in $\mathrm{pos}_\succ(b)$. Conditioned on any value of $\mathrm{pos}_\succ(A)$ and $\mathrm{pos}_\succ(b)$, due to domination sensitivity (Definition 2.2), it holds that

$$\forall 1\leq\ell<k\leq m,\quad \tau_{A\cup i(\ell)}(b)\geq\tau_{A\cup i(k)}(b). \tag{6}$$

We can express $\mathbb{E}_\mu\left[f_{A\cup c'}(b,\succ)\right]$ and $\mathbb{E}_\mu\left[f_{A\cup c}(b,\succ)\right]$ as follows

$$\begin{aligned}
\mathbb{E}_\mu\left[f_{A\cup c'}(b,\succ)\right]&=\mathbb{E}_{\mathcal{W}}\left[\sum_{\ell,m\in[m-2]:\ell<k}\left(\Pr\left[\mathscr{E}_{\ell,k}\right]\cdot\tau_{A\cup i(\ell)}(b)+\Pr\left[\mathscr{E}_{k,\ell}\right]\cdot\tau_{A\cup i(k)}(b)\right)\right]\\
&\quad+\mathbb{E}_{\mathcal{W}}\left[\sum_{\ell\in[m-2]}\Pr\left[\mathscr{E}_{\ell,\ell}\right]\left(\Pr\left[\mathscr{F}\mid\mathscr{E}_{\ell,\ell}\right]\cdot\tau_{A\cup i(\ell)}(b)+\Pr\left[\neg\mathscr{F}\mid\mathscr{E}_{\ell,\ell}\right]\cdot\tau_{A\cup i(\ell+1)}(b)\right)\right],\\
\mathbb{E}_\mu\left[f_{A\cup c}(b,\succ)\right]&=\mathbb{E}_{\mathcal{W}}\left[\sum_{\ell,m\in[m-2]:\ell<k}\left(\Pr\left[\mathscr{E}_{\ell,k}\right]\cdot\tau_{A\cup i(k)}(b)+\Pr\left[\mathscr{E}_{k,\ell}\right]\cdot\tau_{A\cup i(\ell)}(b)\right)\right]\\
&\quad+\mathbb{E}_{\mathcal{W}}\left[\sum_{\ell\in[m-2]}\Pr\left[\mathscr{E}_{\ell,\ell}\right]\left(\Pr\left[\mathscr{F}\mid\mathscr{E}_{\ell,\ell}\right]\cdot\tau_{A\cup i(\ell+1)}(b)+\Pr\left[\neg\mathscr{F}\mid\mathscr{E}_{\ell,\ell}\right]\cdot\tau_{A\cup i(\ell)}(b)\right)\right].
\end{aligned}$$

Hence, it follows that

$$\begin{aligned}
&\mathbb{E}_\mu\left[f_{A\cup c'}(b,\succ)\right]-\mathbb{E}_\mu\left[f_{A\cup c}(b,\succ)\right]\\
&\quad\geq\mathbb{E}_{\mathcal{W}}\left[\sum_{\ell,m\in[m-2]:\ell<k}\left(\Pr\left[\mathscr{E}_{\ell,k}\right]-\Pr\left[\mathscr{E}_{k,\ell}\right]\right)\cdot\left(\tau_{A\cup i(\ell)}(b)-\tau_{A\cup i(k)}(b)\right)\right]\\
&\quad+\mathbb{E}_{\mathcal{W}}\left[\sum_{\ell\in[m-2]}\Pr\left[\mathscr{E}_{\ell,\ell}\right]\cdot\left(\Pr\left[\mathscr{F}\mid\mathscr{E}_{\ell,\ell}\right]-\Pr\left[\neg\mathscr{F}\mid\mathscr{E}_{\ell,\ell}\right]\right)\cdot\left(\tau_{A\cup i(\ell)}(b)-\tau_{A\cup i(\ell+1)}(b)\right)\right]
\end{aligned} \tag{7}$$

Note that $\mathrm{pos}(A)$ and $\mathrm{pos}(b)$ are deterministic conditioned on any specific value of $\mathcal{W}$ and either (1) the event $\mathscr{E}_{\ell,k}$ for $\ell\neq k$ or (2) the event $\mathscr{E}_{\ell,\ell}$ and $\mathscr{F}$. Hence, Equation (6) holds. Further, we have the following claims using the two properties in Definition B.3.

$$\forall_{1\leq\ell<k\leq m},\qquad \Pr\left[\mathscr{E}_{\ell,k}\right]-\Pr\left[\mathscr{E}_{k,\ell}\right]\geq 0, \tag{8}$$

$$\forall_{1\leq\ell\leq m},\quad \Pr\left[\mathscr{F}\mid\mathscr{E}_{\ell,\ell}\right]-\Pr\left[\neg\mathscr{F}\mid\mathscr{E}_{\ell,\ell}\right]\geq 0. \tag{9}$$

Substituting Equations (6) to (9) in Equation (7), it follows that

$$\mathbb{E}_\mu\left[f_{A\cup c'}(b,\succ)\right]-\mathbb{E}_\mu\left[f_{A\cup c}(b,\succ)\right]\leq 0.$$

The above is equivalent to Equation (5). Since Equation (5), itself, is equivalent to Property B, Property B also follows.

B.2.2. ORDER PRESERVATION BETWEEN μ AND $\widehat{\mu}$

We first define a parameter $\sigma(f)$ used in Lemma B.7.

Definition B.6. Let $F = \sum_{v\in V} f(\cdot, \succ_v)$ be a multiwinner score function. For each $j \in [m]$, let $i_{\succ,j}$ be the j-th candidate in $\succ$. For each $j \in [m]$ and $S \subseteq S$, define $\tau_{j,S}(f)$ to be the marginal score of $i_{\succ,j}$ with respect to S, i.e.,

$$\tau_{j,S}(f) \coloneqq f_S(\{i_{\succ,j}\}, \succ)$$

which is independent of $\succ \in \mathcal{L}(C)$. Let $\tau_{m+1,S} = 0$ for any $S \subseteq C$ and $\tau_{\min} \coloneqq \min_{S\subseteq C} \tau_{m-1,S}$. $\sigma(f)$ is defined as follows.

$$\sigma(f) \coloneqq \min_{S\subseteq C} \min_{\ell\in[m-1]} \frac{\tau_{\ell,S}(f) + \tau_{\ell+2,S}(f) - 2\tau_{\ell+1,S}(f)}{\tau_{\ell,S}(f) - \tau_{\ell+2,S}(f)}, \tag{10}$$

where we use the convention $\frac{0}{0} = 1$.

Intuitively, $\sigma(f)$ measures a certain notion of convexity measure of f. Concretely, $\sigma(f) \geq 0$ if and only if there is a convex function g that takes value s_ℓ at ℓ (for each $\ell \in [m+1]$). Moreover, strict inequality holds (i.e., $\sigma(f) > 0$) if and only if g is strictly convex. The normalization in $\sigma(f)$ ensures that it is invariant to a multiplicative scaling of F. As for examples: $\sigma(f) = 1$ for SNTV and $\sigma(f) = 0$ for Borda rule or ℓ_1-CC.

Lemma B.7 (Order-preservation between latent and biased utility-based models). *Let $F = \sum_{v\in V} f(\cdot, \succ_v)$ be a multiwinner score function. Let $(\mu, \widehat{\mu})$ be a utility-based generative model defined in Definitions 2.5 and 2.6. The following holds for any $0 \leq \lambda \leq m^{-1/2}$.*

1. *If $\sigma(f) > 0$, then F is $\left(1-\lambda, 1-\sigma(f)\cdot m^{-1}\cdot \Omega(\lambda)\right)$ order preserving between μ and $\widehat{\mu}$; and*
2. *If $\sigma(f) = 0$ and $\tau_{\min} > 0$ and $\tau_{m,\emptyset} = 0$, then F is $\left(1-\lambda, 1-m^{-2}\cdot \frac{\tau_{\min}}{\tau}\cdot \Omega(\lambda)\right)$ order preserving between μ and $\widehat{\mu}$.*

In particular, this result implies the following:

1. The SNTV rule is $\left(1-O(m^{-1/2}), 1-\Omega\left(m^{-3/2}\right)\right)$ order preserving between μ and $\widehat{\mu}$;
2. The ℓ_1-CC rule is $\left(1-O(m^{-1/2}), 1-\Omega\left(m^{-3/2}\right)\right)$ order preserving between μ and $\widehat{\mu}$; and
3. The Borda rule is $\left(1-O(m^{-1/2}), 1-\Omega\left(m^{-3.5}\right)\right)$ order preserving between μ and $\widehat{\mu}$.

Recall that in the utility-based model (Definition 2.5), the variable η is drawn from the uniform distribution on $[0, 1]$. We will, in fact, prove a more general version of the above lemma that holds for any distribution of η that satisfies certain properties (Definition B.8). With some abuse of notation, we use η to denote both the distribution and a value drawn from distribution η (independent of all other randomness).

We will bound the parameters (β, γ) for any η from Definition 2.6 that satisfies the following properties.

Definition B.8 (Properties of the utility distribution η from Definition 2.6). Let η be the distribution on $\mathbb{R}_{\geq 0}$ from Definition 2.6 that parameterizes the generative model μ. Let $\mathrm{cdf}_\eta \colon \mathbb{R} \to [0, 1]$ be the cumulative distribution function of η. We define the following properties of η.

- (Log-Lipshictzness) We say that cdf_η is Log-lipschitz there exists a constant $\pi > 0$ such that, for all $0 < x < y$, $\frac{\mathrm{cdf}_\eta(x)}{\mathrm{cdf}_\eta(y)} \geq 1 - \frac{\pi x}{y}$; and
- (Order preservation) We say η is order preserving if there exists a constant $\pi > 0$ such that, for all $\varepsilon \in [0, 1]$ and $t \geq 0$, if $\Pr_{X,Y\sim\eta}\left[X > Y(1-\varepsilon) \mid X, Y \geq t\right] \geq \frac{1}{2}\cdot\left(1+\frac{\varepsilon}{\pi}\right)$.

Several distributions on $\mathbb{R}_{\geq 0}$ including the uniform distribution on $[0, 1]$ and the exponential distributions satisfy the properties in Definition B.8. Roughly, log-lipshitzness requires that the $\log\left(\mathrm{cdf}_\eta(\cdot)\right)$ is Lipshictz. It guarantees that if x and y are multiplicatively close to each other, then $\mathrm{cdf}_\eta(x)$ and $\mathrm{cdf}_\eta(y)$ are multiplicative close to each other. Order preservation guarantees that if $\omega_c > (1+\varepsilon)\cdot\omega_{c'}$ (for any $c, c' \in C$ and $\varepsilon \in [0,1]$), then $\Pr[w_{v,c} > w_{v,c'}] > 0.5\cdot(1+\Omega(\varepsilon))$. Both of these guarantees are required to establish the multiplicative guarantees in (β,γ) order preservation.

To prove that F is (β,γ) order preserving between μ and $\widehat{\mu}$, we need to show that the following implication holds

$$\beta\cdot\mathbb{E}_\mu\left[f_S(c',\succ)\right] > \mathbb{E}_\mu\left[f_S(c,\succ)\right] > 0. \implies \gamma\cdot\mathbb{E}_{\widehat{\mu}}\left[f_S(c',\not\succ)\right] \geq \mathbb{E}_{\widehat{\mu}}\left[f_S(c,\not\succ)\right]. \tag{11}$$

We divide the proof of Lemma B.7 into two parts, corresponding to the two conditions in Lemma B.7. Both parts rely on the following lemma, whose proof appears later.

Lemma B.9. *Let $F = \sum_{v\in V} f(\cdot,\succ_v)$ be a multiwinner score function. Let $(\mu,\widehat{\mu})$ be a utility-based generative model defined in Definitions 2.5 and 2.6. For any $\beta\in[1-m^{-1/2},1]$, any candidates $c,c'\in C$ in the same group (G_1 or G_2), and any set $S\subseteq C\setminus\{c,c'\}$, the following implication holds*

$$\beta\cdot\mathbb{E}_\mu\left[f_S(c',\succ)\right] \geq \mathbb{E}_\mu\left[f_S(c,\succ)\right] \implies \omega_{c'}\left(1-m^{-1}\cdot\Theta\left(1-\beta\right)\right) > \omega_c.$$

Part 1 ((β,γ) order preservation with $\sigma(f)>0$): Suppose $\sigma(f)>0$. The following is the main lemma in this part.

Lemma B.10. *Let $F = \sum_{v\in V} f(\cdot,\succ_v)$ be a multiwinner score function. Let $(\mu,\widehat{\mu})$ be a utility-based generative model defined in Definitions 2.5 and 2.6. Suppose candidates $c,c'\in C$ are in the same group (G_1 or G_2) and set $S\subseteq C\setminus\{c,c'\}$. If there exists a $\rho>0$ such that $\omega_{c'}(1-\rho)\geq\omega_c$ and $\mathbb{E}_\mu\left[f_S(c',\succ)\right], \mathbb{E}_\mu\left[f_S(c,\succ)\right] > 0$, then there is a constant $\varepsilon>0$ such that*

$$\frac{\mathbb{E}_{\widehat{\mu}}\left[f_S(c',\not\succ)\right]}{\mathbb{E}_{\widehat{\mu}}\left[f_S(c,\not\succ)\right]} \geq \frac{1+\varepsilon\rho\sigma(f)}{1-\varepsilon\rho\sigma(f)}.$$

(β,γ) order preservation follows from Lemmas B.9 and B.10. Concretely, the proof is as follows.

Proof of (β,γ) order preservation assuming Lemmas B.9 and B.10. From the LHS in Equation (11) and Lemma B.9, it follows that $\omega_{c'}\left(1-m^{-1}\cdot\Theta\left(1-\beta\right)\right) > \omega_c$. Hence, Lemma B.10 applicable with $\rho = m^{-1}\cdot\Theta\left(1-\beta\right)$. From Lemma B.10, it follows that the RHS of Equation (11) holds with $\gamma = 1-\Theta(m^{-1}\cdot\Theta\left(1-\beta\right)\cdot\sigma(f))$

In the remainder of this section, we prove Lemma B.10.

Proof of Lemma B.10. Fix any voter v and consider $\succ_v := \succ$ where $\succ\sim\mu$. Define $w_{(i)}$ as the i-th largest order statistic among the utilities of candidates in $C\setminus\{c,c'\}$ for each $i\in[m-1]$. In other words, $w_{(i)}$ is the i-th largest value in the set

$$\mathcal{W} \coloneqq \left\{w_{v,d} \mid d\in C\setminus\{c,c'\}\right\}.$$

By this definition, we have the inequalities

$$w_{(1)} \geq w_{(2)} \geq \cdots \geq w_{(m)}.$$

Define the interval $I_\ell \coloneqq \left[w_{(\ell)},\infty\right)$ for each $\ell\in[m-2]$. Define $\mathscr{E}_\ell$, for each $\ell\in[m-2]$, as the event that

$$w_{v,c'} \in I_\ell. \tag{Event $\mathscr{E}_\ell$}$$

Define $\mathscr{F}$ as the event that

$$w_{v,c'} > w_{v,c}. \tag{Event $\mathscr{F}$}$$

Note that $\mathscr{F}$ is equivalent to the event $\mathrm{pos}_\succ(c') < \mathrm{pos}_\succ(c)$. For each $j\in[m]$, let $i_{\succ,j}$ be the j-th candidate in $\succ$. Finally, for each $\ell\in[m]$, define the random variable

$$\tau_S(\ell) \coloneqq f_S(i_{\succ,j},\succ).$$

The randomness in $\tau_S(\ell)$ due to the randomness in $\mathrm{pos}(S)$. Conditioned on any value of $\mathrm{pos}(S)$, due to domination sensitivity (Definition 2.2), it holds that

$$\forall 1\leq\ell<k\leq m,\quad \tau_S(\ell)\geq\tau_S(k). \tag{12}$$

Using this notation, we can express $\mathbb{E}_\mu\left[f_S(c,\succ)\right]$ and $\mathbb{E}_\mu\left[f_S(c',\succ)\right]$ as follows

$$\begin{aligned}
&\frac{\mathbb{E}_\mu\left[f_S(c',\succ)\right]}{\mathbb{E}_\mu\left[f_S(c,\succ)\right]}\\
&= \frac{\mathbb{E}_{\mathcal{W}}\left[\sum_{\ell\in[m-2]}\Pr\left[\mathscr{E}_\ell\right]\cdot\left(\Pr\left[\mathscr{F}\mid\mathscr{E}_\ell\right]\cdot\left(\tau_S(\ell)-\tau_S(\ell+1)\right)+\Pr\left[\neg\mathscr{F}\mid\mathscr{E}_\ell\right]\cdot\left(\tau_S(\ell+1)-\tau_S(\ell+2)\right)\right)\right]}{\mathbb{E}_{\mathcal{W}}\left[\sum_{\ell\in[m-2]}\Pr\left[\mathscr{E}_\ell\right]\cdot\left(\Pr\left[\mathscr{F}\mid\mathscr{E}_\ell\right]\cdot\left(\tau_S(\ell+1)-\tau_S(\ell+2)\right)+\Pr\left[\neg\mathscr{F}\mid\mathscr{E}_\ell\right]\cdot\left(\tau_S(\ell)-\tau_S(\ell+1)\right)\right)\right]}\\
&= \frac{\mathbb{E}_{\mathcal{W}}\left[\sum_{\ell\in[m-2]}\Pr\left[\mathscr{E}_\ell\right]\cdot\left(\tau_S(\ell)-\tau_S(\ell+2)\right)\cdot\left(1+\varepsilon\rho\cdot\frac{\tau_{\ell,S}(f)+\tau_{\ell+2,S}(f)-2\tau_{\ell+1,S}(f)}{\tau_{\ell,S}(f)-\tau_{\ell+2,S}(f)}\right)\right]}{\mathbb{E}_{\mathcal{W}}\left[\sum_{\ell\in[m-2]}\Pr\left[\mathscr{E}_\ell\right]\cdot\left(\tau_S(\ell)-\tau_S(\ell+2)\right)\cdot\left(1-\varepsilon\rho\cdot\frac{\tau_{\ell,S}(f)+\tau_{\ell+2,S}(f)-2\tau_{\ell+1,S}(f)}{\tau_{\ell,S}(f)-\tau_{\ell+2,S}(f)}\right)\right]} \qquad (13)\\
&= \min_{T\subseteq C\colon |T|=|S|}\ \min_{\ell\in[m]}\frac{1+\varepsilon\rho\cdot\frac{\tau_{\ell,S}(f)+\tau_{\ell+2,S}(f)-2\tau_{\ell+1,S}(f)}{\tau_{\ell,S}(f)-\tau_{\ell+2,S}(f)}}{1-\varepsilon\rho\cdot\frac{\tau_{\ell,S}(f)+\tau_{\ell+2,S}(f)-2\tau_{\ell+1,S}(f)}{\tau_{\ell,S}(f)-\tau_{\ell+2,S}(f)}} \qquad \text{(Using that } \tau_S(\ell)-\tau_S(\ell+2)\geq 0 \text{ for all } \ell)\\
&\geq \frac{1+\varepsilon\rho\sigma(f)}{1-\varepsilon\rho\sigma(f)}. \qquad \text{(Using Definition B.6 and that } \varepsilon,\rho,\sigma(f)\geq 0)
\end{aligned}$$

□

Part 2 ((β,γ) order preservation with $\sigma(f)=0$): Suppose $\sigma(f)=0$ and $\min_{S\subseteq C}\tau_{m-1,S}>0$. The following is the main lemma in this part.

Lemma B.11. *Let $F=\sum_{v\in V}f(\cdot,\succ_v)$ be a multiwinner score function. Let $(\mu,\widehat{\mu})$ be a utility-based generative model defined in Definitions 2.5 and 2.6. Suppose candidates $c,c'\in C$ are in the same group (G_1 or G_2) and set $S\subseteq C\setminus\{c,c'\}$. If there exists a $\rho>0$ such that $\omega_{c'}(1-\rho)\geq\omega_c$ and $\mathbb{E}_\mu\left[f_S(c',\succ)\right],\mathbb{E}_\mu\left[f_S(c,\succ)\right]>0$, then*

$$\frac{\mathbb{E}_{\widehat{\mu}}\left[f_S(c',\not\succ)\right]}{\mathbb{E}_{\widehat{\mu}}\left[f_S(c,\not\succ)\right]}\geq 1+\rho\varepsilon\cdot m^{-1}\cdot\frac{\tau_{\min}}{\tau}\cdot\Omega(1).$$

(β,γ) order preservation follows from Lemmas B.9 and B.11. Concretely, the proof is as follows.

Proof of (β,γ) order preservation assuming Lemmas B.9 and B.10. From the LHS in Equation (11) and Lemma B.9, it follows that $\omega_{c'}\left(1-m^{-1}\cdot\Theta\left(1-\beta\right)\right)>\omega_c$. Hence, Lemma B.11 applicable with $\rho=m^{-1}\cdot\Theta\left(1-\beta\right)$. From Lemma B.11, it follows that the RHS of Equation (11) holds with $\gamma=1+m^{-2}\cdot\frac{\tau_{\min}}{\tau}\cdot\Theta(1-\beta)$.

In the remainder of this section, we prove Lemma B.11.

Proof of Lemma B.11. We borrow notation from Lemma B.10. In addition, for each $\ell\in[m-1]$, define

$$\sigma_\ell(f)\coloneqq\frac{\tau_{\ell,S}(f)+\tau_{\ell+2,S}(f)-2\tau_{\ell+1,S}(f)}{\tau_{\ell,S}(f)-\tau_{\ell+2,S}(f)}. \qquad (14)$$

Note that $\sigma(f)=\min_{\ell\in[m-1]}\sigma_\ell(f)$. Since $\sigma(f)=0$, it follows that for each $\ell\in[m-1]$,

$$\sigma_\ell(f)\geq 0. \qquad (15)$$

Moreover, since $\tau_{m-1,S}\geq\tau_{\min}>0$ (for all $S\subseteq C\setminus i_{m-1}$) and $\tau_{m,S}=\tau_{m+1,S}=0$ (for all $S\subseteq C$) it follows that

$$\sigma_{m-1}(f)=1. \qquad (16)$$

Now we are ready to lower bound $\frac{\mathbb{E}_\mu\left[f_S(c',\succ)\right]}{\mathbb{E}_\mu\left[f_S(c,\succ)\right]}$: Equation (13) shows that

$$
\begin{aligned}
&\frac{\mathbb{E}_\mu\left[f_S(c',\succ)\right]}{\mathbb{E}_\mu\left[f_S(c,\succ)\right]}\\
&\quad\geq \frac{\mathbb{E}_\mathcal{W}\left[\sum_{\ell\in[m-1]}\Pr\left[\mathscr{E}_\ell\right]\cdot\left(\tau_S(\ell)-\tau_S(\ell+2)\right)\cdot\left(1+\varepsilon\rho\cdot\frac{\tau_{\ell,S}(f)+\tau_{\ell+2,S}(f)-2\tau_{\ell+1,S}(f)}{\tau_{\ell,S}(f)-\tau_{\ell+2,S}(f)}\right)\right]}{\mathbb{E}_\mathcal{W}\left[\sum_{\ell\in[m-1]}\Pr\left[\mathscr{E}_\ell\right]\cdot\left(\tau_S(\ell)-\tau_S(\ell+2)\right)\cdot\left(1-\varepsilon\rho\cdot\frac{\tau_{\ell,S}(f)+\tau_{\ell+2,S}(f)-2\tau_{\ell+1,S}(f)}{\tau_{\ell,S}(f)-\tau_{\ell+2,S}(f)}\right)\right]}\\
&\quad\geq \frac{\mathbb{E}_\mathcal{W}\left[\sum_{\ell\in[m-1]}\Pr\left[\mathscr{E}_\ell\right]\cdot\left(\tau_S(\ell)-\tau_S(\ell+2)\right)\cdot\left(1+\varepsilon\rho\cdot\sigma_\ell(f)\right)\right]}{\mathbb{E}_\mathcal{W}\left[\sum_{\ell\in[m-1]}\Pr\left[\mathscr{E}_\ell\right]\cdot\left(\tau_S(\ell)-\tau_S(\ell+2)\right)\cdot\left(1-\varepsilon\rho\cdot\sigma_\ell(f)\right)\right]} &&\text{(Using Equation (14))}\\
&\quad\geq \frac{\mathbb{E}_\mathcal{W}\left[\sum_{\ell\in[m-2]}\Pr\left[\mathscr{E}_\ell\right]\cdot\left(\tau_S(\ell)-\tau_S(\ell+2)\right)+\Pr\left[\mathscr{E}_{m-1}\right](1+\varepsilon\rho)\cdot\tau_{\min}\right]}{\mathbb{E}_\mathcal{W}\left[\sum_{\ell\in[m-2]}\Pr\left[\mathscr{E}_\ell\right]\cdot\left(\tau_S(\ell)-\tau_S(\ell+2)\right)+\Pr\left[\mathscr{E}_{m-1}\right](1-\varepsilon\rho)\cdot\tau_{\min}\right]} &&\text{(Using Equations (15) and (16))}\\
&\quad\geq \frac{\mathbb{E}_\mathcal{W}\left[(m-2)\tau+\Pr\left[\mathscr{E}_{m-1}\right](1+\varepsilon\rho)\cdot\tau_{\min}\right]}{\mathbb{E}_\mathcal{W}\left[(m-2)\tau+\Pr\left[\mathscr{E}_{m-1}\right](1-\varepsilon\rho)\cdot\tau_{\min}\right]} &&\text{(Using that }\Pr[\mathscr{E}_\ell],\tau_S(\ell)-\tau_S(\ell+2)\geq 0\text{ for all }\ell\in[m-1])\\
&\quad= \frac{(m-2)\tau+(1+\varepsilon\rho)\cdot\tau_{\min}}{(m-2)\tau+(1-\varepsilon\rho)\cdot\tau_{\min}} &&\text{(Using that }\Pr\left[\mathscr{E}_{m-1}\right]=1)\\
&\quad\geq 1+\Omega\left(\varepsilon\rho m^{-1}\cdot\frac{\tau_{\min}}{\tau}\right).
\end{aligned}
$$

□

Now we complete the proof of Lemma B.9.

Proof of Lemma B.9. Let $\lambda \coloneqq 1-\beta$. As $\beta\geq 1-m^{-1/2}$, $\lambda\leq m^{1/2}$. We will prove (a strengthening of) the contra-positive of the above statement: for any candidates $d,d'\in C$ in the same group (G_1 or G_2), the following

$$\frac{\omega_{d'}}{\omega_d}\in 1\pm m^{-1}\cdot\Theta(\lambda)\implies\frac{\mathbb{E}_\mu\left[f_S(d',\succ)\right]}{\mathbb{E}_\mu\left[f_S(d,\succ)\right]}\in 1\pm\Theta(\lambda). \tag{17}$$

From the Log-Lipschitzness in Definition B.8, we have the following result: for any $x,y>0$

$$\frac{x}{y}\in 1\pm\lambda\implies\frac{\mathrm{cdf}_\eta(x)}{\mathrm{cdf}_\eta(y)}\in 1\pm O_\pi(\lambda), \tag{18}$$

where π is a constant in Definition B.8.

Towards proving Equation (17), suppose (without loss of generality) that $\omega_{d'}>\omega_d$. We can express $\mathbb{E}_\mu\left[f(d',\succ)\right]$ and $\mathbb{E}_\mu\left[f(d,\succ)\right]$ as follows

$$\mathbb{E}_\mu\left[f(d',\succ)\right]=\sum_{\ell=1}^m\Pr_\mu\left[\mathrm{pos}_\succ(d')=\ell\right]\tau_{\ell,S}(f)\quad\text{and}\quad\mathbb{E}_\mu\left[f(d,\succ)\right]=\sum_{\ell=1}^m\Pr_\mu\left[\mathrm{pos}_\succ(d)=\ell\right]\tau_{\ell,S}(f). \tag{19}$$

To prove Equation (17), it suffices to show that $\Pr_\mu\left[\mathrm{pos}_\succ(d')=\ell\right]$ and $\Pr_\mu\left[\mathrm{pos}_\succ(d)=\ell\right]$ are multiplicatively close for all $\ell\in[m]$. From Equation (18), for any $i\neq d,d'$, we have the following

$$\forall x\geq 0,\quad\frac{\mathrm{cdf}_\eta\left(\frac{\omega_{d'}\cdot x}{\omega_i}\right)}{\mathrm{cdf}_\eta\left(\frac{\omega_d\cdot x}{\omega_i}\right)}\in 1\pm m^{-1}\cdot\Theta_\pi(\lambda).\qquad\text{(Using that }\frac{\omega_{d'}}{\omega_i}\cdot\frac{\omega_i}{\omega_d}\in 1\pm m^{-1}\Theta(\lambda))$$

Moreover, we also have

$$\forall x>0,\quad\mathrm{cdf}_\eta\left(\frac{\omega_{d'}\cdot x}{\omega_d\cdot x}\right)\in 1\pm m^{-1}\cdot\Theta_\pi(\lambda). \tag{20}$$

We can show this as follows:

$$\frac{\Pr_\mu\left[\mathrm{pos}_\succ(d)=\ell\right]}{\Pr_\mu\left[\mathrm{pos}_\succ(d')=\ell\right]}=\frac{\mathbb{E}_{\eta_{d'}}\left[\sum_{S\subseteq C\setminus\{d'\}:\,|S|=\ell-1}\prod_{i\in S}\left(1-\mathrm{cdf}_\eta\left(\frac{\omega_{d'}\eta_{d'}}{\omega_i}\right)\right)\prod_{i\notin S}\mathrm{cdf}_\eta\left(\frac{\omega_{d'}\eta_{d'}}{\omega_i}\right)\right]}{\mathbb{E}_{\eta_d}\left[\sum_{S\subseteq C\setminus\{d\}:\,|S|=\ell-1}\prod_{i\in S}\left(1-\mathrm{cdf}_\eta\left(\frac{\omega_d\eta_d}{\omega_i}\right)\right)\prod_{i\notin S}\mathrm{cdf}_\eta\left(\frac{\omega_d\eta_d}{\omega_i}\right)\right]} \tag{21}$$

Toward bounding the RHS of the above equality, fix any value $\eta_{d'}=\eta_d=\eta>0$. We have the following

$$\begin{aligned}
&\frac{\sum_{S\subseteq C\setminus\{d'\}:\,|S|=\ell-1}\prod_{i\in S}\left(1-\mathrm{cdf}_\eta\left(\frac{\omega_{d'}}{\omega_i}\right)\right)\prod_{i\notin S}\mathrm{cdf}_\eta\left(\frac{\omega_{d'}}{\omega_i}\right)}{\sum_{S\subseteq C\setminus\{d\}:\,|S|=\ell-1}\prod_{i\in S}\left(1-\mathrm{cdf}_\eta\left(\frac{\omega_d}{\omega_i}\right)\right)\prod_{i\notin S}\mathrm{cdf}_\eta\left(\frac{\omega_d}{\omega_i}\right)}\\
&\quad\geq\ \frac{\mathrm{cdf}_\eta\left(\frac{\omega_{d'}}{\omega_d}\right)}{\mathrm{cdf}_\eta\left(\frac{\omega_d}{\omega_{d'}}\right)}\cdot\frac{\sum_{S\subseteq C\setminus\{d',d\}:\,|S|=\ell-1}\prod_{i\in S}\left(1-\mathrm{cdf}_\eta\left(\frac{\omega_{d'}}{\omega_i}\right)\right)\prod_{i\notin S}\mathrm{cdf}_\eta\left(\frac{\omega_{d'}}{\omega_i}\right)}{\sum_{S\subseteq C\setminus\{d',d\}:\,|S|=\ell-1}\prod_{i\in S}\left(1-\mathrm{cdf}_\eta\left(\frac{\omega_d}{\omega_i}\right)\right)\prod_{i\notin S}\mathrm{cdf}_\eta\left(\frac{\omega_d}{\omega_i}\right)}\\
&\quad\overset{(20)}{\geq}\ \left(1-m^{-1}\Theta(\lambda)\right)\cdot\frac{\sum_{S\subseteq C\setminus\{d',d\}:\,|S|=\ell-1}\prod_{i\in S}\left(1-\mathrm{cdf}_\eta\left(\frac{\omega_{d'}}{\omega_i}\right)\right)\prod_{i\notin S}\mathrm{cdf}_\eta\left(\frac{\omega_{d'}}{\omega_i}\right)}{\sum_{S\subseteq C\setminus\{d',d\}:\,|S|=\ell-1}\prod_{i\in S}\left(1-\mathrm{cdf}_\eta\left(\frac{\omega_d}{\omega_i}\right)\right)\prod_{i\notin S}\mathrm{cdf}_\eta\left(\frac{\omega_d}{\omega_i}\right)}\\
&\quad\overset{(19)}{\geq}\ \left(1-m^{-1}\Theta(\lambda)\right)\cdot\left(1-m^{-1}\Theta(\lambda)\right)^{m-\ell+1}\\
&\quad\geq\ 1-\Theta(\lambda). \qquad\qquad \text{(Using that } \ell\geq 0 \text{ and } 0\leq\lambda\leq 1)
\end{aligned} \tag{22}$$

Substituting Equation (22) in Equation (21) and using the fact that if $\frac{x_1}{y_1},\frac{x_2}{y_2}\geq r$, then $\frac{x_1+x_2}{y_1+y_2}\geq r$ (for any $x_1,x_2,y_1,y_2,r\geq 0$), it follows that

$$\frac{\Pr_\mu\left[\mathrm{pos}_\succ(d)=\ell\right]}{\Pr_\mu\left[\mathrm{pos}_\succ(d')=\ell\right]}\geq 1-\Theta(\lambda). \tag{23}$$

An analogous argument, also shows that

$$\frac{\Pr_\mu\left[\mathrm{pos}_\succ(d)=\ell\right]}{\Pr_\mu\left[\mathrm{pos}_\succ(d')=\ell\right]}\leq 1+\Theta(\lambda). \tag{24}$$

The result follows by substituting Equations (23) and (24) into Equation (19).

□

B.3. Swapping-Based Biased Generative Model

In this section, we introduced a "swapping-based" biased model. Let μ be a generative model of latent preference lists $\succ$, e.g., Definition 2.5. We propose the following generative model for biased preferences which can be seen as a biased variant of the popular Mallows model (Mallows, 1957).

Definition B.12 (Swapping-based generative model of biased preference lists). Let $\phi\in[0,1]$ and $t\geq 1$ be a bias parameter and number of swaps respectively. For any $\succ$, let $A(\succ)\subseteq[m]\times[m]$ be the collection of all pairs (i,j) such that there exist $c\in G_1$ and $c'\in G_2$ with $\mathrm{pos}_\succ(c)=i>j=\mathrm{pos}_\succ(c')$. Let $\succ_1$ be a preference list drawn from μ. Given $\succ_i$ (for any $i\in[t]$), $\succ_{i+1}$ is generated as follows.

1. Sample a pair $(i,j)\in A(\succ_i)$ with probability $\frac{\phi^{i-j}}{Z(\succ_i)}$, where, for any $\succ$, $Z(\succ)\geq 0$ is a normalization factor defined as $Z(\succ)=\sum_{(i',j')\in A(\succ)}\phi^{i'-j'}$; and
2. Swap the candidates at positions i and j in $\succ_i$, and obtain $\succ_{i+1}$.

Define $\not\succ:=\succ_{t+1}$. Let $\widehat{\mu}$ denote the generative model of $\not\succ$ that depends on $\succ$ and ϕ.

Intuitively, we randomly improve the ranking of a candidate from the advantaged group and lower the ranking of a candidate from the disadvantaged group, where the probability is proportional to their ranking difference in $\succ$. As ϕ comes closer 1, the average distance between the positions of swapped candidates increases.

Our next result shows that all multiwinner score functions satisfying Definition 2.2 are order preserving between μ and $\widehat{\mu}$, where $\widehat{\mu}$ arises from the swapping-based generative model of biased preference lists and μ satisfies the following condition for some parameter $\rho > 0$

$$\min_{c\in C} \mathbb{E}_\mu \left[f_{C\setminus\{c\}}(c, \succ)\right] \geq \rho. \tag{25}$$

Lemma B.13 (Order-preserving properties of a swapping-based model). *Let $F = \sum_{v\in V} f(\cdot, \succ_v)$ be a latent multiwinner score function satisfying Definition 2.2. Let μ be any generative model such that F is order-preserving with respect to μ (Definition 2.8) and satisfies Equation (25). For any numbers $t \geq 1$ and $\phi \in (0, t^{-1})$ and the generative model $\widehat{\mu}$ in Definition B.12 with parameters μ, ϕ, and $t = 1$, F is $\left(1-\lambda, 1-\frac{\lambda}{2}\right)$ order preserving between μ and $\widehat{\mu}$ where*

$$\lambda \coloneqq \mathbb{E}_\mu \left[\frac{1}{Z(\succ)}\right] \cdot \Theta\left(\frac{t\phi}{1-\phi} \cdot \frac{\tau}{\rho}\right),$$

and $Z(\succ)$ is the normalizing constant corresponding to preference list $\succ$, as defined in Definition B.12.

Note that Lemma B.13 does not fix a specific generative model of latent preference lists μ. The bound on the parameter γ depends on μ via the term $\mathbb{E}_\mu\left[\frac{1}{Z(\succ)}\right]$. In general, we expect this term to be of the order $\Omega\left(m^{-1}\phi^{-1}\right)$. To see why, note that for any preference list $\succ$ where there are at least r candidates $c' \in G_2$ (for any $r \geq 1$) who are placed before $r^{-1}m$ candidates $c \in G_1$, then $Z(\succ) \geq \phi r$. When $\mathbb{E}_\mu\left[\frac{1}{Z(\succ)}\right] = \Omega\left(m^{-1}\phi^{-1}\right)$, then Lemma B.13 implies that F is $\left(1, 1-O\left(m^{-1}t\right)\right)$ order preserving between μ and $\widehat{\mu}$ for any $\phi \in (0, t^{-1})$.

Additionally, we have the following example that shows that this order preservation between μ and $\widehat{\mu}$ does not hold for all $\beta \in [0,1]$ when $\widehat{\mu}$ the swapping-based bias generative model.

Example B.14 (**Order preservation does not hold for all** $\beta \in [0,1]$). Suppose $C = \{d_1, d_2, a_1\}$, $G_1 = \{a_1\}$, and $G_2 = \{d_1, d_2\}$. Let $F\colon 2^C \to \mathbb{R}_{\geq 0}$ be the 2-Bloc rule. Let $\succ_1, \succ_2$, and $\succ_3$ be the following preference lists

$$\succ_1 \coloneqq (d_2 \succ d_1 \succ a_1), \quad \succ_2 \coloneqq (d_1 \succ a_1 \succ d_2), \quad \succ_3 \coloneqq (d_2 \succ a_1 \succ d_1).$$

For some small $\delta > 0$, let μ be a distribution such that

$$\Pr_{\succ\sim\mu}[\succ=\succ_1] = 1-3\delta, \quad \Pr_{\succ\sim\mu}[\succ=\succ_2] = 2\delta, \quad \Pr_{\succ\sim\mu}[\succ=\succ_3] = \delta.$$

In this case, on the one hand, the following holds

$$\frac{\mathbb{E}_\mu\left[f(d_1, \succ)\right]}{\mathbb{E}_\mu\left[f(d_2, \succ)\right]} = \frac{1-\delta}{1-2\delta} = \frac{1}{1-\varepsilon} > 1 \quad \text{for } \varepsilon \approx \delta.$$

On the other hand,

$$\frac{\mathbb{E}_{\widehat{\mu}}\left[f(d_1, \not\succ)\right]}{\mathbb{E}_{\widehat{\mu}}\left[f(d_2, \not\succ)\right]} = \frac{\Theta(\phi+\delta)}{1 \pm \Theta(\phi+\delta)} = \Theta(\phi+\delta) < 1.$$

B.4. Proof of Lemma B.13: Order-Preserving Properties of the Swapping-Based Model

In this section, we prove Lemma B.13.

To show that F (β, γ) *order preserving between μ and $\widehat{\mu}$*, we need to show that for any two candidates $c, c' \in C$ belonging to the same group (G_1 or G_2) and any $S \subseteq C \setminus \{c, c'\}$, the following holds

$$\beta \cdot \mathbb{E}_\mu\left[f_S(c', \succ)\right] \geq \mathbb{E}_\mu\left[f_S(c, \succ)\right] > 0 \implies \gamma \cdot \mathbb{E}_{\widehat{\mu}}\left[f_S(c', \not\succ)\right] \geq \mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ)\right]. \tag{26}$$

At a high level, we will prove the above implication by bounded $\mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ) \mid \succ\right]$ with an $1 \pm O(\phi t)$ multiplicative factor of $f_S(c, \succ)$ (for any $\succ \in \mathcal{L}(C)$ and $c \in C$) and then taking an expectation with respect to $\succ \sim \mu$. Concretely, we prove the following lemma.

Lemma B.15. *Suppose $t \geq 1$, $\phi \in (0, t^{-1})$, and $\succ \in \mathcal{L}(C)$. Let $\lambda \coloneqq \mathbb{E}_\mu\left[\frac{1}{Z(\succ)}\right] \cdot O\left(\frac{t\tau\phi}{\rho(1-\phi)}\right)$. For any $c \in C$, and $S \subseteq C \setminus \{c\}$, it holds that*

$$\mathbb{E}_\mu\left[f_S(c, \succ)\right] \cdot (1-\lambda) \leq \mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ)\right] \leq \mathbb{E}_\mu\left[f_S(c, \succ)\right] \cdot (1+\lambda).$$

Lemma B.13 follows from the above result by taking an expectation with respect to $\succ\sim\mu$.

Proof of Lemma B.13 assuming Lemma B.15. Fix any two candidates $c, c' \in C$ belonging to the same group (G_1 or G_2) and any $S \subseteq C \setminus \{c, c'\}$. To prove Lemma B.13 it suffices to prove Equation (26). Suppose the following is true

$$\beta \cdot \mathbb{E}_{\mu}\left[f_S(c', \succ)\right] \geq \mathbb{E}_{\mu}\left[f_S(c, \succ)\right] > 0. \tag{27}$$

(If this is not true, Equation (26) vacuously holds.) Using Lemma B.15, for any $\succ \in \mathcal{L}(C)$, we have the following inequalities

$$\begin{aligned}
\mathbb{E}_{\mu}\left[f_S(c', \succ)\right] \cdot (1-\lambda) \quad &\leq \quad \mathbb{E}_{\widehat{\mu}}\left[f_S(c', \not\succ)\right] \quad \leq \quad \mathbb{E}_{\mu}\left[f_S(c', \succ)\right] \cdot (1+\lambda), && (28)\\
\mathbb{E}_{\mu}\left[f_S(c, \succ)\right] \cdot (1-\lambda) \quad &\leq \quad \mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ)\right] \quad \leq \quad \mathbb{E}_{\mu}\left[f_S(c, \succ)\right] \cdot (1+\lambda). && (29)
\end{aligned}$$

Now, we are ready to complete the proof

$$\begin{aligned}
\mathbb{E}_{\widehat{\mu}}\left[f_S(c', \not\succ)\right] &\overset{(28)}{\geq} (1-\lambda) \cdot \mathbb{E}_{\mu}\left[f_S(c', \succ)\right]\\
&\overset{(27)}{\geq} \frac{1-\lambda}{\beta} \cdot \mathbb{E}_{\mu}\left[f_S(c, \succ)\right]\\
&\overset{(29)}{\geq} \frac{1-\lambda}{(1+\lambda)\cdot\beta} \cdot \mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ)\right].
\end{aligned}$$

Lemma B.13 follows by choosing $r := \frac{4\lambda}{1+3\lambda} = \mathbb{E}_{\mu}\left[\frac{1}{Z(\succ)}\right] \cdot O\left(\frac{t\tau\phi}{\rho(1-\phi)}\right)$,

$$\beta = 1 - r, \quad \text{and} \quad \gamma = \frac{\beta(1+\lambda)}{1-\lambda} = 1 - \frac{r}{2}.$$

□

It remains to prove Lemma B.15.

Notation. Fix $c \in C$ and $S \subseteq C \setminus \{c\}$. For each $j \in [m]$, let $i_{\succ,j}$ be the j-th candidate in $\succ$. For each $j \in [m]$, define $\tau_{j,S}(f)$ to be the marginal score of $i_{\succ,j}$ with respect to S, i.e.,

$$\tau_{j,S}(f) := f_S(\{i_{\succ,j}\}, \succ)$$

which is independent of $\succ \in \mathcal{L}(C)$. Fix any draw $\succ$ from μ. For each $1 \leq \ell \leq m$, let $A(\ell) \subseteq [m] \setminus [\ell]$ be the set of indices among $[m] \setminus [\ell]$ where G_1 candidates appear in $\succ$. For each $1 \leq \ell \leq m$, let $B(\ell) \subseteq [\ell]$ be the set of indices among $[\ell]$ where G_2 candidates appear in $\succ$.

We divide the proof into two cases depending on the group of c.

Case A ($c \in G_1$): Suppose $c \in G_1$. Let

$$j := \mathrm{pos}_{\succ}(c).$$

We will first consider the case where there is only one swap, i.e., $t = 1$, and later generalize to multiple swaps. We can express $\mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ) \mid \succ\right]$ as follows.

$$\begin{aligned}
\mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ) \mid \succ\right] &= \sum_{\ell \in [j]} \Pr\left[c \text{ is swapped from } j \text{ to } \ell\right] \cdot \tau_{\ell,S} + \left(1 - \sum_{\ell \in [j]} \Pr\left[c \text{ is swapped from } j \text{ to } \ell\right]\right) \tau_{j,S} \\
&\qquad \text{(Using the facts that } c \in G_1\text{, the positions of candidates in } G_1 \text{ only reduces, and the definition of } \tau_{\ell,S}) \\
&= \sum_{\ell \in B(j)} \Pr\left[c \text{ is swapped from } j \text{ to } \ell\right] \cdot \tau_{\ell,S} + \left(1 - \sum_{\ell \in B(j)} \Pr\left[c \text{ is swapped from } j \text{ to } \ell\right]\right) \tau_{j,S} \\
&\qquad \text{(Using the fact that } c \in G_1 \text{ and candidates in } G_1 \text{ only swap positions with candidates in } G_2 \text{ in one swap)} \\
&= \frac{1}{Z(\succ)} \sum_{\ell \in B(j)} \phi^{j-\ell} \cdot \tau_{\ell,S} + \left(1 - \frac{1}{Z(\succ)} \sum_{\ell \in B(j)} \phi^{j-\ell}\right) \tau_{j,S} \\
&\qquad \text{(By construction in the swapping-based bias model; Definition B.12)} \qquad (30)
\end{aligned}$$

We can upper bound the above expression as follows

$$\begin{aligned}
\mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ) \mid \succ\right] &= \frac{1}{Z(\succ)} \sum_{\ell \in B(j)} \phi^{j-\ell} \cdot \tau_{\ell,S} + \left(1 - \frac{1}{Z(\succ)} \sum_{\ell \in B(j)} \phi^{j-\ell}\right) \tau_{j,S} \\
&\leq \frac{1}{Z(\succ)} \left(\tau_{1,S} \cdot \phi^{j-1} + \tau_{2,S} \cdot \phi^{j-2} + \cdots + \tau_{j-1,S} \cdot \phi\right) + \left(1 - \frac{1}{Z(\succ)} \cdot \left(\phi + \phi^2 + \cdots + \phi^{j-1}\right)\right) \tau_{j,S} \\
&\leq \frac{1}{Z(\succ)} \left(\tau_{1,S} \cdot \phi^{j-1} + \tau_{2,S} \cdot \phi^{j-2} + \cdots + \tau_{j-1,S} \cdot \phi\right) + \tau_{j,S} && \text{(Using that } \phi, Z(\succ) \geq 0) \\
&\leq \frac{\tau_{1,S}}{Z(\succ)} \cdot \left(\phi + \phi^2 + \cdots + \phi^{j-1}\right) + \tau_{j,S} && \text{(Using that } \tau_{1,S} \geq \tau_{2,S} \geq \cdots \geq \tau_m) \\
&\leq \frac{\tau}{Z(\succ)} \cdot \frac{\phi}{1-\phi} + \tau_{j,S} && \text{(Using that } \phi \geq 0 \text{ and } \tau_{1,S} \leq \tau) \\
&= f_S(c, \succ) + \frac{\tau}{Z(\succ)} \cdot \frac{\phi}{1-\phi}.
\end{aligned}$$

Taking the expectation over $\succ \sim \mu$, we get that

$$\begin{aligned}
\mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ)\right] &\leq \mathbb{E}_{\mu}\left[f_S(c, \succ)\right] + \mathbb{E}_{\mu}\left[\frac{1}{Z(\succ)}\right] \cdot \frac{\tau\phi}{1-\phi} \\
&\overset{(25)}{\leq} \mathbb{E}_{\mu}\left[f_S(c, \not\succ)\right] \cdot \left(1 + \mathbb{E}_{\mu}\left[\frac{1}{Z(\succ)}\right] \cdot \frac{\tau\phi}{\rho(1-\phi)}\right).
\end{aligned}$$

Since $c \in G_1$, the position of candidates in G_1 only reduces, and reducing the position increases the score, it follows that

$$\mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ) \mid \succ\right] \geq \tau_{j,S} = f_S(c, \not\succ).$$

Taking the expectation over $\succ \sim \mu$, we get that

$$\mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ)\right] \geq \mathbb{E}_{\mu}\left[f_S(c, \not\succ)\right].$$

Case B ($c \in G_2$): Suppose $c \in G_2$. Let

$$j \coloneqq \mathrm{pos}_{\succ}(c).$$

We will first consider the case where there is only one swap, i.e., $t = 1$, and later generalize to multiple swaps. We can express $\mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ) \mid \succ\right]$ as follows.

$$\mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ) \mid \succ\right] = \sum_{\ell \in [m] \setminus [j]} \Pr\left[c \text{ is swapped from } j \text{ to } \ell\right] \cdot \tau_{\ell,S} + \left(1 - \sum_{\ell \in [m] \setminus [j]} \Pr\left[c \text{ is swapped from } j \text{ to } \ell\right]\right) \tau_{j,S}$$

(Using the facts that $c \in G_2$, the positions of candidates in G_2 only increases, and the definition of $\tau_{\ell,S}$)

$$= \sum_{\ell \in A(j)} \Pr\left[c \text{ is swapped from } j \text{ to } \ell\right] \cdot \tau_{\ell,S} + \left(1 - \sum_{\ell \in A(j)} \Pr\left[c \text{ is swapped from } j \text{ to } \ell\right]\right) \tau_{j,S}$$

(Using the fact that $c \in G_2$ and candidates in G_2 only swap positions with candidates in G_1 in one swap)

$$= \frac{1}{Z(\succ)} \sum_{\ell \in A(j)} \phi^{\ell - j} \cdot \tau_{\ell,S} + \left(1 - \frac{1}{Z(\succ)} \sum_{\ell \in A(j)} \phi^{\ell - j}\right) \tau_{j,S} \tag{31}$$

(By construction in the swapping-based bias model; Definition B.12)

We can lower bound the above expression as follows

$$\begin{aligned}
\mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ) \mid \succ\right] &= \frac{1}{Z(\succ)} \sum_{\ell \in A(j)} \phi^{\ell - j} \cdot \tau_{\ell,S} + \left(1 - \frac{1}{Z(\succ)} \sum_{\ell \in A(j)} \phi^{\ell - j}\right) \tau_{j,S} \\
&\geq \left(1 - \frac{1}{Z(\succ)} \sum_{\ell \in A(j)} \phi^{\ell - j}\right) \tau_{j,S} && \text{(Using that } \phi, \tau_{\ell,S}, Z(\succ) \geq 0 \text{ for all } \ell \in [m]) \\
&\geq \left(1 - \frac{1}{Z(\succ)} \left(\phi + \phi^2 + \dots\right)\right) \tau_{j,S} && \text{(Using that } \phi, \tau_{\ell,S}, Z(\succ) \geq 0 \text{ for all } \ell \in [m]) \\
&= \left(1 - \frac{1}{Z(\succ)} \cdot \frac{\phi}{1 - \phi}\right) \tau_{j,S}.
\end{aligned}$$

Taking the expectation over $\succ \sim \mu$, we get that

$$\begin{aligned}
\mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ)\right] &\geq \mathbb{E}_{\mu}\left[\tau_{j,S}\right] - \frac{\phi}{1 - \phi} \cdot \mathbb{E}_{\mu}\left[\frac{\tau_{j,S}}{Z(\succ)}\right] \\
&\geq \mathbb{E}_{\mu}\left[\tau_{j,S}\right] - \frac{\phi}{1 - \phi} \cdot \mathbb{E}_{\mu}\left[\frac{\tau}{Z(\succ)}\right] && \text{(Using that } \tau_{j,S} \leq \tau) \\
&\geq \mathbb{E}_{\mu}\left[f_S(c, \succ)\right] - \frac{\phi\tau}{1 - \phi} \cdot \mathbb{E}_{\mu}\left[\frac{1}{Z(\succ)}\right] && \text{(Using the definition of } j) \\
&\overset{(25)}{\geq} \mathbb{E}_{\mu}\left[f_S(c, \succ)\right] \left(1 - \frac{\phi\tau}{\rho(1 - \phi)} \cdot \mathbb{E}_{\mu}\left[\frac{1}{Z(\succ)}\right]\right).
\end{aligned}$$

Since $c \in G_2$, the position of candidates in G_2 only increases, and increasing the position does not increase the score, it follows that

$$\mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ) \mid \succ\right] \leq \tau_{j,S} = f_S(c, \not\succ).$$

Taking the expectation over $\succ \sim \mu$, we get that

$$\mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ)\right] \leq \mathbb{E}_{\mu}\left[f_S(c, \not\succ)\right]$$

Completing the proof. Thus, across both cases, the following inequalities hold when $t = 1$:

$$\left(1 - \mathbb{E}_{\mu}\left[\frac{1}{Z(\succ)}\right] \cdot \frac{\tau\phi}{\rho(1 - \phi)}\right) \cdot \mathbb{E}_{\mu}\left[f_S(c, \not\succ)\right] \leq \mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ)\right] \leq \left(1 + \mathbb{E}_{\mu}\left[\frac{1}{Z(\succ)}\right] \cdot \frac{\tau\phi}{\rho(1 - \phi)}\right) \cdot \mathbb{E}_{\mu}\left[f_S(c, \not\succ)\right].$$

Consider a draw of $\not\succ$, say $\not\succ_1$, obtained after performing one swap on $\succ$. Replacing $\succ$ by $\not\succ_1$ in the above proof, we get bounds on utility with the swapping based model with $t = 2$ swaps. Repeating this argument t times, when $\phi \in (0, t^{-1})$, we get the following bounds.

$$\left(1 - \mathbb{E}_\mu\left[\frac{1}{Z(\succ)}\right] \cdot \frac{O(t\tau\phi)}{\rho(1-\phi)}\right) \cdot \mathbb{E}_\mu\left[f_S(c, \not\succ)\right] \leq \mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ)\right] \leq \left(1 + \mathbb{E}_\mu\left[\frac{1}{Z(\succ)}\right] \cdot \frac{O(t\tau\phi)}{\rho(1-\phi)}\right) \cdot \mathbb{E}_\mu\left[f_S(c, \not\succ)\right].$$

C. Missing Proofs From Section 3

C.1. Proof of Theorem 3.2: Main Algorithmic Result

We first provide the algorithm for Theorem 3.2, say Algorithm 1, which is a simple greedy algorithm that first selects ℓ candidates from G_2 in Line 1, then selects $k - \ell$ candidates from G_1 in Line 2, and finally outputs their union S.

Algorithm 1 A greedy algorithm with an intervention constraint

1: **Input:** Numbers $k, \ell \in \mathbb{N}$, sets $G_1, G_2 \subseteq C$, and a value oracle $\mathcal{O}$ for $\widehat{F}(\cdot)$
2: **Output:** A subset $S \subseteq \mathcal{K}(\ell)$ of size k
3: Select $S_2 \coloneqq$ GREEDY$(\ell, G_2, \mathcal{O}, B = \emptyset)$
4: Select $S_1 \coloneqq$ GREEDY$(k, G_1, \mathcal{O}, B = S_2)$
5: **return** $S \coloneqq S_1 \cup S_2$

Algorithm 2 GREEDY(Oracle for F, C, k) ((Nemhauser et al., 1978))

Input: A number $\ell \in \mathbb{N}$, two sets B and G, and a value oracle $\mathcal{O}$ for $\widehat{F}(\cdot)$
Output: A subset $S \subseteq G \cup B$ with $|S| = k$
Initialize $S = B$
while $|S| < k$ **do**
  Set $S = S \cup \arg\max_{i \in G} F_S(i)$
end while
return S

To prove Theorem 3.2, we need to show that for any score function of multiwinner voting $F : 2^C \to \mathbb{R}_{\geq 0}$ that is (α, β, γ)-smooth with respect to generative models $(\mu, \widehat{\mu})$ the following holds. If the number of voters is at least $n \geq \Omega\left(k(\alpha\varepsilon_0)^{-2} \cdot \log\frac{2}{\delta_0}\right)$ (for any $0 < \varepsilon_0, \delta_0 < 1$), then there exists an integer $0 \leq \ell \leq m$ specifying the lower bound constraint such that

$$\Pr_{\mu, \widehat{\mu}}\left[F(S) \geq (\beta - \varepsilon_0) \cdot \mathrm{OPT}\right] \geq 1 - \delta_0.$$

Where S is the subset output by Algorithm 1, given the number ℓ, an oracle $\mathcal{O}$ for $\widehat{F}(\cdot)$, and other parameters (namely, k, G_1, and G_2) as input. Moreover, Algorithm 1 makes $O(mk)$ calls to $\mathcal{O}$ and performs $O(m \log m)$ arithmetic operations.

Fix any $\alpha, \beta, \gamma \in (0, 1]$. Let $F(\cdot) = \sum_{v \in V} f(\cdot, \succ_v)$ be any score function of multiwinner voting that is (α, β, γ)-smooth with respect to generative models $(\mu, \widehat{\mu})$. Recall that $M \subseteq C$ is the set of k candidates c with the highest value $\mathbb{E}_\mu\left[f(c, \succ)\right]$. Define ℓ as the following value

$$\ell \coloneqq |M \cap G_2|$$

We use τ to represent $\tau_1(f)$ for simplicity. We claim that this ℓ satisfies the claim in Theorem 3.2. To simplify the notation, we define the following parameters

$$\varepsilon \coloneqq \frac{\min\{\varepsilon_0, 1-\gamma\}}{4\beta}, \quad \delta \coloneqq \frac{\delta_0}{3}, \quad \text{and} \quad n_0(\varepsilon_0, \delta_0) \coloneqq \frac{51k}{(\alpha \min\{\varepsilon_0, 1-\gamma\})^2} \cdot \log\frac{2}{\delta_0}.$$

We divide the proof of Theorem 3.2 into the following two lemmas.

Lemma C.1. *For any $0 < \varepsilon_0, \delta_0 < 1$, if $n \geq n_0(\varepsilon_0, \delta_0)$, then it holds that*

$$\Pr_{\mu, \widehat{\mu}}\left[F(S) \geq \beta \cdot (1 - \varepsilon) \cdot F(M)\right] \geq 1 - \delta. \tag{32}$$

Lemma C.2. *For any* $0 < \varepsilon_0, \delta_0 < 1$*, if* $n \geq n_0(\varepsilon_0, \delta_0)$*, then it holds that*

$$\Pr_{\mu,\widehat{\mu}}\left[F(M) \geq (1-\varepsilon) \cdot F(S^\star)\right] \geq 1 - 2\delta. \tag{33}$$

Theorem 3.2 follows from the above lemmas as follows.

Proof of Theorem 3.2 assuming Lemmas C.1 and C.2. Due to the lower bound on n in Theorem 3.2, it holds that $n \geq n_0(\varepsilon_0, \delta_0)$. Hence, from Lemmas C.1 and C.2 the Equations (32) and (33) hold. Since $3\delta \leq \delta_0$, taking a union bound over Equations (32) and (33), it follows that

$$\Pr_{\mu,\widehat{\mu}}\left[F(S) \geq \beta \cdot (1-\varepsilon)^2 \cdot F(S^\star)\right] \geq 1 - \delta.$$

Since $(1-\varepsilon)^2 \geq 1 - 2\varepsilon$ (for any $\varepsilon \in \mathbb{R}$) and $2\varepsilon \cdot \beta \leq \varepsilon_0$, it follows, as required, that

$$\Pr_{\mu,\widehat{\mu}}\left[F(S) \geq (\beta - \varepsilon_0) \cdot F(S^\star)\right] \geq 1 - \delta.$$

It remains to bound the number of calls and the number of arithmetic operations in Algorithm 1. Note that Algorithm 2 is called as a subroutine from Algorithm 1 in Steps 1 and 2. Each run of Algorithm 2 makes exactly $k\,|G|$ calls to $\mathcal{O}$ and does $O(|G| \log |G|)$ arithmetic operations (to sort the marginal scores). Algorithm 2 is called twice in Algorithm 1, once with parameters $(k = \ell, |B| = 0, G = G_1)$, and once with $(k = k, |B| = \ell, G = G_2)$. Since $|G_1|, |G_2| \leq m$, the total number oracle calls is $O(mk)$ and the total number of arithmetic operations are $O(m \log m)$. □

In the remainder of this section, we prove Lemmas C.1 and C.2. The proof of both Lemmas C.1 and C.2 uses the following concentration result.

Lemma C.3 (Concentration of marginal utilities). *For any generative models of preference lists* $\mu, \widehat{\mu}$ *and any score function* $F(\cdot) = \sum_{v \in V} f(\cdot, \succ_v)$ *from Definition 2.2 the following holds. For any* $\delta > 0$

$$\Pr_{\mu}\left[\exists_{T=C \text{ or } (T\subseteq C\colon |T|\leq k)},\ \ \exists_{c\in C},\quad \left|F_T(c) - \mathbb{E}_\mu\left[F_T(c)\right]\right| \geq \tau\sqrt{nk\log\frac{m}{\delta}}\right] \leq \delta,$$
$$\Pr_{\widehat{\mu}}\left[\exists_{T=C \text{ or } (T\subseteq C\colon |T|\leq k)},\ \ \exists_{c\in C},\quad \left|\widehat{F}_T(c) - \mathbb{E}_{\widehat{\mu}}\left[\widehat{F}_T(c)\right]\right| \geq \tau\sqrt{nk\log\frac{m}{\delta}}\right] \leq \delta.$$

Here, we slightly abuse the notation and denote singleton sets $\{c\}$ by c. We also note in passing that concentration inequality holds for any generative models of preference lists $(\mu, \widehat{\mu})$ and not just the generative models that satisfy Definition 2.8.

Proof. We first prove the first inequality. Since F is a separable function (Definition 2.2) and for each voter $v \in V$, their preference list $\succ_v$ is drawn iid from μ, for any $c \in C$, $\{f(c, \succ_v)\}_{v \in V}$ is a set of iid and bounded random variables. The concentration inequality follows from the Hoeffding's inequality and the union bound (Motwani & Raghavan, 1995) as shown next.

Fix any $T \subseteq C$ and any $c \in C$. For each $v \in V$, define the random variable $Z_v \coloneqq f_T(c, \succ_v)$. As discussed, Z_u and Z_v are independent for any $u \neq v$. Moreover, for all $v \in V$, $0 \leq Z_v \leq \tau$ with probability 1 (by the non-negativity of f and the definition of τ). Hence, Hoeffding's inequality is applicable on $F(c) = \sum_{v \in V} f(c, \succ_v)$ (Motwani & Raghavan, 1995). From the Hoeffding's inequality (Motwani & Raghavan, 1995), it holds that

$$\begin{aligned}
\Pr_{\mu}\left[\left|F_T(c) - \mathbb{E}_\mu\left[F_T(c)\right]\right| \geq \tau\sqrt{nk\log\frac{m}{\delta}}\right] &\leq \exp\left(-\frac{2}{n\tau^2} \cdot \tau^2 \cdot n \cdot k \cdot \log\left(\frac{m}{\delta}\right)\right) \\
&\leq \frac{\delta^{2k}}{m^{2k}} \\
&\leq \frac{\delta}{m^{2k}}. && \text{(Using that } 0 \leq \delta \leq 1\text{)}
\end{aligned}$$

The first concentration inequality in Lemma C.3 follows by taking the union bound over all choices of $T \subseteq C$ of either (1) size at most k or (2) $T = C$, and any $c \in C$, as there are at most $2^k \cdot m + 1 \leq m^{2k}$ choices of (T, c). The proof of the second inequality follows by replacing μ and F by $\widehat{\mu}$ and $\widehat{F}$ in the above proof. □

C.1.1. Proof of Lemma C.1

The proof is divided into multiple steps. We begin by defining the additional notation used in this proof.

Notation. Recall that $\ell = |M \cap G_2|$. Define the following sets

$$S \cap G_1 \coloneqq \{a_1, a_2, \ldots, a_{k-\ell}\} \quad \text{and} \quad S \cap G_2 \coloneqq \{b_1, b_2, \ldots, b_\ell\},$$
$$M \cap G_1 \coloneqq \{x_1, x_2, \ldots, x_{k-\ell}\} \quad \text{and} \quad M \cap G_2 \coloneqq \{y_1, y_2, \ldots, y_\ell\}.$$

Where the elements in $S \cap G_1$ and $S \cap G_2$ are in the order they are selected in Algorithm 1. The elements in $M \cap G_1$ and $M \cap G_2$ are ordered in non-increasing order by $\mathbb{E}_\mu[F(\cdot)]$: for all $i \in [k-\ell]$ and $j \in [\ell]$,

$$\mathbb{E}_\mu[F(x_i)] \geq \mathbb{E}_\mu[F(x_{i+1})] \quad \text{and} \quad \mathbb{E}_\mu[F(y_j)] \geq \mathbb{E}_\mu[F(y_{j+1})].$$

To simplify the notation, for each $i \in [k-\ell]$ and $j \in [\ell]$, define the following prefixes:

$$A(i) \coloneqq \{a_1, a_2, \ldots, a_i\} \quad \text{and} \quad B(j) \coloneqq \{b_1, b_2, \ldots, b_j\},$$
$$X(i) \coloneqq \{x_1, x_2, \ldots, x_i\} \quad \text{and} \quad Y(j) \coloneqq \{y_1, y_2, \ldots, y_j\}.$$

Define $A(0)$, $B(0)$, $X(0)$, and $Y(0)$ as empty sets. Since $B(j-1)$ has $j-1$ elements by the Pigeonhole principle (for any $j \in [\ell]$), there exists a $y \in Y(j)$ such that $y \notin B(j-1)$. Label this y as $y_{(j)}$. Similarly, there exists an $x_{(i)} \in X(i)$ (for any $i \in [k-\ell]$) such that $x_{(i)} \notin A(i-1)$. Let $\mathscr{E}$ be the following event

$$\forall_{T=C \text{ or } (T\subseteq C:\ |T|\leq k)},\ \forall_{c\in C},\qquad |F_T(c) - \mathbb{E}_\mu[F_T(c)]| \leq \tau\sqrt{nk\log\frac{m}{\delta}} \quad \text{and} \quad \left|\widehat{F}_T(c) - \mathbb{E}_{\widehat{\mu}}\left[\widehat{F}_T(c)\right]\right| \leq \tau\sqrt{nk\log\frac{m}{\delta}}.$$

Lemma C.3 shows that $\Pr[\mathscr{E}] \geq 1 - 2\delta$.

Lemma C.4. *Fix any $i \in [k-\ell]$ and $j \in [\ell]$. Conditioned on the event $\mathscr{E}$, the following inequalities hold*

$$F_{B(\ell)\cup A(i-1)}(a_i) \geq (1-\varepsilon)\cdot\beta\cdot F_{B(\ell)\cup A(i-1)}(x_{(i)}) \quad \textit{and} \quad F_{B(j-1)}(b_j) \geq (1-\varepsilon)\cdot\beta\cdot F_{B(j-1)}(y_{(j)}).$$

Proof. Fix any $i \in [k-\ell]$ and $j \in [\ell]$. Suppose the event $\mathscr{E}$ holds.

Step 1 (Lower bound on $\widehat{F}_{B(j-1)(b_j)}$ and $\widehat{F}_{B(j-1)(y_{(j)})}$): Consider the step where the set of items selected so far is $B(j-1)$. At this step, Algorithm 1 selects the item with the largest value with respect to $\widehat{F}_{B(j-1)}(\cdot)$. Since Algorithm 1 selects b_j instead of $y_{(j)}$. it must hold that

$$\widehat{F}_{B(j-1)}(b_j) \geq \widehat{F}_{B(j-1)}(y_{(j)}). \tag{34}$$

Since $\mathscr{E}$ holds, it follows that

$$\widehat{F}_{B(j-1)}(y_{(j)}) \geq \mathbb{E}_{\widehat{\mu}}\left[\widehat{F}_{B(j-1)}(y_{(j)})\right] - \tau\sqrt{nk\log\frac{m}{\delta}}. \tag{35}$$

Since $y_{(j)} \in M$ and $y_{(j)} \notin B(j-1)$, one can use the definition of α (Definition 3.1) to get the following lower bound

$$\mathbb{E}_{\widehat{\mu}}\left[\widehat{F}_{B(j-1)}(y_{(j)})\right] \geq \mathbb{E}_{\widehat{\mu}}\left[\widehat{F}_{C\setminus\{y_{(j)}\}}(y_{(j)})\right]$$
$$\text{(Using that for any submodular function } F,\ F_R(c) \geq F_{R\cup T}(c) \text{ for any sets } R \text{ and } T \text{ and element } c) \tag{36}$$
$$\geq \alpha\tau n. \qquad \text{(Using the definition of } \alpha \text{ and that } y_{(j)} \in M) \tag{37}$$

Using Inequalities (34), (35), and (37) and the fact that $n \geq n(\varepsilon_0, \delta_0) \geq 4\alpha^{-2}k\log\frac{m}{\delta}$, it follows that

$$\widehat{F}_{B(j-1)}(b_j),\ \ \widehat{F}_{B(j-1)}(y_{(j)}) \geq \frac{\alpha\tau n}{2}. \tag{38}$$

Step 2 (Lower bound on $\frac{\mathbb{E}_\mu\left[F_{B(j-1)}(b_j)\right]}{\mathbb{E}_\mu\left[F_{B(j-1)}(y_{(j)})\right]}$): From Inequalities (35) and (37), and the fact that $n \geq n(\varepsilon_0, \delta_0) \geq 4\varepsilon^{-2}\alpha^{-2}k \log \frac{m}{\delta}$, it follows that

$$\widehat{F}_{B(j-1)}(y_{(j)}) \geq (1-\varepsilon) \cdot \mathbb{E}_{\widehat{\mu}}\left[\widehat{F}_{B(j-1)}(y_{(j)})\right]. \tag{39}$$

Since the event $\mathscr{E}$ holds, it also follows that

$$\begin{aligned}\mathbb{E}_{\widehat{\mu}}\left[\widehat{F}_{B(j-1)}(b_j)\right] &\geq \widehat{F}_{B(j-1)}(b_j) - \tau\sqrt{nk\log\frac{m}{\delta}} \\ &\overset{(34)}{\geq} \widehat{F}_{B(j-1)}(y_{(j)}) - \tau\sqrt{nk\log\frac{m}{\delta}} \\ &\geq (1-\varepsilon)\cdot \widehat{F}_{B(j-1)}(y_{(j)}). \\ &\text{(Using Equation (38) and the fact that } n \geq n(\varepsilon_0,\delta_0) \geq 4\varepsilon^{-2}\alpha^{-2}k\log\tfrac{m}{\delta}\text{)}\end{aligned} \tag{40}$$

Chaining Inequalities (39) and (40), it follows that

$$\mathbb{E}_{\widehat{\mu}}\left[\widehat{F}_{B(j-1)}(b_j)\right] \overset{(40)}{\geq} (1-\varepsilon)\cdot\widehat{F}_{B(j-1)}(y_{(j)}) \overset{(39)}{\geq} (1-\varepsilon)^2\cdot\mathbb{E}_{\widehat{\mu}}\left[\widehat{F}_{B(j-1)}(y_{(j)})\right]. \tag{41}$$

If $\frac{\mathbb{E}_\mu\left[F_{B(j-1)}(b_j)\right]}{\mathbb{E}_\mu\left[F_{B(j-1)}(y_{(j)})\right]} \leq \beta$, then using (β,γ) order-preservation between μ and $\widehat{\mu}$, it follows that

$$\frac{\mathbb{E}_{\widehat{\mu}}\left[\widehat{F}_{B(j-1)}(b_j)\right]}{\mathbb{E}_{\widehat{\mu}}\left[\widehat{F}_{B(j-1)}(y_{(j)})\right]} \leq \gamma. \tag{42}$$

Since $\mathscr{E}$ holds, the above inequality implies that

$$\begin{aligned}\widehat{F}_{B(j-1)}(y_{(j)}) &\geq \gamma^{-1}\cdot\widehat{F}_{B(j-1)}(b_j) - 2\tau\sqrt{nk\log\frac{m}{\delta}} && \text{(Using that } \gamma \leq 1\text{)} \\ &\geq \gamma^{-1}\cdot(1-\varepsilon)\cdot\widehat{F}_{B(j-1)}(b_j) \\ &\qquad\text{(Using Equation (38) and the fact that } n \geq n(\varepsilon_0,\delta_0) \geq 4\gamma^2(\varepsilon\alpha)^{-2}k\log\tfrac{m}{\delta}\text{)} \\ &\geq \gamma^{-1}\cdot(1-\varepsilon)\cdot\widehat{F}_{B(j-1)}(b_j).\end{aligned}$$

Since $\varepsilon < 1-\gamma$, the above equation is a contradiction to Equation (34). Hence,

$$\frac{\mathbb{E}_\mu\left[F_{B(j-1)}(b_j)\right]}{\mathbb{E}_\mu\left[F_{B(j-1)}(y_{(j)})\right]} \geq \beta. \tag{43}$$

Step 3 (Completing the proof of the claim): Since $y_{(j)} \in M$ and $y_{(j)} \notin B(j-1)$, the definition of α (Definition 3.1) implies that

$$\mathbb{E}_\mu\left[F_{B(j-1)}(y_{(j)})\right] \geq \alpha\tau n. \tag{44}$$

Substituting this in Equation (43) gives the following inequality

$$\mathbb{E}_\mu\left[F_{B(j-1)}(b_j)\right] \geq \beta\cdot\mathbb{E}_\mu\left[F_{B(j-1)}(y_{(j)})\right]. \tag{45}$$

Which, as $\mathscr{E}$ holds and $\beta, (1-\varepsilon)^2 \leq 1$, implies the following inequality

$$F_{B(j-1)}(b_j) \geq \beta\cdot F_{B(j-1)}(y_{(j)}) - 2\tau\sqrt{nk\log\frac{m}{\delta}} \tag{46}$$

Equation (44) and the fact that $\mathscr{E}$ holds, also gives us that

$$\begin{aligned}F_{B(j-1)}(y_{(j)}) &\geq \alpha\tau n - \tau\sqrt{nk\log\frac{m}{\delta}} \\ &\geq \alpha\tau n\cdot\left(1 - \frac{1}{\alpha\tau n}\cdot\tau\sqrt{nk\log\frac{m}{\delta}}\right) \\ &\geq \frac{\alpha n\tau}{2}. && \text{(Using that } n \geq n_0(\varepsilon_0,\delta_0) \geq 4\cdot\alpha^{-2}\cdot k\log\tfrac{m}{\delta}\text{)}\end{aligned} \tag{47}$$

Substituting this lower bound in Equation (46) gives us the following multiplicative lower bound on $F_{B(j-1)}(b_j)$

$$\begin{aligned} F_{B(j-1)}(b_j) &\overset{(47)}{\geq} \beta \cdot F_{B(j-1)}(y_{(j)}) \cdot \left(1 - \frac{4}{\beta\alpha n\tau} \cdot \tau\sqrt{nk\log\frac{m}{\delta}}\right) \\ &\geq \beta \cdot (1-\varepsilon) \cdot F_{B(j-1)}(y_{(j)}). \qquad \text{(Using that } n \geq n(\varepsilon_0, \delta_0) \geq 16 \cdot (\alpha\varepsilon \cdot \beta)^{-2} \cdot k \log\tfrac{m}{\delta}\text{)} \end{aligned}$$

Replacing $B(j-1), b_j$ and $y_{(j)}$ by $B(\ell) \cup A(i-1)$, a_i, and $x_{(i)}$ in Steps 1, 2, and 3 shows that

$$F_{B(\ell)\cup A(i-1)}(a_i) \geq \beta \cdot (1-\varepsilon) \cdot F_{B(\ell)\cup A(i-1)}(x_{(i)}).$$

□

Completing the proof of Lemma C.1. Now we are ready to complete the proof of Lemma C.1.

Proof of Lemma C.1. Suppose the event $\mathscr{E}$ holds. $F(S)$ can be lower bounded as follows

$$\begin{aligned} F(S) &= \sum_{j=1}^{\ell} F_{B(j-1)}(b_j) + \sum_{i=1}^{k-\ell} F_{B(\ell)\cup A(i-1)}(a_i) \\ &\geq \beta \cdot (1-\varepsilon) \cdot \left(\sum_{j=1}^{\ell} F_{B(j-1)}(y_{(j)}) + \sum_{i=1}^{k-\ell} F_{B(\ell)\cup A(i-1)}(x_{(i)})\right). \end{aligned}$$

(Using Lemma C.4 and the fact that $\mathscr{E}$ holds) (48)

Next, we lower bound the expected value of each term in the parenthesis in the RHS. Fix any $j \in [\ell]$. Let $c_1, c_2, \ldots, c_{j-1} \subseteq B(j-1)$ be a rearrangement of the elements of $B(j-1)$ such that $\mathbb{E}_\mu\left[F(c_r)\right] \geq \mathbb{E}_\mu\left[F(c_{r+1})\right]$ for each $r \in [j-2]$. Since $Y(j-1)$ is the set $j-1$ candidates in G_2 corresponding to the $j-1$ largest values in $\{\mathbb{E}_\mu\left[F(c)\right] \mid c \in G_2\}$, it follows that

$$\forall_{r\in[j-1]}, \quad \mathbb{E}_\mu\left[F(y_r)\right] \geq \mathbb{E}_\mu\left[F(c_r)\right]. \tag{49}$$

We consider two cases depending on the value of $y_{(j)} \in Y(j)$ to compute a lower bound.

Case A ($y_{(j)} \notin Y(j-1)$): Since $y_{(j)} \in Y(j)$ and $y_{(j)} \notin Y(j-1)$, in this case $y_{(j)} = y_j$. The following lower bound holds

$$\mathbb{E}_\mu\left[F_{\{c_1,c_2,\ldots,c_{j-1}\}}(y_{(j)})\right] = \mathbb{E}_\mu\left[F_{\{c_1,c_2,\ldots,c_{j-1}\}}(y_j)\right] \qquad \text{(Using that, in this case, } y_{(j)} = y_j\text{)} \tag{50}$$

Let $R \subseteq T \subseteq C$ be the following sets

$$R \coloneqq \{c_2, \ldots, c_{j-1}\} \quad \text{and} \quad T \coloneqq \{c_2, \ldots, c_{j-1}, y_j\}.$$

Substituting these in the above equation, implies the following equation

$$\begin{aligned} \mathbb{E}_\mu\left[F_{\{c_1,c_2,\ldots,c_{j-1}\}}(y_j)\right] &= \mathbb{E}_\mu\left[F\left(c_1 \cup T\right)\right] - \mathbb{E}_\mu\left[F\left(c_1 \cup R\right)\right] \\ &= \mathbb{E}_\mu\left[F_T\left(c_1\right)\right] - \mathbb{E}_\mu\left[F_R\left(c_1\right)\right] + \mathbb{E}_\mu\left[F(T)\right] - \mathbb{E}_\mu\left[F(R)\right] \\ &\geq \mathbb{E}_\mu\left[F_T\left(y_1\right)\right] - \mathbb{E}_\mu\left[F_R\left(y_1\right)\right] + \mathbb{E}_\mu\left[F(T)\right] - \mathbb{E}_\mu\left[F(R)\right] \\ &\qquad \text{(Using } \mathbb{E}_\mu\left[F\left(y_1\right)\right] \geq \mathbb{E}_\mu\left[F\left(c_1\right)\right] \text{ and the fact that } F \text{ is order-preserving with respect to } \mu\text{)} \\ &= \mathbb{E}_\mu\left[F\left(y_1 \cup T\right)\right] - \mathbb{E}_\mu\left[F\left(y_1 \cup R\right)\right] \\ &= \mathbb{E}_\mu\left[F_{\{y_1,c_2,\ldots,c_{j-1}\}}(y_j)\right]. \end{aligned}$$

Similarly, using that $\mathbb{E}_\mu\left[F\left(y_2\right)\right] \geq \mathbb{E}_\mu\left[F\left(c_2\right)\right]$, it follows that

$$\mathbb{E}_\mu\left[F_{\{y_1,c_2,\ldots,c_{j-1}\}}(y_j)\right] \geq \mathbb{E}_\mu\left[F_{\{y_1,y_2,\ldots,c_{j-1}\}}(y_j)\right].$$

More generally, it holds that for each $r \in [j-2]$

$$\mathbb{E}_\mu\left[F_{\{y_1,\ldots,y_{r-1},c_r,c_{r+1},\ldots,c_{j-1}\}}(y_j)\right] \geq \mathbb{E}_\mu\left[F_{\{y_1,\ldots,y_{r-1},y_r,c_{r+1},\ldots,c_{j-1}\}}(y_j)\right].$$

Chaining these $j-2$ inequalities and Equality (50), it follows that

$$\mathbb{E}_\mu\left[F_{\{c_1,c_2,\ldots,c_{j-1}\}}(y_{(j)})\right] \overset{(50)}{\geq} \mathbb{E}_\mu\left[F_{\{c_1,c_2,\ldots,c_{j-1}\}}(y_j)\right] \geq \mathbb{E}_\mu\left[F_{\{y_1,y_2,\ldots,y_{j-1}\}}(y_j)\right]. \tag{51}$$

Case B ($y_{(j)} \in Y(j-1)$): Suppose $y_{(j)} = y_s \in Y(j-1)$. In this case, the following lower bound holds

$$\begin{aligned}
\mathbb{E}_\mu\left[F_{\{c_1,c_2,\ldots,c_{j-1}\}}(y_{(j)})\right] &= \mathbb{E}_\mu\left[F_{\{c_1,c_2,\ldots,c_{j-1}\}}(y_s)\right] && \text{(Using that, in this case, } y_{(j)} = y_s\text{)}\\
&\geq \mathbb{E}_\mu\left[F_{\{c_1,c_2,\ldots,c_{j-1}\}}(y_j)\right]\\
&\qquad \text{(Using } \mathbb{E}_\mu\left[F\left(y_s\right)\right] \geq \mathbb{E}_\mu\left[F\left(y_j\right)\right] \text{ and the fact that } F \text{ is order-preserving with respect to } \mu\text{)}\\
&\geq \mathbb{E}_\mu\left[F_{\{y_1,y_2,y_3,\ldots,y_{j-1}\}}\left(y_j\right)\right]. && \text{(Using Equation (51))}
\end{aligned}$$

Hence, in either case, the following holds

$$\mathbb{E}_\mu\left[F_{\{c_1,c_2,\ldots,c_{j-1}\}}(y_{(j)})\right] \overset{\{c_1,c_2,\ldots,c_{j-1}\}=B(j-1)}{=} \mathbb{E}_\mu\left[F_{B(j-1)}(y_{(j)})\right] \geq \mathbb{E}_\mu\left[F_{Y(j-1)}\left(y_j\right)\right]. \tag{52}$$

Replacing $B(j-1)$, $Y(j-1)$, $y_{(j)}$, G_2, and j by $B(\ell) \cup A(i-1)$, $X(i-1)$, $x_{(i)}$, G_1, and i gives the following lower bound

$$\mathbb{E}_\mu\left[F_{B(\ell)\cup A(i-1)}(x_{(i)})\right] \geq \mathbb{E}_\mu\left[F_{Y(\ell)\cup X(i-1)}\left(x_i\right)\right]. \tag{53}$$

Substituting Equations (52) and (53) in Equation (48), we get that

$$\begin{aligned}
F(S) &\geq \beta \cdot (1-\varepsilon) \cdot \left(\sum_{j=1}^{\ell} F_{Y(j-1)}(y_j) + \sum_{i=1}^{k-\ell} F_{Y(\ell)\cup X(i-1)}(x_i)\right)\\
&= \beta \cdot (1-\varepsilon) \cdot \left(F\left(Y(\ell) \cup X(k-\ell)\right)\right)\\
&= \beta \cdot (1-\varepsilon) \cdot F\left(M\right). && \text{(Using that } Y(\ell) \cup X(k-\ell) = M\text{)}
\end{aligned}$$

□

C.1.2. Proof of Lemma C.2

Proof. Let $\mathscr{E}$ be the following event

$$\forall_{T=C \text{ or } (T\subseteq C:\ |T|\leq k)},\ \forall_{c\in C}, \qquad \left|F_T(c) - \mathbb{E}_\mu\left[F_T(c)\right]\right| \leq \tau\sqrt{nk\log\frac{m}{\delta}} \quad \text{and} \quad \left|\widehat{F}_T(c) - \mathbb{E}_{\widehat{\mu}}\left[\widehat{F}_T(c)\right]\right| \leq \tau\sqrt{nk\log\frac{m}{\delta}}.$$

From Lemma C.3, it follows that $\Pr[\mathscr{E}] \geq 1-2\delta$. Suppose the event $\mathscr{E}$ holds. Since M is the set of k candidates with the largest values of $\mathbb{E}_\mu\left[F(\cdot)\right]$ and F is order preserving with respect to μ, it holds that

$$\forall\, d \in M,\ \forall\, c \notin M, \quad \mathbb{E}_\mu\left[F_S(d)\right] \geq \mathbb{E}_\mu\left[F_S(c)\right].$$

Further, as $\mathscr{E}$ holds, the following inequality also holds

$$\forall\, d \in M,\ \forall\, c \notin M, \quad F_S(d) \geq F_S(c) - 2\tau\sqrt{nk\log\frac{m}{\delta}}. \tag{54}$$

Suppose $|M \cap S^\star| = a$. Since both M and $S^\star$ have size k, it holds that $|M \backslash S^\star| = |S^\star \backslash M| = k-a$. Let

$$M\backslash S^\star \coloneqq \{d_1, d_2, \ldots, d_{k-a}\} \quad \text{and} \quad S^\star\backslash M \coloneqq \{c_1, c_2, \ldots, c_{k-a}\}.$$

For each $i \in [k-a]$, define

$$D(i) \coloneqq \{d_1, d_2, \ldots, d_i\} \quad \text{and} \quad C(i) \coloneqq \{c_1, c_2, \ldots, c_i\}$$

For $i=0$, define $D(0)$ and $C(0)$ to be the emptyset. Conditioned on $\mathscr{E}$, it holds that

$$F(M) = F(M \cap S^\star) + F_{M\cap S^\star}(M \backslash S^\star) = F(M \cap S^\star) + F_{M\cap S^\star}\left(\{d_1, d_2, \ldots, d_{k-a}\}\right). \tag{55}$$

Next, we lower bound $F_{M\cap S^\star}\ \{d_1, d_2, \ldots, d_{k-a}\})$. The following inequality holds

$$\begin{aligned}
F_{M\cap S^\star}\left(\{d_1, d_2, \ldots, d_{k-a}\}\right) &= F_{M\cap S^\star}\left(\{d_2, \ldots, d_{k-a}\}\right) + F_{(M\cap S^\star)\cup\{d_2,\ldots,d_{k-a}\}}\left(d_1\right) \\
&\overset{(54)}{\geq} F_{M\cap S^\star}\left(\{d_2, \ldots, d_{k-a}\}\right) + F_{(M\cap S^\star)\cup\{d_2,\ldots,d_{k-a}\}}\left(c_1\right) - 2\tau\sqrt{nk\log\frac{m}{\delta}} \\
&= F_{M\cap S^\star}\left(\{c_1, d_2, \ldots, d_{k-a}\}\right) - 2\tau\sqrt{nk\log\frac{m}{\delta}}.
\end{aligned}$$

Similarly, for any $r \in [k-a]$, it holds that

$$F_{M\cap S^\star}\left(\{c_1, \ldots, c_{r-1}, d_r, d_{r+1}\ldots, d_{k-a}\}\right) \geq F_{M\cap S^\star}\left(\{c_1, \ldots, c_{r-1}, c_r, d_{r+1}\ldots, d_{k-a}\}\right) - 2\tau\sqrt{nk\log\frac{m}{\delta}}.$$

Chaining these $k-a$ inequalities, we get that

$$F_{M\cap S^\star}\left(\{d_1, d_2, \ldots, d_{k-a}\}\right) \geq F_{M\cap S^\star}\left(\{c_1, c_2, \ldots, c_{k-a}\}\right) - 2(k-a)\cdot\tau\sqrt{nk\log\frac{m}{\delta}}.$$

Substituting this in Equation (55) and upper bounding the coefficient of the second term, $2(k-a)$, by $2k$ implies that

$$\begin{aligned}
F(M) &\geq F\left(M \cap S^\star\right) + F_{M\cap S^\star}\left(\{c_1, c_2, \ldots, c_{k-a}\}\right) - 2k\cdot\tau\sqrt{nk\log\frac{m}{\delta}} \\
&\geq F\left(S^\star\right) - 2k\cdot\tau\sqrt{nk\log\frac{m}{\delta}}.
\end{aligned} \tag{56}$$

To convert this to a multiplicative guarantee, we need to lower bound $F(S^\star)$. Since $S^\star$ maximizes F among all sets of size at most k and M has size k, a lower bound on $F(S^\star)$ is as follows

$$\begin{aligned}
F(S^\star) &\geq F(M) \\
&\geq \sum_{c\in M} F_{C\backslash\{c\}}(c) \\
&\geq k\cdot\min_{c\in M} F_{C\backslash\{c\}}(c) \\
&\geq k\cdot\min_{c\in M}\mathbb{E}_\mu\left[F_{C\backslash\{c\}}(c)\right] - k\tau\sqrt{nk\log\frac{m}{\delta}} && \text{(Using that the event } \mathscr{E} \text{ holds)} \\
&\geq kn\cdot\min_{c\in M}\mathbb{E}_\mu\left[f_{C\backslash\{c\}}(c,\succ)\right] - k\tau\sqrt{nk\log\frac{m}{\delta}} && \text{(Using the separability of } F\text{; see Definition 2.2)} \\
&\geq \alpha kn\tau - k\tau\sqrt{nk\log\frac{m}{\delta}} && \text{(Using the definition of } \alpha\text{; see Definition 3.1)} \\
&\geq \frac{\alpha kn\tau}{2} && \text{(Using that } n \geq n_0(\varepsilon_0,\delta_0) \geq \alpha^{-2}k\log\tfrac{m}{\delta}\text{)}
\end{aligned} \tag{57}$$

Using this inequality we can get a multiplicative lower bound on $F(M)$ as follows

$$\begin{aligned}
F(M) &\overset{(56)}{\geq} F\left(S^\star\right)\left(1 - \frac{2k}{F\left(S^\star\right)}\cdot\tau\sqrt{nk\log\frac{m}{\delta}}\right) \\
&\overset{(57)}{\geq} F\left(S^\star\right)\left(1 - \frac{4k}{\alpha kn\tau}\cdot\tau\sqrt{nk\log\frac{m}{\delta}}\right) \\
&\geq (1-\varepsilon)\cdot F\left(S^\star\right). && \text{(Using that } n \geq n_0(\varepsilon_0,\delta_0) \geq 16\alpha^{-2}k\log\tfrac{m}{\delta}\text{)}
\end{aligned} \tag{58}$$

it follows that $F(S^\star) \geq \min_{c\in M} F(c)$. From the definition of α, we have that $\min_{c\in M} \mathbb{E}_\mu [F(c)] \geq n\alpha\tau$. Conditioned on $\mathscr{E}$, it holds that $\min_{c\in M} F(c) \geq \min_{c\in M} \mathbb{E}_\mu [F(c)] - \tau\sqrt{nk\log\frac{m}{\delta}}$. Chaining the last three inequalities, shows that conditioned on $\mathscr{E}$

$$F(S^\star) \geq n\alpha\tau - \tau\sqrt{nk\log\frac{m}{\delta}} \geq \frac{n\alpha\tau}{2} \qquad \text{(Since } n \geq n(\varepsilon_0, \delta_0) \geq \alpha^{-2}\cdot k\cdot\log\tfrac{m}{\delta}) \quad (59)$$

Finally, as $\Pr[\mathscr{E}] \geq 1 - 2\delta$, the above inequality implies the desired result:

$$\Pr_{\mu,\widehat{\mu}}\left[F(M) \geq F(S^\star)\cdot(1-\varepsilon)\right] \geq 1 - 2\delta.$$

□

C.2. Additional Remarks About Theorem 3.2

Our definition of smoothness is also relevant to works that study the robustness of scoring functions to noise in the "ground truth" preference list. These works assess the capabilities of scoring functions to uncover some ground truth from noisy signals, which is a popular question in the field of epistemic social choice (Procaccia et al., 2012; Caragiannis et al., 2016; Caragiannis & Micha, 2017a; Caragiannis et al., 2022). Our results and smoothness definitions extend to this setting as follows. Suppose there is a ground truth preference list $\succ_{\texttt{truth}}$ of all candidates. The generative model μ of latent preferences returns $\succ_{\texttt{truth}}$ with probability one, and all candidates are part of the disadvantaged group (i.e., G_1 is empty). Now, the noise model one is interested in simply becomes our generative model $\hat{\mu}$ for biased preference lists.

The smoothness of a function then gives insights into its robustness to such noise generalizing some concepts from the literature: For instance, in a closely related setting, Caragiannis et al. (2022) study the robustness of approval-based multiwinner scoring functions and introduce the notion of "accurate in the limit," which roughly speaking means that if the number of voters is high enough, then the underlying best committee maximizes the score function with respect to the noisy votes with high probability. This can be captured by our smoothness definition. If for some noise model, the scoring function F is (α, β, γ)-smooth with $\beta = 1$, then our results imply that F is accurate in the limit (Theorem 3.2). Moreover, unlike existing work, Theorem 3.2 also gives a bound on how many noisy votes are required to achieve a $(1-\varepsilon)$-"approximately" optimal committee (Caragiannis et al., 2022).

Remark C.5 (**Comparing α with other notations**). Further, our work is related to a line of works on predicting the outcome of an election by sampling some of the voters (Bhattacharyya & Dey, 2021; Dey et al., 2022). These works bound the number of voters required to accurately predict the outcome of the election. Such sampling bounds are often parameterized by the margin of victory (Magrino et al., 2011; Xia, 2012), i.e., the lead of the election winner in the full election. Conceptually, the margin of victory is related to α in the definition of smoothness (Definition 3.1), as α captures the quality of the weakest candidate of the winning committee and thereby in some sense its "lead" against the remaining ones. Also, note that α has a similar form as the curvature of submodular functions. However, there are some significant differences between them implying that the curvature does not measure the effectiveness of representational constraints, as, e.g., the curvature is unable to distinguish modular functions; see discussions in Remark C.6.

Remark C.6 (**Comparison of α and the curvature of submodular functions**). The curvature of a submodular function often shows up in approximation-ratios of (constrained) monotone submodular maximization algorithms (Conforti & Cornuéjols, 1984; Vondrák, 2010). The curvature $\lambda(f) \in [0,1]$ of $\mathbb{E}_\mu[f(\cdot,\succ)]$ is defined as $1 - \min_{S\subseteq C, c\notin S} \frac{\mathbb{E}_\mu[f_S(c,\succ)]}{\mathbb{E}_\mu[f(c,\succ)]}$.[9] Hence, $\frac{1}{1-\lambda(f)} = \min_{S\subseteq C, c\notin S} \frac{\mathbb{E}_\mu[f_S(c,\succ)]}{\mathbb{E}_\mu[f(c,\succ)]}$. If $\mu = \widehat{\mu}$, then this has a similar form as α in Definition 3.1, but there are a few important differences: the main difference is the denominator in $\frac{1}{1-\lambda(f)}$ depends both on c and f, whereas the denominator in α depends f but not c. Because of this difference, while $\lambda(f)$ measures the "closeness" of $\mathbb{E}_\mu[f(\cdot)]$ to being modular, α is not related to the "closeness" to being modular. Many common multiwinner voting functions (including the SNTV and Borda rules) are modular. Hence, the curvature $\lambda(f)$ of all of these functions (including, both the SNTV and the Borda rule) are the same (equal to 0). However, as our results show, in some cases, representational constraints have significantly higher effectiveness for the Borda rule compared to the SNTV rule (Theorem 3.3). Thus, the curvature is not the right parameter to measure the effectiveness of lower-bound constraints. In contrast to the curvature, α varies across modular functions (e.g., Theorem 3.3).

[9] Note that the curvature of $\mathbb{E}_\mu[f(\cdot,\succ)]$ is well-defined as it is a submodular function.

C.3. Proof of Theorem 3.3: Bounding α For the Utility-Based Model

In this section, we prove Theorem 3.3–which lower bounds α for the utility-based latent and biased generative models.

Recall that in the utility-based model (Definition 2.5), the variable η is drawn from the uniform distribution on $[0, 1]$. We will, in fact, prove a more general version of the above lemma that holds for any distribution of η that satisfies certain properties (Definition C.7). With some abuse of notation, we use η to denote both the distribution and a value drawn from distribution η (independent of all other randomness).

Definition C.7 (**Properties of the utility distribution η from Definition 2.6**). Let η be the distribution on $\mathbb{R}_{\geq 0}$ from Definition 2.6 that parameterizes the generative model μ. Let $\mathrm{cdf}_\eta\colon \mathbb{R} \to [0, 1]$ be the cumulative distribution function of η. We define the following properties of η.

- (Linear scaling) We say that the cumulative distribution function of η scales linearly if there exist constants $\lambda_0, \pi \in (0, 1)$ such that for all $\lambda \in (0, \lambda_0)$ and $\rho > 0$, $\mathrm{cdf}_\eta(\lambda\rho) \in \left[\pi\lambda \cdot \mathrm{cdf}_\eta(\rho), \pi^{-1}\lambda \cdot \mathrm{cdf}_\eta(\rho)\right]$;
- (Exponential tail) We say that η has an exponential tail if there exists a constant $\pi > 0$ such that, for all $t \geq \pi^{-1}$, $\Pr_{X\sim\eta}\left[X \geq t\right] \leq \exp\left(-\pi t\right)$.

Several distributions satisfy the above properties including the uniform distribution on $[0, 1]$, the exponential distribution, and the normal distribution truncated to $\mathbb{R}_{\geq 0}$. Roughly, the linear scaling property ensures that $\mathrm{cdf}_\eta(\lambda\rho)$ behaves as $\lambda \cdot \mathrm{cdf}_\eta(\rho)$ for λ and ρ close to 0. The exponential tail property ensures that with high probability the utilities of all candidates are bounded above. These are relevant as for two candidates $c \in G_1$ and $d \in G_2$ and any user $v \in V$, the probability that $\Pr\left[\widehat{w}_{v,d} > \widehat{w}_{v,c} \mid w_{v,c}\right] = \mathrm{cdf}_\eta\left(\theta \cdot \frac{w_{v,d}}{\omega_c}\right)$. The exponential tail ensures that $\frac{w_{v,d}}{\omega_c}$ is bounded and, hence, we can invoke the linear scaling property. The linear scaling property, itself, ensures that the probability that a disadvantaged candidate $d \in G_2$ is placed before an advantaged candidate $c \in G_1$ scales with θ.

For each $j \in [m]$, let $i_{\succ,j}$ be the j-th candidate in $\succ$. For each $j \in [m]$, define $\sigma_j(f)$ to be the marginal score of $i_{\succ,j}$ with respect to $C \setminus \{i_{\succ,j}\}$, i.e., $\sigma_j(f) \coloneqq f_{C\setminus\{i_{\succ,j}\}}(\{i_{\succ,j}\}, \succ)$ for all $\succ \in \mathcal{L}(C)$.

For any η which has the properties in Definition C.7, we show the following bounds on α.

Lemma C.8 (**Bound on α for the utility-based generative model**). *Let $F = \sum_{v\in V} f(\cdot, \succ_v)$ be a latent multiwinner score function such that $\sigma_j > 0$ and $\sigma_{j+1} = \sigma_{j+2} = \cdots = 0$. Let $(\mu, \widehat{\mu})$ be generative models in Definitions 2.5 and 2.6 with any parameters $\omega \in \mathbb{R}^C_{\geq 0}$ and $\theta \in (0, 1]$. Let μ be such that $M \cap G_2 \neq \emptyset$. There exists a $\theta_0 > 0$ (which is a function of η and ω) such that for all $\theta \in (0, \theta_0)$ the following bounds holds for α.*

$$\Theta\left(\theta^{|G_1|-j} \times \alpha_0 \times \frac{\sigma_j}{\tau}\right) \leq \alpha \leq \Theta\left(\theta^{|G_1|-j}\right).$$

Where $\alpha_0 \coloneqq (\tau(f))^{-1} \cdot \min_{c\in M} \mathbb{E}_\mu\left[f_{C\setminus\{c\}}(c, \succ)\right]$. If $\ell = m - 1$, then the upper bound improves to 1. If $M \cap G_2 = \emptyset$, the lower bound improves to α_0 (the best possible).

Recall the closer α is to 1 the "easier" it is to distinguish candidates in M and in $C \setminus M$. The above result shows that for any multiwinner score function F, α is characterized by the value b, which is the smallest position such that $f_{C\setminus c}(c, \succ) = 0$ for any c with $\mathrm{pos}_\succ(c) > b$. Substituting the values of b for functions F from Section B.1 gives us the following bounds on α for any $\theta \in (0, \theta_0)$.

- If F is the ℓ_1 CC rule, then $\Theta\left(\alpha_0 \cdot \theta^{|G_1|-1}\right) \leq \alpha \leq \Theta\left(\theta^{|G_1|-1}\right)$;
- If F is the b-Bloc rule, then $\Theta\left(\alpha_0 \cdot \theta^{|G_1|-b}\right) \leq \alpha \leq \Theta\left(\theta^{|G_1|-b}\right)$; and
- If F is the Borda rule, then $\Theta\left(\alpha_0 \cdot \theta\right) \leq \alpha \leq \Theta(1)$.

Substituting the lower bounds on α into Theorem 3.2, we recover Theorem 3.3. It remains to prove Lemma C.8.

Order entries of ω, to get the following values $\omega_{(1)} \geq \omega_{(2)} \geq \cdots \geq \omega_{(m)}$. Define $\rho_0 \coloneqq \frac{\omega_{(1)}}{\omega_{(m)}}$. Let π, λ_0 be the constants corresponding to ρ_0 in the linear scaling property (Definition C.7). Define

$$\theta_0 \coloneqq \min\left\{\rho_0, \lambda_0, \pi\right\}.$$

Suppose $\theta \in (0, \theta_0)$. Fix any candidate $c \in G_2 \cap M$ and any voter $v \in V$. To prove Lemma C.8, we need to prove that

$$\Theta\left(\theta^{m-j} \times \alpha_0 \times \frac{\sigma_j}{\tau}\right) \quad \leq \quad \frac{1}{\tau} \cdot \mathbb{E}_{\widehat{\mu}}\left[f_{C\setminus\{c\}}(c, \succ)\right] \quad \leq \quad \Theta\left(\theta^{|G_1|-j}\right). \tag{60}$$

We can express $\mathbb{E}_{\widehat{\mu}}\left[f_{C\setminus\{c\}}(c, \succ)\right]$ as follows

$$\begin{aligned}\mathbb{E}_{\widehat{\mu}}\left[f_{C\setminus\{c\}}(c, \not\succ)\right] &= \sum_{\ell=1}^{m} \Pr_{\mu}\left[\mathrm{pos}_{\not\succ}(c) = \ell\right] \sigma_\ell \\ &= \sum_{\ell=1}^{m} \Pr_{\widehat{\mu}}\left[\mathrm{pos}_{\not\succ}(c) \leq \ell\right] (\sigma_\ell - \sigma_{\ell+1}). \end{aligned} \tag{61}$$

We will bound $\Pr_\mu\left[\mathrm{pos}_{\not\succ}(i) = \ell\right]$. Let $\mathrm{cdf}_\eta(\cdot)$ be the cumulative distribution function of distribution η. For each $d \in C$, let η_d be a draw from distribution η such that $w_{v,d} = \omega_d \cdot \eta_d$. We can upper bound $\Pr_\mu\left[\mathrm{pos}_{\not\succ}(i) = \ell\right]$ as follows

$$\Pr_{\widehat{\mu}}\left[\mathrm{pos}_{\not\succ}(c) \leq \ell\right] = \mathbb{E}_{\eta_c}\left[\sum_{S\subseteq C\setminus\{c\}\colon\, |S|\leq \ell} \prod_{d\in S}\left(1 - \mathrm{cdf}_\eta\left(\frac{\omega_c \cdot \theta \cdot \eta_c}{w_d \cdot \theta^{\mathbb{I}[d\in G_2]}}\right)\right) \prod_{d\notin S} \mathrm{cdf}_\eta\left(\frac{\omega_c \cdot \theta \cdot \eta_c}{w_d \cdot \theta^{\mathbb{I}[d\in G_2]}}\right)\right]$$
$$\text{(Using the fact for all } i, j \in C,\ i \text{ appears before } j \text{ in } \succ \text{ if and only if } \omega_i \cdot \theta^{\mathbb{I}[i\in G_2]} \cdot \eta_i > \omega_j \cdot \theta^{\mathbb{I}[j\in G_2]} \cdot \eta_j) \tag{62}$$

We can separate the second product as follows

$$\prod_{d\notin S} \mathrm{cdf}_\eta\left(\frac{\omega_c \cdot \theta \cdot \eta_c}{w_d \cdot \theta^{\mathbb{I}[d\in G_2]}}\right) = \prod_{d\in G_1\setminus S} \mathrm{cdf}_\eta\left(\frac{\omega_c \cdot \theta \cdot \eta_c}{w_d}\right) \cdot \prod_{d\in G_2\setminus S} \mathrm{cdf}_\eta\left(\frac{\omega_c \cdot \eta_c}{w_d}\right) \tag{63}$$

From the linear scaling property of η (Definition C.7) and the facts that $\theta \leq \theta_0 \leq \lambda_0$ and $\frac{\omega_c \cdot \eta_c}{w_d} > 0$, we have the following bounds for each $d \in G_1$

$$\pi\theta \cdot \mathrm{cdf}_\eta\left(\frac{\omega_c \cdot \eta_c}{w_d}\right) \leq \mathrm{cdf}_\eta\left(\frac{\omega_c \cdot \theta \cdot \eta_c}{w_d}\right) \leq \frac{\theta}{\pi} \cdot \mathrm{cdf}_\eta\left(\frac{\omega_c \cdot \eta_c}{w_d}\right) \leq \frac{\theta}{\pi}. \qquad \text{(Using that } \mathrm{cdf}_\eta(x) \leq 1 \text{ for all } x \in \mathbb{R})$$

Substituting these bounds in Equation (63), we get

$$\begin{aligned}\prod_{d\notin S} \mathrm{cdf}_\eta\left(\frac{\omega_c \cdot \theta \cdot \eta_c}{w_d \cdot \theta^{\mathbb{I}[d\in G_2]}}\right) &\leq \left(\frac{\theta}{\pi}\right)^{|G_1\setminus S|} \cdot \prod_{d\in G_2\setminus S} \mathrm{cdf}_\eta\left(\frac{\omega_c \cdot \eta_c}{w_d}\right) \\ &\leq \left(\frac{\theta}{\pi}\right)^{|G_1\setminus S|}, \qquad \text{(Using that } \mathrm{cdf}_\eta(x) \leq 1 \text{ for all } x \in \mathbb{R}) \\ \prod_{d\notin S} \mathrm{cdf}_\eta\left(\frac{\omega_c \cdot \theta \cdot \eta_c}{w_d \cdot \theta^{\mathbb{I}[d\in G_2]}}\right) &\geq (\pi\theta)^{|G_1\setminus S|} \cdot \prod_{d\notin S} \mathrm{cdf}_\eta\left(\frac{\omega_c \cdot \theta \cdot \eta_c}{w_d \cdot \theta^{\mathbb{I}[d\in G_2]}}\right)\end{aligned}$$

Substituting these in Equation (62), implies the following bounds

$$\begin{aligned}
\Pr_{\widehat{\mu}}\left[\mathrm{pos}_{\not\succ}(c)\leq \ell\right] &\leq \mathbb{E}_{\eta_c}\left[\sum_{S\subseteq C\setminus\{c\}\colon\, |S|\leq \ell}\ \prod_{d\in S}\left(1-\mathrm{cdf}_\eta\left(\frac{\omega_c\cdot\theta\cdot\eta_c}{w_d\cdot\theta^{\mathbb{I}[d\in G_2]}}\right)\right)\cdot\left(\frac{\theta}{\pi}\right)^{|G_1\setminus S|}\right]\\
&\leq \mathbb{E}_{\eta_c}\left[\left(\frac{\theta}{\pi}\right)^{|G_1|-\ell}\right] \qquad \text{(Using that } |S|\leq \ell,\ \theta\leq\theta_0\leq\pi, \text{ and } \mathrm{cdf}_\eta(x)\leq 1 \text{ for all } x\in\mathbb{R})\\
&\leq \left(\frac{\theta}{\pi}\right)^{|G_1|-\ell},\\
\Pr_{\widehat{\mu}}\left[\mathrm{pos}_{\not\succ}(c)\leq \ell\right] &\geq \mathbb{E}_{\eta_c}\left[\sum_{S\subseteq C\setminus\{c\}\colon\, |S|\leq \ell}(\pi\theta)^{|G_1\setminus S|}\cdot\prod_{d\in S}\left(1-\mathrm{cdf}_\eta\left(\frac{\omega_c\cdot\theta\cdot\eta_c}{w_d\cdot\theta^{\mathbb{I}[d\in G_2]}}\right)\right)\cdot\prod_{d\notin S}\mathrm{cdf}_\eta\left(\frac{\omega_c\cdot\theta\cdot\eta_c}{w_d\cdot\theta^{\mathbb{I}[d\in G_2]}}\right)\right]\\
&\qquad \text{(Using that } |S|\leq \ell,\ \theta\pi\leq\theta_0\leq 1, \text{ and } \mathrm{cdf}_\eta(x)\leq 1 \text{ for all } x\in\mathbb{R})\\
&\geq (\pi\theta)^{|G_1|-\ell}\cdot\Pr_{\mu}\left[\mathrm{pos}_{\succ}(c)\leq \ell\right].
\end{aligned}$$

Substituting this in Equation (61), we get

$$\begin{aligned}
\mathbb{E}_{\widehat{\mu}}\left[f_{C\setminus\{c\}}(c,\not\succ)\right] &= \sum_{\ell=1}^{m}\Pr_{\widehat{\mu}}\left[\mathrm{pos}_{\not\succ}(c)\leq \ell\right](\sigma_\ell-\sigma_{\ell+1})\\
&\leq \sum_{\ell=1}^{m}\left(\frac{\theta}{\pi}\right)^{|G_1|-\ell}(\sigma_\ell-\sigma_{\ell+1})\\
&= \sum_{\ell=1}^{j}\left(\frac{\theta}{\pi}\right)^{|G_1|-\ell}(\sigma_\ell-\sigma_{\ell+1}) \qquad \text{(Using that } \sigma_{j+1}=\sigma_{j+2}=\cdots=\sigma_m=0)\\
&\leq \tau\cdot\sum_{\ell=1}^{j}\left(\frac{\theta}{\pi}\right)^{|G_1|-\ell} \qquad \text{(Using that } \sigma_\ell\leq\tau \text{ for all } \ell\in[m])\\
&\leq \tau\cdot\left(\frac{\theta}{\pi}\right)^{|G_1|-j}. \qquad \text{(Using that } 0\leq\tfrac{\theta}{\pi}\leq 1)
\end{aligned}$$

We also have the following lower bound

$$\begin{aligned}
\mathbb{E}_{\widehat{\mu}}\left[f_{C\setminus\{c\}}(c,\not\succ)\right] &= \sum_{\ell=1}^{m}\Pr_{\widehat{\mu}}\left[\mathrm{pos}_{\not\succ}(c)\leq \ell\right](\sigma_\ell-\sigma_{\ell+1})\\
&\geq \sum_{\ell=1}^{m}(\pi\theta)^{|G_1|-\ell}\cdot\Pr_{\mu}\left[\mathrm{pos}_{\succ}(c)\leq \ell\right]\cdot(\sigma_\ell-\sigma_{\ell+1})\\
&= \sum_{\ell=1}^{j}(\pi\theta)^{|G_1|-\ell}\cdot\Pr_{\mu}\left[\mathrm{pos}_{\succ}(c)\leq \ell\right]\cdot(\sigma_\ell-\sigma_{\ell+1}) \qquad \text{(Using that } \sigma_{j+1}=\sigma_{j+2}=\cdots=\sigma_m=0)\\
&\geq (\pi\theta)^{|G_1|-j}\cdot\Pr_{\mu}\left[\mathrm{pos}_{\succ}(c)\leq j\right]\cdot(\sigma_j-\sigma_{j+1})\\
&\qquad \text{(Using that } \sigma_\ell\geq\sigma_{\ell+1} \text{ and } \Pr_\mu\left[\mathrm{pos}_{\succ}(c)\leq \ell\right]\geq 0 \text{ for all } \ell\in[m])\\
&\geq (\pi\theta)^{|G_1|-j}\cdot\Pr_{\mu}\left[\mathrm{pos}_{\succ}(c)\leq j\right]\cdot\sigma_j \qquad \text{(Using that } \sigma_{j+1}=0.)\\
&\geq (\pi\theta)^{|G_1|-j}\cdot\frac{\alpha_0}{\tau}.\\
&\qquad \text{(Using that } \alpha_0=\textstyle\sum_{\ell=1}^{m}\Pr_\mu\left[\mathrm{pos}_{\succ}(c)\leq \ell\right]\cdot(\sigma_\ell-\sigma_{\ell+1})\leq \Pr_\mu\left[\mathrm{pos}_{\succ}(c)\leq j\right]\cdot\tau)
\end{aligned}$$

This completes the proof of Equation (60) and, hence, Lemma C.8.

C.4. Bounding α For the Swapping-Based Model

In this section, we bound the parameter α for any multiwinner scoring function when the biased generative model is the swapping-based model (Definition B.12) and the latent generative model is μ is any generative model that satisfies the following condition (for some parameter $\rho > 0$)

$$\min_{c \in C} \mathbb{E}_{\mu}\left[f_{C\setminus\{c\}}(c, \succ)\right] \geq \rho. \tag{64}$$

Concretely, we prove the following bound on α.

Lemma C.9 (**Bounds on α for the swapping-based model**). *Let $F = \sum_{v\in V} f(\cdot, \succ_v)$ be a latent multiwinner score function. Suppose μ be a generative model such that F is order-preserving with respect to μ (Definition 2.8) and that satisfies Equation (64) for some $\rho > 0$. For any numbers $\phi \in (0, t^{-1})$ and the generative model $\widehat{\mu}$ in Definition B.12 with parameters μ and ϕ, α satisfies the following bound*

$$\alpha \in \alpha_0 \cdot \left(1 \pm \mathbb{E}_{\mu}\left[\frac{1}{Z(\succ)}\right] \cdot \frac{O(t\phi)}{1-\phi} \cdot \frac{\tau}{\rho}\right)$$

Where $\alpha_0 := (\tau(f))^{-1} \cdot \min_{c\in M} \mathbb{E}_{\mu}\left[f_{C\setminus\{c\}}(c, \not\succ)\right]$ and $Z(\succ)$ is the normalizing constant corresponding to preference list $\succ$, as defined in Definition B.12.

Similar to Lemma B.13, the above lemma also does not fix a specific generative model of latent preference lists μ. The parameter α depends on μ via the terms α_0 and $\mathbb{E}_{\mu}\left[\frac{1}{Z(\succ)}\right]$. As discussed in Section 2.2.3, in general, we expect $\mathbb{E}_{\mu}\left[\frac{1}{Z(\succ)}\right]$ to be of the order of $\Omega\left(m\phi\right)$. If this holds, then Lemma C.9, implies that $\alpha \in \alpha_0\left(1 \pm O\left(m^{-1}t\right)\right)$. The proof of Lemma C.9 appears in Section C.4.

Proof of Lemma C.9. Lemma C.9 follows as a corollary of Lemma B.15, which was proved in Section B.3. Consider any $c \in M$, selecting the set S in Lemma B.15 as $C \setminus \{c\}$, we get that

$$\mathbb{E}_{\mu}\left[f_{C\setminus\{c\}}(c, \succ)\right] \cdot (1-\lambda) \leq \mathbb{E}_{\widehat{\mu}}\left[f_{C\setminus\{c\}}(c, \not\succ)\right] \leq \mathbb{E}_{\mu}\left[f_{C\setminus\{c\}}(c, \succ)\right] \cdot (1+\lambda).$$

where $\lambda := \mathbb{E}_{\mu}\left[\frac{1}{Z(\succ)}\right] \cdot O\left(\frac{t\phi}{1-\phi} \cdot \frac{\tau}{\rho}\right)$. The lemma follows by taking the minimum over all $c \in M$. $\square$

D. Impossibility Results for Problem 2

In this section, we provide a necessary condition for Problem 2. Roughly speaking, we show that if n is not sufficiently large, representational constraints cannot recover an (approximate) optimal solution. In combination with our algorithmic result, we observe that under the utility-based model, the number of voters necessary for SNTV or ℓ_1-CC to find a close-to-optimal committee can be much larger than for Borda (Remark D.4).

Again, we let $F = \sum_{v\in V} f(\cdot, \succ_v)$ be a latent score function of multiwinner voting, and let $(\mu, \widehat{\mu})$ be a generative model defined in Definitions 2.7 and 2.8. We propose the following definition.

Definition D.1 (**Contribution of $G_2 \cap S^\star$**). Let $0 \leq \ell \leq k$ be an integer. Define $r_\ell := \mathbb{E}_{\mu}\left[\max_{S\subseteq C\setminus(G_2\cap S^\star):|S|=k,|S\cap G_2|=\ell} \frac{F(S)}{\mathrm{OPT}}\right]$ to be the expected maximum ratio between the scores of an optimal committee S in $C \setminus (G_2 \cap S^\star)$ with $|S \cap G_2| = \ell$ and $S^\star$.

$1 - r_\ell$ measures the contribution of candidates in $G_2 \cap S^\star$ to F for all subsets S with $|S \cap G_2| = \ell$. If r_ℓ is close to 1, we can safely ignore candidates in $G_2 \cap S^\star$ and can still find a near-optimal solution from the remaining candidates, even with constraint $\mathcal{K}(\ell)$. Otherwise, if r_ℓ is not close to 1, we may need the candidates $G_2 \cap S^\star$ for a high-score committee. Then if all candidates in G_2 are indistinguishable under $\widehat{\mu}$, it is likely that $\widehat{S}_\ell$ loses a lot.

Remark D.2 (**Discussion on the scale of r_ℓ**). Usually, r_ℓ becomes smaller as the difference $|\ell - |G_2 \cap S^\star||$ becomes larger. For instance, for SNTV under Definition 2.5 with μ being the uniform distribution on interval $[0, 1]$, suppose $\omega_c = 1$ for all $c \in S^\star$ and $\omega_c = 0.5$ for all $c \in C \setminus S^\star$. We can verify that $\mathbb{E}_{\mu}[F(c)] \geq 2\mathbb{E}_{\mu}[F(c')]$ for any $c \in S^\star$ and $c' \in C \setminus S^\star$, which implies that $r_\ell \leq 1 - \frac{|\ell - |G_2\cap S^\star||}{2k}$ when n, m are sufficiently large. This observation matches the intuition that a representational constraint $\mathcal{K}(\ell)$ with $\ell \approx |G_2 \cap S^\star|$ may be better for debiasing, also observed in (Celis et al., 2020; Mehrotra et al., 2022).

Another consequence of this example is that we find $F(\widehat{S}) \ll \mathrm{OPT}$ is possible. If the bias parameter $\theta < 0.5$, we know that $\widehat{S}$ is likely to contain only candidates in G_1, which results in $F(\widehat{S}) \le r_0 \cdot \mathrm{OPT} \le (1 - \frac{|G_2 \cap S^\star|}{2k}) \cdot \mathrm{OPT}$, i.e., $\widehat{S}$ is far from optimal.

For each $j \in [m]$, let $i_{\succ,j}$ be the j-th candidate in $\succ$. For each $j \in [m]$, define $\tau_j(f)$ to be the score of $i_{\succ,j}$, i.e., $\tau_j(f) := f(\{i_{\succ,j}\}, \succ)$ for all $\succ \in \mathcal{L}(C)$. Let $\succ \in \mathcal{L}(C)$ be a fixed preference list and define $\tau_{\min}(f) := \min\{\tau_j(f) \colon j \in [m],\ \tau_j(f) > 0\}$ to be the smallest non-zero candidate value of f. Our main impossibility result is summarized as follows.

Theorem D.3 (**Impossibility result for Problem 2**). *Let $F \colon 2^C \to \mathbb{R}_{\ge 0}$ be a score function of multiwinner voting. Let $\mu, \widehat{\mu}$ be generative models of latent/biased preference lists such that every $\succ_v / \not\succ_v$ is i.i.d. drawn from $\mu/\widehat{\mu}$. Let $\zeta := \max_{c \in G_2} \mathbb{E}_{\widehat{\mu}}[f(c, \not\succ)]$. If $\zeta < \tau_{\min}(f)$, $k = o(|G_2|)$, and $n = o(\frac{\tau_{\min}(f)}{m\zeta})$, with probability at least 0.9, $F(\widehat{S}_\ell) \le (r_\ell + o(1)) \cdot \mathrm{OPT}$ holds for every $0 \le \ell \le k$.*

Proof. By the definition of $s(f)$ and ζ, we know that for every $c \in G_2$,

$$\Pr_{\widehat{\mu}}[f(c, \not\succ) > 0] \le \frac{\zeta}{s(f)}.$$

Consequently, we have that

$$\Pr_{\widehat{\mu}}[\widehat{F}(c) = 0, \forall c \in G_2] \ge \left(1 - \frac{\zeta}{s(f)}\right)^{mn}.$$

Then since $n = o(= \left(\frac{s(f)}{m\zeta}\right)$, we have

$$\Pr_{\widehat{\mu}}[\widehat{F}(c) = 0, \forall c \in G_2] \ge 1 - o(1).$$

Conditioned on the event that $\widehat{F}(c) = 0$ holds for all $c \in G_2$, $\widehat{S}_\ell$ exactly selects ℓ candidates from G_2 since G_2 does not contribute to $\widehat{F}$. Moreover, we can not distinguish candidates in G_2 and can only randomly select ℓ candidates from G_2 in $\widehat{S}_\ell$, which results in $o(1)$ probability to select any candidate from $G_2 \cap S^\star$ since $|G_2 \cap S^\star| \le k = o(|G_2|)$. Overall, with probability at least $1 - o(1)$, we have that

$$F(\widehat{S}_\ell) \le \max_{S \subseteq C \setminus (G_2 \cap S^\star) : |S| = k, |S \cap G_2| = \ell} F(S).$$

Thus, we conclude the theorem by the definition of r_ℓ.

□

Observe that the above theorem considers a more general generative model $(\mu, \widehat{\mu})$ that does not require order preserving properties in Definitions 2.7 and 2.8. Roughly, Theorem D.3 indicates that the required number of voters is $n = \Omega(\frac{\tau_{\min}(f)}{m\zeta})$ which is non-trivial when $\frac{\tau_{\min}(f)}{\zeta} \gg m$. Note that $\tau_{\min}(f) \le \tau_1(f)$ always holds but usually has at most a $\mathrm{poly}(m)$ gap. For instance, $\tau_{\min}(f) = \tau_1(f)$ for SNTV and Bloc, and $\tau_{\min}(f) = \frac{\tau_1(f)}{m-1}$ for Borda and ℓ_1-CC. Also, note that ζ is comparable to $\min_{c \in M} \mathbb{E}_{\widehat{\mu}}\left[f_{C \setminus \{c\}}(c, \not\succ)\right]$, specifically, $\zeta \le \min_{c \in M} \mathbb{E}_{\widehat{\mu}}\left[f_{C \setminus \{c\}}(c, \not\succ)\right]$ for SNTV and ℓ_1-CC. Thus, the scale of $\frac{\tau_{\min}(f)}{\zeta}$ and $\frac{1}{\alpha}$ in Definition 3.1 could be the same order, e.g., $\frac{\tau_{\min}(f)}{\zeta} \approx \Omega(\frac{1}{\alpha \cdot \mathrm{poly}(m)})$, which leads to a required number of voters $n = \Omega(\frac{1}{\alpha \cdot \mathrm{poly}(m)})$. Combining with the bound on α in Section 3.3, this required voter number is interesting; see the following comparison.

Remark D.4 (**Comparison of robustness between different rules**). Specifically, under the utility-based generative model, we have that $n = \Omega(\frac{\theta^{-m+1}}{\mathrm{poly}(m)})$ for SNTV and ℓ_1-CC and $n = \Omega(\frac{\theta^{-1}}{\mathrm{poly}(m)})$ for Borda to achieve a near-optimal solution $\widehat{S}_\ell$ with a representational constraint, say $F(\widehat{S}_\ell) \approx \mathrm{OPT}$. Combining with Theorem 3.3, we know that the required voting number of Problem 2 for SNTV and ℓ_1-CC is at least $\Omega(\theta^{-m+1})$, which is much larger than the sufficient voting number $O(\theta^{-2} \cdot \mathrm{poly}(m))$ for Borda. Thus, we may conclude that Borda is "more robust" than SNTV and ℓ_1-CC under the utility-based model.

E. Summary of Bounds on Smoothness Parameters

In this section, we summarize all bounds on smoothness parameters proved in this work.

First, we summarize results for the utility-based model (see Definition 2.6), which also apply to variants where η is drawn from an exponential or normal distribution (instead of the uniform distribution) on $[0,1]$.

Multiwinner scoring function	α (Lemma C.8)	β (Lemma B.7)	γ (Lemma B.7)
SNTV	$\Theta\left(\theta^{-2(m-1)}\right)$	$1-\Theta(m^{-1/2})$	$1-\Theta(m^{-3/2})$
ℓ_1-CC	$\Theta\left(\theta^{-2(m-1)}\right)$	$1-\Theta(m^{-1/2})$	$1-\Theta(m^{-3/2})$
Borda	$\Theta\left(\theta^{-2}\right)$	$1-\Theta(m^{-1/2})$	$1-\Theta(m^{-5/2})$

Table 2. Smoothness parameters for the utility-based model Definition 2.6. The formal statements of the results appear as Lemmas B.7 and C.8. Note that these results hold for the utility-based model in Definition 2.6 as well as its variants where η is not uniformly distributed on $[0,1]$ but is instead drawn from, say, an exponential distribution.

Next, we summarize the results for the swapping-based model (Definition B.12).

Multiwinner scoring function	α (Lemma C.9)	β (Lemma B.13)	γ (Lemma B.13)
SNTV	$1-\Theta(\phi t)$	$1-\Theta(\phi t)$	$1-\Theta(\phi t)$
ℓ_1-CC	$1-\Theta(\phi t)$	$1-\Theta(\phi t)$	$1-\Theta(\phi t)$
Borda	$1-\Theta(\phi t)$	$1-\Theta(\phi t)$	$1-\Theta(\phi t)$

Table 3. Smoothness parameters for the swapping-based model (Definition B.12). The formal statements of the results appear as Lemmas B.13 and C.9.

F. Case Study: Utility-Based Generative Model of Latent and Biased Preferences

In this section, we present a self-contained discussion of our main result (Theorem 3.2) within the utility-based model of preferences (Definitions 2.5 and 2.6). In the utility-based model, each candidate c has an *intrinsic* utility $\omega_c \geq 0$. The true or *latent* utility of candidate c for voter v is

$$w_{v,c} = \eta_{v,c} \cdot \omega_c, \tag{65}$$

where $\eta_{v,c}$ is a random variable drawn uniformly at random from $[0,1]$ and independent of $\eta_{v',c'}$ for any $v' \neq v$ and $c' \neq c$. These latent utilities, in turn, define the latent preference list of voters: for each voter v, $\succ_v$ is the list of candidates c arranged in decreasing order of $w_{v,c}$. The voters, however, do not observe their latent utilities for the candidates. Instead, they observe a (possibly) biased version of these latent utilities. These observed utilities are modeled by a bias parameter $\theta \in [0,1]$: for each candidate c and voter v, the observed utility of c for v is

$$\widehat{w}_{v,c} := \begin{cases} w_{v,c} & \text{if } c \in G_1, \\ \theta \cdot w_{v,c} & \text{if } c \in G_2. \end{cases} \tag{66}$$

Greedy is optimal without bias. Recall that our goal is to study the latent quality of solutions produced by algorithms that satisfy representational constraints in the presence of bias. For simplicity, suppose that the exact expected observed, biased marginal contributions $\mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ)\right]$ are known for each $c \in C$ and $S \subseteq C$ (we relax this assumption later in this section). Consider the greedy algorithm that, in each iteration $t \in [k]$, selects the candidate c that maximizes the expected marginal contribution to the set S_t selected so far (i.e., $\arg\max_{c \in C \setminus S_t} \mathbb{E}_{\widehat{\mu}}\left[f_{S_t}(c, \succ)\right]$). Without bias (i.e., $\theta = 1$), it can be shown that this algorithm outputs the set S_k that has the optimal latent utility among sets of size k (i.e., $\mathbb{E}_{\mu}\left[f(S_k, \succ)\right] = \max_{S \subseteq C\colon |S|=k} \mathbb{E}_{\mu}\left[f(S, \succ)\right]$). Intuitively, this is because of the following invariant: for any $c, c' \in C$

$$\text{if } \omega_c > \omega_{c'}, \quad \text{then,} \quad \text{for any } S \subseteq C \setminus \{c, c'\}, \quad \mathbb{E}_{\mu}\left[f_S(c, \succ)\right] > \mathbb{E}_{\mu}\left[f_S(c', \succ)\right]. \tag{67}$$

This invariant itself holds because of the facts that: (1) if $\omega_c > \omega_{c'}$ then with probability strictly larger than $\frac{1}{2}$, c appears before c' in $\succ_v$ (for any v) and (2) f is domination sensitive (Definition 2.2).

Constrained-greedy is optimal with bias. However, when there is bias (i.e., $\theta < 1$), the above invariant does not hold and the greedy algorithm may achieve a low latent utility. Nevertheless, because the latent utilities of all candidates in the same group are uniformly reduced (Equation (66)), a group-wise version of the invariant holds: for any two candidates c, c' in the *same* group (G_1 or G_2) for any $c, c' \in C$

$$\text{if } \omega_c > \omega_{c'}, \quad \text{then,} \quad \text{for any } S \subseteq C \setminus \{c, c'\}, \quad \mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ)\right] > \mathbb{E}_{\widehat{\mu}}\left[f_S(c', \not\succ)\right]. \tag{68}$$

This property enables us to design an algorithm that, given k and the expected biased marginal contributions utilities of all candidates (i.e., $\mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ)\right]$ for all $c \in C$ and $C \subseteq S$) outputs a size-k subset satisfying representational constraints that has optimal latent utility. Let $S^\star$ be any size-k subset maximizing the expected latent utility. Let $\ell = |S^\star \cap G_2|$. It can be shown that the algorithm that first greedily selects ℓ candidates from G_2 and, then, greedily selects $k - \ell$ candidates from G_1 achieves optimal expected latent utility (Algorithm 1). This is because (1) for any S, the top t candidates in G_1 (respectively G_2) by expected observed scores $\mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ)\right]$ is the same as the top t candidates in G_1 (respectively G_2) by expected latent scores $\mathbb{E}_{\mu}\left[f_S(c, \succ)\right]$ and (2) $S^\star$ consists of the top $k - \ell$ (respectively ℓ) candidates from G_1 (respectively G_2) by expected latent scores. Where the first fact is implied by Equations (67) and (68) and the second fact holds because of Equation (67).

Thus, if one has access to the expected marginal contributions of the candidates $\{\mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ)\right] : c \in C \setminus S\}$ (for all S) and $\ell = |S^\star \cap G_2|$, then one can execute the aforementioned algorithm, which outputs a size-k subset satisfying the representational constraints specified by ℓ and maximizing the expected latent utility. As mentioned before in Section 3, ℓ can be estimated under natural assumptions on ω. Meanwhile, it remains to discuss how one has access to accurate *expected* observed marginal contributions.

Effect of n on the effectiveness of representational constraints. In Problem 2, the input is n samples of observed preference lists $\{\not\succ_v\}_v$ generated as described above. For any $S \subseteq C$ and $c \in C \setminus S$, a natural estimate of $\mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ)\right]$ is the sample mean $\frac{1}{n}\sum_{v \in V} f_S(c, \not\succ_v)$. One can prove an additive concentration bound for the resulting estimate (e.g., see Lemma C.3 for a similar result). For any $\delta > 0$, it holds that:

$$\Pr\left[\forall_{c \in C}, \ \forall_{S \subseteq C \setminus \{c\} : \, |S| = k-1} \left| \frac{1}{n}\sum_{v \in V} f_S(c, \not\succ_v) - \mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ)\right] \right| \geq \tau\sqrt{nm \log \frac{m}{\delta}}\right] \leq \delta. \tag{69}$$

Note that, for the above bound to be meaningful, the expected observed utilities $\mathbb{E}_{\widehat{\mu}}\left[f_S(c, \not\succ)\right]$ must be large compared to the error term $\tau\sqrt{nm \log \frac{m}{\delta}}$. We parameterize the scale of expected observed utilities using a parameter α°:[10]

$$\alpha^\circ(\omega, \theta, f) = \frac{1}{\tau_1(f)} \min_{c \in C} \mathbb{E}_{\widehat{\mu}}\left[f_{C \setminus \{c\}}(c, \not\succ)\right],$$

where $\tau_1(f)$ is a normalizing constant, defined as the maximum possible expected score $\mathbb{E}_{\mu}\left[f(c, \succ)\right]$ of a candidate. Instead of taking the minimum over each candidate $c \in C$, one can show that it suffices to take a minimum over only certain candidates. We do so in Definition 3.1 and Theorem 3.2.

For the utility-based model, a lower bound on α° is sufficient to ensure that, when n is large enough, representational constraints achieve near-optimal latent utility.

Theorem F.1 (Informal version of Theorem 3.2 specialized to then utility-based model). *Let $F \colon 2^C \to \mathbb{R}_{\geq 0}$ be a multiwinner score function and $\mu, \widehat{\mu}$ be generative models corresponding to the utility-based model specified by ω and θ. For any $0 < \varepsilon, \delta < 1$, if*

$$n \geq \frac{\mathrm{poly}(m) \cdot \log(1/\delta)}{\mathrm{poly}\left(\varepsilon \cdot \alpha^\circ(\omega, \theta, f)\right)},$$

there is an algorithm that given $\ell = |S^\star \cap G_2|$ and observed preferences $\{\not\succ_v\}_{v \in V}$, outputs a size-$k$ subset $S \in \mathcal{K}(\ell)$ such that

$$\Pr_{\mu, \widehat{\mu}}\left[F(S) \geq (1 - \varepsilon) \cdot \mathrm{OPT}\right] \geq 1 - \delta.$$

Generalizations. To derive Theorem F.1, we used two main properties: (1) in the absence of bias, there is a greedy algorithm that achieves the optimal latent utility, and (2) that the order of candidates in any one group (G_1 or G_2) by their expected

[10] We use the superscript in α° to differentiate it from α in Definition 3.1.

latent utilities is the same as their order by expected observed utilities (Equations (67) and (68)). Neither of these conditions may be true beyond the utility-based model.

Order-preservation with respect to μ ensures that the greedy algorithm is optimal. In more general models, the first condition in order-preservation with respect to μ (Definition 2.7) ensures that the greedy algorithm achieves the optimal latent utility in the absence of bias. The second condition in order-preservation with respect to μ (Definition 2.7) enables us to show that, even the constrained version of the greedy algorithm (which first selects $k-\ell$ candidates from G_1 and then selects ℓ candidates from G_2) achieves the optimal latent utility.

Order-preservation between μ and $\widehat{\mu}$ bounds distance between orderings of candidates by observed and latent preferences. Beyond the utility-based model, we may not be able to guarantee that the order of candidates in each group (G_1 and G_2) by their expected latent utilities is the same as their order by expected observed utilities. Indeed, this is not true in the swapping-based model, where there is constant β such that if two candidates c, c' in the same group have latent scores $\mathbb{E}_\mu\left[f_S(c,\succ)\right]$ and $\mathbb{E}_\mu\left[f_S(c',\succ)\right]$ within a multiplicative factor β (i.e., $\beta\mathbb{E}_\mu\left[f_S(c',\succ)\right] \leq \mathbb{E}_\mu\left[f_S(c,\succ)\right] \leq \frac{1}{\beta}\mathbb{E}_\mu\left[f_S(c',\succ)\right]$), then it is possible that (see Example B.14)

$$\mathbb{E}_\mu\left[f_S(c,\succ)\right] > \mathbb{E}_\mu\left[f_S(c',\succ)\right], \quad \text{but}, \quad \mathbb{E}_{\widehat{\mu}}\left[f_S(c,\not\succ)\right] < \mathbb{E}_{\widehat{\mu}}\left[f_S(c',\not\succ)\right].$$

Order-preservation between μ and $\widehat{\mu}$ (Definition 2.8) bounds the "distance" between the orderings of candidates in each group (G_1 and G_2) by their expected latent utilities and their order by expected observed utilities: it requires the relative order of two candidates c, c' in the same group to be the same if their latent utilities are at least a factor of β apart, i.e., if either $\mathbb{E}_\mu\left[f_S(c',\succ)\right] \leq \beta \cdot \mathbb{E}_\mu\left[f_S(c,\succ)\right]$ or $\beta \cdot \mathbb{E}_\mu\left[f_S(c,\succ)\right] \geq \mathbb{E}_\mu\left[f_S(c',\succ)\right]$.

G. Tool to Study Smoothness and Effectiveness of Representational Constraints With Novel Bias Models

In this section, we illustrate how one could use the code as a tool for preliminary studies of the effectiveness of representational constraints with respect to novel bias models and multiwinner score functions, for which theoretical bounds may not be readily available.

The code is available at the following link: `https://github.com/AnayMehrotra/Selection-with-Multiple-Rankings-with-Bias`

G.1. Implementation Details

The code takes as input oracles that (1) evaluate a multiwinner score function F and (2) sample from generative models $(\mu, \widehat{\mu})$. These oracles are specified by (python) functions. The code also takes m, n, and k as input.

First, for the specified m and k, it outputs estimates $(\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\gamma}) \in [0,1]$ along with corresponding confidence intervals, which follow from a concentration inequality (Lemma C.3). This allows for theoretical estimates of the capabilities of representational constraints using our main result Theorem 3.2. (Note that α, β, γ are independent of n and ℓ). Concretely, we numerically estimate (α, β, γ) as follows: since the empirical averages are concentrated around the expectations $\mathbb{E}_\mu[f_S(c,\succ)]$ and $\mathbb{E}_{\hat{\mu}}[f_S(c,\not\succ)]$, we use this to compute α from its definition (Definition 3.1). As β and γ, given any $0 \leq \beta \leq 1$, we compute $0 \leq \gamma \leq 1$ that satisfies the condition in Definition 2.8 via binary search.

Second, given values of n, m, and k, the code estimates the fraction of the optimal score recovered by representational constraints for the given F with respect to the given $(\mu, \widehat{\mu})$.

In the next section, we illustrate the code using a set of latent generative models provided by (Szufa et al., 2020) and the corresponding swapping-based biased generative model (Definition B.12).

G.2. Illustration of the Code: Models of Latent Preferences by (Szufa et al., 2020) and the Swapping-Based Bias Model

In this section, to illustrate use cases of our code, we analyze the effectiveness of the representational constraints with (1) the family of generative models proposed by (Szufa et al., 2020) and (2) using the swapping-based bias model (Definition B.12).

Note that for any pair of generative models $(\mu, \widehat{\mu})$ one can study the effectiveness of representational constraints with a multiwinner score function F using our theoretical results, by computing the smoothness parameters for Definition 3.1.

Concretely, for the swapping-based model, one can use the bounds in Lemmas B.13 and C.9. The usefulness of the code is in doing preliminary analysis on novel bias models, for which theoretical bounds may not be available.

G.2.1. SETUP

We vary the generative model μ across a subset of generative models provided by (Szufa et al., 2020); namely, we consider the Single-Peaked model by Conitzer, the Mallows model, the Polya-Eggenberger Urn model, and the Impartial culture model. We fix $\widehat{\mu}$ to be the generative model corresponding to the swapping-based bias model (Definition B.12). We fix the number of voters to be $m = 50$, the size of the committee to be $k = 10$, and vary the number of voters $n \in \{25, 50, 100\}$. We fix the groups G_1 and G_2 to have equal size (i.e., $|G_1| = |G_2|$). We select the partitioning into groups uniformly at random. The fraction of possible executed swaps is specified by a parameter $\lambda \in [0, 1]$.

In each simulation, we fix a generative model μ, a value of $n \in \{25, 50, 100\}$, and a voting rule F from SNTV and Borda. For fixed μ, n, and F, we vary λ over $[0, 1]$. Given a μ, n, F, and λ, we draw latent preferences $\{\succ_v \mid v \in V\}$ i.i.d. from μ. For each $v \in V$, we compute the maximum number of swaps $t_{\max}(v)$ that can be performed on the preference list $\succ_v$ before all candidates in G_1 are placed before all candidates in G_2. Concretely, $t_{\max}(v)$ is the Kendall-Tau distance between preference lists $\succ_v$ and $\succ_v^\star$, where $\succ_v^\star$ is the unique preference list that (1) ranks all candidates in G_1 before any candidate in G_2 and (2) satisfies for any c, c' in the same group (G_1 or G_2) $c \succ_v c'$ if and only if $c \succ_v^\star c'$.

In the swapping-based model, we arbitrarily fix $\phi = 0.5$ for illustration. $\widehat{\mu}$ is the generative model defined by the swapping-based biased model specified by $\phi = 0.5$ and which, for each $v \in V$, performs t swaps, where

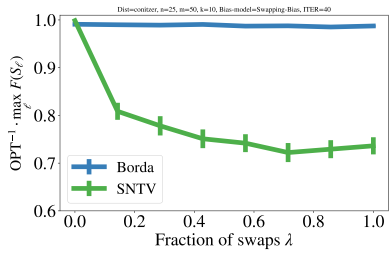
$$t = \lambda \cdot t_{\max}(v).$$

Recall that ϕ controls the average difference in the positions of swapped candidates. When ϕ is close to 0, with high probability, all swapped candidates are "neighbours" in the preference lists. At the other extreme, when ϕ is close to 1, candidates who are "far" in the preference lists are also swapped.

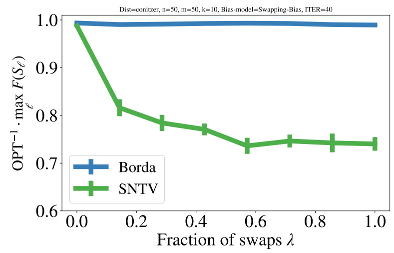
G.2.2. OBSERVATIONS

The results appear in Figure 1. In We observe that, across all choices of μ, n, and fraction of swaps λ: representational constraints recover a higher fraction of the optimal score with the Borda rule compared to the SNTV rule: for Borda, the fraction recovered is > 0.99 across all simulations, whereas, for SNTV, it can be as low as 0.75 Further, for the SNTV rule, the fraction of the optimal score recovered by representational constraints increases with n (for a given λ). For the Borda rule, since the fraction of the optimal score recovered by representational constraints is already larger than 0.99 differences across n are small.

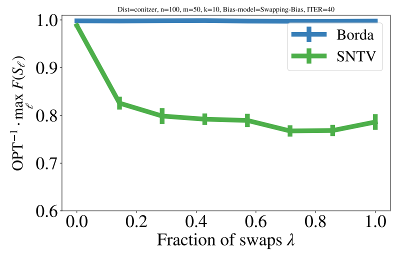
These observations align with our theoretical results: From Theorem 3.3, we expect representational constraints to have a higher effectiveness with the Borda rule compared to the SNTV rule. Similarly, from Theorems 3.2 and D.3, we expect the effectiveness of the representational constraints to increase with n.



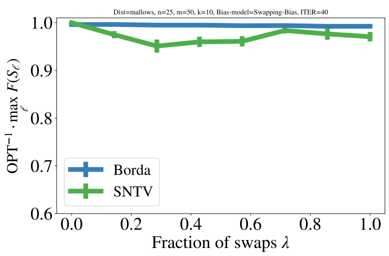
(a) $n = 25$, $m = 50$ and μ is the Single-Peaked generative model by Conitzer



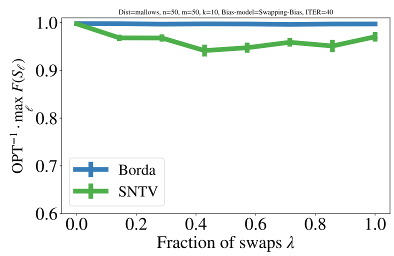
(b) $n = 50$, $m = 50$ and μ is the Single-Peaked generative model by Conitzer



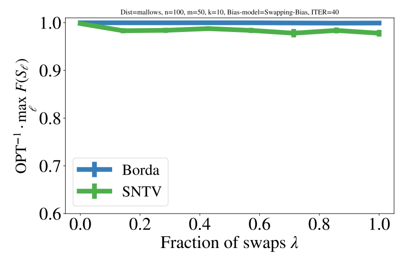
(c) $n = 100$, $m = 50$ and μ is the Single-Peaked generative model by Conitzer



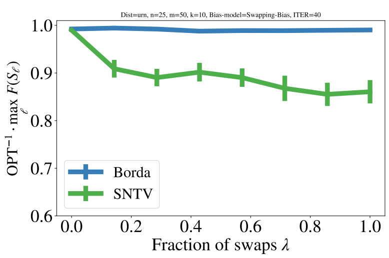
(d) $n = 25$, $m = 50$ and μ is the Mallows generative model



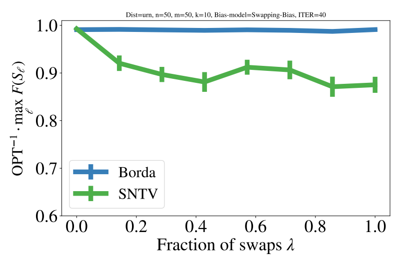
(e) $n = 50$, $m = 50$ and μ is the Mallows generative model



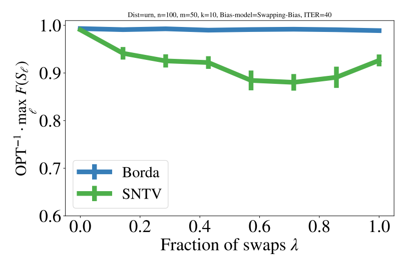
(f) $n = 100$, $m = 50$ and μ is the Mallows generative model



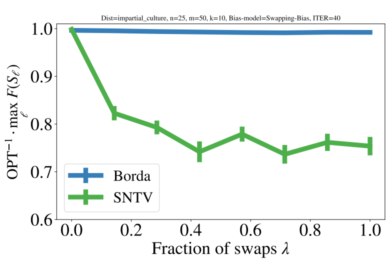
(g) $n = 25$, $m = 50$ and μ is the Polya-Eggenberger Urn generative model



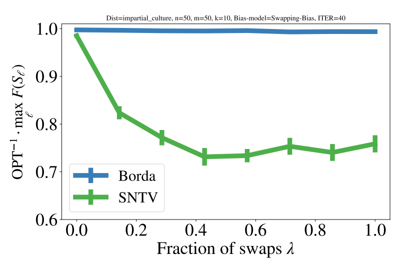
(h) $n = 50$, $m = 50$ and μ is the Polya-Eggenberger Urn generative model



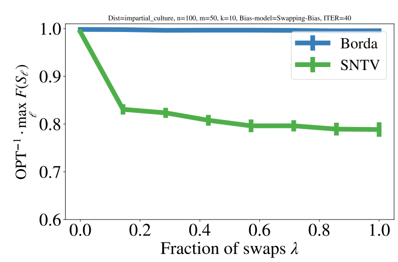
(i) $n = 100$, $m = 50$ and μ is the Polya-Eggenberger Urn generative model



(j) $n = 25$, $m = 50$ and μ is the Mallows generative model



(k) $n = 50$, $m = 50$ and μ is the Impartial culture generative model



(l) $n = 100$, $m = 50$ and μ is the Impartial culture generative model

Figure 1. Simulations results with different families of generative models μ: The plots show the fraction of the score recovered by representational constraints with different preference aggregation functions F and generative models μ, under the swapping-based bias model with $\phi = 0.5$ (Definition B.12). In all of these simulations, the number of candidates is $m = 50$ and the size of the output committee is $k = 10$. The number of candidates n and the generative model μ vary and are specified with the sub-figures. The y-axis shows the fraction of the optimal score recovered by representational constraints. The x-axis shows the number of swaps t allowed in the swapping-based model.