Robustness and Feasibility of Voting in Linear Social Choice

Alina Chadwick, Anson Kahng, and Euan Seow

University of Rochester, Rochester, NY, USA
achadwic@cs.rochester.edu, anson.kahng@rochester.edu,
eseow2@u.rochester.edu

Abstract. We study robustness and structural properties of rank aggregation rules, also called social welfare functions (SWFs) or voting rules, in the recently-codified *linear social choice* setting. We prove that a wide range of voting methods, including positional scoring rules, approval voting, pairwise-majority consistent (PMC) rules, and instant-runoff voting (IRV), are not robust to errors in a linear social choice setting. Our empirical results support our theoretical findings, but also demonstrate that PMC rules and the Borda rule are fairly robust to errors in synthetic data. We also explore feasibility and attainability in linear social choice; we find that, for many non-dictatorial voting rules, attainability does not imply feasibility.

Keywords: Computational Social Choice · Voting · Robustness

1 Introduction

Linear social choice (Ge et al., 2024a) provides a framework for collective decision-making in settings where alternatives are represented by feature vectors, an almost-ubiquitous occurrence in many modern-day applications. Each alternative a_j is represented by a feature vector, x_j, each agent i has a weight vector β_i, and the utility of agent i for alternative a_j is $\beta_i \cdot x_j$. Unlike classical social choice, in which preferences are assumed to be arbitrary given rankings, linear social choice provides an explicit structure that underlies agent preferences.

The linear social choice framework has two common application domains, both motivated by modern machine learning settings where alternatives have rich feature representations. The first is Reinforcement Learning from Human Feedback (RLHF), a prominent paradigm for aligning large language models (LLMs) with human preferences (Ouyang et al., 2022; Stiennon et al., 2020; Ziegler et al., 2019; Lambert, 2026). In RLHF, human annotators provide pairwise comparisons between model outputs, and these comparisons are aggregated into a reward signal, often via the Bradley-Terry-Luce (BTL) model (Bradley and Terry, 1952; Ouyang et al., 2022; Christiano et al., 2017). Linear social choice offers a natural interpretation of this process: Instead of assuming that annotators are noisily sampled from a single ground-truth reward function, annotators

instead may have heterogeneous weight vectors that capture fundamental disagreement about which features of outputs are desirable. Importantly, Siththaranjan et al. (2024) establish that the BTL model corresponds to the Borda rule in the linear social choice setting, directly connecting the dominant RLHF aggregation method to a canonical voting rule. However, Ge et al. (2024a) observe that rankings produced by voting rules in the linear social choice setting may fail to correspond to any underlying weight vector β. In other words, some rankings are *attainable* through voting but not *feasible* under the linear social choice model itself.

The second motivating domain is *virtual democracy* (Kahng et al., 2019; Lee et al., 2019), a paradigm for algorithmic decision-making in which explicit models of voter preferences are elicited and used to collectively evaluate alternatives represented as feature vectors. Virtual democracy is particularly relevant to our work because it aggregates individual predicted votes, which therefore means that it treats individual voter weight vectors (i.e., the β_i's) as explicit objects that must be learned. Because no learning method can perfectly learn individual β_i vectors, robustness to noise is a central concern in virtual democracy (Kahng et al., 2019). This perspective is also applicable to RLHF, where estimates of latent preference vectors may be inferred from noisy pairwise comparison data. More broadly, recent work has pushed for the adoption of voting-based aggregation principles in AI alignment domains (Conitzer et al., 2024), suggesting that linear social choice offers a general framework for studying collective preference aggregation in large machine learning systems.

In this paper, we study structural properties of linear social choice itself. In particular, we focus on settings in which latent voter weight vectors are learned, and the ordinal rankings they induce are then aggregated. Our first question is one of robustness: *How sensitive are voting outcomes to noise in the underlying preference vectors*? In RLHF, since the BTL model corresponds to Borda in the linear social choice setting, robustness guarantees or impossibility results directly translate into implications for standard RLHF aggregation pipelines. Beyond robustness, we also study a complementary structural question motivated by Ge et al. (2024a): *For various canonical voting rules, are all attainable rankings feasible*? Together, our results shed light on both the reliability and the expressive limitations of voting rules in linear social choice.

1.1 Related Work

Several recent papers argue that social choice provides a principled framework for aggregating diverse human feedback in AI alignment in the context of RLHF (Conitzer et al., 2024) and apply it accordingly (Ge et al., 2024a; Park et al., 2024; Kim et al., 2026; Dai and Fleisig, 2024; Mishra, 2023; Siththaranjan et al., 2024; Ge et al., 2024b; Chakraborty et al., 2024; Swamy et al., 2024; Zhong et al., 2024; Xiao et al., 2025). A few works further highlight that heterogeneous voter preferences call for principled aggregation methods beyond a single reward model (Park et al., 2024; Chakraborty et al., 2024); our work contributes to

this agenda by studying the robustness and structural properties of voting rules specifically within the linear social choice model.

We also assume a level of noise arising from modeling preferences of individuals rather than explicitly calculating the preferences of voters at each point in time. Our approach of modeling voter preferences specifically employs the framework of virtual democracy (Noothigattu et al., 2018; Kahng et al., 2019; Lee et al., 2019). A complementary model of per-voter weight uncertainty appears in the robust ordinal regression literature (Greco et al., 2008), which infers a set of compatible weight vectors from elicited pairwise comparisons rather than assuming additive noise.

Numerous works study the robustness of voting rules to noise (Procaccia et al., 2007; Shiryaev et al., 2013; Shockey, 2008; Salehi-Abari and Larson, 2021), with measures of robustness including outputting the same winner or ranking after introducing noise. A related line of work studies when noisy pairwise preferences reveal an underlying ground-truth ranking, dating back to Condorcet's original noisy-voting model (Young, 1988) and including algorithmic and sample-complexity results for recovering or approximating this ranking (Coppersmith et al., 2006; Conitzer et al., 2006; Caragiannis et al., 2016). However, few of these existing works (noise robustness or ground truth recovery) study robustness to noise in a linear model. Ge et al. (2024a) consider axiomatic properties of voting rules in linear social choice particularly in the RLHF setting. They also introduce the tension between feasibility and attainability, which we explore further in this paper. Specifically, Ge et al. (2024a) find that there exists a construction in the linear social choice setting that violates feasibility under PMC rules.

Among our findings is an observation that aggregating preferences using utilities from the linear social choice model directly is more robust to errors than using voting rules in a theoretical worst-case. This result is related to the concept of distortion (Procaccia and Rosenschein, 2006; Anshelevich et al., 2021), which can be defined as a loss in efficiency attributable to lacking perfect information about agents' preferences. Recent work (Ge et al., 2026) also studies distortion specifically within the linear social choice model, showing how ordinal mechanisms can be optimized under structured utility assumptions. Our results complement this idea by demonstrating that ordinal aggregation can be non-robust to noise in learned utilities, whereas direct cardinal aggregation avoids this concern.

1.2 Our Contributions

Our contributions are threefold. First, we demonstrate that, under a general noise model in the linear social choice setting, standard voting rules like positional scoring rules, pairwise-majority consistent (PMC) rules, instant-runoff voting, and approval voting are not robust; in fact, both the Borda count and the class of PMC rules can result in a noisy ranking that is exactly the reverse of the true ranking. Second, we explore structural elements of the linear social choice setting, such as the relationship between attainable and feasible rankings, including conditions under which attainability does not imply feasibility

for Borda, plurality, general positional scoring rules, IRV, and approval voting. Third, we empirically demonstrate that, on a wide range of synthetic data, voting rules often only marginally underperform direct utility-based approaches, suggesting that worst-case instances are often far from realistic.

2 Model and Notation

We begin by restating the linear social choice model introduced by Ge et al. (2024a) in our context. Throughout, let $[x] := \{1, \ldots, x\}$.

Consider a set of m distinct prompts/responses referred to as *alternatives* $A := \{a_1, \ldots, a_m\}$, each of which is objectively evaluated according to a set of k *features*, $F := \{f_1, \ldots, f_k\}$. Alternatively, we may think of each alternative a_i as a point in $\mathbb{R}^k$. We will often present the m alternatives, along with their feature values, in a *feature matrix* $M \in \mathbb{R}^{m \times k}$ depicted in Figure 1a, where the (i, j) entry in M, $\alpha_{ij} \in [0, 1]$, corresponds to the prominence of feature f_j in alternative a_i, with 1 being the highest ("f_j is very prominent in a_i") and 0 being the lowest ("f_j is not prominent in a_i"). In the RLHF setting, feature values may correspond to measurable properties of model outputs, such as helpfulness or safety scores. Slightly overloading notation, let $\alpha_j := (\alpha_{j1}, \ldots, \alpha_{jk})$ represent the feature vector for alternative a_j for all $j \in [m]$. Throughout the paper, we will present matrices in table format to be explicit about the alternative (row) and feature (column) labels. Furthermore, we have n *voters* $N := [n]$. Each voter i is associated with a *(feature) weight vector* $\beta_i := (\beta_{i1}, \ldots, \beta_{ik}) \in \mathbb{R}^k$ in which, for all $j \in [k]$, β_{ij} denotes voter i's weight for feature f_j. These weight vectors induce linear utility functions for each voter in the sense that we assume voter i's (true) utility for alternative a_j is exactly the dot product of i's weight vector and alternative a_j's feature vector, or $u_i(a_j) = \beta_i \cdot \alpha_j$. We will often be interested in voter-specific *rankings* obtained by ordering alternatives by utility.

Noise Model. We draw inspiration from the framework of virtual democracy to learn participants' weight vectors. However, no machine learning procedure, no matter how advanced, is error-free, and, for each voter $i \in [n]$, the system only has access to a noisy estimate of β_i, which we term $\hat{\beta}_i := (\hat{\beta}_{i1}, \ldots, \hat{\beta}_{ik})$. We assume that each $\hat{\beta}_{ij} = \beta_{ij} + \epsilon_{ij}$, where each ϵ_{ij} is independently drawn from some symmetric and unbiased noise distribution X that has finite quantiles, i.e., for all $p \in (0, 1)$, the CDF $F(x)$ reaches or exceeds p at a finite value of x. For our purposes, it is enough to guarantee that $F(x)$ is continuous from the left and either (1) X is bounded between $[-T_1, T_2]$ for finite $T_1, T_2 \in \mathbb{R}^+$ or (2) X is sub-Gaussian (i.e., has strong tail decay).

Given noisy estimates of voter weights, $\hat{\beta}_i$, we let $\hat{u}_i(a_j) := \hat{\beta}_i \cdot \alpha_j$ denote the estimated utility that voter i has for alternative a_j. Additionally, let $\hat{\sigma}_i$ denote the noisy ranking for voter i obtained by ordering alternatives in terms of decreasing $\hat{u}_i(a_j)$, and let σ_i^* denote the (unobservable) true ranking for voter i obtained by ordering alternatives in terms of decreasing $u_i(a_j)$. We will let $\sigma^* = (\sigma_1^*, \ldots, \sigma_n^*)$ and $\hat{\sigma} = (\hat{\sigma}_1, \ldots, \hat{\sigma}_n)$ represent the true and noisy profiles.

Voting Rules. We now define the four (families of) voting rules we consider throughout the paper, all of which are common in the computational social choice literature (Brandt et al., 2016). Let $L(A)$ denote the set of linear orders (rankings) over A. Positional scoring rules, PMC rules, and instant-runoff voting are social welfare functions $f : L(A)^n \to L(A)$ that receive a preference profile as input and return a ranking as output. Finally, approval voting is a function $f : \{0,1\}^{n\times m} \to L(A)$ that takes subsets of approved alternatives and returns a ranking.

Definition 1 (Positional scoring rules (PSRs)). *A positional scoring rule takes as input a score vector $s = (s_1, s_2, \ldots, s_m)$, where $s_i \geq s_{i+1}$ for all $i \in [m-1]$ and $s_1 > s_m$. Each alternative x receives a score from voters, where each voter who ranks alternative x in position p gives s_p points to x. Alternatives are then ordered in terms of decreasing total score, with ties broken arbitrarily.*

Definition 2 (Borda count). *The Borda count is the PSR with score vector $s = (m-1, m-2, \ldots, 0)$.*

Definition 3 (Plurality). *Plurality voting is the PSR with score vector $s = (1, 0, \ldots, 0)$.*

We refer to the setting where plurality only outputs a winner as *winner-only plurality* and the setting where plurality outputs a ranking over all alternatives corresponding to their plurality scores as *plurality in the full-ranking setting.*

Definition 4 (Pairwise-majority consistent (PMC) rules). *A pairwise-majority consistent (PMC) rule satisfies the following (weak) property, which can be understood as a generalization of Condorcet consistency: If there exists a ranking such that a majority of all voters agrees with every pairwise comparison in the ranking, then that ranking must be returned.*

A PMC ranking may not always exist; however, when a PMC ranking does exist, it is unique.

Definition 5 (Approval voting). *In approval voting, each voter designates a subset of alternatives to be "approved." Alternatives are ordered by decreasing numbers of approvals, and ties are broken uniformly at random.*

We also consider a canonical iterative voting rule, instant-runoff voting (IRV).

Definition 6 (Instant-runoff voting). *Instant-runoff voting proceeds in rounds as follows. In each of at most $m-1$ rounds, the alternative with the fewest first-place votes is eliminated*[1] *and all voters who had the eliminated candidate in first place move on to their next-most preferred candidate. If a voter runs out of candidates to support, their vote is removed from the election. The order of elimination naturally determines the final ranking.*

[1] Ties are broken in each round.

Finally, we also consider dictatorial voting rules as we examine the tension between feasibility and attainability.

Definition 7 (Dictatorial Voting Rules). *A voting rule is dictatorial if there exists some voter $i \in N$ (the dictator) such that for all profiles, the voting rule outputs i's ranking.*

Throughout, we also use the following definition of the *Kendall tau distance* between two rankings of equal length. As a note, the maximum Kendall tau distance between two rankings of length m is $\binom{m}{2}$, which occurs when the two rankings are exact reversals of each other.

Definition 8 (Kendall tau distance). *The Kendall tau distance, denoted d_{KT}, between two rankings is the number of pairs of alternatives on which the two rankings disagree. Equivalently, it is the "bubble sort" distance between the rankings. Formally, $d_{KT}(\sigma, \sigma') = |\{(x, y) \in A \times A | x \succ_\sigma y \text{ and } y \succ_{\sigma'} x\}|$, where $x \succ_\sigma y$ means that x is ranked ahead of y in σ.*

Feasibility and Attainability.

In linear social choice, we must differentiate between so-called *feasible* and *attainable* rankings (Ge et al., 2024a). On a high level, feasible rankings are possible rankings induced by a single β, and attainable rankings (under a social welfare function f) are possible rankings after running f on a profile of feasible rankings. We are not aware of prior work that systematically studies the relationship between feasibility and attainability in the linear social choice model, although Ge et al. (2024a) present a result for PMC rules.

Definition 9 (Feasible rankings). *Given feature matrix M, an aggregate ranking $\sigma^* = a_1 \succ \cdots \succ a_m$ is feasible if and only if there exists a β^* such that $\beta^* \cdot \alpha_{a_1} \geq \beta^* \cdot \alpha_{a_2} \geq \cdots \geq \beta^* \cdot \alpha_{a_m}$.*

Definition 10 (Attainable rankings). *Given feature matrix M, ranking σ is attainable under social welfare function f if and only if there exists some $\beta = (\beta_1, ..., \beta_n)$ such that*

$$f(r(\beta_1, M), r(\beta_2, M), ..., r(\beta_n, M)) = \sigma,$$

where $r(\beta_i, M)$ is the ranking induced by β_i on M.

3 Theoretical Results

We begin by presenting our main results about the robustness of various voting rules in linear social choice. Our most surprising result is the non-robustness of the Borda rule to errors in our setting. Because BTL corresponds to Borda in the linear social choice setting, this non-robustness result applies directly to RLHF aggregation. This result diverges from the findings of Kahng et al. (2019), which state that, in the general case, the Borda count is robust to prediction errors.

3.1 Non-Robustness of Voting Rules

Definition 11 (Robustness). *Fix the noise model of Section 2. For a voting rule f, a feature matrix M, and a weight profile $\beta = (\beta_1, \ldots, \beta_n)$ inducing true profile $\sigma^* = \sigma(M, \beta)$ and noisy realization $\hat{\sigma}$ (via $\hat{\beta}$), f is:*

- ***robust*** *on (M, β) if $f(\hat{\sigma}) = f(\sigma^*)$ with high probability, and*
- ***non-robust*** *if there exists some (M, β) for which $f(\hat{\sigma}) \neq f(\sigma^*)$ with high probability.*

Unlike in the findings of Kahng et al. (2019), the Borda count and, additionally, PMC rules are not robust to errors in our model. In fact, we show a strong form of non-robustness: There exist cases in which the noisy ranking is, with high probability,[2] a complete reversal of the true ranking, i.e., the Kendall tau distance (d_{KT}) between the two rankings is the maximum value of $\binom{m}{2}$.

	f_1	f_2	$\ldots$	f_k
a_1	α_{11}	α_{12}	$\ldots$	α_{1k}
a_2	α_{21}	α_{22}	$\ldots$	α_{2k}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
a_m	α_{m1}	α_{m2}	$\ldots$	α_{mk}

(a) An example of a feature matrix.

	f_1	f_2
a_1	ϵ	$1-\epsilon$
a_2	2ϵ	$1-2\epsilon$
$\vdots$	$\vdots$	$\vdots$
a_m	$m\epsilon$	$1-m\epsilon$

(b) Matrix for Theorem 1.

Fig. 1: Feature matrices for robustness counterexamples.

Theorem 1. *Under our noise model, for all $m \geq 2$ and $k \geq 2$, there exists a feature matrix M and weight profile β inducing true profile σ such that, with high probability, (1) $d_{KT}(Borda(\sigma^*), Borda(\hat{\sigma})) = \binom{m}{2}$ and (2) $d_{KT}(PMC(\sigma^*), PMC(\hat{\sigma})) = \binom{m}{2}$; i.e., both Borda and PMC rules result in a noisy ranking that is exactly the opposite of the true ranking.*

Proof (Proof sketch). Consider the feature matrix in Figure 1b. Suppose that there are two groups, group A consisting of a $1/2 + \delta$ fraction of voters with $\beta_A = (1/2+\epsilon, 1/2-\epsilon)$ and true ranking $\sigma_A = a_m \succ a_{m-1} \succ \ldots \succ a_1$, and group B consisting of a $1/2-\delta$ fraction of voters with $\beta_B = (0, L)$ for L sufficiently large relative to the noise scale and true ranking $\sigma_B = a_1 \succ a_2 \succ \ldots \succ a_m$. In this case, it is easy to verify that $\text{PMC}(\sigma^*) = \text{Borda}(\sigma^*) = a_m \succ a_{m-1} \succ \ldots \succ a_1$.

Due to the structure of the feature matrix, every voter's ranking is fully determined by the sign of a single scalar (the difference between their two weight coordinates), so the only possible noisy rankings are σ_A, σ_B, or a tie (which occurs

[2] Throughout, "with high probability" (whp) means with probability $1 - \frac{1}{poly(n)}$, where n is the number of voters.

with vanishing probability). When applying noise to this profile, a constant fraction of group A voters will switch their ranking to σ_B ; by choosing L large relative to the noise scale ($L > T_1 + T_2$ for bounded noise, or $L = \Theta(\sqrt{\log n})$ for sub-Gaussian noise), the probability that a group B voter switches to σ_A can be made zero or negligible, respectively. With high probability, strictly more than half the voters have noisy ranking σ_B, so $\text{PMC}(\hat{\sigma}) = \text{Borda}(\hat{\sigma}) = a_1 \succ a_2 \succ \cdots \succ a_m$, a complete reversal of the true ranking.

Remark 1. Leximax Copeland subject to PO (LCPO), a voting rule introduced by Ge et al. (2024a), is not robust to errors in a linear social choice setting.

A weaker version of the above result can be extended beyond Borda to all positional scoring rules (PSRs).

Corollary 1. *Under our noise model, for all $m \geq 2$ and $k \geq 2$, there exists a feature matrix M and associated profile σ such that, with high probability, PSR(σ^*) $\neq$ PSR($\hat{\sigma}$) for any positional scoring rule.*

Additionally, the counterexample from Theorem 1 also applies to some iterative voting rules, such as instant-runoff voting (IRV).

Observation 2 *Under our noise model, for all $m \geq 2$ and $k \geq 2$, there exists a feature matrix M and associated profile σ such that, with high probability, IRV(σ^*) $\neq$ IRV($\hat{\sigma}$).*

We also show that approval voting suffers from the same non-robustness, although here we use more features in the family of constructions.

Theorem 3. *Under our noise model, for values of $m \geq 2$, $k = m$, there exists a feature matrix M and corresponding approval profile σ such that, with high probability, $d_{KT}(\text{Approval}(\hat{\sigma}), \text{Approval}(\sigma)) = \binom{m}{2}$, i.e., approval voting results in a noisy ranking that is exactly the opposite of the true ranking.*

3.2 Robustness of Noisy Utility Estimates

Finally, although the positional scoring rules, PMC rules, instant-runoff voting, and approval voting are not robust to errors in our model, we show that direct usage of estimated utilities is robust to errors in our model. This finding suggests that moving from cardinal to ordinal information in the linear social choice model is a point of fragility; summing utilities lets noise across voters average out, whereas the same noise can flip a single pairwise comparison once utilities are discretized into a ranking.

Theorem 4. *Let Sum($\hat{\sigma}$) denote the ranking of alternatives in decreasing order of noisy utility, i.e., in decreasing order of $\hat{u}(a_j) := \sum_{i \in [n]} \hat{u}_i(a_j)$, where $\hat{u}_i(a_j) := \hat{\beta}_i \cdot \alpha_j$. Then, for any ranking σ^* with a minimum distance of $\upsilon > 0$ between any consecutive pair of alternatives, Sum($\hat{\sigma}$) = Sum(σ^*) with high probability.*

In addition to our first question about the robustness of voting rules in the linear social choice setting, we also further explore the relationship between feasibility and attainability, complementing the findings of Ge et al. (2024a).

3.3 Feasibility and Attainability

Inspired by Ge et al. (2024a), we are interested in further investigating the implications between attainability and feasibility in the realm of linear social choice. One direction is immediate: Any feasible ranking is attainable by any voting rule that satisfies unanimity.[3] However, for each voting rule, are all attainable rankings feasible? We answer this question in the negative, meaning that it is possible to apply a voting rule to a set of linear social choice preferences and obtain an infeasible ranking. In other words, for all non-dictatorial voting rules we consider[4], feasibility implies attainability, but attainability does not imply feasibility. First, however, we consider dictatorial voting rules (such as a randomized dictatorship):

Theorem 5. *For any dictatorial voting rule, attainability implies feasibility.*

Proof. By definition, a voting rule is dictatorial if there is a specific voter (say i^*) whose ranking is always the outcome; voter i^*'s ranking is induced by β_{i^*}.

A similar proof sketch applies to plurality in the winner-only setting:

Theorem 6. *For winner-only plurality, attainability implies feasibility.*

Proof. For an alternative a^* to be a plurality winner, it must appear as a winner in some voters' ballots. Therefore, there must be a β that results in a^* winning the single-winner election. This implies feasibility.

We now move to our negative results, starting with positional scoring rules.

Theorem 7. *For any $m \geq 4$, there exists some construction over m alternatives that violates feasibility under positional scoring rules, implying that for positional scoring rules, attainability does not imply feasibility.*

Our next theorem restates a result from Ge et al. (2024a) in our setting.

Theorem 8. *For any $m \geq 6$, there exists some construction over m alternatives that violates feasibility under PMC rules, implying that for PMC rules, attainability does not imply feasibility.*

We next analyze attainability and feasibility for instant-runoff voting.

[3] A voting rule f is *unanimous* if, when applied to a profile $\boldsymbol{\sigma}$ in which all voters have the same preference σ, $f(\boldsymbol{\sigma}) = \sigma$.

[4] When we consider non-dictatorial voting rules, we examine those which output a complete ranking.

Theorem 9. *For any $m \geq 4$, there exists some construction over m alternatives that violates feasibility under instant-runoff voting, implying that for IRV, attainability does not imply feasibility.*

Finally, we consider feasibility and attainability for approval voting. Because approval votes segment alternatives into two classes, approved and not approved, we must define feasibility (of a ranking) differently.

Definition 12 (Approval-feasibility). *A ranking σ is approval-feasible if it is possible to decompose σ into non-empty $\sigma_1 \succ \sigma_2$ such that there exists a β and threshold t such that all of the alternatives in σ_1 are approved and all of the alternatives in σ_2 are not approved, where "approved" means that they have score at least t under β.*

Note that this is a very weak definition of feasibility, so any infeasibility results are very strong. In particular, it is easy to show infeasibility results for a stronger definition where *every* possible decomposition of σ into non-empty $\sigma_1 \succ \sigma_2$ must be feasible; however, we are able to extend this result significantly to the case where no decomposition of σ into $\sigma_1 \succ \sigma_2$ is feasible.

Theorem 10. *For any $m \geq 5$, there exists some construction over m alternatives that violates approval-feasibility, implying that, for approval voting, attainability does not imply feasibility.*

4 Empirical Results

To complement the theoretical results from the previous section, we also examine how aggregation methods we have studied in this setting — the Borda count, PMC rules (with Copeland, Ranked Pairs, and Schulze as representatives of the greater class of rules), and cardinal utility summation ("Sum") — perform on synthetically generated linear social choice instances.

In particular, we measure the Kendall tau distance between the output of each aggregation method on noisy preferences and the output of each aggregation method on true preferences, or $d_{KT}(f_x(\hat{\sigma}), f_x(\sigma^*))$.

4.1 Methodology

Given n voters, m alternatives, and k features, we randomly generate feature matrices, i.e., the α values of each alternative and feature, β values for the n voters, and ϵ error values. These values are then used to calculate the u and $\hat{u}$ values of the alternatives for each voter. Each aggregation method, f_x, is applied to the values of $\hat{u}$ and $\hat{\beta}$ for each circumstance to derive the output of the aggregation method, $f_x(\hat{\sigma})$. Simultaneously, the value of σ^* (the true ranking) is generated from the values of the true utilities. The Kendall tau distance between these two outputs is then calculated and normalized to a value between 0 and 1 by dividing the total Kendall tau distance by $\binom{m}{2}$, the total number of pairs of alternatives.

Sampling			Kendall tau Distance				
α	β	ϵ	Borda	Cop.	RP	Sch.	Sum
U	HU	U	0.094	0.106	0.106	0.102	**0.048**
U	HU	N	0.126	0.143	0.143	0.142	**0.076**
U	BN	U	0.114	0.117	0.117	0.112	**0.046**
U	BN	N	0.152	0.155	0.155	0.151	**0.074**
U	HN	U	0.120	0.116	0.115	0.106	**0.043**
U	HN	N	0.156	0.160	0.159	0.151	**0.072**
U	BU	U	0.093	0.104	0.104	0.102	**0.047**
U	BU	N	0.125	0.144	0.144	0.142	**0.078**

(a) Uniform α.

Sampling			Kendall tau Distance				
α	β	ϵ	Borda	Cop.	RP	Sch.	Sum
N	HU	U	0.094	0.106	0.105	0.103	**0.046**
N	HU	N	0.126	0.138	0.139	0.138	**0.077**
N	BN	U	0.117	0.116	0.116	0.112	**0.046**
N	BN	N	0.151	0.154	0.154	0.150	**0.074**
N	HN	U	0.118	0.116	0.116	0.105	**0.043**
N	HN	N	0.157	0.159	0.160	0.152	**0.070**
N	BU	U	0.096	0.103	0.103	0.102	**0.048**
N	BU	N	0.127	0.143	0.143	0.141	**0.074**

(b) Normal α.

Fig. 2: Normalized Kendall tau distances for various parameter combinations. Lower is better. U = uniform, N = normal, HU = uniformly drawn across bounds, BN = bimodal normal, HN = normal with same mean, BU = uniform from two bounds.

We run three sets of experiments using these features. In the first set, we examine how varying our sampling methods for values of α, β, and ϵ affect the accuracy of results, measured by normalized Kendall tau distance. We initialize our values to $m = 8, n = 10, k = 5$. The different sampling methods used for each parameter are described below and each sampling method is used to generate 10,000 simulations, where we ultimately calculate the mean normalized Kendall tau distance from all of the simulations. During aggregation, (rare) ties are broken arbitrarily. The α and ϵ values are generated in two ways: sampling from a normal distribution and sampling from a uniform distribution.[5] The values of β are generated in three ways, as we aim to capture "extremes" of voter preferences, namely voters having approximately the same β values (homogeneous sampling), and voters having opposite preferences (bimodal preferences): (1) Homogeneous uniform (HU) sampling: Each value within β was drawn independently from $\mathcal{U}(0, 1)$. (2) Bimodal normal (BN) sampling: The values of β are drawn from two normal distributions, $\mathcal{N}(0.25, 1.0)$ and $\mathcal{N}(0.75, 1.0)$, with half drawn from each distribution. (3) Homogeneous normal (HN) sampling: Each value within β is drawn independently from $\mathcal{N}(0.5, 1.0)$.

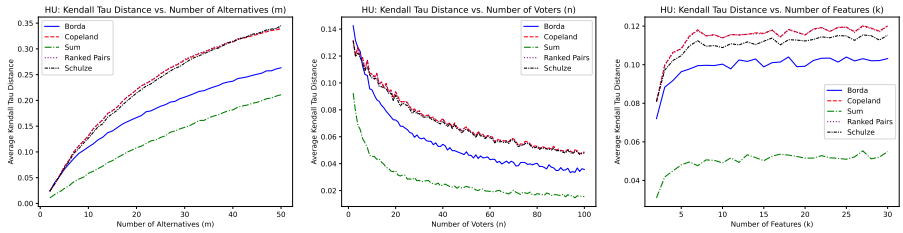
For our second set of experiments, we observe how the normalized Kendall tau distance varies as our parameter values change. We consider $m \in [2, 50]$, $n \in [2, 100]$, and $k \in [2, 30]$, with results computed for every different β sampling method and for every different aggregation method.[6] Values of ϵ and α are drawn uniformly from the distributions $\mathcal{U}(-0.1, 0.1)$ and $\mathcal{U}(0.0, 1.0)$ respectively.

The third set of experiments explores the performance of the different aggregation methods when α values are drawn from correlated distributions. We

[5] For α values, the uniform distribution sampled from is $U(0, 1)$ and the normal distribution is $\mathcal{N}(0.5, 1)$. For ϵ values, the uniform distribution sampled from is $\mathcal{U}(-0.1, 0.1)$ and the normal distribution is $\mathcal{N}(0, 0.1)$.

[6] The parameter values when they are not being varied are $m = 8, n = 10, k = 5$.

include the full description of these experiments as well as the results in Appendix 6.2.



(a) d_{KT} vs. m (b) d_{KT} vs. n (c) d_{KT} vs. k

Fig. 3: KT distance vs. parameter size with HU β generation. Graphs for other β sampling methods have been included in the Supplementary Material.

4.2 Results

The results collected from our first experiment on the Kendall tau distances for Borda, Copeland, Ranked Pairs, Schulze, and Sum are shown in Figure 2.[7] Unsurprisingly, computing rankings by summing utilities consistently outperformed other aggregation methods. PMC rules (i.e., Copeland, Ranked Pairs, and Schulze) and Borda performed comparably in all cases; this result is in line with our theoretical results which show that the probability of swapping alternatives is low when aggregating summed utilities, while Borda and PMC rules are both more prone to swaps.

In our second experiment, illustrated in Figure 3, we further demonstrate that directly summing utility values outperforms Borda and PMC rules in terms of robustness to errors where the numbers of features, alternatives, and voters vary for different β value generation methods described above. As k and m increase, the Kendall tau distance tends to increase, indicating that, as expected, considering more features and alternatives with the same number of voters results in worse accuracy. On the other hand, increasing the number of voters n increases the accuracy of Borda and PMC rules, with Borda outperforming PMC rules for sufficiently large n. Throughout, summing utility values directly is consistently most accurate.

In our last experiment, we show that when α values are correlated, Borda, PMC rules (Copeland, Ranked Pairs, and Schulze), and Sum all exhibit few pairwise swaps, but Sum continues to perform the best. Full results are presented in the Supplementary Material.

[7] In this section, we omit results from approval voting, as it performed far worse than all other aggregation methods. Full results for approval voting are presented in the Supplementary Material.

5 Discussion

The central takeaways of this paper are twofold: (1) unlike in the general case (Kahng et al., 2019), the Borda count is not robust to errors in a linear social choice setting, and (2) especially in linear social choice, it is vital to use all the (cardinal) information available when making accurate decisions. Encouragingly, voting rules appear to perform better in practice than their worst-case properties suggest, and our simulation results suggest that standard voting rules like Borda and Copeland can reasonably approximate decision frameworks that involve cardinal information.

Limitations: In our model, even though our noise assumptions are fairly generous, we do assume that noise is symmetric, unbiased, and either a) bounded or b) sub-Gaussian. While our model applies to many settings due to these assumptions, there are cases of asymmetric noise that our model does not cover. Additionally, our empirical results are based on synthetic data with either normal or uniform noise, which may not be sufficiently representative of real-world stakeholder preferences over features.

Bibliography

Elliot Anshelevich, Aris Filos-Ratsikas, Nisarg Shah, and Alexandros A Voudouris. 2021. Distortion in social choice problems: The first 15 years and beyond. *International Joint Conference on Artificial Intelligence* (2021).

Ralph Allan Bradley and Milton E Terry. 1952. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika* 39, 3/4 (1952).

Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D Procaccia. 2016. *Handbook of computational social choice*. Cambridge University Press.

Ioannis Caragiannis, Ariel D Procaccia, and Nisarg Shah. 2016. When do noisy votes reveal the truth? *ACM Transactions on Economics and Computation (TEAC)* 4, 3 (2016), 1–30.

Souradip Chakraborty, Jiahao Qiu, Hui Yuan, Alec Koppel, Dinesh Manocha, Furong Huang, Amrit Bedi, and Mengdi Wang. 2024. MaxMin-RLHF: Alignment with Diverse Human Preferences. In *Proceedings of the 41st International Conference on Machine Learning*. PMLR, 6116–6135.

Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2017. Deep reinforcement learning from human preferences. *Advances in neural information processing systems* 30 (2017).

Vincent Conitzer, Andrew Davenport, and Jayant Kalagnanam. 2006. Improved bounds for computing Kemeny rankings. In *AAAI*, Vol. 6. 620–626.

Vincent Conitzer, Rachel Freedman, Jobst Heitzig, Wesley H. Holliday, Bob M. Jacobs, Nathan Lambert, Milan Mossé, Eric Pacuit, Stuart Russell, Hailey

Schoelkopf, Emanuel Tewolde, and William S. Zwicker. 2024. Position: Social choice should guide AI alignment in dealing with diverse human feedback. In *Proceedings of the 41st International Conference on Machine Learning* (Vienna, Austria) *(ICML'24)*.

Don Coppersmith, Lisa Fleischer, and Atri Rudra. 2006. Ordering by weighted number of wins gives a good ranking for weighted tournaments. In *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*. 776–782.

Jessica Dai and Eve Fleisig. 2024. Mapping social choice theory to RLHF. *arXiv preprint arXiv:2404.13038* (2024).

Luise Ge, Daniel Halpern, Evi Micha, Ariel D Procaccia, Itai Shapira, Yevgeniy Vorobeychik, and Junlin Wu. 2024a. Axioms for ai alignment from human feedback. *Advances in Neural Information Processing Systems* 37 (2024), 80439–80465.

Luise Ge, Brendan Juba, and Yevgeniy Vorobeychik. 2024b. Learning linear utility functions from pairwise comparison queries. *arXiv preprint arXiv:2405.02612* (2024).

Luise Ge, Gregory Kehne, and Yevgeniy Vorobeychik. 2026. Optimized distortion in linear social choice. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 40. 16930–16937.

Salvatore Greco, Vincent Mousseau, and Roman Słowiński. 2008. Ordinal regression revisited: multiple criteria ranking using a set of additive value functions. *European Journal of Operational Research* 191, 2 (2008), 416–436.

Anson Kahng, Min Kyung Lee, Ritesh Noothigattu, Ariel Procaccia, and Christos-Alexandros Psomas. 2019. Statistical foundations of virtual democracy. In *International Conference on Machine Learning*. PMLR, 3173–3182.

Kihyun Kim, Jiawei Zhang, Asuman E Ozdaglar, and Pablo A Parrilo. 2026. Beyond RLHF and NLHF: Population-Proportional Alignment under an Axiomatic Framework. In *The Fourteenth International Conference on Learning Representations*.

Nathan Lambert. 2026. *Reinforcement Learning from Human Feedback: Alignment and post-training of LLMs*. Manning Publications.

Min Kyung Lee, Daniel Kusbit, Anson Kahng, Ji Tae Kim, Xinran Yuan, Allissa Chan, Daniel See, Ritesh Noothigattu, Siheon Lee, Alexandros Psomas, et al. 2019. WeBuildAI: Participatory framework for algorithmic governance. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019).

Abhilash Mishra. 2023. AI Alignment and Social Choice: Fundamental Limitations and Policy Implications. *SSRN Electronic Journal* (01 2023).

Ritesh Noothigattu, Snehalkumar Gaikwad, Edmond Awad, Sohan Dsouza, Iyad Rahwan, Pradeep Ravikumar, and Ariel Procaccia. 2018. A voting-based system for ethical decision making. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems* 35 (2022), 27730–27744.

Chanwoo Park, Mingyang Liu, Dingwen Kong, Kaiqing Zhang, and Asuman Ozdaglar. 2024. RLHF from heterogeneous feedback via personalization and preference aggregation. *arXiv preprint arXiv:2405.00254* (2024).

Ariel D Procaccia and Jeffrey S Rosenschein. 2006. The distortion of cardinal preferences in voting. In *International Workshop on Cooperative Information Agents*. Springer, 317–331.

Ariel D Procaccia, Jeffrey S Rosenschein, and Gal A Kaminka. 2007. On the robustness of preference aggregation in noisy environments. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*. 1–7.

Amirali Salehi-Abari and Kate Larson. 2021. Group recommendation with noisy subjective preferences. *Computational Intelligence* 37, 1 (2021), 210–225.

Dmitry Shiryaev, Lan Yu, and Edith Elkind. 2013. On elections with robust winners. In *Proceedings of the 2013 International Conference on Autonomous agents and Multi-agent Systems*. 415–422.

Derek M Shockey. 2008. *A comparative study of the robustness of voting systems under various models of noise*. Rochester Institute of Technology.

Anand Siththaranjan, Cassidy Laidlaw, and Dylan Hadfield-Menell. 2024. Distributional preference learning: Understanding and accounting for hidden context in rlhf. In *International Conference on Learning Representations*, Vol. 2024.

Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul F Christiano. 2020. Learning to summarize with human feedback. *Advances in neural information processing systems* 33 (2020), 3008–3021.

Gokul Swamy, Christoph Dann, Rahul Kidambi, Zhiwei Steven Wu, and Alekh Agarwal. 2024. A minimaximalist approach to reinforcement learning from human feedback. In *Proceedings of the 41st International Conference on Machine Learning* (Vienna, Austria) *(ICML'24)*.

Jiancong Xiao, Zhekun Shi, Kaizhao Liu, Qi Long, and Weijie J Su. 2025. Theoretical tensions in RLHF: Reconciling empirical success with inconsistencies in social choice theory. *arXiv preprint arXiv:2506.12350* (2025).

H Peyton Young. 1988. Condorcet's theory of voting. *American Political science review* 82, 4 (1988), 1231–1244.

Huiying Zhong, Zhun Deng, Weijie J Su, Zhiwei Steven Wu, and Linjun Zhang. 2024. Provable multi-party reinforcement learning with diverse human feedback. *arXiv preprint arXiv:2403.05006* (2024).

Daniel M Ziegler, Nisan Stiennon, Jeffrey Wu, Tom B Brown, Alec Radford, Dario Amodei, Paul Christiano, and Geoffrey Irving. 2019. Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593* (2019).

6 Technical Appendix

6.1 Deferred Proofs

Lemma 1. *Assume the noise model illustrated in Section 2, and let $D_i = \epsilon_{i1} - \epsilon_{i2}$ for a voter i. Then, for every constant $p \in (0, 1/2)$, there exists some $\epsilon_0 > 0$ such that for all $0 < \epsilon < \epsilon_0$, $p_A(\epsilon) := \Pr[D_i < -2\epsilon] \geq p$. Further, for every $L > 0$, let $p_B(L) = \Pr[D_i > L]$. Then,*

1. ***Bounded noise**: If each ϵ_{ij} is bounded by $[-T_1, T_2]$, then $D_i = \epsilon_{i1} - \epsilon_{i2} \in [-(T_1 + T_2), T_1 + T_2]$, so $p_B(L) = 0$ for any $L > T_1 + T_2$.*
2. ***Sub-Gaussian noise**: If each ϵ_{ij} is sub-Gaussian with a mean of 0 and parameter s, then D_i is sub-Gaussian with mean 0 and parameter $\sqrt{2}s$. Then $p_B(L) \leq \exp(-\frac{L^2}{4s^2})$. In particular, setting $L = L(n) := 2s\sqrt{2\log(n)}$ means $p_B(L) \leq 1/n^2$.*

Proof. Since each ϵ_{ij} is symmetric about 0 and has a continuous distribution, $D_i = \epsilon_{i1} - \epsilon_{i2}$ is symmetric about 0 and has a continuous CDF G with $G(0) = 1/2$. As $\epsilon \to 0^+$, $p_A(\epsilon) = G(-2\epsilon) \to G(0) = \frac{1}{2}$. Therefore, for any $p < 1/2$, there exists $\epsilon_0 > 0$ such that $p_A(\epsilon) \geq p$ for all $0 < \epsilon < \epsilon_0$.

Then, in the bounded-noise case, $\epsilon_{i1}, \epsilon_{i2} \in [-T_1, T_2]$ almost surely, and hence $D_i = \epsilon_{i1} - \epsilon_{i2} \in [-(T_1+T_2), T_1+T_2]$ almost surely. Therefore, for any $L > T_1+T_2$, $p_B(L) = \Pr[D_i > L] = 0$.

In the sub-Gaussian case, ϵ_{i1} and $-\epsilon_{i2}$ are independent, mean-zero, sub-Gaussian random variables with parameter s. Hence $D_i = \epsilon_{i1} + (-\epsilon_{i2})$ is sub-Gaussian with parameter $\sqrt{s^2 + s^2} = \sqrt{2}\, s$. Therefore, the standard sub-Gaussian tail bound gives $p_B(L) = \Pr[D_i > L] \leq \exp(-\frac{L^2}{4s^2})$. Setting $L = 2s\sqrt{2\log n}$ gives $p_B(L) \leq 1/n^2$.

Theorem 1. *Under our noise model, for all $m \geq 2$ and $k \geq 2$, there exists a feature matrix M and weight profile β inducing true profile σ such that, with high probability, (1) $d_{KT}(Borda(\sigma^*), Borda(\hat{\sigma})) = \binom{m}{2}$ and (2) $d_{KT}(PMC(\sigma^*), PMC(\hat{\sigma})) = \binom{m}{2}$; i.e., both Borda and PMC rules result in a noisy ranking that is exactly the opposite of the true ranking.*

Proof. Consider the following feature matrix:[8]

	f_1	f_2
a_1	ϵ	$1-\epsilon$
a_2	2ϵ	$1-2\epsilon$
$\vdots$	$\vdots$	$\vdots$
a_m	$m\epsilon$	$1-m\epsilon$

[8] For $k > 2$, pad M with $k-2$ all-zero columns and extend every β with zeros in those coordinates; this leaves every utility comparison unchanged, so it suffices to construct the $k = 2$ case below.

Suppose that there are two groups, group A consisting of a $1/2+\delta$ fraction of voters with $\beta_A = (1/2+\epsilon, 1/2-\epsilon)$ and a true ranking $\sigma_A = a_m \succ a_{m-1} \succ \ldots \succ a_1$, and group B consisting of a $1/2-\delta$ fraction of voters with $\beta_B = (0, L)$ for a value L (fixed below) and a true ranking $\sigma_B = a_1 \succ a_2 \succ \ldots \succ a_m$. In this case, it is easy to verify that $\mathrm{PMC}(\sigma^*) = \mathrm{Borda}(\sigma^*) = a_m \succ a_{m-1} \succ \ldots \succ a_1$.

We aim to show that, with high probability, strictly more than half the voters have noisy ranking σ_B, resulting in $\mathrm{PMC}(\hat{\sigma}) = \mathrm{Borda}(\hat{\sigma}) = a_1 \succ a_2 \succ \ldots \succ a_m$.

Because each $a_i = (i\epsilon, 1-i\epsilon)$ lies on a single line in feature space, for any weight vector $\beta = (\beta_1, \beta_2)$ the utility gap between a_{i+1} and a_i is $u(a_{i+1}) - u(a_i) = \epsilon(\beta_1 - \beta_2)$, independent of i. Therefore, every voter's induced ranking is determined entirely by the sign of $\beta_1 - \beta_2$: it is σ_A if $\beta_1 > \beta_2$ and σ_B if $\beta_1 < \beta_2$ (ties occur only on a measure-zero event). In particular, voter i's noisy ranking is σ_A-type iff $\hat{\beta}_{i1} - \hat{\beta}_{i2} > 0$ and σ_B-type iff $\hat{\beta}_{i1} - \hat{\beta}_{i2} < 0$. For a group-$A$ voter, $\hat{\beta}_{i1} - \hat{\beta}_{i2} = 2\epsilon + D_i$, so the voter's ranking flips to σ_B-type iff $D_i < -2\epsilon$, an event of probability $p_A(\epsilon)$. For a group-B voter, $\hat{\beta}_{i1} - \hat{\beta}_{i2} = -L + D_i$, so the voter's ranking flips to σ_A-type iff $D_i > L$, an event of probability $p_B(L)$ (Lemma 1).

Fix any constant $p \in (0, 1/2)$ and let $\delta < p/(2(1-p))$; set $c := (1/2+\delta)p - \delta > 0$. By Lemma 1, we can choose $\epsilon \leq 1/m$ small enough that $p_A(\epsilon) \geq p$ (the constraint $\epsilon \leq 1/m$ ensures every feature value $i\epsilon \in [0,1]$), and choose a fixed constant L (exceeding $T_1 + T_2$ in the bounded case, so that $p_B(L) = 0$, or large enough that $(1/2-\delta)p_B(L) \leq c/2$ in the sub-Gaussian case, using the tail bound of Lemma 1).

Let $n_A = \lceil (1/2+\delta)n \rceil$, $n_B = \lfloor (1/2-\delta)n \rfloor$, and let S_A and S_B denote the number of group-A voters who flip to σ_B-type and the number of group-B voters who flip to σ_A-type, respectively. The number of voters with noisy ranking σ_B is $w_B = (n_B - S_B) + S_A$, and $\mathrm{PMC}(\hat{\sigma}) = \mathrm{Borda}(\hat{\sigma}) = \sigma_B$ whenever $w_B > n/2$ (a strict majority is required for the PMC property), i.e. whenever $S_A - S_B > n/2 - n_B = \delta n$. Let $\Delta := n_A p_A(\epsilon) - n_B p_B(L) - \delta n$. Since $p_A(\epsilon) \geq p$ and $(1/2-\delta)p_B(L) \leq c/2$ by our choice of L, $\Delta \geq nc - n(1/2-\delta)p_B(L) \geq nc/2 =: \Delta_0$. Setting $t = \Delta_0/2 = cn/8$, a union bound with Hoeffding's inequality gives

$$\begin{aligned}\Pr[S_A - S_B \leq \delta n] &\leq \Pr[S_A < n_A p_A - t] + \Pr[S_B > n_B p_B + t] \\ &\leq 2\exp(-2t^2/n) \leq 2\exp(-c^2 n/32),\end{aligned}$$

which is exponentially small in n. Therefore, with probability at least $1 - 2e^{-\Omega(n)}$, $\mathrm{PMC}(\hat{\sigma}) = \mathrm{Borda}(\hat{\sigma}) = a_1 \succ a_2 \succ \cdots \succ a_m$, the exact reversal of σ^*.

Remark 1. Leximax Copeland subject to PO (LCPO), a voting rule introduced by Ge et al. (2024a), is not robust to errors in a linear social choice setting.

Proof. We will start by restating the definition of the *Leximax Copeland subject to PO* voting rule from Ge et al. (2024a). Leximax Copeland always chooses a feasible ranking by ranking the first alternative with the highest Copeland score that can be feasibly ranked first under some weight vector, then ranking second the candidate with the highest Copeland score which can be feasibly ranked second, and so on. Leximax Copeland subject to PO incorporates the

Pareto optimality criterion; for every pair of alternatives where one dominates the other, this rule restricts rankings to place the dominating alternative above the dominated one.

We can apply the counterexample from Theorem 1 to LCPO. Consider the feature matrix and voter group breakdown established in Theorem 1. Recall that group A consists of a $1/2+\delta$ fraction of voters with $\beta_A = (1/2+\epsilon, 1/2-\epsilon)$ and a true ranking $\sigma_A = a_m \succ a_{m-1} \succ \ldots \succ a_1$, and group B consists of a $1/2-\delta$ fraction of voters with $\beta_B = (0, L)$ and a true ranking $\sigma_B = a_1 \succ a_2 \succ \ldots \succ a_m$. In this case, $\text{LCPO}(\sigma^*) = a_m \succ a_{m-1} \succ \cdots \succ a_1$, and it is easy to verify that this ranking is feasible, pairwise-majority consistent, and Pareto optimal.

By the same argument as in the proof of Theorem 1, with probability at least $1-2\exp(-\Omega(n))$ strictly more than half the voters have noisy ranking σ_B, and this ranking is likewise feasible, pairwise-majority consistent, and Pareto optimal, giving $\text{LCPO}(\hat{\sigma}) = a_1 \succ \cdots \succ a_m$.

Corollary 1. *Under our noise model, for all $m \geq 2$ and $k \geq 2$, there exists a feature matrix M and associated profile σ such that, with high probability, PSR(σ^*) $\neq$ PSR($\hat{\sigma}$) for any positional scoring rule.*

Proof. Given a score vector $s_1 \geq \ldots \geq s_k > s_{k+1} \geq \ldots \geq s_m$, consider the following feature matrix:

	f_1	f_2
M_1	1	1
M_2	$\frac{m-2}{m-1}$	$\frac{m-2}{m-1}$
$\vdots$	$\vdots$	$\vdots$
M_{k-1}	$\frac{m-k+1}{m-1}$	$\frac{m-k+1}{m-1}$
M_k	$\frac{m-k}{m-1}$	$\frac{m-k-1}{m-1}$
M_{k+1}	$\frac{m-k-1}{m-1}$	$\frac{m-k}{m-1}$
M_{k+2}	$\frac{m-k-2}{m-1}$	$\frac{m-k-2}{m-1}$
$\vdots$	$\vdots$	$\vdots$
M_m	0	0

Further, suppose that there are two groups, group A consisting of a $1/2+\delta$ fraction of voters with $\beta_A = (1/2+\epsilon, 1/2-\epsilon)$ and a true ranking $\sigma_A = a_1 \succ a_2 \succ \ldots \succ a_k \succ a_{k+1} \succ \ldots \succ a_m$, and group B consisting of a $1/2-\delta$ fraction of voters with $\beta_B = (0, L)$ for L as in Lemma 1 , giving true ranking σ_B with $M_{k+1} \succ M_k$ (all other pairwise comparisons agree between σ_A and σ_B, and are unaffected by the value of L). By the strict inequality $s_k > s_{k+1}$, $\text{PSR}(\sigma^*)$ places M_k above M_{k+1}.

As in Theorem 1, a group-A voter's comparison between M_k, M_{k+1} flips if and only if $D_i < -2\epsilon$ (with probability $p_A(\epsilon)$), and a group-B voter's comparison flips if and only if $D_i > L$ (with probability $p_B(L)$). Applying the concentration inequalities from the proof of Theorem 1 to these flip events, with probability at least $1-2e^{(-\Omega(n))}$ strictly more than half of the voters rank M_{k+1} above M_k, so $\text{PSR}(\hat{\sigma})$ places M_{k+1} above M_k, resulting in $\text{PSR}(\sigma^*) \neq \text{PSR}(\hat{\sigma})$.

Theorem 3. *Under our noise model, for values of $m \geq 2$, $k = m$, there exists a feature matrix M and corresponding approval profile σ such that, with high probability, $d_{KT}(Approval(\hat{\sigma}), Approval(\sigma)) = \binom{m}{2}$, i.e., approval voting results in a noisy ranking that is exactly the opposite of the true ranking.*

Proof. Consider the following feature matrix with ones down the diagonal and zeros everywhere else.

	f_1	f_2	$\dots$	f_m
a_1	1	0	$\dots$	0
a_2	0	1	$\dots$	0
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
a_m	0	0	$\dots$	1

Now, consider the following m groups of voters, $G_1, G_2, \dots, G_m$. Sizes of the groups, respective β values, thresholds t, and approval sets for each group are given below. Here, each δ_i for all $i \in \{2, \dots, m\}$ is chosen such that with probability $c \cdot (m - i + 1)/m$, the noise model returns a value at most $1 - \delta_i$ for some constant c. This can be done by, e.g., inspecting the CDF of the noise function and choosing appropriate quantiles.

Group	Size	β	t	Approval
G_1	$1/m + \frac{m(m-1)}{2}\epsilon$	$(1, 0, \dots, 0)$	1	$\{a_1\}$
G_2	$1/m - \epsilon$	$(0, 1, 0, \dots, 0)$	$1 - \delta_2$	$\{a_2\}$
G_3	$1/m - 2\epsilon$	$(0, 0, 1, 0, \dots, 0)$	$1 - \delta_3$	$\{a_3\}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
G_m	$1/m - (m-1)\epsilon$	$(0, \dots, 0, 1)$	$1 - \delta_m$	$\{a_m\}$

It is easy to verify that the true ranking will be $a_1 \succ a_2 \succ \dots \succ a_m$. However, after noise, G_i will stop approving a_i with decreasing probability in i. By choosing values of ϵ and δ_i carefully, we can ensure that the final noisy ranking is $a_m \succ a_{m-1} \succ \dots \succ a_1$ with high probability.

Assume for ease of exposition that all groups have equal size, i.e., have size n/m. Further, define the "dropout rate" for a group G_i as the fraction of that group's voters who, after noise, no longer approve candidate a_i (and in fact approve no candidates at all). We would like to prove that, with high probability, the noisy approval voting process leads to the order $a_1 \succ a_2 \succ \dots \succ a_m$, which implies that the dropout rate is highest for G_1, then G_2, and so on down to G_m. Let us define a series of random variables, $X_1, X_2, \dots, X_m$, where X_i represents the number of dropouts in group G_i. Each X_i is a binomial random variable with success (i.e., dropout) rate $c \cdot (m - i + 1)/m$. These success probabilities are separated by a constant factor, so as n goes to infinity, we can show that with high probability the number of dropouts will indeed be in the correct order with high probability using Hoeffding's inequality.

To show that, we aim to first provide a lower bound on the probability that two consecutive random variables will not satisfy the desired order then use a union bound to provide a lower bound on the probability that the number of

dropouts will indeed be in the right order. To bound $Pr[X_i \leq X_{i+1}]$, we first need to calculate $E[X_i] = \frac{n}{m} \cdot \frac{c(m-i+1)}{m}$. Then,

$$E[X_i - X_{i+1}] = \frac{cn}{m}\left(\frac{m-i+1}{m} - \frac{m-(i+1)+1}{m}\right) = \frac{cn}{m^2}.$$

Then, we can bound

$$\begin{aligned} Pr[X_i \leq X_{i+1}] &= Pr[X_i - X_{i+1} - (E[X_i - X_{i+1}]) \leq -E[X_i - X_{i+1}]] \\ &\leq e^{-\frac{2(E[X_i - X_{i+1}])^2}{2n/m}} = e^{-\frac{m(E[X_i - X_{i+1}])^2}{n}} = e^{-\frac{c^2 n}{m^3}}. \end{aligned}$$

Therefore, the probability that two consecutive random variables will satisfy the correct order is $1 - e^{-\frac{c^2 n}{m^3}}$. After doing a union bound over all $(m-1)$ consecutive pairs, we get that the probability that the number of dropouts will be in the right order is $1 - (m-1)e^{-\frac{c^2 n}{m^3}}$. For all $n \gg \frac{2m^3 ln(m-1)}{c^2}$, this holds with high probability.

Corollary 2. *Under our noise model, for values of $k \geq 2$, there exists a feature matrix M and corresponding approval profile σ such that, with high probability, Approval(σ^*) $\neq$ Approval($\hat{\sigma}$) even for $k \neq m$.*

Proof. Consider the following feature matrix:

	f_1	f_2
a^*	1	0
a_1	0	$1-\epsilon$
$\vdots$	$\vdots$	$\vdots$
a_{m-1}	0	$1-(m-1)\epsilon$

Further, suppose there are two groups of voters, where group A consists of $\frac{n}{2}+\xi$ voters with $\beta_A = (1/2+\epsilon, 1/2-\epsilon)$ and threshold $t_A = 0.5$, and group 2 consists of $\frac{n}{2} - \xi$ voters with $\beta_B = (\epsilon, 1-\epsilon)$ and threshold $t_B = 1 - 2\epsilon$. Then, we have that group A only approves of a^* and group B approves of a_1, resulting in Approval(σ) $= a^* \succ a_1 \succ a_2 = \ldots = a_{m-1}$. However, when applying noise to this profile, a constant fraction of group A voters will switch their approval set to contain only a_1, while no group B voters will switch their approval set.

Let AV be the number of approval votes an alternative gets in the noisy setting and let $\omega = E[AV(a_1)] - E[AV(a^*)] = -2\xi$ be the difference in expected number of approvals for a_1 and a^* in the noisy setting. We aim to lower bound $Pr[AV(a^*) < AV(a_1)]$, so we can apply Hoeffding's inequality to get an upper bound on $Pr[AV(a_1) \leq AV(a^*)]$. We can rewrite $Pr[AV(a_1) \leq AV(a^*)]$ as

$$\begin{aligned} &Pr[AV(a_1) - E[AV(a_1)]] \leq AV(a^*) - E[AV(a^*] + \omega] \\ &= Pr[AV(a_1) - E[AV(a_1)] \leq \omega/2] + Pr[AV(a^*)] - E[AV(a^*)] \geq \omega/2]. \end{aligned}$$

We can bound these two terms separately, bounding $Pr[AV(a^*) - E[AV(a^*)] \geq \omega/2] \leq e^{-\frac{\omega^2}{2n}}$ and rewriting $Pr[AV(a_1) - E[AV(a_1)] \leq \omega/2]$ as $Pr[|AV(a_1) -$

$E[AV(a_1)]| \geq \omega/2] \leq 2e^{-\frac{\omega^2}{2n}}$. Therefore $Pr[AV(a_1) \leq AV(a^*)] \leq 3e^{-\frac{4\xi}{2n}}$, and thus, $Pr[AV(a^*) < AV(a_1)] = 1 - 3e^{-\frac{4\xi}{2n}}$. For sufficiently large n, this holds with high probability.

Theorem 4. *Let Sum($\hat{\sigma}$) denote the ranking of alternatives in decreasing order of noisy utility, i.e., in decreasing order of $\hat{u}(a_j) := \sum_{i\in[n]} \hat{u}_i(a_j)$, where $\hat{u}_i(a_j) := \hat{\beta}_i \cdot \alpha_j$. Then, for any ranking σ^* with a minimum distance of $\upsilon > 0$ between any consecutive pair of alternatives, Sum($\hat{\sigma}$) = Sum(σ^*) with high probability.*

Proof. We will break this proof into two cases: one where the noise is bounded (between $[-T_1, T_2]$) and one where the noise is sub-Gaussian.
Case 1: Bounded noise
Suppose, without loss of generality, that $a_a \succ a_b \in \sigma^*$. We aim to show that, in this case, $Pr[\hat{u}(a_b) > \hat{u}(a_a)] \approx 0$. Let $\hat{u}_i(a_a) - \hat{u}_i(a_b) = \hat{\beta}_i(\alpha_a - \alpha_b)$ be the difference in noisy utilities between alternatives a and b for an individual i. Then, let $\hat{d}_{ab} = \sum_{i=1}^n \hat{u}_i(a_a) - \hat{u}_i(a_b) = \sum_{i=1}^n \hat{\beta}_i(\alpha_a - \alpha_b)$ be the total difference in noisy utilities between alternatives a and b. Further, let $d_{ab} = \mathbb{E}[\hat{d}_{ab}] = \sum_{i=1}^n \beta_i(\alpha_a - \alpha_b)$ be the total difference in true utilities between alternatives a and b. This value $d_{ab} > 0$, by our assumption that $a_a \succ a_b \in \sigma^*$ and that there exists some value υ of minimum separation between a_a and a_b (i.e., a_a and a_b are not tied). Then, since each $\hat{\beta}_i$ differs from the true β_i by at most $T = \max(|T_1|, |T_2|)$, we can bound each noisy utility difference $(\hat{u}_i(a_a) - \hat{u}_i(a_b))$ in a range of $2\psi_{ab}$ where

$$\psi_{ab} = |\epsilon_i \cdot (\alpha_a - \alpha_b)| \leq T||\alpha_a - \alpha_b||_1.$$

Then, we can apply Hoeffding's inequality to get

$$Pr[\hat{d}_{ab} \leq 0] = Pr[\hat{u}(a_b) > \hat{u}(a_a)] \leq e^{-\frac{2(d_{ab})^2}{\sum_{i=1}^n (2\psi_{ab})^2}}$$

We can rewrite the variable d_{ab} as $n \cdot \chi_{ab}$ where we define χ_{ab} to be the average $\beta_i(\alpha_a - \alpha_b)$ across all voters n. Therefore, we can rewrite $Pr[\hat{d}_{ab} \leq 0]$ as

$$Pr[\hat{d}_{ab} \leq 0] = Pr[\hat{u}(a_b) > \hat{u}(a_a)] \leq e^{-\frac{2(d_{ab})^2}{\sum_{i=1}^n (2\psi_{ab})^2}} = e^{-\frac{2n^2\chi_{ab}^2}{4n\psi_{ab}^2}} = e^{-\frac{n\chi_{ab}^2}{(2\psi_{ab}^2)}}.$$

Therefore, the probability of keeping the "correct" order for any two alternatives in the noisy ranking is $1 - e^{-\frac{n\chi_{ab}^2}{\psi_{ab}^2}}$. Even after doing a union bound over the $\binom{m}{2}$ alternatives, we still get the correct ranking with $1 - \frac{m(m-1)}{2e^{-n\chi_{\min}^2/2T^2\Psi_{\max}^2}}$, where $\chi_{\min} := \min_{a\succ b} \chi_{ab} > 0$ and $\Psi_{\max} := \max_{a\succ b} ||\alpha_a - \alpha_b||_1$. If $n \gg \ln(m^2)$, our "with high probability" statement holds.
Case 2: Sub-Gaussian noise
Suppose, as before, that $a_a \succ a_b \in \sigma^*$. Let $\hat{d}_{ab}, d_{ab}$ be defined as above, with ϵ as a mean-zero sub-Gaussian random variable with parameter s^2. We can apply

a sub-Gaussian bound (i.e., $Pr[S \leq -t] \leq e^{\frac{-t^2}{3v^2}}$ where v^2 is a variance proxy) to get

$$\Pr[\hat{d}_{ab} \leq 0] = \Pr[\sum_{i=1}^{n}(\epsilon_i(\alpha_a - \alpha_b)) \leq -d_{ab}] \leq e^{-\frac{d_{ab}^2}{2ns^2||\alpha_a - \alpha_b||_2^2}}.$$

Defining χ_{ab} as before, we get

$$Pr[\hat{d}_{ab} \leq 0] \leq e^{-\frac{n^2\chi_{ab}^2}{2ns^2||\alpha_a - \alpha_b||_2^2}} = e^{-\frac{n\chi_{ab}^2}{2s^2||\alpha_a - \alpha_b||_2^2}}.$$

Similarly to in the Case 1 proof, even after doing a union bound over the $\binom{m}{2}$ alternatives, we still get the correct ranking with probability $1 - \frac{m(m-1)}{2e^{-n\chi_{\min}^2/2s^2\Psi_{\max}^2}}$, where $\chi_{\min} := \min_{a \succ b} \chi_{ab} > 0$ and $\Psi_{\max} := \max_{a \succ b} ||\alpha_a - \alpha_b||_2$. If $n \gg \ln(m^2)$, our "with high probability" statement holds.

	f_1	f_2	f_3
a_1	0.6	0.2	0.1
a_2	$0.3-\epsilon$	$0.3-\epsilon$	$0.3-\epsilon$
a_3	0.1	0.6	0.2
a_4	0.2	0.1	0.6

(a) Counterexample for Theorem 12.

	f_1	f_2
a_1	$0.2-\epsilon$	$0.2-\epsilon$
a_2	0.3	0.1
a_3	0.1	0.3
a_4	0.19	$0.21+\epsilon$
a_5	$0.19+\epsilon$	$0.21+2\epsilon$

(b) Counterexample for Theorem 10.

Fig. 4: Feature matrices for feasibility counterexamples.

Theorem 7. *For any $m \geq 4$, there exists some construction over m alternatives that violates feasibility under positional scoring rules, implying that for positional scoring rules, attainability does not imply feasibility.*

Proof. We will break this proof into cases on the scoring vector s associated with positional scoring rules, first starting with the case where there exists a value x followed by a value y in the first three positions of s where $x > y$.

Case 1. In this case, suppose that s can be defined as $s = [s_1, s_2, s_3, \dots]$ where values after s_3 in the positional scoring vector are non-increasing and at least one of $s_1 > s_2$ and $s_2 > s_3$ is true. Then, we construct three flavors of counterexample, one where $2s_2 > s_1 + s_3$, another where $2s_2 < s_1 + s_3$, and finally the case where $2s_2 = s_1 + s_3$.

Case 1.1: $2s_2 > s_1 + s_3$
Consider the following feature matrix:

	f_1	f_2
a_1	1	0
a_2	$0.5-\epsilon$	$0.5-\epsilon$
a_3	0	1

Suppose there are two groups where some group of size $n/2$ of voters M has $\beta_A = (1,0)$ resulting in ranking $\sigma_A = a_1 \succ a_2 \succ a_3$ and another group of size $n/2$ of voters B of voters has $\beta_B = (0,1)$ resulting in ranking $\sigma_B = a_3 \succ a_2 \succ a_1$. Then, from the scoring vector a_1 receives $\frac{n}{2}(s_1+s_3)$ points, a_2 receives $\frac{n}{2}(2s_2)$ points, and a_3 receives $\frac{n}{2}(s_1+s_3)$ points. Therefore, because $2s_2 > s_1+s_3$, the final ranking is $a_2 \succ a_1 = a_3$. However, there does not exist any β under which a_2 can win.

Case 1.2: $2s_2 < s_1+s_3$

Consider the following feature matrix:

	f_1	f_2
a_1	1	0
a_2	$0.5+\epsilon$	$0.5+\epsilon$
a_3	0	1

Suppose there are two groups where some group of size $n/2$ of voters M has $\beta_A = (1,0)$ resulting in ranking $\sigma_A = a_1 \succ a_2 \succ a_3$ and another group of size $n/2$ of voters B of voters $(A+B=n)$ has $\beta_B = (0,1)$ resulting in ranking $\sigma_B = a_3 \succ a_2 \succ a_1$. Then, from the scoring vector a_1 receives $\frac{n}{2}(s_1+s_3)$ points, a_2 receives $\frac{n}{2}(2s_2)$ points, and a_3 receives $\frac{n}{2}(s_1+s_3)$ points. Therefore, since $2s_2 < s_1+s_3$, the final ranking is $a_1 = a_3 \succ a_2$. However, there does not exist any β under which a_2 can lose.

Case 1.3: $2s_2 = s_1+s_3$

Consider the following feature matrix:

	f_1	f_2	f_3
a_1	1	ϵ	2ϵ
a_2	$1/3-\epsilon$	$1/3-\epsilon$	$1/3-\epsilon$
a_3	2ϵ	1	ϵ
a_4	ϵ	2ϵ	1

We will extend the positional scoring vector to $s = [s_1, s_2, s_3, s_4, \dots]$, where $s_4 \leq s_3$ to allow for at least four alternatives in the feature matrix. Suppose there exist three voter groups, all of size $n/3$, where group A has $\beta_A = (1,0,0)$, group B has $\beta_B = (0,1,0)$, and group C has $\beta_C = (0,0,1)$, resulting in rankings $\sigma_A = a_1 \succ a_2 \succ a_3 \succ a_4$, $\sigma_B = a_3 \succ a_2 \succ a_4 \succ a_1$, and $\sigma_C = a_4 \succ a_2 \succ a_1 \succ a_3$. In this case, a_2 receives $3(n/3) \times s_2 = ns_2$ points from the scoring vector, while all other alternatives receive $(n/3)(s_1+s_3+s_4)$ points. In this case, to prove that attainability does not imply feasibility, we must simply show that a_2 wins (while no β can result in a_2 winning), i.e., $ns_2 > (n/3)(s_1+s_3+s_4)$. We can rewrite this inequality as $3s_2 > s_1+s_4+s_3$. Using the fact that $2s_2 = s_1+s_3$, we know that the above inequality holds as long as $s_2 > s_4$. This is true unless

$s_2 = s_3 = s_4$, however, if $s_2 = s_3$, then from $2s_2 = s_1 + s_3$, we have that $s_1 = s_2$, which cannot hold, since we assume at least one strict inequality in the first three positions of the scoring vector.

Case 2. Next, we address scoring vectors where there is a strict inequality ($>$) after the second position in the scoring vector (i.e., there exists a value x followed by a value y where x is in the second position of s or later where $x > y$). Suppose that we have a positional scoring rule where the first inequality ($>$) in the scoring vector associated with that rule is after the i^*-th position (where $i^* \geq 2$) in the scoring vector. Then, consider the following feature matrix, where each alternative a_j for $j \neq i^*$ has a 1 in feature f_j, exactly $i^* - 3$ entries of $\frac{1}{k}$ in the subsequent features $f_{j+1}, \ldots, f_{j+i^*-3}$ (indices taken mod k), and 0 elsewhere. The alternative a_{i^*} has $\frac{1}{k} - \epsilon$ in every feature. Formally:

$$\alpha_{j\ell} = \begin{cases} 1 & \text{if } \ell = j \\ \frac{1}{k} & \text{if } \ell \in \{j+1, \ldots, j+i^*-3\} \pmod{k} \\ 0 & \text{otherwise} \end{cases} \quad \text{for } j \neq i^*$$

$$\alpha_{i^*} = \left(\frac{1}{k} - \epsilon, \frac{1}{k} - \epsilon, \ldots, \frac{1}{k} - \epsilon \right)$$

	f_1	f_2	f_3	$\cdots$	f_k
a_1	1	$\frac{1}{k}$	$\frac{1}{k}$	$\cdots$	0
a_2	0	1	$\frac{1}{k}$	$\cdots$	0
a_3	0	0	1	$\cdots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
a_{i^*}	$\frac{1}{k} - \epsilon$	$\frac{1}{k} - \epsilon$	$\frac{1}{k} - \epsilon$	$\cdots$	$\frac{1}{k} - \epsilon$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
a_{m-1}	$\frac{1}{k}$	0	0	$\cdots$	$\frac{1}{k}$
a_m	$\frac{1}{k}$	$\frac{1}{k}$	0	$\cdots$	1

Given that the inequality in the positional scoring vector happens after the i^*-th position, we can represent the scoring vector as $s = [x, \ldots x, y, \ldots]$ where the final x is in the i^*-th position. Then, if we have separate voter groups for each feature (i.e., k voter groups for all $j \in [k]$ with each $\beta_j = (0, \ldots 1, 0, \ldots)$ where the 1 is at the j-th position), each voter group j has the ranking of $\sigma_j = a_j \succ \{\text{the } i^* - 3 \text{ rotating alternatives}\} \succ a_{i^*} \succ \{\text{all other alternatives}\}$, where a_{i^*} is always ranked in the i^*-th position.

Therefore, a_{i^*} gets $k \times x$ points from the scoring vector and each a_j ($j \neq i^*$) gets at most $(i^* - 2)x + (k - i^* + 2)y$ points from the scoring vector ($1 \times x$ for the group for which it is ranked first, $(i^* - 3)x$ for the $i^* - 3$ groups for which it is a rotating alternative, and at most $y(k - (i^* - 3) - 1)$ for the groups for which it is neither a rotating alternative nor the winner). By the definition of the scoring vector ($x > y$), a_{i^*} wins, however it is infeasible for a_{i^*} to win, as there exists no β for which a_{i^*} can win.

Theorem 8. *For any $m \geq 6$, there exists some construction over m alternatives that violates feasibility under PMC rules, implying that for PMC rules, attainability does not imply feasibility.*

Proof. Consider the following feature matrix:

	f_1	f_2	f_3
a^*	2.1	2.1	2.1
a_1	3	2	2
a_2	2	3	2
a_3	2	2	3
a_4	1	2	2
a_5	2	1	2

Suppose that $n/3$ voters in group A have weight vector $\beta_A = (1, 2\epsilon, \epsilon)$, $n/3$ voters in group B have weight vector $\beta_B = (2\epsilon, 1, \epsilon)$, and $n/3$ voters in group C have the preferences $\beta_C = (2\epsilon, \epsilon, 1)$ for some small ϵ. In this case, group A's ranking is $a_1 \succ a^* \succ a_2 \succ a_3 \succ a_5 \succ a_4$, group B's ranking is $a_2 \succ a^* \succ a_1 \succ a_3 \succ a_4 \succ a_5$, and population C's ranking is $a_3 \succ a^* \succ a_1 \succ a_2 \succ a_5 \succ a_4$. Any PMC voting rule returns $a^* \succ a_1 \succ a_2 \succ a_3 \succ a_5 \succ a_4$. However, this ranking is infeasible because there is no β that produces a ranking such that a^* is the winner given this feature matrix.
This counterexample can be extended beyond 6 alternatives by adding alternatives that are weakly Pareto dominated by any of $\{a_1, a_2, a_3\}$.

	f_1	f_2	f_3
a^*	2.4	2.4	2.4
a_1	3	2	2
a_2	2	3	2
a_3	2	2	3

Table 1: Feature matrix for Theorems 9 and 11

Theorem 9. *For any $m \geq 4$, there exists some construction over m alternatives that violates feasibility under instant-runoff voting, implying that for IRV, attainability does not imply feasibility.*

Proof. Consider the following feature matrix in Table 1.

Suppose that $n/3 + \xi$ voters in group A have weight vector $\beta_A = (1, 2\epsilon, \epsilon)$, $n/3$ voters in group B have weight vector $\beta_B = (2\epsilon, 1, \epsilon)$, and $n/3 - \xi$ voters in group C have the preferences $\beta_C = (2\epsilon, \epsilon, 1)$ for some small ϵ. In this case, group A's ranking is $a_1 \succ a^* \succ a_2 \succ a_3$, group B's ranking is $a_2 \succ a^* \succ a_1 \succ a_3$, and population C's ranking is $a_3 \succ a^* \succ a_1 \succ a_2$. In instant-runoff voting, a^* would be eliminated first, followed by a_3 and then a_2, resulting in the ranking

$a_1 \succ a_2 \succ a_3 \succ a^*$. However, this ranking is infeasible because there is no β that produces a ranking such that a^* is the loser given this feature matrix. This example can be extended to $m > 4$ by adding copies of alternatives a_1, a_2, or a_3.

Theorem 10. *For any $m \geq 5$, there exists some construction over m alternatives that violates approval-feasibility, implying that, for approval voting, attainability does not imply feasibility.*

Proof. We first begin by describing the counterexample for $m = 5$. Consider the feature matrix in Figure 4b. Suppose that there are two major blocs of voters that have the same β vectors: Group G has $\beta_G = (1, 0)$ and Group B has $\beta_B = (0, 1)$. However, suppose that not everyone in the major blocs agree on the same threshold. Suppose that we have the following breakdown of the aforementioned blocs into subgroups:

Group	Size	β	t	Approval
G_1	$n/3$	$(1, 0)$	0.195	$\{a_2, a_1\}$
G_2	$n/18$	$(1, 0)$	0.25	$\{a_2\}$
B_1	$n/3$	$(0, 1)$	0.15	$\{a_3, a_5, a_4, a_1\}$
B_2	$n/6$	$(0, 1)$	$0.21 + 1.5\epsilon$	$\{a_3, a_5\}$
B_3	$n/9$	$(0, 1)$	0.295	$\{a_3\}$

In this case, approval voting returns the ranking $\sigma^* = a_1 \succ a_3 \succ a_5 \succ a_2 \succ a_4$. However, for any decomposition of σ^* into an approved set of alternatives and a non-approved set of alternatives, there exists no singular β and t such that, given this feature matrix, will yield the same approved and non-approved sets. By nature of how the matrix is constructed: a_1 can never win with a single β and t because it does not perform the best in either feature, nor does any equal or otherwise weighing of the features allow for a_1 to be the winner. Similarly, a_4 can never lose under any single β and t because it doesn't perform the worst in either of the features, nor does any equal or otherwise weighing of the features allow a_4 to be the loser. For a_3 to beat a_5 and a_2 in any ranking with a single β and t, the weight on f_2 of the β vector must be higher than the weight on f_1. However, for a_1 to beat a_5 or for a_2 to beat a_4 in the ranking with a single β and t, the weight on f_1 of the β vector must be higher than the weight on f_2, which is a contradiction. Therefore, the ranking $a_1 \succ a_3 \succ a_5 \succ a_2 \succ a_4$ is not approval-feasible.

This counterexample can be extended to more than 5 alternatives by considering alternatives that fall between a_3 and a_5 or a_5 and a_2 in the aggregate ($m = 5$) ranking. Specifically, for $i \geq 3$, $M(a_{2i}) = (0.3 - (2i-5)\epsilon, 0.1 + (2i-5)\epsilon)$ and $M(a_{2i+1}) = (0.3 + (2i-4)\epsilon, 0.1 - (2i-4)\epsilon)$. In other words, each successive even-numbered alternative has features near a_2 and each successive odd-numbered alternative has features near a_3. This is illustrated in the general

matrix construction below.

	f_1	f_2
a_1	$0.2-\epsilon$	$0.2-\epsilon$
a_2	0.3	0.1
a_3	0.1	0.3
a_4	0.19	$0.21+\epsilon$
a_5	$0.19+\epsilon$	$0.21+2\epsilon$
$\vdots$	$\vdots$	$\vdots$
a_{2i}	$0.3-(2i-5)\epsilon$	$0.1+(2i-5)\epsilon$
a_{2i+1}	$0.1+(2i-4)\epsilon$	$0.3-(2i-4)\epsilon$

For the breakdowns of "subgroups" of the A and B voter blocs for extensions beyond $m = 5$, this process requires the same total number of voter groups (between $A.1, B.1, A.2, B.2, \ldots$) as the number of alternatives. This is to ensure that there are no ties in number of approvals between alternatives (i.e., each subgroup has an unique approval set, and all alternatives belong to at least one subgroup; this requires m subgroups). To illustrate the breakdown of subgroups, recall the breakdown from the $m = 5$ example:

Group	Size	β	t	Approval
G_1	$n/3$	$(1,0)$	0.195	$\{a_2, a_1\}$
G_2	$n/18$	$(1,0)$	0.25	$\{a_2\}$
B_1	$n/3$	$(0,1)$	0.15	$\{a_3, a_5, a_4, a_1\}$
B_2	$n/6$	$(0,1)$	$0.21+1.5\epsilon$	$\{a_3, a_5\}$
B_3	$n/9$	$(0,1)$	0.295	$\{a_3\}$

If we add one "even" candidate (e.g., a_{2i}) according to the aforementioned procedure, the subgroup breakdown would look as follows:

Group	Size	β	t	Approval
G_1	$n/3$	$(1,0)$	0.195	$\{a_2, \mathbf{a_{2i}}, a_1\}$
G_2	w	$(1,0)$	$0.3-(2i-5)\epsilon$	$\{a_2, \mathbf{a_{2i}}\}$
B_1	$n/3$	$(0,1)$	0.15	$\{a_3, a_5, a_4, a_1\}$
B_2	x	$(0,1)$	$0.21+1.5\epsilon$	$\{a_3, a_5\}$
B_3	y	$(0,1)$	0.295	$\{a_3\}$
$\mathbf{B_{2i}}$	z	$(0,1)$	$0.1+(2i-5)\epsilon$	$\{a_3, a_5, a_4, a_1, \mathbf{a_{2i}}\}$

where the values of w, x, y, z must sum up to $n/3$, the relationship between w, x, y must stay the same as in the base example (without a_{2i}), and $z < \min\{w, x, y\}$. In other words, the new subgroup gets the smallest size, and the sizes of the other (non-$n/3$-sized) subgroups shrink slightly while still respecting their previous relative sizes with respect to one another. To summarize, if we include an "even" candidate, we include it in the existing group A approval subgroups, and additionally add a B group with a threshold that barely approves the new a_{2i}. This results in a cumulative approval voting ranking of $a_1 \succ a_3 \succ a_{2i} \succ a_5 \succ a_2 \succ a_4$.

Similarly, if we add one "odd" candidate according to the aforementioned procedure, the subgroup breakdown would look as follows:

Group	Size	β	t	Approval
G_1	$n/3$	$(1,0)$	1.95	$\{a_2, a_1\}$
G_2	w	$(1,0)$	2.5	$\{a_2\}$
$\mathbf{G_{2i+1}}$	z	$(1,0)$	$0.1 + (2i-4)\epsilon$	$\{a_2, a_1, \mathbf{a_{2i+1}}\}$
B_1	$n/3$	$(0,1)$	1.5	$\{a_3, \mathbf{a_{2i+1}}, a_5, a_4, a_1\}$
B_2	x	$(0,1)$	$2.1 + 1.5\epsilon$	$\{a_3, \mathbf{a_{2i+1}}, a_5\}$
B_3	y	$(0,1)$	$0.3 - (2i-4)\epsilon$	$\{a_3, \mathbf{a_{2i+1}}\}$

where the values of w, z, x, y must sum up to $n/3$, the relationship between w, x, y must stay the same base example (without a_{2i+1}), and $z < \min\{w, x, y\}$. In other words, the new subgroup gets the smallest size, and the sizes of the other (non-$n/3$-sized) subgroups shrink slightly while still respecting their previous relative sizes with respect to one another. This group breakdown can be done iteratively for any number of candidates.

In this general case, we will construct a profile in which approval voting returns the ranking

$$\sigma^* = a_1 \succ a_3 \succ \underbrace{a_m \succ a_{m-2} \succ \ldots \succ a_6}_{\text{even candidates (decreasing)}} \succ a_5 \succ \underbrace{a_7 \succ a_9 \succ \ldots \succ a_{m-1}}_{\text{odd candidates (increasing)}} \succ a_2 \succ a_4$$

in the case where m is even or

$$\sigma^* = a_1 \succ a_3 \succ \underbrace{a_{m-1} \succ a_{m-3} \succ \ldots \succ a_6}_{\text{even candidates (decreasing)}} \succ a_5 \succ \underbrace{a_7 \succ a_9 \succ \ldots \succ a_m}_{\text{odd candidates (increasing)}} \succ a_2 \succ a_4$$

when m is odd.

For any i, there exists no feasible breakdown of σ^* into approved and non-approved sets with a singular β and t. As established in the $m = 5$ counterexample, it is impossible to get approval sets of $\{a_1\}$ and $\{a_1, a_3\}$ given this setup and also impossible for the sets $\{a_4\}$ and $\{a_2, a_4\}$ to be the non-approved sets, but we still must establish that no other separations of σ^* into approval/non-approval sets are possible. Now, assume m is even; the following argument can naturally be extended for m odd. Also, we write $\beta := (\beta_1, \beta_2)$ in the following argument. For any other separation of σ^* into non-empty $\sigma_1 \succ \sigma_2$, at least a_m must be approved and a_{m-1} must be disapproved. This means that $\beta_1 > \beta_2$. However, in order for a_3 to be approved while a_2 is not approved, we need $\beta_2 > \beta_1$, which is a contradiction. Therefore, σ^* is not approval-feasible.

Example with $m = 7$: We will walk through an extension of the $m = 5$ example via the process mentioned in the proof. Consider the following feature

matrix (where $m = 7$):

	f_1	f_2
a_1	$0.2 - \epsilon$	$0.2 - \epsilon$
a_2	0.3	0.1
a_3	0.1	0.3
a_4	0.19	$0.21 + \epsilon$
a_5	$0.19 + \epsilon$	$0.21 + 2\epsilon$
a_6	$0.3 - \epsilon$	$0.1 + \epsilon$
a_7	$0.1 + \epsilon$	$0.3 - \epsilon$

Further, suppose that there are two major blocs of voters that have the same β vectors: Group G has $\beta_G = (1, 0)$ and Group B has $\beta_B = (0, 1)$. However, suppose that not everyone in the major blocs agree on the same threshold. Suppose that we have the following breakdown of the aforementioned blocs into subgroups:

Group	Size	β	t	Approval
G_1	$n/3$	$(1, 0)$	0.195	$\{a_2, a_6, a_1\}$
G_2	$n/15$	$(1, 0)$	0.25	$\{a_2, a_6\}$
G_3	$n/75$	$(1, 0)$	$0.19 + 0.5\epsilon$	$\{a_2, a_6, a_1, a_5\}$
B_1	$n/3$	$(0, 1)$	0.15	$\{a_3, a_7, a_5, a_4, a_1\}$
B_2	$4n/75$	$(0, 1)$	$0.21 + 1.5\epsilon$	$\{a_3, a_7, a_5\}$
B_3	$2n/15$	$(0, 1)$	0.295	$\{a_3\}$
B_4	$n/15$	$(0, 1)$	0.101	$\{a_3, a_7, a_5, a_4, a_1, a_6\}$

In this case, approval voting returns the ranking $\sigma^* = a_1 \succ a_3 \succ a_6 \succ a_5 \succ a_7 \succ a_2 \succ a_4$. Similarly to the $m = 5$ case, a_1 can never win with a single β and t because it does not perform the best in either feature, nor does any equal or otherwise weighing of the features allow for a_1 to be the winner. Similarly, a_4 can never lose under any single β and t because it doesn't perform the worst in either of the features, nor does any equal or otherwise weighing of the features allow a_4 to be the loser. For a_3 to beat a_5, a_6, a_7 and a_2 in any ranking with a single β and t, the weight on f_2 of the β vector must be higher than the weight on f_1. However, for a_1 to beat a_5, a_5, and a_3 or for a_2 to beat a_4 in the ranking with a single β and t, the weight on f_1 of the β vector must be higher than the weight on f_2, which is a contradiction. Therefore, the ranking $a_1 \succ a_3 \succ a_6 \succ a_5 \succ a_7 \succ a_2 \succ a_4$ is not approval-feasible.

Theorem 11. *For any $m \geq 4$, there exists some construction over m alternatives that violates feasibility under plurality in the full-ranking setting.*

Proof. Consider the feature matrix in Table 1.

Suppose that $n/3 + \xi$ voters in group A have weight vector $\beta_A = (1, 2\epsilon, \epsilon)$, $n/3$ voters in group B have weight vector $\beta_B = (2\epsilon, 1, \epsilon)$, and $n/3 - \xi$ voters in group C have the preferences $\beta_C = (2\epsilon, \epsilon, 1)$ for some small ϵ. In this case, group A's ranking is $a_1 \succ a^* \succ a_2 \succ a_3$, group B's ranking is $a_2 \succ a^* \succ a_1 \succ a_3$, and population C's ranking is $a_3 \succ a^* \succ a_1 \succ a_2$.

The ranking output by plurality is $a_1 \succ a_2 \succ a_3 \succ a^*$. However, this ranking is infeasible because there is no β that produces a ranking such that a^* is the loser given this feature matrix. This example can be extended to $m > 4$ by adding copies of alternatives a_1, a_2, or a_3.

Theorem 12. *For any $m \geq 4$, there exists some construction over m alternatives that violates feasibility under Borda.*

Proof. Consider the feature matrix in Figure 4a. Suppose that there are 3 groups of voters, where $n/3$ voters in group A have weights $\beta_A = (1, 0, 0)$, $n/3$ voters in group B have weights $\beta_B = (0, 1, 0)$, and $n/3$ voters in group C have weights $\beta_C = (0, 0, 1)$. Then, population A's ranking is $a_1 \succ a_2 \succ a_4 \succ a_3$, population B's ranking is $a_3 \succ a_2 \succ a_1 \succ a_4$, and population C's ranking is $a_4 \succ a_2 \succ a_3 \succ a_1$. The Borda rule returns $a_2 \succ a_1 = a_3 = a_4$. However, note that it is impossible to choose a value of β that ranks a_2 first. This means that the Borda count can return an infeasible ranking. This example naturally extends to $m > 4$ by adding alternatives that have values of 0 for all features.

6.2 Additional Empirical Results

Supplemental Methodology In our third set of experiments, we explore the performance of the different aggregation methods when α values are drawn from correlated distributions. Specifically, given a row α_i, each of its feature values (each α_{ij}) will be drawn from either a uniform or a normal distribution. These distributions are varied by two parameters : *inter-* distance (d_{inter}, the distance *between* the distributions) and *intra-* distance (d_{intra}, the *size* of the distributions). For each alternative i, its feature values are then drawn from the distribution $\mathcal{N}(d_{inter} \cdot i + \frac{d_{intra}}{2}, \frac{d_{intra}}{2})$ in the normal case, and $\mathcal{U}(d_{inter} \cdot i, d_{inter} \cdot i + d_{intra})$ in the uniform case. For all simulations in the last experiment, $m = 8, n = 50, k = 5$, $\beta \in \mathcal{U}(0, 1)$, and $\epsilon \in \mathcal{U}(-0.1, 0.1)$. This experimental set-up is further clarified through the visualization of the relationship between inter and intra distance shown in Figure 5, where the periods which uniform distributions are drawn from are marked alongside intra/inter distance, and the corresponding mean, μ, and standard deviation, σ, for a normal distribution with those intra/inter distance values are shown.

Robustness of Approval Voting The robustness of approval voting can be seen in Figure 7. The two primary challenges faced by approval voting are (1) a large quantity of ties and (2) poor accuracy even in the absence of ties. In cases where the threshold is large (e.g., $t = 0.9$) while the unpenalized Kendall Tau distance (where ties are unpenalized) would indicate good accuracy, the penalized Kendall Tau distance is often close to 0.5, indicating that the low Kendall Tau distance results from the produced ranking being exclusively ties. There are cases where the number of ties is relatively lower, such as the $t = 0.5$ case when α is generated normally (Figure 7b); the penalized Kendall Tau distance in this

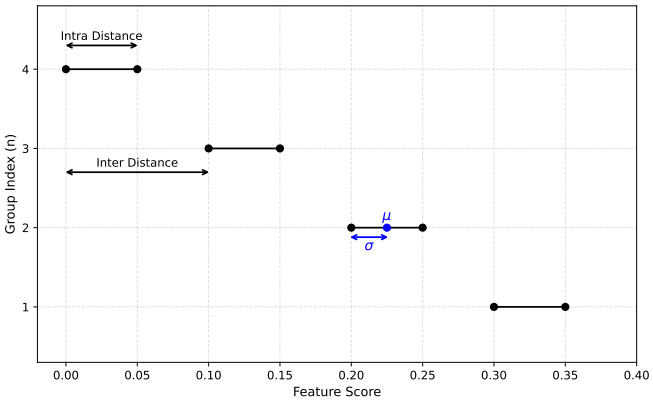
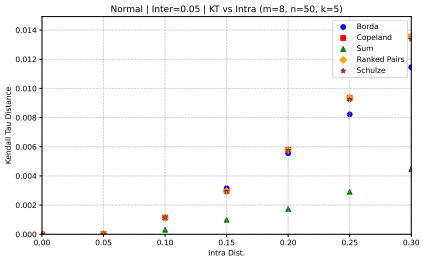
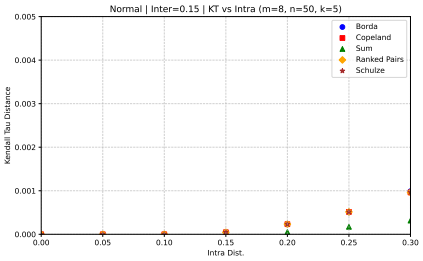


Fig. 5: Illustration of inter/intra distance parameters.



(a) d_{KT} vs. intra distance (inter distance = 0.05)



(b) d_{KT} vs. intra distance (inter distance = 0.15)

Fig. 6: KT distance vs. inter/intra-distribution distance drawn from normal distributions

case can be observed to be relatively close to the unpenalized value. However, under these conditions the unpenalized value performs (at best) on par with Borda and the PMC rules we empirically study. Given that ties, which reduce useful discernible information from the ranking, still exist in these cases there would appear to be no reason to consider approval voting within this context over Borda or PMC rules, let alone over aggregation through utility sums. Approval voting is therefore both difficult to compare with other aggregation methods as a result of the prevalence of ties and also seems to offer no improvement over other aggregation methods. For these reasons it has been excluded from comparisons in the main body of the paper.

Additional Results To further our investigation on how Kendall Tau distance changes as parameter values varied, this relationship is explored across different methods of β generation. While in the main paper Homogeneous Uniform (HU) sampling of β values was shown in graphs, Bimodal Normal (BN), Homogeneous

Sampling			$t = 0.1$		$t = 0.3$		$t = 0.5$		$t = 0.7$		$t = 0.9$	
α	β	ϵ	**Pen.**	**Unpen.**	**Pen.**	**Unpen.**	**Pen.**	**Unpen.**	**Pen.**	**Unpen.**	**Pen.**	**Unpen.**
U	HU	U	0.4085	0.1665	0.3043	0.2510	0.4522	0.0532	0.4997	0.0008	0.5000	0.0000
U	HU	N	0.3999	0.1535	0.2910	0.2350	0.4509	0.0463	0.4998	0.0005	0.5000	0.0000
U	BN	U	0.3824	0.2819	0.3154	0.2496	0.3112	0.1740	0.4262	0.0505	0.4897	0.0040
U	BN	N	0.3700	0.2650	0.3005	0.2294	0.2977	0.1555	0.4223	0.0472	0.4878	0.0031
U	HN	U	0.3453	0.2529	0.3041	0.2387	0.2882	0.1841	0.3426	0.0976	0.3942	0.0338
U	HN	N	0.3333	0.2356	0.2902	0.2206	0.2724	0.1644	0.3333	0.0832	0.3863	0.0307
U	BU	U	0.4077	0.1719	0.3017	0.2482	0.4515	0.0556	0.4998	0.0009	0.5000	0.0000
U	BU	N	0.4004	0.1550	0.2908	0.2350	0.4517	0.0467	0.4998	0.0004	0.5000	0.0000

(a) Approval KT distances for α = Uniform

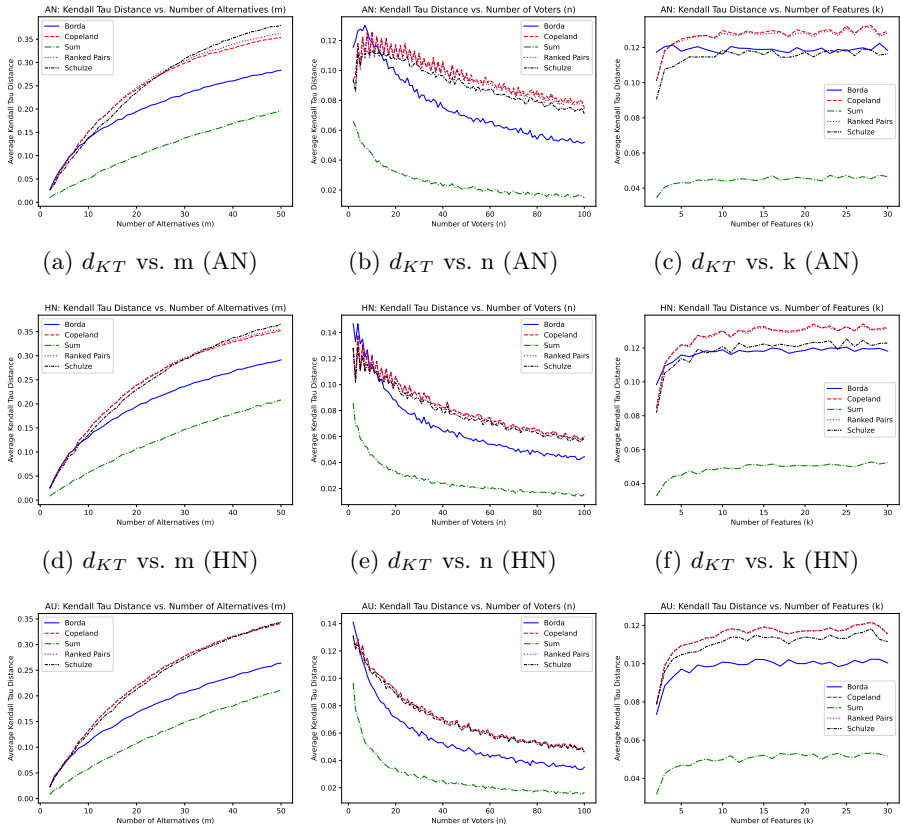
Sampling			$t = 0.1$		$t = 0.3$		$t = 0.5$		$t = 0.7$		$t = 0.9$	
α	β	ϵ	**Pen.**	**Unpen.**	**Pen.**	**Unpen.**	**Pen.**	**Unpen.**	**Pen.**	**Unpen.**	**Pen.**	**Unpen.**
N	HU	U	0.2705	0.1900	0.2514	0.1960	0.2870	0.1382	0.3741	0.0741	0.4402	0.0204
N	HU	N	0.2568	0.1748	0.2354	0.1769	0.2779	0.1240	0.3697	0.0616	0.4385	0.0169
N	BN	U	0.2871	0.2390	0.2585	0.2139	0.2485	0.1711	0.2757	0.1231	0.3448	0.0615
N	BN	N	0.2738	0.2237	0.2418	0.1952	0.2340	0.1529	0.2649	0.1073	0.3352	0.0551
N	HN	U	0.2693	0.2185	0.2501	0.2019	0.2387	0.1835	0.2356	0.1534	0.2894	0.0949
N	HN	N	0.2455	0.1900	0.2339	0.1841	0.2206	0.1636	0.2234	0.1382	0.2752	0.0861
N	BU	U	0.2713	0.1921	0.2488	0.1938	0.2868	0.1384	0.3743	0.0729	0.4406	0.0200
N	BU	N	0.2545	0.1720	0.2330	0.1738	0.2759	0.1213	0.3704	0.0642	0.4388	0.0176

(b) Approval KT distances for α = Normal

Fig. 7: Normalized Kendall tau distances for various parameter combinations for approval voting. Lower is better. U = uniform, N = normal, HU = uniformly drawn from across bounds, BN = normally drawn from a bimodal distribution, HN = normally drawn from same mean, BU = uniformly drawn from two bounds. Penalized Kendall Tau distances are labeled "Pen.", unpenalized are labeled "Unpen."

Normal (HN) and Bimodal Uniform (BU) sampling methods were also applied; these results can be seen in Figure 8. These results exhibit similar trends to the Homogeneous Uniform results which were included in the main body, affirming the validity of the observed trends, even across different methods of β value generation.

In our last set of experiments, we show that when α values are correlated, Borda, PMC rules (Copeland, Ranked Pairs, and Schulze), and Sum all exhibit few pairwise swaps, but Sum continues to perform the best. As seen in Figure 6, as the intra-distribution distance increases, the KT distance worsens. This is to be expected: When distributions overlap more, there is a higher chance of swaps, and as the inter-distribution distance increases, the distributions overlap less. In the extreme case when the intra-distribution distance is much smaller than the inter-distribution distance, swaps never occur. Even with significant overlap, the KT distance remains small across aggregation methods. To support our results in our empirical section on the effects of varying inter and intra values on the Kendall Tau distance of aggregation methods, further simulations were conducted with varying parameter sizes. Graphs of these results are shown below. The following additional parameter combinations were tested: $(m = 20, n = 10, k = 5)$, $(m = 8, n = 50, k = 5)$, and $(m = 20, n = 50, k = 5)$. These results are shown in Figure 9 and Figure 10. It can be observed from these figures that the



(a) d_{KT} vs. m (AN) (b) d_{KT} vs. n (AN) (c) d_{KT} vs. k (AN)

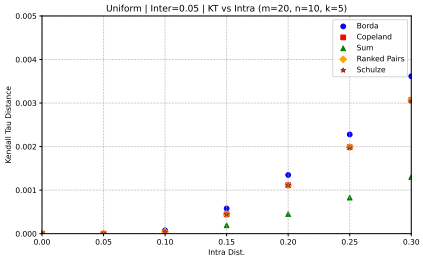
(d) d_{KT} vs. m (HN) (e) d_{KT} vs. n (HN) (f) d_{KT} vs. k (HN)

(g) d_{KT} vs. m (AU) (h) d_{KT} vs. n (AU) (i) d_{KT} vs. k (AU)

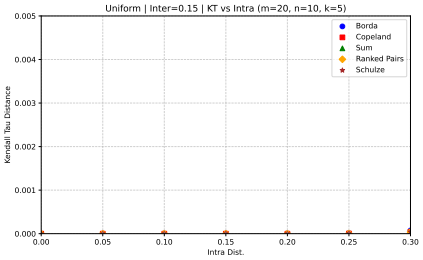
Fig. 8: KT distance vs. parameter size for additional β generation methods (AN, HN, AU). HU results are shown in Figure 3.

trends mentioned in our main paper are further affirmed. Particularly, as d_{inter} increases the Kendall Tau distance seems to decrease. Similarly, increasing d_{intra} seems to increase the Kendall Tau distance. Finally, increasing the number of alternatives also increases the Kendall Tau distance. This is in line with our prior experiments showing that increasing m also increased the Kendall Tau distance of output rankings.

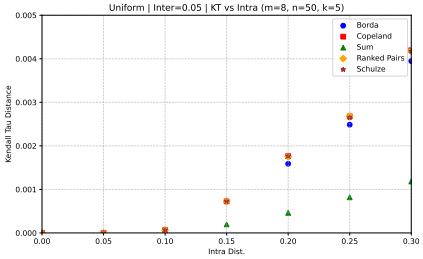
All code for the simulations in this paper is available at the following repository: `https://anonymous.4open.science/r/VD-LSC777/`



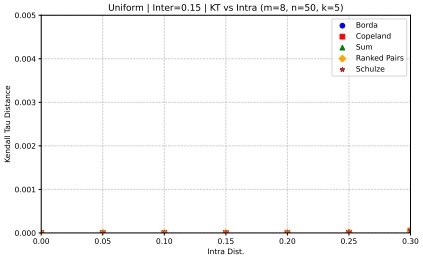
(a) ($m = 20$, $n = 10$, $k = 5$), inter distance = 0.05



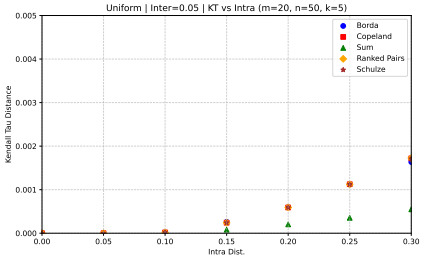
(b) ($m = 20$, $n = 10$, $k = 5$), inter distance = 0.15



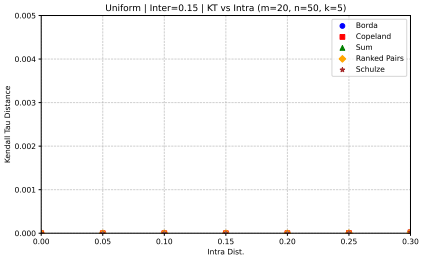
(c) ($m = 8$, $n = 50$, $k = 5$), inter distance = 0.05



(d) ($m = 8$, $n = 50$, $k = 5$), inter distance = 0.15

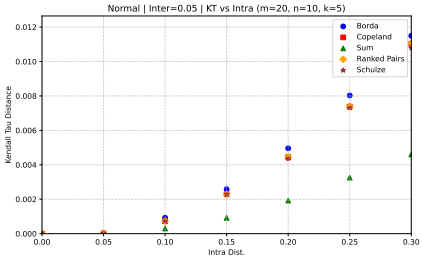


(e) ($m = 20$, $n = 50$, $k = 5$), inter distance = 0.05

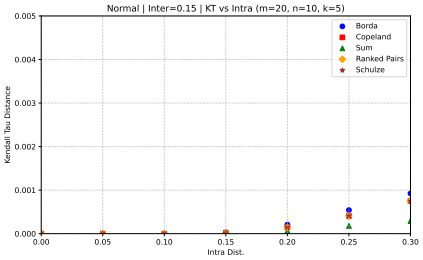


(f) ($m = 20$, $n = 50$, $k = 5$), inter distance = 0.15

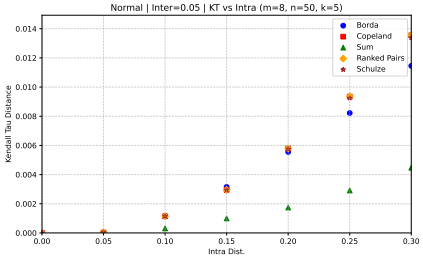
Fig. 9: KT distance vs. inter-distribution distance drawn from uniform distributions.



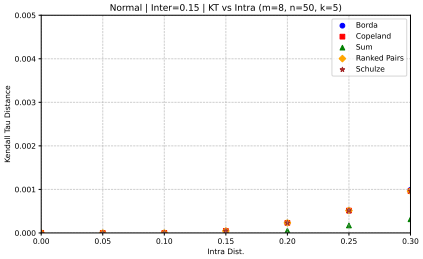
(a) ($m = 20$, $n = 10$, $k = 5$), inter distance = 0.05



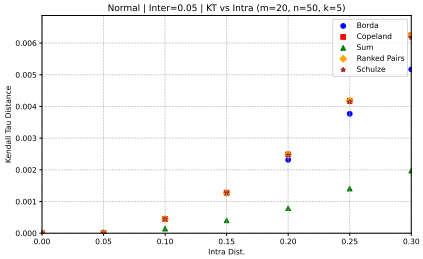
(b) ($m = 20$, $n = 10$, $k = 5$), inter distance = 0.15



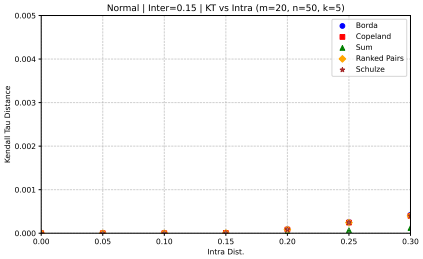
(c) ($m = 8$, $n = 50$, $k = 5$), inter distance = 0.05



(d) ($m = 8$, $n = 50$, $k = 5$), inter distance = 0.15



(e) ($m = 20$, $n = 50$, $k = 5$), inter distance = 0.05



(f) ($m = 20$, $n = 50$, $k = 5$), inter distance = 0.15

Fig. 10: KT distance vs. inter-distribution distance drawn from normal distributions.