AXIOM SATISFIABILITY OF LINEAR REWARDS IN ALIGNMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning from human preference data is the dominant route to aligning language models with human values. In the linear social choice model, where rewards are linear in a fixed feature representation of prompt-response pairs, Ge et al. (2024) show that fitting such a reward by minimizing any non-decreasing convex loss, including Bradley-Terry-Luce (BTL), fails Pareto Optimality (PO) and Pairwise Majority Consistency (PMC). Moreover, no rule that reads only the majority relation can satisfy PO once the output is required to be linearly induced. We ask what it costs to enforce these axioms anyway. To this end, we relax the linear model to allow per-candidate additive slack. We compute the relaxed linear reward with the smallest total slack that satisfies the axioms with a margin η, the minimum required difference between two reward values. Our solution satisfies the axioms under no assumptions about the voters or how comparisons were collected. We bound the optimal total slack by $O(1)$ when η is at most $O(\frac{1}{m^2})$ for m candidates. Furthermore, we exhibit an instance that forces this bound, concluding that the rate is tight up to constants. In practice, the number of candidates far exceeds the feature dimension, and only a linear reward can be evaluated on unseen responses. We therefore introduce a new method that charges the linear part for each comparison it gets wrong. It simultaneously minimizes the total slack and the number of violations, with a parameter λ that trades off between them. We show that the total slack is monotone but saturating in λ: raising it reduces the violations of the deployed linear reward and increases the slack, yet the slack stays below $(\eta+\Delta\sqrt{d})\lfloor m^2/4 \rfloor$, where Δ and d are the diameter and dimension of the features, respectively. Experiments on both synthetic and real-life preference data with per-annotator labels corroborate our theoretical result and show that the linear reward output by our method beats linear BTL.

1 INTRODUCTION

AI Alignment is the problem of ensuring that language models act in ways consistent with human intentions, preferences, and normative constraints (Halpern et al., 2026). Reward models trained from human feedback are the interface through which aggregated human preferences enter a language model (Christiano et al., 2017; Stiennon et al., 2020; Ouyang et al., 2022; Bai et al., 2022). Such models are typically fit by maximum likelihood under a Bradley-Terry model (Bradley & Terry, 1952) on pairwise comparisons, which produces a single scalar reward from the judgments of many annotators. When those judgments disagree, the reward model is performing an act of aggregation. How can we identify the best way to aggregate preferences? Social choice theory (Arrow, 2012; Brandt et al., 2016) has developed a rich axiomatic framework for addressing this question. The axiomatic approach characterizes aggregation rules through precisely defined, normatively meaningful properties. Thus, it is natural to ask which axioms a reward model satisfies. A recent line of work argues that this question should be central to alignment (Conitzer et al., 2024; Dai & Fleisig, 2024).

Given a prompt, an annotator ranks a set of responses. We view each annotator as a voter and each prompt–response pair as a candidate. Comparisons between every two candidates induce binary preference labels, which are then used to learn a reward model. A loss-based rule assigns rewards by minimizing a measure that penalizes disagreements with voters over candidate pairs.Ge et al. (2024) and, subsequently, Hollender & Kraiczy (2026) show, surprisingly, an optimally trained linear reward violates a unanimous preference, known as Pareto optimality (PO), i.e, even when

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

everyone selects candidate a over candidate b, the learned reward of b is higher. It also violates the preference of a majority, known as pairwise majority consistency (PMC), i.e., even when a majority of voters selects candidate a over candidate b, the reward of b is higher. Needless to say, a final ranking representing the human feedback should agree with the unanimous and majority preferences. Thus, in search of an aggregation method, Ge et al. (2024) studied the linear reward models. They gave a linear rule based on Copeland (1951), a well-known rule in social choice theory, that satisfies both of the above constraints. However, a desired ranking need not always be representable by a simple linear function, or even by a polynomial function, of the candidates' feature vectors.

This work asks what it costs to impose the axioms anyway. We relax the linear model of Ge et al. (2024) by a per-candidate *slack*, ask for the smallest slack consistent with the axioms, and study it as a measure of how far the preference profile lies from what the feature geometry can express.

The linear social choice model (Ge et al., 2024). Let $\mathcal{C}$ be the set of m distinct prompt-response pairs, referred to as *candidates*, and $\mathcal{V}$ be the set of n human annotators, the *voters*. Each candidate $c \in \mathcal{C}$ is associated with a distinct feature vector $\boldsymbol{x}_c \in \mathbb{R}^d$. Let $\mathcal{X}$ denote the set of feature vectors of the candidates $\mathcal{C}$. We abuse notation and use $\mathcal{C}$ to denote the set of candidates along with their feature vectors. A parameter vector $\boldsymbol{\theta} \in \mathbb{R}^d$ induces a linear reward function $r_{\boldsymbol{\theta}} : \mathcal{C} \mapsto \mathbb{R}$ defined by $r_{\boldsymbol{\theta}}(c) = \langle \boldsymbol{\theta}, \boldsymbol{x}_c \rangle$ for each candidate $c \in \mathcal{C}$. We are interested in how any such $r_{\boldsymbol{\theta}}$ ranks the candidates. For a ranking $\succ_{\boldsymbol{\theta}}$ over the candidates, we say that $\boldsymbol{\theta}$ *induces* $\succ_{\boldsymbol{\theta}}$ if

$$a \succ_{\boldsymbol{\theta}} b \quad \iff \quad r_{\boldsymbol{\theta}}(a) \geq r_{\boldsymbol{\theta}}(b). \tag{1}$$

We say $\pi = (\succ_1, \succ_2, \ldots, \succ_n)$ a *preference profile*, where ranking $\succ_i$ of $\mathcal{C}$ is given by voter $i \in \mathcal{V}$. A *linear aggregation rule* takes as input a profile π and outputs a parameter vector $\boldsymbol{\theta}^*$. The goal is to design a rule that finds a $\boldsymbol{\theta}^*$ whose induced ranking respects some desirable properties (the *axioms*) with respect to π. In alignment, the model size (e.g., the feature dimension d) is much smaller than the prompt/response space (e.g., the number of candidates m). In this case ($d < m$), some ranking $\succ$ might not be realized by any $\boldsymbol{\theta}$, and hence, the linear aggregation rule might not output a $\boldsymbol{\theta}^*$ satisfying the axioms. The example below shows this.

Example 1. *Consider three candidates a, b, c in $\mathbb{R}$ with $x_a = 0$, $x_b = 1$, and $x_c = 2$. All three lie on a line, so r_θ is monotone along the line. If $\theta > 0$, then θ induces the ordering $c \succ b \succ a$, and if $\theta < 0$, then it induces the ordering $a \succ b \succ c$. Out of the 3! rankings of $\{a, b, c\}$, only two are realizable by a linear reward. The remaining four cannot be induced by any θ.*

A linear reward orders the candidates by projecting onto a single direction. So the rankings available to it are constrained by the features $\{\boldsymbol{x}_c\}_{c \in \mathcal{C}}$, and a profile aggregated from voters whose disagreements carry no obligation to respect that constraint. Two improvements follow immediately: enrich the feature representation until the desired ranking becomes realizable, or relax the reward model. We take the second approach by allowing small perturbations to the linear rewards.

Our contributions. We relax the linear reward by a per-candidate slack ε_c, yielding

$$r_{\boldsymbol{\theta}, \boldsymbol{\varepsilon}}(c) := \langle \boldsymbol{\theta}, \boldsymbol{x}_c \rangle + \varepsilon_c, \qquad \textit{for } c \in \mathcal{C}. \tag{2}$$

In the linear model, the slack vector $\boldsymbol{\varepsilon} = \mathbf{0}$; allowing $\boldsymbol{\varepsilon} \neq \mathbf{0}$ admits a deviation from linearity, and $\|\boldsymbol{\varepsilon}\|_1$ measures the total absolute slack. By allowing slack in linear rewards, we ensure that all rankings of $\mathcal{C}$ are realized. e.g., in Example 1, the ranking $b \succ c \succ a$ isn't realized by linear rewards. But, with the slacks, we can realize it with, say, $\theta = 1$ and $\boldsymbol{\varepsilon} = (0, 1.001, 0)$.

We want each slack to be bounded and as small as possible, since otherwise, one can realize any ranking of $\mathcal{C}$ by taking $\boldsymbol{\theta} = \mathbf{0}$ and only using $\boldsymbol{\varepsilon}$. Our goal is to measure *how much* slack a given profile forces to satisfy the axioms, i.e., how far the axioms push the reward away from the linear class. Formally, we ask for the slack vector $\boldsymbol{\varepsilon} \in \mathbb{R}^m$ of least ℓ_1-norm under which Pareto Optimality (PO) and Pairwise Majority Consistency (PMC) hold. We make the following contributions:

(i) We formulate minimizing total slack as a linear program (LP), explicitly enforcing PO and PMC as margin constraints. Condorcet cycles are handled by removing a feedback arc set that is disjoint from the unanimous pairs, thereby preserving as many majority constraints as the cycles allow. Apart from PO and PMC, the reward $r_{\boldsymbol{\theta}, \boldsymbol{\varepsilon}}$ returned by the program satisfies Condorcet consistency, and hence, majority consistency, under no assumptions on the profile and feature vectors.

(ii) In practice, the number of candidates m is much larger than the parameter dimension d. So, only the linear model $\boldsymbol{\theta} \in \mathbb{R}^d$ is deployable, not the slack $\boldsymbol{\varepsilon} \in \mathbb{R}^m$. However, $\boldsymbol{\theta}$ returned from the

LP formulated above may exhibit both PO and Majority violations. We introduce a second program that jointly minimizes the total number of violations by $r_{\boldsymbol{\theta}}$ and the total slack, with a parameter λ controlling their relative importance. The trade-off it governs is monotone: increasing λ reduces the violations of $r_{\boldsymbol{\theta}}$ and increases $\|\boldsymbol{\varepsilon}\|_1$. However, $\|\boldsymbol{\varepsilon}\|_1$ saturates at $(\eta+\Delta\sqrt{d})\lfloor m^2/4 \rfloor$ for $\lambda \geq \Delta\sqrt{d}/\eta$, where Δ is the ℓ_2 diameter of the candidate features and η is the margin, i.e., the minimum required difference between two reward values. We construct instances, where the total slack $\|\boldsymbol{\varepsilon}\|_1$ is at least $\frac{\eta}{2}\lfloor m^2/4 \rfloor$, and show that the upper bound is tight on a class of (infinitely many) instances.

(iii)We characterize when linear rewards can model the preferences. We show that a positive slack does not always imply that the axioms are not realizable with a linear reward $r_{\boldsymbol{\theta}}$: for η not larger than the highest attainable margin by $r_{\boldsymbol{\theta}}$, a zero-slack certifies realizability. Otherwise, the certificate is the non-existence of a zero net feature-weighted flow on the majority graph.

(iv) Finally, we compare PO and Majority violations of our deployed linear reward against those of a linear model obtained by minimizing the Bradley-Terry loss on both synthetically generated instances and real-life *Machine Deliberation* dataset. We observe a significant reduction in total axiom violations compared to linear BTL, e.g., on the Deliberation dataset, we see a 16.12% reduction.

Related Work. Ge et al. (2024) introduces the linear social choice model and shows that every loss-based rule with a nondecreasing weakly convex or strictly convex loss (e.g., BTL) fails both PO and PMC, and that the failure of PO extends to every $C1$ rule. The proof of the former contains an incorrect value of $\boldsymbol{\theta}$, which can be corrected to establish the statement. They answer this with Leximax Copeland subject to PO (LCPO), a rule that fills the ranking positions in Copeland order, admitting a candidate when a linear reward can place it there consistently with Pareto dominance, and prove it satisfies PO and PMC. Hollender & Kraiczy (2026) shows that enlarging the reward class does not help, since loss-based rules over bounded-degree polynomial rewards still fail PO and PMC. PO is satisfied under balanced data coverage where comparison directions are uniformly distributed over a sphere, though PMC still fails. Both works assume that the axioms can be met: LCPO requires a PO-respecting feasible ranking to exist, which Ge et al. (2024) secures by assuming every voter's ranking is itself induced by a parameter vector, and Hollender & Kraiczy (2026) requires control over how comparisons are sampled. Instead, we take the features and comparisons as given and ask how far the profile lies from what the geometry can express: the slack is zero exactly when a linear reward satisfies the axioms unaided, and, otherwise, it measures the shortfall. Xiao et al. (2025) weight the BTL loss by majority indicators, implement the Copeland rule, thereby satisfying PO and PMC. Their reward class, however, is unrestricted. Dong et al. (2026) study this trade-off between representation and rationalizability of axioms. Procaccia et al. (2025) shows that weighted MLE is robust to cloning candidates, preserving other desirable properties of MLE, while Siththaranjan et al. (2024) showed that MLE using regularized BTL loss outputs Borda (de Borda, 1784) ranking. Halpern et al. (2026) provides a comprehensive survey of recent related works.

2 Social Choice Axioms

In social choice theory, an axiom constrains an aggregation rule, which maps a profile of individual rankings to a collective ranking. In this work, we examine several axioms from social choice theory to evaluate the reasonableness of our aggregation rule, which is done using a reward function. The rule takes the profile π and returns a reward r, and the collective ranking is the one induced by ranking candidates in decreasing reward. We therefore state each axiom as a condition on r.

As in Ge et al. (2024), we primarily consider $C1$ rules, which only need to know, for each pair, which candidate the majority prefers. These axioms constrain r only through pairwise comparisons on which the voters agree to some degree, so we record two sets of pairs: the *unanimous* pairs $\mathcal{U}$, on which every voter ranks a above b, and the *strict majority* pairs $\mathcal{M}$, on which more than half do. To this end, let $\pi = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ be a preference profile, where σ_i is a ranking of $\mathcal{C}$ given by voter $i \in \mathcal{V}$. For $a, b \in \mathcal{C}$, write $w_{a\succ b}(\pi) = \frac{1}{n}\big|\{i \in \mathcal{V} : a \succ_i b\}\big|$ for the fraction of voters ranking a above b, so that $w_{a\succ b} + w_{b\succ a} = 1$. Define

$$\mathcal{U} := \{(a,b) \in \mathcal{C} \times \mathcal{C} : w_{a\succ b}(\pi) = 1\} \qquad \text{(unanimous pairs)}, \tag{3}$$

$$\mathcal{M} := \{(a,b) \in \mathcal{C} \times \mathcal{C} : w_{a\succ b}(\pi) > \tfrac{1}{2}\} \qquad \text{(strict majority pairs)}. \tag{4}$$

The set $\mathcal{M}$ defines a (weighted) directed graph $G_{\mathcal{M}}$, called the weighted majority graph, with vertices $\mathcal{V}$, edges $\mathcal{M}$ with weights $w_{a\succ b}(\pi)$. For an odd number of voters, the majority graph is com-

plete, i.e., for all $a, b \in \mathcal{C}$, either $(a, b) \in \mathcal{M}$ or $(b, a) \in \mathcal{M}$ (called a tournament (Laslier, 1997)).

Definition 1 (PO and PMC)**.** *On a preference profile π, a reward function $r : \mathcal{C} \to \mathbb{R}$ satisfies (i) Pareto Optimality (PO) if $r(a) > r(b)$ for every $(a, b) \in \mathcal{U}$, and (ii) Pairwise Majority Consistency (PMC) if the graph $G_{\mathcal{M}}$ has no cycle and $r(a) > r(b)$ for every $(a, b) \in \mathcal{M}$.*

Observe that if there is a cycle, no ranking can satisfy PMC for all pairs, and the axiom is vacuous. Moreover, a PMC ranking for a profile need not exist, but when it does, it is unique. A cycle in the graph $G_{\mathcal{M}}$ is called a *Condorcet cycle*.A weaker version of PMC is *Condorcet consistency*: if a candidate c^* is preferred by a majority of voters over every other candidate, then c^* must be ranked first. Equivalently, if $w_{c^* \succ c} > \frac{1}{2}$ for each $c \in \mathcal{C} \setminus \{c^*\}$, then $r(c^*) > r(c)$. The candidate c^* is called the Condorcet winner. A further relaxed version is majority consistency. A rank aggregation rule satisfies *majority consistency (MC)* if, when a candidate a is ranked first by a majority of voters in the input profile, a is ranked first in the output ranking.

Neither PO nor PMC is satisfied with the rules ordinarily used. Every loss-based rule with a convex loss fails, including BTL. Moreover, no $C1$ rule can satisfy PO once the output is required to be induced by a linear reward. Leximax Copeland subject to PO (LCPO) works when the preference profile is linearly realizable, and LCPO satisfies both PO and PMC (Ge et al., 2024). A profile π is *linearly realizable* if there exists a $\boldsymbol{\theta} \in \mathbb{R}^d$ s.t $r_{\boldsymbol{\theta}}(a) > r_{\boldsymbol{\theta}}(b)$ for all $(a, b) \in \mathcal{M}$ when the PMC ranking exist, otherwise the condition holds for each $(a, b) \in \mathcal{U}$. A linear reward may not satisfy these axioms. As Example 1 shows, the rankings a linear function can induce are limited by the geometry of the feature vectors, and the axioms may demand one that lies outside that set – for any candidate set with $m > d + 1$, most rankings are unavailable. We therefore relax the model with per-candidate slack equation 2 and ask for the smallest slack under which the axioms hold.

3 A Solution via Linear Programming

The axioms of Section 2 translate directly into margin constraints. For a margin $\eta > 0$, PO asks for $\boldsymbol{\theta}^\top(\boldsymbol{x}_a - \boldsymbol{x}_b) + (\varepsilon_a - \varepsilon_b) \geq \eta$ on every unanimous pair, and PMC asks for the same on every strict majority pair. The latter cannot always be imposed: when $\mathcal{M}$ contains a Condorcet cycle, no reward orders all of it consistently (Liu et al., 2026), and the natural remedy is to drop the majority constraints entirely in that case, enforcing them only when $\mathcal{M}$ is acyclic. Writing $\mathbb{1}_{\text{PMC}}$ for the indicator of acyclicity, this gives the constraint family $\mathcal{U} \cup \mathbb{1}_{\text{PMC}}\mathcal{M}$.

The relation $\succ$ of equation 1 is defined with a weak inequality, under which $\boldsymbol{\theta} = \mathbf{0}$ induces every ranking. Thus, we ask for strict separation at a fixed margin $\eta > 0$, which also keeps the feasible set closed. Moreover, without a bound on $\boldsymbol{\theta}$, one could scale $\boldsymbol{\theta} \mapsto \alpha\boldsymbol{\theta}$ with $\alpha \to \infty$ to meet every margin while driving $\boldsymbol{\varepsilon}$ to zero, so we constrain $\boldsymbol{\theta}$ in the box $\|\boldsymbol{\theta}\|_\infty \leq 1$, which fixes the scale and makes $\boldsymbol{\varepsilon}$ a meaningful measure of deviation from linearity.[1]

Minimizing the total slack $\|\boldsymbol{\varepsilon}\|_1$ subject to these constraints is a linear program:

$$
\begin{aligned}
&(\mathsf{LP1}) \qquad \min_{\boldsymbol{\theta}, \boldsymbol{\varepsilon}, \boldsymbol{t}} \ \sum_{a \in \mathcal{C}} t_a \\
&\text{subject to} \quad -t_a \leq \varepsilon_a \leq t_a, \quad t_a \geq 0 \qquad \forall a \in \mathcal{C}, \\
&\qquad\qquad \langle \boldsymbol{\theta}, \boldsymbol{x}_a - \boldsymbol{x}_b \rangle + (\varepsilon_a - \varepsilon_b) \geq \eta \quad \forall (a, b) \in \mathcal{U} \cup \mathbb{1}_{\text{PMC}}\mathcal{M}, \qquad \text{(PO+PMC)} \\
&\qquad\qquad -1 \leq \theta_j \leq 1 \qquad \forall j \in \{1, \ldots, d\}.
\end{aligned}
$$

Only $(\boldsymbol{\theta}, \boldsymbol{\varepsilon})$ is of interest in what follows: the auxiliary variables satisfy $t_a = |\varepsilon_a|$ at any optimum, so they carry no information beyond $\boldsymbol{\varepsilon}$ and serve only to write $\|\boldsymbol{\varepsilon}\|_1$ linearly.

Lemma 1 (LP1 satisfies PO and PMC)**.** *Every feasible solution of* LP1 *induces a reward function $r_{\boldsymbol{\theta}, \boldsymbol{\varepsilon}}$ (equation 2) that satisfies Pareto optimality and pairwise majority consistency, but need not satisfy Condorcet consistency.*

The proof is given in Appendix A. Intuitively, when $\mathcal{M}$ has a cycle, the constraint family collapses to $\mathcal{U}$, and a candidate ranked first by a majority of voters need not be preferred to every other candidate

[1] Hollender & Kraiczy (2026) observe that the Bradley–Terry loss already imposes an implicit budget on $\|\boldsymbol{\theta}\|$, as unbounded scaling drives the loss to infinity.

unanimously. A single cycle anywhere in $\mathcal{M}$ therefore costs every majority constraint. To recover them by deleting some pairwise constraints, we need the following definition.

Definition 2. *Let $G = (V, E)$ be a directed graph. A* feedback arc set (FAS) *of G is a set $\mathcal{F} \subseteq E$ such that $(V, E \setminus \mathcal{F})$ acyclic; it is* minimal *if no proper subset of it is a feedback arc set.*

The (weighted) majority graph $G_{\mathcal{M}}$ defined by $\mathcal{M}$ is a directed graph on $\mathcal{C}$. We take $\mathcal{F}$ to be a feedback arc set of $G_{\mathcal{M}}$ such that $\mathcal{F} \cap \mathcal{U} = \emptyset$, and set

$$\mathcal{P} \;:=\; \mathcal{U} \cup (\mathcal{M} \setminus \mathcal{F}). \tag{5}$$

The omission of the pairs in $\mathcal{F}$ ensures a valid program, where the cardinality of $\mathcal{F}$ affects how many of the majority constraints survive. One can compute a 5-approximation of the minimum FAS in time $O(|\mathcal{C}|^2)$ in (weighted) tournaments by simply ordering the vertices in non-decreasing order of their (weighted) in-degrees (Coppersmith et al., 2010). With this, we obtain a new linear program LP2 by replacing the (PO+PMC) constraint of LP1 with

$$(\mathsf{LP2}): \qquad \langle \boldsymbol{\theta}, \boldsymbol{x}_a - \boldsymbol{x}_b \rangle + (\varepsilon_a - \varepsilon_b) \geq \eta \qquad \forall (a, b) \in \mathcal{P}. \tag{6}$$

PO and PMC are the axioms we set out to enforce, and the program enforces them by construction, each being written as an explicit family of margin constraints. We show in Theorem 1 that Condorcet consistency, consequently majority consistency, follows without further constraints, as a consequence of the structure of $\mathcal{P}$. The proof is in Appendix A.

Theorem 1 (LP2 satisfy all axioms)**.** *Every feasible solution of* LP2 *induces a reward function $r_{\boldsymbol{\theta},\boldsymbol{\varepsilon}}$ (equation 2) that satisfies Pareto optimality, pairwise majority consistency, Condorcet consistency, and majority consistency.*

Proof Sketch. When $G_{\mathcal{M}}$ is acyclic, then $\mathcal{F} = \emptyset$, so $\mathcal{P} = \mathcal{M} \supseteq \mathcal{U}$ and equation 6 enforce PO and PMC outright. When $G_{\mathcal{M}}$ is cyclic, only the pairs of $\mathcal{M} \setminus \mathcal{F}$ are constrained, and the majority constraint holds on those by construction; PO holds as $\mathcal{F} \cap \mathcal{U} = \emptyset$, and thus $\mathcal{U}$ in $\mathcal{P}$. For Condorcet consistency, suppose a candidate a is ranked before candidate b by a majority of voters, for every candidate $b \in \mathcal{C}$. Then a is a source vertex in the majority graph defined by $\mathcal{M}$. Consequently, no arc out of a lies on a directed cycle, and a will be ranked first with the highest reward. Therefore, Condorcet consistency holds and, by definition, satisfies majority consistency as well.

Comparison with LCPO. LCPO (Ge et al., 2024) achieves the same axioms by filling the m positions of the output ranking one at a time. For each position, they test candidates in the order of their Copeland score until a feasible one is found. LCPO solves $O(m^2)$ linear programs in the worst case. In contrast, LP2 is a single linear program whose constraints are indexed by $\mathcal{P}$, which is supported on the comparisons actually observed. Computing $\mathcal{P}$ costs $O(m^2)$ time by ordering candidates on in-degree and deleting the resulting back arcs, after which a single linear program in $d + 2m$ variables and $O(|\mathcal{P}|)$ constraints is solved – as opposed to $O(m^2)$ programs for LCPO, each in d variables and carrying all of $\mathcal{U}$. Moreover, LCPO needs a parameter vector that separates every unanimous pair, since each of its feasibility tests yields all Pareto-dominances. When one is absent, it cannot fill even the first position and returns nothing. Such a vector is obtained by assuming that each voter's ranking is itself induced by a $\boldsymbol{\theta}'$. This assumption is not testable without eliciting complete rankings, and it restricts each voter to $m^{O(d)}$ linearly induced orderings out of $m!$ possible ones. In contrast, LP2 is feasible on every instance, returning a reward that satisfies the axioms together with an optimal slack $\boldsymbol{\varepsilon}^*$. So, we lose nothing in instances where LCPO applies, but we have a measurement available when it does not.

4 Moving Closer to Linear Rewards

In practice, we may have $m \gg d$ since the number of prompt-response pairs can be large. Thus, it might not be practical to release both $\boldsymbol{\theta}$ and $\boldsymbol{\varepsilon}$ for model evaluation. We therefore ask whether we can get a "good" ranking of the candidates based on the linear reward using only using the parameter vector $\boldsymbol{\theta}$ obtained from LP2? Let the ranking of candidates induced by the linear reward $r_{\boldsymbol{\theta}}$ be denoted by $\succ_{\boldsymbol{\theta}}$. Then, we want to find a $\boldsymbol{\theta}$ such that the number of unanimous and strict majority pairs (as defined in Equations (3) and (4)) that are inverted in $\succ_{\boldsymbol{\theta}}$ is minimized. Each inverted pair is either a PO violation or violates a majority pair that is not PO (majority violation), indicating a violated constraint in the LP. Since we have no assumption on the preference profile being linearly realizable, we may have some violations for every $\boldsymbol{\theta}$ (as in Example 1). Formally, we ask:

> Given a profile π and features $\mathcal{X}$, can we obtain a $\boldsymbol{\theta} \in \mathbb{R}^d$ efficiently such that the number of PO and Majority violations is minimized?

Given a preference profile π, let $\succ^*$ be the optimal one that minimizes the number of violations with respect to the PO and majority pairs in $\mathcal{P}$. A standard measure of distance between two rankings is the *Kendall-Tau (KT) distance*[2]. Given two permutations $\succ^*$ and $\succ$ of $[m]$, the KT distance $\mathrm{KT}(\succ^*, \succ)$ is defined as the number of pairs (i, j) such that $i \succ^* j$ and $j \succ i$ for $1 \leq i < j \leq m$. Since $\mathcal{P}$ can be a partial ordering, we define *KT distance of $\succ$ from $\mathcal{P}$* to be the number of pairs $(i, j) \in \mathcal{P}$ such that $j \succ i$. Let $(\boldsymbol{\theta}^*, \boldsymbol{\varepsilon}^*)$ denote an optimal solution of LP2. We first show that $\succ_{\boldsymbol{\theta}^*}$ does not minimize the KT distance from a ranking that orders the pairs according to $\mathcal{P}$.

Example 2. *Let $\mathcal{C} = \{c_1, c_2, c_3, c_4\}$, with 1-D features $x_{c_1} = -3, x_{c_2} = -2, x_{c_3} = -1$, and $x_{c_4} = 60$. Suppose that there is only one voter in the profile π with preference $\succ = (c_3 \succ c_2 \succ c_4 \succ c_1)$. Then, the set of unanimous pairs $\mathcal{U} = \{(c_3, c_2), (c_3, c_4), (c_3, c_1), (c_2, c_4), (c_2, c_1), (c_4, c_1)\}$. The majority graph constructed from the preference $\succ$ contains no cycles. So $\mathcal{F} = \emptyset$ and $\mathcal{P} = \mathcal{U}$ in LP2. We run LP2 with $\eta = 1$ and obtain $\|\boldsymbol{\varepsilon}^*\|_1 = 3.032$ and $\theta^* = -0.01612$. Now, for any $\boldsymbol{\theta} < 0$, the ranking according to $r_{\boldsymbol{\theta}}$ is $\succ' = c_1 \succ c_2 \succ c_3 \succ c_4$. Therefore, $\mathsf{KT}(\succ', \succ) = 4$ for θ^*. Whereas, if we had $\boldsymbol{\theta} > 0$, then the ranking according to $r_{\boldsymbol{\theta}}$ would be $\widehat{\succ} = c_4 \succ c_3 \succ c_2 \succ c_1$ and $\mathsf{KT}(\widehat{\succ}, \succ) = 2$. So, while LP2 does minimize the L_1 norm of $\boldsymbol{\varepsilon}$, it does not give us the $\boldsymbol{\theta}$ which minimizes the KT distance from the desired profile.*

4.1 Kendall-Tau Minimizing Model

We want to find a $\boldsymbol{\theta}$ that minimizes the number of PO and Majority violations. In order to do this, we minimize the KT distance from $\mathcal{P}$. In the objective function, we can first minimize the KT distance, then minimize the ℓ_1-norm of $\boldsymbol{\varepsilon}$. Towards this, we introduce a variable u_{ab} for each pair $(a, b) \in \mathcal{P}$. If $\langle \boldsymbol{\theta}, \boldsymbol{x}_a - \boldsymbol{x}_b \rangle < \eta$, then $u_{ab} = 1$ and it is zero otherwise. That is, the variables u_{ab} for $(a, b) \in \mathcal{P}$ count the number of violated constraints when we only consider $\boldsymbol{\theta}$ to compute the reward. Recall that there are no inversions when the reward is computed using $(\boldsymbol{\theta}, \boldsymbol{\varepsilon})$. Thus, $\sum_{(a,b)\in\mathcal{P}} u_{ab}$ is the total number of violations by $r_{\boldsymbol{\theta}}$. Equivalently, it is the total number of inversions when we obtain a ranking $\succ_{\boldsymbol{\theta}}$ using $r_{\boldsymbol{\theta}}$. Thus, this term is the least KT distance between $\succ_{\boldsymbol{\theta}}$ from $\mathcal{P}$. Therefore, we design the following integer LP.

$$
\begin{aligned}
(\mathsf{ILP}) \qquad & \min_{\boldsymbol{\theta}, \boldsymbol{\varepsilon}, \boldsymbol{t}, \boldsymbol{u}} \quad \sum_{a \in \mathcal{C}} t_a + \lambda \sum_{(a,b)\in\mathcal{P}} u_{ab} \\
\text{subject to} \quad & -t_a \leq \varepsilon_a \leq t_a, \quad t_a \geq 0 && \forall a \in \mathcal{C}, \\
& \langle \boldsymbol{\theta}, \boldsymbol{x}_a - \boldsymbol{x}_b \rangle + (\varepsilon_a - \varepsilon_b) \geq \eta && \forall (a, b) \in \mathcal{P}, \\
& u_{ab} L \geq \eta - \langle \boldsymbol{\theta}, \boldsymbol{x}_a - \boldsymbol{x}_b \rangle, \quad u_{ab} \in \{0, 1\} && \forall (a, b) \in \mathcal{P}, \\
& -1 \leq \theta_j \leq 1 && \forall j \in \{1, \ldots, d\}.
\end{aligned} \tag{7}
$$

We set L to be a large constant with $L \geq \eta + \Delta\sqrt{d}$. When λ is a large enough, e.g., $O(m^2)$, we realize our desired goal of first minimizing the KT distance and then minimizing the ℓ_1-norm of $\boldsymbol{\varepsilon}$. Only $\boldsymbol{\theta}$ is of our interest in this ILP, and the auxiliary variables u_{ab} for each $(a, b) \in \mathcal{P}$ indicate the violations and are not useful for constructing the reward function. As the ILP cannot be efficiently solved, we relax the constraints $u_{ab} \in \{0, 1\}$ by $u_{ab} \geq 0$ to obtain a linear program (LP3). In the LP we do not need the large constant L, so we remove it from Equation (7).

We begin by proving the correctness of LP3. Using Theorem 1 and the fact that the feasible solutions of LP3 are also feasible solutions of LP2, we get the next lemma. All the complete proofs from this section are in Appendix B.

Lemma 2 (LP3 satisfy all axioms)**.** *Every feasible solution of* LP3 *induces a reward function $r_{\boldsymbol{\theta}, \boldsymbol{\varepsilon}}$ (equation 2) that satisfies all three of Pareto optimality, pairwise majority consistency, Condorcet Consistency, and majority consistency.*

Now, we upper bound the total slack of LP3 and provide a corresponding lower bound.

[2]KT distance is well known in social choice literature, and the voting rule that produces a ranking minimizing the KT distance from $\succ_i$ for each voter $i \in \mathcal{V}$ is called the Kemeny rule (Kemeny, 1959).

Theorem 2 (Bounds on total slack)**.** *Let* $\Delta := \max_{(a,b)\in\mathcal{P}} \|\boldsymbol{x}_a - \boldsymbol{x}_b\|_2$*. For every* $\lambda \geq 0$*, the optimal slack* $\boldsymbol{\varepsilon}^*_\lambda$ *of* LP3 *satisfies*

$$\|\boldsymbol{\varepsilon}^*_\lambda\|_1 \leq \min\left\{\eta\lfloor\tfrac{m^2}{4}\rfloor + \lambda\,\eta\,|\mathcal{P}|\,,\ \left(\eta + \Delta\sqrt{d}\right)\lfloor\tfrac{m^2}{4}\rfloor\right\}.$$

For every m, d, η*, there exist features in* $\mathbb{R}^d$ *and a profile, for which* $\|\boldsymbol{\varepsilon}^*_\lambda\|_1 \geq \frac{\eta}{2}\lfloor\frac{m^2}{4}\rfloor$.

The first upper bound $\eta\lfloor m^2/4\rfloor + \lambda\eta|\mathcal{P}|$ is a feature-free ($\boldsymbol{\theta} = \mathbf{0}$) baseline and serves as a loose bound that always holds. Setting the margin $\eta < \min\{\frac{1}{m^2}, \frac{1}{\lambda|\mathcal{P}|}\}$, we can drive $\|\boldsymbol{\varepsilon}^*_\lambda\|_1$ to $O(1)$. Moreover, in all lower bound instances, $\Delta \leq \eta/(2\sqrt{d})$. On these instances, infinitely many of them, $\|\boldsymbol{\varepsilon}^*_\lambda\|_1 \leq \frac{3\eta}{2}\lfloor\frac{m^2}{4}\rfloor$, which matches the lower bound within a factor of 3.

Setting $\lambda = 0$ reduces LP3 to LP2, and our bound gives $\|\boldsymbol{\varepsilon}^*\|_1 \leq \eta\lfloor m^2/4\rfloor$, where $\boldsymbol{\varepsilon}^*$ is an optimal solution of LP2. Thus, the total slack of LP2 is less than that of LP3 in the worst-case, but this comes at a cost of more PO and Majority violations, considering only the linear part of the reward (see Figure 1). As we increase λ until it reaches $\Delta\sqrt{d}/\eta$, the total number of constraint violations decreases, and (upper bound on) total slack increases linearly with rate $\eta|\mathcal{P}|$. Beyond this point, total slack remains capped at $\left(\eta + \Delta\sqrt{d}\right)\lfloor\frac{m^2}{4}\rfloor$ (see Figure 3).

Next, we show the dependence of the objective function of the ILP on λ. Using an upper bound on the total slack for the ILP (similar to Theorem 2), we can compute the required value of λ.

Lemma 3 (Increasing λ decreases violated pairs)**.** *Let* $\boldsymbol{\theta}$ *denote a solution of* ILP *for a given* λ *and* $\succ_{\boldsymbol{\theta}}$ *denote the ranking induced by the reward* $r_{\boldsymbol{\theta}}$*. Then, the KT distance of* $\succ_{\boldsymbol{\theta}}$ *from* $\mathcal{P}$ *is non-increasing in* λ*. After a certain cut-off* λ^**, the number of violated pairs remains the same.*

Due to Equation (7) in the ILP, variable u_{ab} acts as an indicator of a violation by the linear reward for each pair (a, b) in $\mathcal{P}$. We next show that when λ satisfies the cut-off shown in Lemma 3, i.e., $\lambda > \left(\eta + \Delta\sqrt{d}\right)\lfloor\frac{m^2}{4}\rfloor$, then the second term in the objective function is minimized.

Lemma 4 (ILP minimizes KT distance)**.** *Let* $\boldsymbol{\theta}$ *be a solution of* (ILP) *and* $\succ_{\boldsymbol{\theta}}$ *denote the ranking induced by the reward* $r_{\boldsymbol{\theta}}$*. Then, there exists a* λ *such that the KT distance of* $\succ_{\boldsymbol{\theta}}$ *from* $\mathcal{P}$ *is minimized.*

Example 3. *In continuation of Example 2, we see that for* LP3*, if we focus on the Kendall-Tau distance minimization in the objective by setting* $\lambda = 30$*, we obtain* $\boldsymbol{\theta} = 0.0158$ *and* $\|\boldsymbol{\varepsilon}\|_1 = 4.95$*. Therefore, the ranking based on* $r_{\boldsymbol{\theta}}$ *is* $\widehat{\sigma} = c_4 \succ c_3 \succ c_2 \succ c_1$ *with KT distance* 2 *as opposed to* 4 *as in* LP2*. Thus, we get an optimal* $\boldsymbol{\theta}$ *at the cost of increasing the* ℓ_1 *norm of* $\boldsymbol{\varepsilon}$*.*

4.2 Certificate of Linear Realizability

Given a set of candidates $\mathcal{C}$ and their feature vectors $\mathcal{X} = \{\boldsymbol{x}_a\}_{a\in\mathcal{C}}$, recall that a profile π on $\mathcal{C}$ is *linearly realizable* if there exists a $\boldsymbol{\theta} \in \mathbb{R}^d$ such that $\langle\boldsymbol{\theta}, \boldsymbol{x}_a - \boldsymbol{x}_b\rangle > 0$ for all $(a, b) \in \mathcal{P}$. Define the *best attainable margin* by a normalized linear reward

$$\kappa(\pi, \mathcal{X}) \ := \ \max_{\|\boldsymbol{\theta}\|_\infty \leq 1} \ \min_{(a,b)\in\mathcal{P}} \ \left\langle \boldsymbol{\theta},\, \boldsymbol{x}_a - \boldsymbol{x}_b \right\rangle. \tag{8}$$

Proposition 1 (Realizability $\not\Longrightarrow$ zero-slack always)**.** *If* π *is linearly realizable, then* $\kappa(\pi, \mathcal{X}) > 0$ *and* $\boldsymbol{\varepsilon}^*_\lambda = \mathbf{0}$ *if and only if* $\eta \leq \kappa(\pi, \mathcal{X})$ *for every* $\lambda \geq 0$*.*

A non-zero optimal slack $\boldsymbol{\varepsilon}^*_\lambda$, therefore, doesn't certify non-realizability. We give a characterization of linear realizability based on the feature vectors. Collect the feature differences over constrained pairs into the rows of $\boldsymbol{A} := \left[(\boldsymbol{x}_a - \boldsymbol{x}_b)^\top\right]_{(a,b)\in\mathcal{P}}$. A profile is linearly realizable precisely when some $\boldsymbol{\theta}$ has $\boldsymbol{A}\boldsymbol{\theta} > \mathbf{0}$ component-wise. Gordan's theorem (Vandenberghe & Boyd, 2004) states that this fails if and only if the rows of $\boldsymbol{A}$ admit a non-trivial nonnegative combination summing to zero, i.e., for some $\boldsymbol{\mu} \in \mathbb{R}^{|\mathcal{P}|}, \boldsymbol{\mu} \gneq \mathbf{0}$, we have that $\boldsymbol{A}^\top\boldsymbol{\mu} = \mathbf{0}$.

Let μ_{ab} denotes the component of $\boldsymbol{\mu}$ for the pair $(a, b) \in \mathcal{P}$. The weights $\mu_{ab} \geq 0$ have a natural reading as a *flow* in the majority graph with edges $\mathcal{P}$. It is the amount of flow along the edge (a, b). Regrouping the sum by candidates, we get

$$\boldsymbol{A}^\top\boldsymbol{\mu} = \sum_{(a,b)\in\mathcal{P}} \mu_{ab}(\boldsymbol{x}_a - \boldsymbol{x}_b) = \sum_{a\in\mathcal{C}} \nu_a(\boldsymbol{\mu})\,\boldsymbol{x}_a, \ \text{ where } \ \nu_a(\boldsymbol{\mu}) := \sum_{b:(a,b)\in\mathcal{P}} \mu_{ab} - \sum_{b:(b,a)\in\mathcal{P}} \mu_{ba}.$$

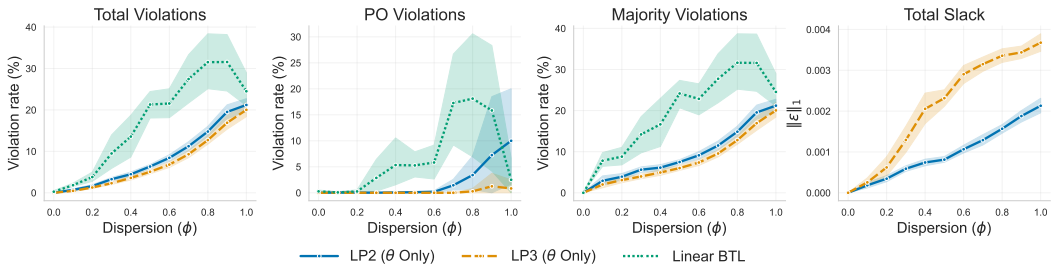

Figure 1: Axiom violations and candidate-level slack as a function of normalized Mallows dispersion ϕ. Curves show mean violation rates across trials, with shaded regions denoting 95% confidence intervals. Panels report (a) Total violations, (b) PO violations, (c) Majority violations, and (d) Total $\|\boldsymbol{\varepsilon}\|_1$ for LP2 and LP3. Experiments use 11 voters, 50 candidates, 25-dimensional features and 20 seeds.

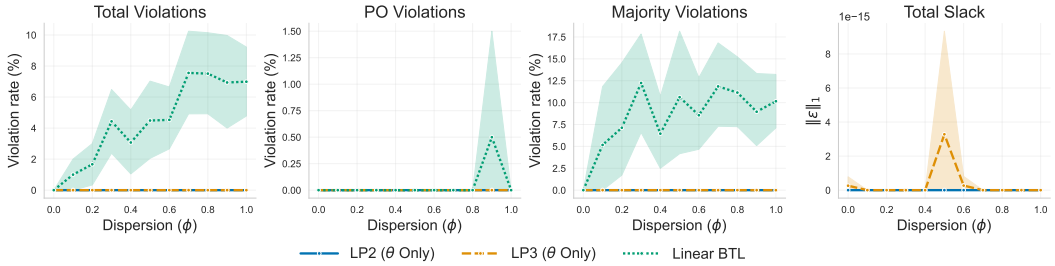

Figure 2: Axiom violations and candidate-level slack as a function of normalized Mallows dispersion ϕ. Curves show mean violation rates across trials, with shaded regions denoting 95% confidence intervals. Panels report (a) Total violations, (b) PO violations, (c) Majority violations, and (d) The $\|\boldsymbol{\varepsilon}\|_1$ for LP2 and LP3. Experiments use 3 voters, 6 candidates, 5-dimensional feature embeddings and 20 seeds.

The *net flow* at candidate a is $\nu_a(\boldsymbol{\mu})$. The total flow $\sum_{a\in\mathcal{C}} \nu_a(\boldsymbol{\mu}) = 0$ since each pair $(a,b) \in \mathcal{P}$ contributes $+\mu_{ab}$ to $\nu_a(\boldsymbol{\mu})$. A non-zero $\boldsymbol{\mu}$ with $\sum_a \nu_a(\boldsymbol{\mu})\boldsymbol{x}_a = \mathbf{0}$ is called *feature-balanced*. Therefore, either $\boldsymbol{A\theta} > \mathbf{0}$ component-wise, or $\boldsymbol{\mu}$ is feature balanced.

Proposition 2 (Certificate of realizability)**.** *A profile π is linearly realizable if and only if no non-zero feature balanced flow exists, i.e., $\nexists\, \boldsymbol{\mu} \gneq \mathbf{0}$, with $\sum_{a\in\mathcal{C}} \nu_a(\boldsymbol{\mu})\boldsymbol{x}_a = \mathbf{0}$.*

We now record when a feature-balanced flow exists. To this end, we say the feature vectors $\mathcal{X}$ are in *general position* if every subset $\mathcal{T}$ of at most $d+1$ of them is affinely independent, that is, there exist coefficients $(\zeta_a)_{a\in\mathcal{T}}$, not all zero, with $\sum_{a\in\mathcal{T}} \zeta_a\, \boldsymbol{x}_a = \mathbf{0}$ and $\sum_{a\in\mathcal{T}} \zeta_a = 0$.

Corollary 1. *If $\mathcal{X} \subset \mathbb{R}^d$ is in general position and $m \leq d+1$, then every profile is linearly realizable.*

Figure 2 shows linear realizability when $m = d+1$, as there are no violations of LP2 and LP3.

5 Experiments

We evaluate LP2, LP3, and Linear BTL (Ge et al., 2024) on controlled synthetic and real-world preference datasets. Violations are evaluated using the learned linear reward parameter $\boldsymbol{\theta}$, without applying the candidate-specific slack $\boldsymbol{\varepsilon}$, thereby isolating the agreement between the linear representation and the observed preference structure. We report *PO Violations* and *Majority Violations* as percentages of the corresponding unanimous and majority relations in $\mathcal{P}$, and *Total Violations* as the percentage of all relations in $\mathcal{P}$ reversed by $\boldsymbol{\theta}$.

Synthetic Data. We study the effect of increasing preference heterogeneity under controlled feature geometry. Candidate features are sampled from $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Preferences are generated according to the normalized Mallows model (Boehmer et al., 2021) with $\phi \in \{0, 0.1, \ldots, 1.0\}$, where larger ϕ corresponds to greater dispersion from the reference ranking. The reference ranking is generated using a linear utility of the candidate feature vector and a $\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. LP2 and LP3 use margin $\eta = 10^{-5}$, while $\lambda = 10$ for LP3. Figure 1 reports the violations (PO, Majority, total), as well as the ℓ_1 norm of the slack for LP2 and LP3.

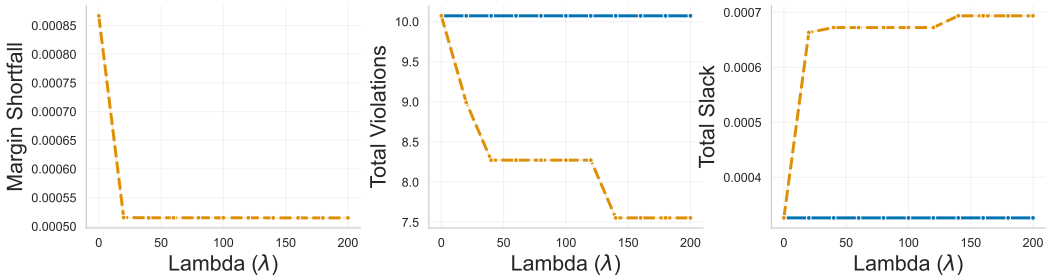

Figure 3: Total violations, total slack, and margin shortfall as a function of λ. The blue line represents LP2, and the orange LP3. Panels report (a) Margin shortfall ($\sum_{(a,b)\in\mathcal{P}} u_{ab}$), (b) Total violations (%), (c) The $\|\boldsymbol{\varepsilon}\|_1$. Experiment uses 3 voters, 25 candidates, and 10-dimensional feature.

Table 1: Machine Deliberation dataset. $K = 500$ questions, $m = 2000, d = 768, \eta = 10^{-5}$.

Method	Total Viol. (%)	PO Viol. (%)	Majority Viol. (%)	$\|\boldsymbol{\varepsilon}\|_1$	Time (s)
LP2	1.59 ± 1.16	0.65 ± 0.63	1.91 ± 1.37	$\mathbf{0.00207 \pm 0.001}$	$\mathbf{308.96}$
LP3	$\mathbf{1.39 \pm 0.98}$	$\mathbf{0.44 \pm 0.35}$	$\mathbf{1.72 \pm 1.21}$	0.00249 ± 0.00123	722.08
Linear BTL	17.51 ± 0.56	2.24 ± 0.53	22.71 ± 0.62	–	2979.11

The results show a clear separation between the methods as preference dispersion increases. LP3 consistently exhibits the lowest observed violations (total, PO, majority) across the dispersion range. LP2 incurs the smallest $\|\boldsymbol{\varepsilon}\|_1$, reflecting its direct minimization of the total slacks. Linear BTL exhibits substantially larger violation rates, with the difference becoming more pronounced as the preference profile becomes more dispersed. These results also illustrate the intended tradeoff introduced by LP3 between candidate-wise slack and pairwise inversions, as slack is more in LP3 than LP2 while all violation rates are less for LP3.

Effect of λ. We perform a study on the effects of the parameter λ in LP3. Preference profile, features, and reference ranking are generated as before. We keep the dispersion parameter $\phi = 0.5$ and vary $\lambda \in \{0, 20, 40, \ldots, 200\}$. Figure 3 reports the total violations and slack of LP2 and LP3, as well as the margin shortfall $\sum_{(a,b)\in\mathcal{P}} u_{ab}$ of LP3. We see that the margin shortfall decreases monotonically as λ increases. The total violations and the total slack also show desired behaviors as we increase the λ, but they are not exactly monotonous as we relax ILP to obtain LP3.

Real-World Preference Data. We evaluate LP2, LP3, and Linear BTL on *Habermas Machine Deliberation* dataset (Tessler et al., 2024). It contains 2,710 separate elections, each consisting of five voters ranking four LLM-generated statements for a deliberation topic. For each trial, we randomly sample K elections and combine their preference DAGs as the disjoint union $P = \bigsqcup_{i=1}^{K} P_i$, yielding a disconnected DAG over $4K$ uniquely identified candidates. Each question-response pair is embedded using `all-mpnet-base-v2` (Song et al., 2020), yielding 768-dimensional representations. Table 1 reports mean violations (total, PO, and majority), total slack for LP2 and LP3, and optimization time. LP3 achieves the lowest violation in all 3 categories, while LP2 has the lowest optimization time and $\|\boldsymbol{\varepsilon}\|_1$. Linear BTL exhibits substantially higher violation rates and considerably larger optimization time. Additional results (varying m, d, n) are in Appendix C.

6 Discussion and Future Work

In this paper, we investigated how to satisfy desirable social choice axioms in linear reward models for AI alignment. We proposed an implementable algorithm that learns a linear reward model, inducing a ranking with fewer violations of PO and PMC. Our algorithm accommodates both synthetic and real-world preference data, including partial preferences and Condorcet cycles, and outperforms the linear BTL model. These findings open several directions for future research: Can our approach be extended to implementable non-linear reward models, such as convex models? What structural properties of the underlying LPs can enable better reward modeling and help formalize, analyze, and enhance LLM behavior? What would be the most efficient approximation of LP to implement it efficiently for large m?

AI USE STATEMENT

We have not used generative AI tools to generate synthetic data sets, help develop theoretical models or conceptual frameworks, assist with writing proofs, formulate mathematical claims, provide critical ingredients for proving mathematical claims, propose or refine hypotheses, design or provide feedback on research methodology or experiments, implement methods, assist with translation, clean and reformat dataset, support qualitative and thematic data analysis, interpret results. We have used generative AI tools to create or edit software code, summarize or analyze existing literature, source or search for information, and edit a paragraph to improve readability. We have reviewed all AI-assisted work. LLM-generated code was verified and tested for correctness. LLM-generated texts are also thoroughly checked and verified for correctness by every author. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work does not have any ethical concerns.

REPRODUCIBILITY STATEMENT

The whole work of this paper is reproducible. An anonymous downloadable source code is submitted as supplementary materials. All the theoretical claims in our paper have a complete proof in Appendix A and Appendix B. All the experimental setup and the dataset description are mentioned in Section 5 and Appendix C

REFERENCES

Akshay Agrawal, Robin Verschueren, Steven Diamond, and Stephen Boyd. A rewriting system for convex optimization problems. *Journal of Control and Decision*, 5(1):42–60, 2018.

Kenneth J Arrow. *Social choice and individual values*, volume 12. Yale university press, 2012.

Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*, 2022.

Niclas Boehmer, Robert Bredereck, Piotr Faliszewski, Rolf Niedermeier, and Stanisław Szufa. Putting a compass on the map of elections. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI 2021)*, pp. 59–65, 2021. doi: 10.24963/ijcai.2021/9. URL https://ijcai.org.

Ralph Allan Bradley and Milton E Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.

Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D Procaccia. *Handbook of computational social choice*, volume 1. Cambridge University Press New York, NY, USA, 2016.

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

Vincent Conitzer, Rachel Freedman, Jobst Heitzig, Wesley H Holliday, Bob M Jacobs, Nathan Lambert, Milan Mossé, Eric Pacuit, Stuart Russell, Hailey Schoelkopf, et al. Social choice should guide ai alignment in dealing with diverse human feedback. *arXiv preprint arXiv:2404.10271*, 2024.

A. H. Copeland. A 'reasonable' social welfare function. Notes from a seminar on applications of mathematics to the social sciences, 1951.

Don Coppersmith, Lisa K. Fleischer, and Atri Rurda. Ordering by weighted number of wins gives a good ranking for weighted tournaments. *ACM Transactions on Algorithms (TALG)*, 6(3):55:1–55:13, July 2010. ISSN 1549-6325. doi: 10.1145/1798596.1798608.

Jessica Dai and Eve Fleisig. Mapping social choice theory to rlhf. *arXiv preprint arXiv:2404.13038*, 2024.

Jean-Charles de Borda. Mémoire sur les élections au scrutin. *Histoire de l'Académie Royale des Sciences*, pp. 657–665, 1784.

Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

Jing Dong, Yaoliang Yu, and Pascal Pourpart. The representation-rationalizability tradeoff in reward learning. *arXiv preprint arXiv:2606.00291*, 2026.

Luise Ge, Daniel Halpern, Evi Micha, Ariel D. Procaccia, Itai Shapira, Yevgeniy Vorobeychik, and Junlin Wu. Axioms for AI Alignment from Human Feedback. In *Advances in Neural Information Processing Systems*, volume 37, pp. 80439–80465. Curran Associates, Inc., 2024. doi: 10.52202/079017-2557.

Daniel Halpern, Evi Micha, Ariel D. Procaccia, Benjamin Schiffer, Itai Shapira, and Shirley Zhang. AI Alignment From Social Choice Perspectives, June 2026. URL http://arxiv.org/abs/2606.21550.

Alexandros Hollender and Sonja Kraiczy. Enforcing axioms for ai alignment under loss-based rules. In *The Fourteenth International Conference on Learning Representations*, 2026.

John G Kemeny. Mathematics without numbers. *Daedalus*, 88(4):577–591, 1959.

Jean-François Laslier. *Tournament Solutions and Majority Voting*, volume 7. Springer, 1997.

Kaizhao Liu, Qi Long, Zhekun Shi, Weijie J Su, and Jiancong Xiao. Statistical impossibility and possibility of aligning llms with human preferences: From condorcet paradox to nash equilibrium. *The Annals of Statistics*, 54(4):2110–2135, 2026.

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

Ariel D. Procaccia, Benjamin Schiffer, and Shirley Zhang. Clone-Robust AI Alignment, January 2025.

Python Software Foundation. *Python Language Reference, Version 3.13*, 2024. URL `https://www.python.org/`.

Anand Siththaranjan, Cassidy Laidlaw, and Dylan Hadfield-Menell. Distributional Preference Learning: Understanding and Accounting for Hidden Context in RLHF. *International Conference on Learning Representations*, 2024:27448–27471, May 2024.

Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. Mpnet: Masked and permuted pre-training for language understanding. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 16857–16867. Curran Associates, Inc., 2020. URL `https://neurips.cc`.

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

Michael Henry Tessler, Michiel A Bakker, Daniel Jarrett, Hannah Sheahan, Martin J Chadwick, Raphael Koster, Georgina Evans, Lucy Campbell-Gillingham, Tantum Collins, David C Parkes, et al. Ai can help humans find common ground in democratic deliberation. *Science*, 386(6719): eadq2852, 2024.

Lieven Vandenberghe and Stephen Boyd. *Convex optimization*, volume 1. Cambridge university press Cambridge, 2004.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

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

A Deferred Proofs from Section 3

Lemma 1 (LP1 satisfies PO and PMC)**.** *Every feasible solution of* LP1 *induces a reward function* $r_{\boldsymbol{\theta},\boldsymbol{\varepsilon}}$ *(equation 2) that satisfies Pareto optimality and pairwise majority consistency, but need not satisfy Condorcet consistency.*

Proof. Let $\mathcal{C} = \{c_1, c_2, \dots c_m\}$ be the set of candidates and let π be a preference profile over $\mathcal{C}$. According to the preference profile, we get the sets $\mathcal{U}$ and $\mathcal{M}$, which contain the unanimous and majority pairs, respectively. Let $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\varepsilon}})$ be any feasible solution of LP1.

Pareto Optimality. Irrespective of the value of $\mathbb{1}_{PMC}$, the set of constraints in LP1 for the unanimous pairs $\mathcal{U}$ is preserved. Therefore, for every Pareto-optimal pair $(c_i, c_j) \in \mathcal{U}$, we have $\langle \hat{\boldsymbol{\theta}}, \boldsymbol{x}_{c_i} - \boldsymbol{x}_{c_j} \rangle + \hat{\varepsilon}_{c_i} - \hat{\varepsilon}_{c_j} \geq \eta$, which implies $r_{\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\varepsilon}}}(c_i) > r_{\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\varepsilon}}}(c_j)$. Thus, the induced reward function satisfies PO.

Pairwise Majority Consistency. From the definition, if there is a Condorcet cycle, then PMC is vacuously true. Otherwise, we have that $\mathbb{1}_{PMC} = 1$. As $\mathcal{U} \subseteq \mathcal{M}$, it implies $\mathcal{U} \cup \mathbb{1}_{PMC}\mathcal{M} = \mathcal{U} \cup \mathcal{M} = \mathcal{M}$. Since $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\varepsilon}})$ is feasible, the constraint for each $(c_i, c_j) \in \mathcal{M}$ is satisfied. That is, $\langle \hat{\boldsymbol{\theta}}, \boldsymbol{x}_a - \boldsymbol{x}_b \rangle + \hat{\epsilon}_{c_i} - \hat{\epsilon}_{c_j} \geq \eta$. Hence $r_{\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\varepsilon}}}(c_i) > r_{\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\varepsilon}}}(c_j)$ for every majority edge, establishing PMC.

Failure of Condorcet Consistency. For a 1 dimensional feature space, let $\mathcal{C} = \{c_1, c_2, c_3, c_4\}$, $x_{c_1} = 0$, $x_{c_2} = 1$, $x_{c_3} = 2$, and $x_{c_4} = 1.5$. Let there be 3 voters ={1,2,3}, whose preferences are $c_4 \succ_1 c_1 \succ_1 c_2 \succ_1 c_3$, $c_4 \succ_2 c_2 \succ_2 c_3 \succ_2 c_1$ and $c_3 \succ_3 c_1 \succ_3 c_2 \succ_3 c_4$, respectively. Based on the preference profile, c_4 is preferred to c_i by a majority of voters for each $i \in \{1, 2, 3\}$. Hence, c_4 is a Condorcet winner. Since there are no unanimous pairs in the preference profile, $\mathcal{U} = \emptyset$. Also, c_1 beats c_2, c_2 beats c_3 and c_3 beats c_1 in majority; this creates a cycle rendering $\mathbb{1}_{PMC} = 0$.

Therefore, the LP1 for this instance has no constraint for (PO+PMC). Then, the following is a solution for LP1: $\hat{\boldsymbol{\varepsilon}} = \mathbf{0}$ and $\hat{\theta} = 1$ or $\hat{\theta} = -1$. In either of the solutions, c_4 is not the winner in the ranking induced by the reward function $r_{\boldsymbol{\theta}}$. Hence, it fails Condorcet consistency. □

Theorem 1 (LP2 satisfy all axioms)**.** *Every feasible solution of* LP2 *induces a reward function* $r_{\boldsymbol{\theta},\boldsymbol{\varepsilon}}$ *(equation 2) that satisfies Pareto optimality, pairwise majority consistency, Condorcet consistency, and majority consistency.*

In order to prove the theorem, we first need the following lemma describing the property of the feedback arc set we delete.

Lemma 5. *The feedback arc set* $\mathcal{F}$ *obtained using the algorithm in Coppersmith et al. (2010) has no Pareto-optimal (unanimous) edge.*

Proof. Let (c_1, c_2) be a unanimous pair. Thus every voter prefers c_1 to c_2, i.e., $n_{c_1 \succ c_2}(\pi) = n$. Consider any candidate $x \in C \setminus \{c_1, c_2\}$. If $c_2 \succ_i x$ for some voter i, transitivity of voter i's preference implies $c_1 \succ_i x$. Hence $n_{c_1 \succ x}(\pi) \geq n_{c_2 \succ x}(\pi)$, which means every majority victory of c_2 is also a majority victory of c_1. Therefore, $\text{out-deg}(c_1) \geq \text{out-deg}(c_2) + 1$ in $G_{\mathcal{M}}$, since c_1 defeats c_2 and every out neighbor of c_2. Equivalently, $\text{in-deg}(c_1) < \text{in-deg}(c_2)$. The CopperSmith algorithm orders vertices by non-decreasing in-degree. Hence c_1 always appears before c_2, making (c_1, c_2) a forward edge. However, the feedback arc set the set of backward edges in the ordering and does not contain the edge (c_1, c_2). Thus, the algorithm never adds a PO edge in $\mathcal{F}$. If $\mathcal{U}$ is the set containing the unanimous pairs, $\mathcal{F} \cap \mathcal{U} = \emptyset$. □

Now we are ready to prove the theorem.

Proof of Theorem 1. Let $\mathcal{C} = \{c_1, c_2, \dots c_m\}$ be the set of candidates and let π be a preference profile over $\mathcal{C}$. Let $G_{\mathcal{M}}$ denote the majority graph corresponding to π. If $G_{\mathcal{M}}$ contains a cycle, then we first apply the CopperSmith algorithm (Coppersmith et al., 2010). We order the vertices in the non-increasing order of their in-degree. The feedback arc set $\mathcal{F}$ is the set of backward arcs in the ordering. If $G_{\mathcal{M}}$ is acyclic, then $\mathcal{F} = \emptyset$. Let the set of pairs $\mathcal{P} = E(G_{\mathcal{M}}) \setminus \mathcal{F}$ be the remaining

edges of $G_{\mathcal{M}}$ after removing the feedback arc set. There is a constraint for each pair of candidates in $\mathcal{P}$ in LP2. Let $(\hat{\boldsymbol{\theta}}, \hat{\varepsilon})$ be any feasible solution of LP2. We prove each property separately.

Pareto Optimality. If $\mathcal{F} = \emptyset$, every Pareto-optimal edge already belongs to $\mathcal{P}$. If $\mathcal{F} \neq \emptyset$, we show in Lemma 5 that the CopperSmith algorithm never removes a Pareto-optimal edge. Hence, every Pareto-optimal pair remains in $\mathcal{P}$. Therefore, for every Pareto-optimal pair (c_i, c_j) it holds that $\langle \hat{\boldsymbol{\theta}}, \boldsymbol{x}_{c_i} - \boldsymbol{x}_{c_j} \rangle + \hat{\varepsilon}_{c_i} - \hat{\varepsilon}_{c_j} \geq \eta$, which implies $r_{\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\varepsilon}}}(c_i) > r_{\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\varepsilon}}}(c_j)$. Thus, the induced reward function satisfies PO.

Pairwise Majority Consistency. From the definition, if there is a Condorcet cycle, then PMC is vacuously true. Otherwise, we have that $G_{\mathcal{M}}$ is acyclic, which implies $\mathcal{F} = \emptyset$, it means that $\mathcal{P} = \mathcal{M}$. Since $(\hat{\boldsymbol{\theta}}, \hat{\varepsilon})$ is feasible, each constraint for $\mathcal{P}$ is satisfied: $\langle \hat{\boldsymbol{\theta}}, \boldsymbol{x}_a - \boldsymbol{x}_b \rangle + \hat{\epsilon}_{c_i} - \hat{\epsilon}_{c_j} \geq \eta$, for each $(c_i, c_j) \in \mathcal{P}$. Hence, $r_{\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\varepsilon}}}(c_i) > r_{\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\varepsilon}}}(c_j)$ for every retained majority edge, establishing PMC.

Condorcet Consistency. Suppose a candidate $c^* \in \mathcal{C}$ is the Condorcet winner in the preference profile π. Then, for every candidate $c \in C \setminus \{c^*\}$, we have $w_{c^*,c}(\pi) > \frac{1}{2}$, that is, c^* defeats every other candidate by majority. Hence, in the majority graph $G_{\mathcal{M}}$, the vertex c^* has an outgoing edge to every other candidate and no incoming edge. Therefore, $\text{in-deg}(c^*) = 0$ and $\text{in-deg}(c) \geq 1$ for each $c \neq c^*$. The Coppersmith algorithm orders vertices by non-decreasing in-degree. Since c^* is the unique vertex with minimum in-degree, it is placed first in the ordering. Consequently, there is no backward arc incident to c^*, and none of its outgoing edges are removed while constructing the feedback arc set $\mathcal{F}$. Thus, for every $c \neq c^*$, the edge (c^*, c) belongs to the constraint set $\mathcal{P}$. Since $(\hat{\boldsymbol{\theta}}, \hat{\varepsilon})$ is feasible, we get $\langle \hat{\boldsymbol{\theta}}, x_{c^*} - x_c \rangle + \hat{\varepsilon}_{c^*} - \hat{\varepsilon}_c \geq \eta$. Therefore, $r_{\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\varepsilon}}}(c^*) > r_{\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\varepsilon}}}(c)$ for each $c \neq c^*$. Hence, c^* is ranked first in the ranking induced by the reward function, proving Condorcet Consistency.

Majority Consistency. If candidate $c^* \in \mathcal{C}$ is ranked first by a majority of the voters, it defeats all other candidates by majority, and c^* becomes the Condorcet winner. It is shown that Condorcet consistency is maintained for every feasible solution of LP2. So, c^* is ranked first by $r_{\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\varepsilon}}}$. By the definition of majority consistency, $r_{\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\varepsilon}}}$ also satisfies majority consistency.

Thus, it is proved that any feasible solution of LP2 induces a reward function using Equation (2) that satisfies PO, PMC, Condorcet Consistency, and Majority Consistency. □

B Deferred proofs from Section 4

Theorem 2 (Bounds on total slack)**.** *Let* $\Delta := \max_{(a,b)\in\mathcal{P}} \|\boldsymbol{x}_a - \boldsymbol{x}_b\|_2$. *For every* $\lambda \geq 0$*, the optimal slack* $\boldsymbol{\varepsilon}^*_\lambda$ *of* LP3 *satisfies*

$$\|\boldsymbol{\varepsilon}^*_\lambda\|_1 \leq \min\left\{\eta \lfloor \tfrac{m^2}{4} \rfloor + \lambda\, \eta\, |\mathcal{P}|\,,\ \left(\eta + \Delta\sqrt{d}\right) \lfloor \tfrac{m^2}{4} \rfloor\right\}.$$

For every m, d, η*, there exist features in* $\mathbb{R}^d$ *and a profile, for which* $\|\boldsymbol{\varepsilon}^*_\lambda\|_1 \geq \frac{\eta}{2} \lfloor \frac{m^2}{4} \rfloor$.

Proof. For each $\boldsymbol{\theta} \in \mathbb{R}^d$, define the minimum total slack satisfying margin constraints as

$$c(\boldsymbol{\theta}) := \min\Big\{ \|\boldsymbol{\varepsilon}\|_1 :\ \varepsilon_a - \varepsilon_b \geq \eta - \langle \boldsymbol{\theta}, \boldsymbol{x}_a - \boldsymbol{x}_b \rangle\ \ \forall (a,b) \in \mathcal{P} \Big\}.$$

Let $(\boldsymbol{\theta}^*_\lambda, \boldsymbol{\varepsilon}^*_\lambda, \boldsymbol{u}^*)$ be an optimal solution of LP3. Then $\|\boldsymbol{\varepsilon}^*_\lambda\|_1 = c(\boldsymbol{\theta}^*_\lambda)$. To see this, suppose $\|\boldsymbol{\varepsilon}^*_\lambda\|_1 > c(\boldsymbol{\theta}^*_\lambda)$, and let some other slack $\widehat{\boldsymbol{\varepsilon}}$ attain $c(\boldsymbol{\theta}^*_\lambda)$. Then $(\boldsymbol{\theta}^*_\lambda, \widehat{\boldsymbol{\varepsilon}}, \boldsymbol{u}^*)$ is a feasible point of LP3. It leaves $\sum \boldsymbol{u}^*_{ab}$ term unchanged, and thus has a strictly smaller objective, a contradiction.

Now we bound $c(\boldsymbol{\theta})$ for any $\boldsymbol{\theta}$ with $\|\boldsymbol{\theta}\|_\infty \leq 1$. For $(a,b) \in \mathcal{P}$, Hölder's inequality gives

$$|\langle \boldsymbol{\theta}, \boldsymbol{x}_a - \boldsymbol{x}_b \rangle| \leq \|\boldsymbol{\theta}\|_\infty \|\boldsymbol{x}_a - \boldsymbol{x}_b\|_1 \leq \|\boldsymbol{x}_a - \boldsymbol{x}_b\|_1 \leq \Delta\sqrt{d} \implies \eta - \langle \boldsymbol{\theta}, \boldsymbol{x}_a - \boldsymbol{x}_b \rangle \leq \eta + \Delta\sqrt{d}. \quad (9)$$

The second inequality follows from the fact that $\|\boldsymbol{\theta}\|_\infty \leq 1$. Since $\mathcal{P}$ is acyclic, without loss of generality, we assume it admits a linear extension $a_1 \succ \ldots \succ a_m$: every pair has the form (a_i, a_j) with $i < j$. Setting $\varepsilon_{a_i} = (\eta + \Delta\sqrt{d})\left(\frac{m+1}{2} - i\right)$, we can satisfy every margin constraint since

$$\varepsilon_{a_i} - \varepsilon_{a_j} = (\eta + \Delta\sqrt{d})(j - i) \geq \eta + \Delta\sqrt{d}.$$

Hence, its ℓ_1-norm upper bounds $c(\boldsymbol{\theta})$, i.e.,

$$\sup_{\|\boldsymbol{\theta}\|_\infty \le 1} c(\boldsymbol{\theta}) \le \|\boldsymbol{\varepsilon}\|_1 = (\eta + \Delta\sqrt{d}) \sum\nolimits_{i=1}^{m} \left| \tfrac{m+1}{2} - i \right| \le (\eta + \Delta\sqrt{d}) \left\lfloor \tfrac{m^2}{4} \right\rfloor.$$

Since $\boldsymbol{\theta}^*_\lambda$ lies in the box, it yields $\|\boldsymbol{\varepsilon}^*_\lambda\|_1 = c(\boldsymbol{\theta}^*_\lambda) \le (\eta + \Delta\sqrt{d})\lfloor m^2/4 \rfloor$.

Now, set $\boldsymbol{\theta} = \mathbf{0}$, $\varepsilon_{a_i} = \eta\big(\frac{m+1}{2} - i\big)$ along the same linear extension, and $u_{ab} = \eta$. Then, in Equation (7), we get $\eta - \langle \boldsymbol{\theta}, \boldsymbol{x}_a - \boldsymbol{x}_b \rangle \le \eta$. Moreover, every margin constraint reads $\varepsilon_{a_i} - \varepsilon_{a_j} = \eta(j-i) \ge \eta$, and all constraints hold. The objective value of LP3 for this feasible point is $\eta\lfloor m^2/4 \rfloor + \lambda\eta|\mathcal{P}|$. Optimality, together with $\lambda \ge 0$ and $\boldsymbol{u}^* \ge 0$, gives

$$\|\boldsymbol{\varepsilon}^*_\lambda\|_1 \le \eta\big\lfloor m^2/4 \big\rfloor + \lambda\big(\eta|\mathcal{P}| - \sum\nolimits_{(a,b)\in\mathcal{P}} u^*_{ab}\big) \le \eta\big\lfloor m^2/4 \big\rfloor + \lambda\eta|\mathcal{P}|.$$

Combining both bounds proves the upper bound.

For the lower bound, take a unanimous profile in which every voter ranks $a_1 \succ \cdots \succ a_m$. Then $G_{\mathcal{M}}$ is the complete order and acyclic, so $\mathcal{F} = \emptyset$ and $\mathcal{P} = \mathcal{U} = \mathcal{M}$. Now, fix $\boldsymbol{x}_0 \in \mathbb{R}^d$, pairwise distinct unit vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$, and place the candidate features at $\boldsymbol{x}_{a_i} := \boldsymbol{x}_0 + \gamma\, \boldsymbol{u}_i$ with $\gamma := \eta/(4\sqrt{d})$. Since $\|\boldsymbol{u}\|_1 \le \sqrt{d}\,\|\boldsymbol{u}\|_2 \le \sqrt{d}$, every $\boldsymbol{\theta}$ in the box $\|\boldsymbol{\theta}\|_\infty \le 1$ satisfies

$$\big|\langle \boldsymbol{\theta},\, \boldsymbol{x}_{a_i} - \boldsymbol{x}_{a_j} \rangle\big| \le \|\boldsymbol{\theta}\|_\infty \, \|\boldsymbol{x}_{a_i} - \boldsymbol{x}_{a_j}\|_1 \le \gamma\big(\|\boldsymbol{u}_i\|_1 + \|\boldsymbol{u}_j\|_1\big) \le 2\gamma\sqrt{d} = \tfrac{\eta}{2}.$$

So, for any feasible slack $\boldsymbol{\varepsilon}$, the margin constraint forces $\varepsilon_{a_i} - \varepsilon_{a_j} \ge \eta/2$ for each $i < j$. To find the slack with minimum ℓ_1-norm, pair i-th candidate with $(m+1-i)$-th for $i \le \lfloor m/2 \rfloor$ and see that $\|\boldsymbol{\varepsilon}\|_1 = \sum_{i=1}^{\lfloor m/2 \rfloor} \big(|\varepsilon_{a_i}| + |\varepsilon_{a_{m+1-i}}|\big)$. The minimum is attained by $\boldsymbol{\varepsilon}^*_{a_i} = \frac{\eta}{2}\big(\frac{m+1}{2} - i\big)$, yielding

$$\|\boldsymbol{\varepsilon}^*\|_1 \ge \sum\nolimits_{i=1}^{\lfloor m/2 \rfloor} \left(\varepsilon^*_{a_i} - \varepsilon^*_{a_{m+1-i}}\right) \ge \tfrac{\eta}{2} \sum\nolimits_{i=1}^{\lfloor m/2 \rfloor} (m + 1 - 2i) \ge \tfrac{\eta}{2} \left\lfloor \tfrac{m^2}{4} \right\rfloor,$$

which completes the proof. □

Lemma 4 (ILP minimizes KT distance)**.** *Let $\boldsymbol{\theta}$ be a solution of* (ILP) *and $\succ_{\boldsymbol{\theta}}$ denote the ranking induced by the reward $r_{\boldsymbol{\theta}}$. Then, there exists a λ such that the KT distance of $\succ_{\boldsymbol{\theta}}$ from $\mathcal{P}$ is minimized.*

Proof. Recall that $r_{\boldsymbol{\theta}}(c) = \langle \boldsymbol{\theta}, \boldsymbol{x}_c \rangle$ for each candidate $c \in \mathcal{C}$. Thus, $r_{\boldsymbol{\theta}}$ violates the constraint for a pair $(a, b) \in \mathcal{P}$ when $\langle \boldsymbol{\theta}, \boldsymbol{x}_a - \boldsymbol{x}_b \rangle < \eta$. Due to the constraints in Equation (7), we have that if $\langle \boldsymbol{\theta}, \boldsymbol{x}_a - \boldsymbol{x}_b \rangle < \eta$, then u_{ab} is non-zero for each $(a, b) \in \mathcal{P}$. Therefore, if (a, b) is a violated pair, then $u_{ab} = 1$ since $u_{ab} \in \{0, 1\}$, and it is zero otherwise. That is, the variables u_{ab} count the number of violated constants when we only consider $\boldsymbol{\theta}$ to compute the reward function. For a large enough λ, the ILP minimizes the sum $\sum_{(a,b)\in\mathcal{P}} u_{ab}$. Therefore, it minimizes the number of violations. □

Lemma 3 (Increasing λ decreases violated pairs)**.** *Let $\boldsymbol{\theta}$ denote a solution of* ILP *for a given λ and $\succ_{\boldsymbol{\theta}}$ denote the ranking induced by the reward $r_{\boldsymbol{\theta}}$. Then, the KT distance of $\succ_{\boldsymbol{\theta}}$ from $\mathcal{P}$ is non-increasing in λ. After a certain cut-off λ^*, the number of violated pairs remains the same.*

Proof. For a feasible solution $s = (\boldsymbol{\theta}, \boldsymbol{\varepsilon}, \boldsymbol{t}, \boldsymbol{u})$ of ILP, let $f(s) = \sum_{a\in\mathcal{C}} t_a$ and $v(s) = \sum_{(a,b)\in\mathcal{P}} u_{ab}$. Then we can write the objective function of ILP as $g(s, \lambda) = f(s) + \lambda v(s)$.

Let $\lambda_1 < \lambda_2$. For a given preference profile π, the parameter η, feature vectors, and $\mathcal{P}$, the ILP gives an optimal solution s_1 for λ_1 and an optimal solution s_2 for λ_2. Since ILP is a minimization problem and s_1 is an optimal solution for λ_1, then we have the following,

$$\begin{aligned} g(s_1, \lambda_1) &\le g(s_2, \lambda_1) \\ f(s_1) + \lambda_1 v(s_1) &\le f(s_2) + \lambda_1 v(s_2) \\ f(s_1) - f(s_2) &\le \lambda_1\big(v(s_2) - (s_1)\big) \end{aligned} \tag{10}$$

Similarly, since s_2 is an optimal solution for λ_2, then we have the following,

$$\begin{aligned} g(s_2, \lambda_2) &\le g(s_1, \lambda_2) \\ f(s_2) + \lambda_2 v(s_2) &\le f(s_1) + \lambda_2 v(s_1) \\ f(s_2) - f(s_1) &\le \lambda_2\big(v(s_1) - (s_2)\big) \end{aligned} \tag{11}$$

Adding Equation (10) and Equation (11), we get the following,

$$\lambda_2\big(v(s_1) - v(s_2)\big) + \lambda_1\big(v(s_2) - v(s_1)\big) \geq 0$$
$$(\lambda_2 - \lambda_1)\big(v(s_1) - v(s_2)\big) \geq 0$$

Since, $\lambda_1 < \lambda_2$, we have $v(s_1) \geq v(s_2)$, which means as λ increases, the number of violations decreases monotonically. This proves the first part of the theorem.

Recall that there is a violation for $(a, b) \in \mathcal{P}$ when $\langle \boldsymbol{\theta}, \boldsymbol{x}_a - \boldsymbol{x}_b \rangle < \eta$. Due to constraint (7), we have that $u_{ab} = 1$ for a violated pair (a, b), and it is zero otherwise. Thus, $v(s)$ represents the number of violated pairs in $\mathcal{P}$. Now, let $p = |\mathcal{P}|$. Then, $v(s) \in \{0, 1, \ldots, p\}$, can take a finite number of values since p is finite for a finite number of candidates. Let v_{min} be the minimum number of violations possible for a problem instance and s_{min} is a feasible solution attaining $v(s_{min}) = v_{min}$. Therefore, the value of the objective function at s_{min} for a given λ is

$$g(s_{min}, \lambda) = f(s_{min}) + \lambda v(s_{min})$$
$$= f(s_{min}) + \lambda v_{min}$$

In Theorem 2, we give an upper bound in f for the relaxed LP3. Now, we want a bound on f for ILP. Thus, in the proof of Theorem 2 we set $u_{ab} = 1$ instead of $u_{ab} = \eta$. Then, we get an upper bound on f and g as follows.

$$\begin{aligned} f(s_{min}) \;&\leq\; \min\{\eta\lfloor\tfrac{m^2}{4}\rfloor + \lambda p,\ (\eta + \Delta\sqrt{d})\lfloor\tfrac{m^2}{4}\rfloor\} \\ f(s_{min}) &\leq^{(\mathrm{i})} (\eta + \Delta\sqrt{d})\lfloor\tfrac{m^2}{4}\rfloor \\ f(s_{min}) + \lambda v_{min} \leq\;& (\eta + \Delta\sqrt{d})\lfloor\tfrac{m^2}{4}\rfloor + \lambda v_{min} \\ g(s_{min}, \lambda) \leq\;& (\eta + \Delta\sqrt{d})\lfloor\tfrac{m^2}{4}\rfloor + \lambda v_{min} \end{aligned} \tag{12}$$

The inequality (i) follows from the fact that $\min\{a, b\} \leq b$.

Next, we show that if we set $\lambda > (\eta + \Delta\sqrt{d})\lfloor\frac{m^2}{4}\rfloor$, then any feasible solution with number of violations more than v_{min} has a strictly larger value for the objective. Let s' be a feasible solution of ILP such that $v(s') \geq v_{min} + 1$. If we set $\lambda > (\eta + \Delta\sqrt{d})\lfloor\frac{m^2}{4}\rfloor$, we get the following as

$$\begin{aligned} \lambda \;&>\; (\eta + \Delta\sqrt{d})\lfloor\tfrac{m^2}{4}\rfloor \\ \lambda + \lambda v_{min} &> (\eta + \Delta\sqrt{d})\lfloor\tfrac{m^2}{4}\rfloor + \lambda v_{min} \\ \lambda(v_{min} + 1) &> (\eta + \Delta\sqrt{d})\lfloor\tfrac{m^2}{4}\rfloor + \lambda v_{min} \\ f(s') + \lambda v(s') &>^{(\mathrm{ii})} (\eta + \Delta\sqrt{d})\lfloor\tfrac{m^2}{4}\rfloor + \lambda v_{min} \\ g(s', \lambda) &> (\eta + \Delta\sqrt{d})\lfloor\tfrac{m^2}{4}\rfloor + \lambda v_{min} \\ g(s', \lambda) &>^{(\mathrm{iii})} g(s_{min}, \lambda) \end{aligned}$$

The inequality (ii) follows from $f(s') + \lambda v(s') \geq \lambda(v_{min} + 1)$ as $f(s') \geq 0$. The inequality (iii) follows from equation 12.

The value of the objective function for any feasible solution containing at least $v_{min}+1$ violations is strictly larger than a feasible solution containing v_{min} violations. So, the minimization problem will reach a solution with v_{min} violations, and increasing λ will not decrease the number of violations any more since v is a monotonically decreasing function in λ and v_{min} is the minimum number of violations possible for the problem instance. Hence, $\lambda^* > (\eta + \Delta\sqrt{d})\lfloor\frac{m^2}{4}\rfloor$ is cut-off value.

□

Lemma 2 (LP3 satisfy all axioms)**.** *Every feasible solution of* LP3 *induces a reward function* $r_{\boldsymbol{\theta},\boldsymbol{\varepsilon}}$ *(equation 2) that satisfies all three of Pareto optimality, pairwise majority consistency, Condorcet Consistency, and majority consistency.*

Proof. Let $\mathcal{I} = (\mathcal{X}, \mathcal{P})$ be a problem instance for LP3 where $\mathcal{X}$ and $\mathcal{P}$ are the feature matrix and the preference DAG, respectively. Let $\mathcal{S}_{\boldsymbol{\theta},\boldsymbol{\varepsilon},\boldsymbol{t},\boldsymbol{u}}(\mathcal{I}, \mathsf{LP3})$ be the set of feasible solutions of $(\boldsymbol{\theta}, \boldsymbol{\varepsilon}, \boldsymbol{t}, \boldsymbol{u})$ for instance $\mathcal{I}$ running LP3.

All the constraints set of LP2 are present in LP3, along with two extra set of constraints present in LP3 to decrease the number of violations, i.e., an non-integral, relaxed version of equation 7. Therefore any feasible solution of LP3 (removing $\boldsymbol{u}$) will be a feasible solution of LP2, $\mathcal{S}_{\boldsymbol{\theta},\boldsymbol{\varepsilon},\boldsymbol{t}}(\mathcal{I}, \mathsf{LP3}) \subseteq \mathcal{S}_{\boldsymbol{\theta},\boldsymbol{\varepsilon},\boldsymbol{t}}(\mathcal{I}, \mathsf{LP2})$.

Let $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\varepsilon}}, \hat{\boldsymbol{t}}, \hat{\boldsymbol{u}})$ be a feasible solution of LP3 on problem instance $\mathcal{I}$, hence, $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\varepsilon}}, \hat{\boldsymbol{t}})$ will be a feasible solution of LP2. From Theorem 1, we know that $r_{\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\varepsilon}}}$ will satisfy PO, PMC, Condorcet Consistency, and majority consistency. Since $r_{\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\varepsilon}}}$ is also the induced reward function from a feasible solution of LP3, every feasible solution of LP3 induces a reward function that also satisfies PO, PMC, Condorcet Consistency, and majority consistency. □

Proposition 1 (Realizability $\not\Longrightarrow$ zero-slack always)**.** *If π is linearly realizable, then $\kappa(\pi, \mathcal{X}) > 0$ and $\boldsymbol{\varepsilon}^*_\lambda = \mathbf{0}$ if and only if $\eta \leq \kappa(\pi, \mathcal{X})$ for every $\lambda \geq 0$.*

Proof. Write $\kappa := \kappa(\pi, \mathcal{X})$. Realizability gives a $\boldsymbol{\theta}$ satisfying $\langle \boldsymbol{\theta}, \boldsymbol{x}_a - \boldsymbol{x}_b \rangle > 0$ on every pair of $\mathcal{P}$. Rescaling it by $1/\|\boldsymbol{\theta}\|_\infty$ puts it in the box, yielding $\kappa > 0$. Fix $\eta \leq \kappa$ and let $\widehat{\boldsymbol{\theta}}$ attain the maximum in Equation (8). Then $\langle \widehat{\boldsymbol{\theta}}, \boldsymbol{x}_a - \boldsymbol{x}_b \rangle \geq \kappa \geq \eta$ for every $(a, b) \in \mathcal{P}$, so the pair $(\widehat{\boldsymbol{\theta}}, \mathbf{0})$ satisfies every margin constraint, and the shortfalls $h_{ab}(\widehat{\boldsymbol{\theta}}) = \eta - \langle \widehat{\boldsymbol{\theta}}, \boldsymbol{x}_a - \boldsymbol{x}_b \rangle \leq 0$ allow $u_{ab} = 0$. This feasible point has the objective value 0, hence it is optimal and thus $\boldsymbol{\varepsilon}^*_\lambda = \mathbf{0}$. Now, fix $\eta > \kappa$ and suppose $\boldsymbol{\varepsilon}^*_\lambda = \mathbf{0}$. Each margin constraint then reads $\langle \boldsymbol{\theta}^*_\lambda, \boldsymbol{x}_a - \boldsymbol{x}_b \rangle \geq \eta$. So $\boldsymbol{\theta}^*_\lambda$ is feasible for equation 8 with value at least η, giving $\kappa \geq \eta$ – a contradiction. Hence $\boldsymbol{\varepsilon}^*_\lambda \neq \mathbf{0}$. □

Proposition 2 (Certificate of realizability)**.** *A profile π is linearly realizable if and only if no non-zero feature balanced flow exists, i.e., $\nexists\, \boldsymbol{\mu} \gneq \mathbf{0}$, with $\sum_{a \in \mathcal{C}} \nu_a(\boldsymbol{\mu}) \boldsymbol{x}_a = \mathbf{0}$.*

Proof. Gordan's theorem (Vandenberghe & Boyd, 2004) says that for any matrix $\boldsymbol{A}$, either (i) $\exists\, \boldsymbol{z}$ with $\boldsymbol{A}\boldsymbol{z} > \mathbf{0}$ componentwise, or, (ii) $\exists\, \boldsymbol{y} \gneq \mathbf{0}$, with $\boldsymbol{A}^\top \boldsymbol{y} = \mathbf{0}$. Take $\boldsymbol{A} := \left[(\boldsymbol{x}_a - \boldsymbol{x}_b)^\top\right]_{(a,b) \in \mathcal{P}}$.

(i) Taking $\boldsymbol{z} = \boldsymbol{\theta}$, the condition $\boldsymbol{A}\boldsymbol{\theta} > \mathbf{0}$ is exactly $\langle \boldsymbol{\theta}, \boldsymbol{x}_a - \boldsymbol{x}_b \rangle > 0$ on every constrained pair and gives realizability. It is also equivalent to realizability at any prescribed margin: taking $\boldsymbol{z} = \lambda \boldsymbol{\theta}$, where $\lambda = \eta / \min_{(a,b) \in \mathcal{P}} \langle \boldsymbol{\theta}, \boldsymbol{x}_a - \boldsymbol{x}_b \rangle$, gives $\langle \boldsymbol{\theta}, \boldsymbol{x}_a - \boldsymbol{x}_b \rangle \geq \eta$

(ii) Taking $\boldsymbol{y} = \boldsymbol{\mu}$, with $\boldsymbol{\mu} \gneq \mathbf{0}$, gives $\boldsymbol{A}^\top \boldsymbol{\mu} = \sum_{(a,b) \in \mathcal{P}} \mu_{ab}(\boldsymbol{x}_a - \boldsymbol{x}_b)$. Regrouping the sum by candidates gives $\boldsymbol{A}^\top \boldsymbol{\mu} = \sum_{a \in \mathcal{C}} \nu_a(\boldsymbol{\mu}) \boldsymbol{x}_a = \mathbf{0}$.

Exactly one alternative holds, giving the claim. □

Lemma 6 (Balanced flow need $d + 2$ candidates)**.** *Let $\boldsymbol{\mu} \gneq \mathbf{0}$ satisfy $\sum_{a \in \mathcal{C}} \nu_a(\boldsymbol{\mu}) \boldsymbol{x}_a = \mathbf{0}$. Set $\mathcal{T} := \{a \in \mathcal{C} : \nu_a(\boldsymbol{\mu}) \neq 0\}$. Then $\{\boldsymbol{x}_a\}_{a \in \mathcal{T}}$ is affinely dependent, with coefficients $(\nu_a(\boldsymbol{\mu}))_{a \in \mathcal{T}}$. Moreover, if the features are in the general position, then $|\mathcal{T}| \geq d + 2$.*

Proof. Taking $\zeta_a = \nu_a(\boldsymbol{\mu})$ gives $\sum_{a \in \mathcal{C}} \zeta_a \boldsymbol{x}_a = 0$, and $\sum_{a \in \mathcal{C}} \zeta_a = 0$. Restricting both sums to $\mathcal{T}$ changes nothing, as $\lambda_a = 0$ off $\mathcal{T}$. It remains to show $\boldsymbol{\lambda} \not\equiv \mathbf{0}$, so that the dependence is nontrivial.

Suppose $\boldsymbol{\zeta} \equiv \mathbf{0}$. Since $\mathcal{P}$ is acyclic, let φ_a be the position of a in a reversed topological order, so $\varphi_a - \varphi_b \geq 1$ on every edge $(a, b) \in \mathcal{P}$. Interchanging sums over candidates and edges, we get

$$0 = \sum_{a \in \mathcal{C}} \zeta_a \varphi_a = \sum_{(a,b) \in \mathcal{P}} \mu_{ab}(\varphi_a - \varphi_b) \geq \sum_{(a,b) \in P} \mu_{ab} \geq 0.$$

Hence $\boldsymbol{\mu} \equiv \mathbf{0}$, a contradiction. So $\boldsymbol{\zeta} \not\equiv \mathbf{0}$, $\mathcal{T} \neq \emptyset$, and $\{\boldsymbol{x}_a\}_{a \in \mathcal{T}}$ is affinely dependent. Under the general position, this forces $|\mathcal{T}| \geq d + 2$ since every affinely dependent subset $\mathcal{T}$ has at least $d + 2$ elements. □

Corollary 1. *If $\mathcal{X} \subset \mathbb{R}^d$ is in general position and $m \leq d+1$, then every profile is linearly realizable.*

Proof of Corollary 1. By Lemma 6, a non-zero feature balanced flow would require $|\mathcal{T}| \geq d + 2 > m$ candidates. Hence, no such flow exists, and Proposition 2 gives linear realizability. □

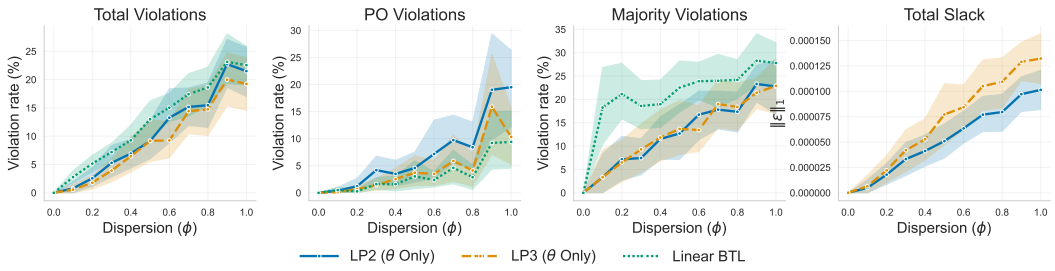

Figure 4: Axiom violations and candidate-level slack as a function of normalized Mallows dispersion ϕ. Curves show mean violation rates across trials, with shaded regions denoting 95% confidence intervals. Panels report (a) Total violations, (b) PO violations, (c) Majority violations, and (d) The $\|\boldsymbol{\varepsilon}\|_1$ for the LP2 and LP3. Experiments use 3 voters, 10 candidates, and 5-dimensional feature embeddings.

C Additional Experimental Results

We conduct additional experiments to examine the robustness of our observations under alternative synthetic and real-world settings. Unless otherwise stated, all experiments use the same evaluation metrics and optimization procedures as described in Section 5.

Real world dataset description. We perform our experiments on the *Habermas Machine Deliberation* dataset, particularly PrefLib dataset 00070 (Tessler et al., 2024). The dataset contains 2,710 separate election files. Each mini-jury consists of five participants who provide rankings over four statements generated by an LLM for a specific deliberation topic; the original data were collected by Google DeepMind and converted to PrefLib format.[3] The resulting files therefore provide a collection of small, independently defined preference profiles.

Experimental setup. Python 3.13 (Python Software Foundation, 2024) is used for running the experiments in this paper. The linear programs LP2 and LP3 are solved using the CVXPY (Diamond & Boyd, 2016; Agrawal et al., 2018) package. Linear BTL model function is trained using a non-linear constrained optimization algorithm, Sequential Least Squares Programming (SLSQP) with a maximum iteration count of 100 and a tolerance of 10^{-6}, on a constraint of $\|\boldsymbol{\theta}\|_\infty \leq 1$. SLSQP is solved using the SciPy package (Virtanen et al., 2020). For the embedding of the real-world data questions and responses, we utilize the `all-mpnet-base-v2` variant of the MPNet model (Song et al., 2020). The model is accessed online from the `huggingface`[4] library. In the real-world data, we sample K questions and get K preference DAGs. If we denote $\mathcal{P}_i$ as the edge set of the ith preference DAG, for the loss calculation of linear BTL, we used the following log-likelihood-based loss function,

$$\mathcal{L}_{\text{Linear-BTL}}(\boldsymbol{\theta}) := -\sum_{i=1}^{K} \sum_{(a,b)\in\mathcal{P}_i} n_{a\succ b} \log(\sigma(\langle \boldsymbol{\theta}, \boldsymbol{x}_a - \boldsymbol{x}_b\rangle)) + n_{b\succ a} \log(\sigma(\langle \boldsymbol{\theta}, \boldsymbol{x}_b - \boldsymbol{x}_a\rangle))$$

We combine the preference graphs to produce the larger graph, which consists only of those smaller preference graphs as disconnected components. So, this loss function is justified.

Additional synthetic experiments. We vary the synthetic problem parameters, including the number of alternatives, the number of voters, and the feature dimension, while keeping the remaining components of the experimental pipeline fixed. Results are reported in Figures 4 and 5.

Additional real-world experiments. We use the same experimental setup as used in Section 5. We perform several experiments for 100 and 500 questions with varying values of $d < 768$. PCA is used to reduce the feature dimension to the desired dimension. The findings are reported in Tables 2 and 3.

[3] `https://preflib.github.io/PrefLib-Jekyll/dataset/00070`

[4] `https://huggingface.co/sentence-transformers/all-mpnet-base-v2`

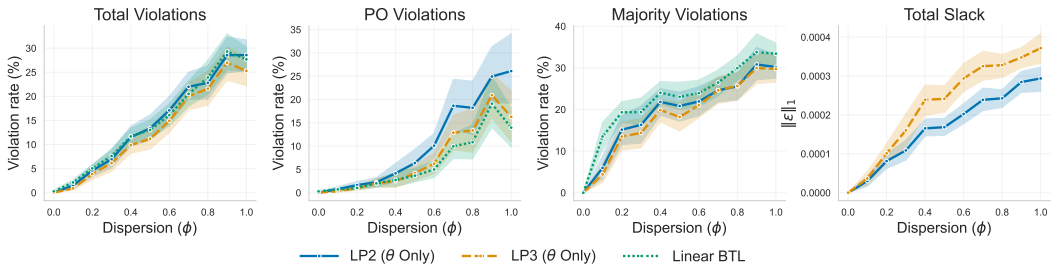

Figure 5: Axiom violations and candidate-level slack as a function of normalized Mallows dispersion ϕ. Curves show mean violation rates across trials, with shaded regions denoting 95% confidence intervals. Panels report (a) Total violations, (b) PO violations, (c) Majority violations, and (d) The $\|\boldsymbol{\varepsilon}\|_1$ for the LP2 and LP3. Experiments use 3 voters, 15 candidates, and 5-dimensional feature embeddings.

Table 2: Performance of LP2, LP3, and Linear BTL on the real-world benchmark of 100 questions comprising 400 candidates, evaluated using PCA-reduced feature representations of dimensions $d \in \{25, 75, 100\}$. Each entry reports the mean $\pm$ standard deviation across the 100 questions. $\eta = 10^{-5}$ for LP2 and LP3.

Model	Total Viol.(%)	PO Viol.(%)	Majority Viol.(%)	$\|\boldsymbol{\varepsilon}\|_1 (\times 10^{-3})$	Time (s)
(a) Feature dimension $d = 25$					
LP2	28.39 ± 2.04	15.60 ± 3.55	$\mathbf{32.48 \pm 2.21}$	2.7 ± 0.155	**1.5**
LP3	$\mathbf{27.95 \pm 2.31}$	$\mathbf{13.69 \pm 3.41}$	32.52 ± 2.43	2.87 ± 0.155	2.95
Linear BTL	31.30 ± 2.54	14.93 ± 2.99	36.53 ± 2.66	–	23.12
(b) Feature dimension $d = 50$					
LP2	23.97 ± 2.20	11.47 ± 2.23	27.98 ± 2.48	2.41 ± 0.136	**1.66**
LP3	$\mathbf{22.76 \pm 2.31}$	$\mathbf{9.97 \pm 2.16}$	$\mathbf{26.84 \pm 2.80}$	2.55 ± 0.144	3.23
Linear BTL	27.96 ± 1.74	11.13 ± 2.48	33.36 ± 2.02	–	45.76
(c) Feature dimension $d = 75$					
LP2	19.85 ± 1.75	9.99 ± 2.26	22.96 ± 2.25	2.07 ± 0.135	**1.73**
LP3	$\mathbf{18.64 \pm 1.78}$	$\mathbf{7.81 \pm 2.34}$	$\mathbf{22.07 \pm 2.07}$	2.24 ± 0.149	3.49
Linear BTL	25.18 ± 1.99	7.92 ± 1.97	30.68 ± 2.32	–	66.19
(d) Feature dimension $d = 100$					
LP2	15.86 ± 1.68	7.58 ± 2.02	18.50 ± 2.09	1.7 ± 0.148	**1.86**
LP3	$\mathbf{14.51 \pm 1.70}$	5.91 ± 2.19	$\mathbf{17.26 \pm 2.11}$	1.89 ± 0.164	4.06
Linear BTL	23.37 ± 1.53	$\mathbf{5.70 \pm 1.99}$	29.03 ± 1.58	–	89.28

Table 3: Performance of LP2, LP3, and Linear BTL on the real-world benchmark of 500 questions comprising 2000 candidates, evaluated using PCA-reduced feature representations of dimensions $d \in \{150, 200, 250, 500\}$. Each entry reports the mean $\pm$ standard deviation across the 500 questions. $\eta = 10^{-5}$ for LP2 and LP3.

	Total Viol.(%)	PO Viol.(%)	Majority Viol.(%)	$\|\varepsilon\|_1$	Time (s)
(a) Feature dimension $d = 150$					
LP2	24.54 ± 1.18	10.90 ± 1.38	29.02 ± 1.21	$1.26 * 10^{-2} \pm 3.57 * 10^{-4}$	**38.73**
LP3	**23.48 ± 0.98**	9.26 ± 0.99	**28.17 ± 1.26**	$1.31 * 10^{-2} \pm 3.48 * 10^{-4}$	109.24
Linear BTL	25.71 ± 1.16	**8.34 ± 0.91**	31.40 ± 1.24	–	863.24
(b) Feature dimension $d = 200$					
LP2	23.26 ± 1.27	9.93 ± 1.44	27.64 ± 1.27	$1.19 * 10^{-2} \pm 3.92 * 10^{-4}$	**48.1**
LP3	**22.12 ± 1.11**	8.44 ± 1.31	**26.61 ± 1.19**	$1.24 * 10^{-2} \pm 4.16 * 10^{-4}$	121.08
Linear BTL	24.40 ± 1.24	**6.70 ± 0.86**	30.20 ± 1.29	–	1093.21
(c) Feature dimension $d = 250$					
LP2	21.69 ± 0.73	9.51 ± 1.82	25.69 ± 0.92	$1.12 * 10^{-2} \pm 3.37 * 10^{-4}$	**60.96**
LP2	**20.60 ± 0.82**	8.04 ± 1.37	**24.73 ± 1.13**	$1.19 * 10^{-2} \pm 3.37 * 10^{-4}$	151.94
Linear BTL	23.91 ± 0.99	**6.29 ± 0.53**	29.69 ± 1.18	–	1271.33
(d) Feature dimension $d = 500$					
LP2	14.57 ± 0.73	6.74 ± 0.75	17.24 ± 0.79	$7.87 * 10^{-3} \pm 2.62 * 10^{-4}$	**275.71**
LP3	**13.31 ± 0.91**	5.07 ± 0.95	**16.12 ± 0.94**	$8.83 * 10^{-3} \pm 2.81 * 10^{-4}$	674.41
Linear BTL	20.50 ± 0.86	**3.34 ± 0.58**	26.35 ± 0.98	–	2368.84