Optimized Distortion in Linear Social Choice

Luise Ge
Washington University in St. Louis
St. Louis, United States
g.luise@wustl.edu

Gregory Kehne
Washington University in St. Louis
St. Louis, United States
kehne@wustl.edu

Yevgeniy Vorobeychik
Washington University in St. Louis
St. Louis, United States
yvorobeychik@wustl.edu

ABSTRACT

Background: One way to design voting rules is to try to maximize hidden voter utilities, but resulting approximation guarantees worsen as the number of candidates grows large. Meanwhile, modern applications like LLM fine-tuning (RLHF) represent candidates using structured embeddings, motivating a new, geometric approach to utility modeling.
Objectives and Research Questions: What structured utility assumptions can create distortion bounds that are completely independent of the number of candidates? Can we do this while generalizing the standard model for implicit utilitarian voting? How can we successfully apply these social choice tools to RLHF?
Methods: Theoretically, we bound distortion using vector inequalities, (information) geometry, and the stable lottery rule. Empirically, we design instance-optimal rules using convex optimization and use dimensionality reduction to evaluate them on real datasets.
Results: Our model generalizes standard distortion and supports guarantees that are independent of the number of candidates. We introduce several provably performant voting rules, including an asymptotically optimal one, which empirically perform significantly better than their worst-case theoretical guarantees.
Conclusions: We establish a geometric model for implicit utilitarian voting that serves as a utility-maximization counterpart to metric distortion. This framework supports meaningful welfare approximation ratios even when there are infinite candidates, and offers a theoretical foundation for RLHF applications.

ACM Reference Format:
Luise Ge, Gregory Kehne, and Yevgeniy Vorobeychik. 2026. Optimized Distortion in Linear Social Choice. In *Appears at the 8th Games, Agents, and Incentives Workshop (GAIW-26). Held as part of the Workshops at the 25th International Conference on Autonomous Agents and Multiagent Systems., Paphos, Cyprus, May 2026*, IFAAMAS, 13 pages.

1 INTRODUCTION

Conventional voting rules map a profile of voter preference rankings over a set of candidates (options, outcomes, choices) to a winner. If, instead of rankings, we were instead to associate candidates with utility values for each voter, in many settings it would be natural to choose a social-welfare-maximizing candidate. But when voters' preference rankings are generated by latent and unknown utilities, ranking-based voting rules can select candidates that are not welfare-maximizing. Thus, the limited information about utilities obtained from rankings can be viewed as an imperfect means of approximating a welfare-maximizing choice, and the quality of this approximation—the worst-case ratio of optimal welfare to welfare obtained by a voting rule—is known as *distortion*. Since its introduction by Procaccia and Rosenschein [21], an extensive literature has used distortion as means to compare voting rules, identify new rules, and understand the inherent difficulty in different settings of identifying desirable outcomes from ordinal and otherwise incomplete information [1]

If we make no assumptions, no deterministic rules have bounded distortion, and (when there are many voters) the best randomized rule chooses a winning candidate uniformly at random. Starting with Procaccia and Rosenschein [21], a prevalent approach is to assume each voter's utilities are nonnegative and sum to 1 (also see Aziz [2] for discussion of this assumption). Another mild but incomparable choice is to assume each voter's utilities span the range $[0, 1]$, which generalizes approval preferences when each voter has at least one approval and one disapproval. In all such settings, even optimal voting rules have distortion polynomial in the number of candidates [10].

What if we wish to make an approximately welfare-maximizing choice from among many candidates, and voters' utilities are more structured than unit-sum or unit-range? Our motivating setting is reinforcement learning from human feedback (RLHF), where options are naturally represented as vector embeddings and voting data consists of rankings. In this context, a voter's utility for an option reflects how well the option aligns with what the voter cares about—mathematically captured as the inner product between the voter's preference vector and the option's embedding. This linear utility model is both a common assumption in RLHF [12, 24] and arises naturally in recommendation systems [20] and multi-objective decision making [19], where candidates have feature representations that voters weight according to their preferences. More broadly, structured representation spaces for both voters and candidates are increasingly relevant as AI clones and artificial agents acting on behalf of humans are studied and deployed [17, 18, 23], motivating the study of parametric utility structures over such spaces. Linear utilities provide a natural starting point for this investigation.

Unlike in many traditional applications of social choice, such as political elections, in RLHF and similar settings the space of options is extremely large. For example, the space of all possible conversational responses to prompts is so vast as to defy enumeration. Consequently, distortion bounds that depend on the number of candidates m in these settings have little bite, since m can be exponential—or larger—in the number of features d.

We investigate distortion when utility functions are linear, and both the candidates' and voters' embeddings into common feature

Appears at the 8th Games, Agents, and Incentives Workshop (GAIW-26). Held as part of the Workshops at the 25th International Conference on Autonomous Agents and Multiagent Systems., Armstrong, Curry, Hosseini, Mattei, Tsang, Wąs (Chairs), May 2026, Paphos, Cyprus.

C Embedding	Randomized		Deterministic	
Known	$O(\sqrt{d})$	(Thm. 6)	$O(d^3)$	(Thm. 3)
	$\Omega(\sqrt{d})$	(Thm. 4)	$\Omega(d^2)$	(Thm. 2)
Unknown	$O(d)$	(Thm. 9)	-	

Table 1: Distortion in d-dimensional linear social choice when candidates and voters are ℓ^1-normalized. Note the unknown setting inherits lower bounds from the known setting.

space $\mathbb{R}^d$ are non-negative and normalized to 1. We focus on ℓ^1 normalization, and defer discussion of ℓ^2 normalization to Appendix C. Notably, under ℓ^1 normalization, our model recovers the unit-sum setting when $m = d$ and candidates embed to the standard basis of $\mathbb{R}^d$.

A significant benefit of measuring distortion over linear utilities in RLHF and alignment settings is its robustness to the introduction of duplicate candidates, i.e. clones. This is a criterion of particular importance in the application of social choice methods to alignment problems [3, 22]. By way of illustration, consider a profile $\vec{\sigma}$ of voters' rankings of C, and construct $\vec{\sigma}_{\cup c'}$ by adding a single duplicate c' of some $c \in C$ which all voters rank (weakly) directly below c. First, can the space of profile-consistent voter utilities over the original candidates C change from $\vec{\sigma}$ to $\vec{\sigma}_{\cup c'}$? For unit-sum and unit-range utilities, yes; for d-dimensional linear utilities, no. Second, m-dependent distortion bounds degrade from m to $m' = m + 1$, while d-dependent bounds do not.

Summary of contributions. On the positive side, we introduce three novel voting rules with distortion bound a function of *only the dimension d*. We summarize these results in Table 1. The first is the deterministic *max coordinate plurality* (MCP) rule. Given the candidate embeddings in $\mathbb{R}^d_{\geq 0}$, it first chooses one candidate that maximizes the value of each coordinate, and then selects a plurality winner among this subset. We show that MCP attains distortion $O(d^3)$ (Theorem 3). The second and third are randomized rules that are adaptations of the stable lottery rule of Ebadian et al. [10] to both the settings where candidate embeddings into $\mathbb{R}^d$ are known and unknown. We consider the case when rules can access candidate embeddings to be the standard setting. Here we construct a *linear stable lotteries* rule, which attains the asymptotically optimal $\Theta(\sqrt{d})$ distortion for linear utilities (Theorem 6). When, as in more traditional social choice settings, access to candidate vectors is unavailable, a version of the stable lotteries rule which only samples candidates from suitably-sized (and randomly chosen) committees attains distortion $O(d)$; we dub this the *pure stable lotteries* rule (Theorem 9). We also show that random dictatorship attains $O(d^3)$ distortion in this setting (Theorem 7).

Alongside rules with provable worst-case guarantees, we develop a linear programming (LP)-based approach for computing instance-optimal distortion-minimizing candidates for a given set of options and voter preferences. While the resulting LP has an infinite set of constraints, we identify an efficient separation oracle that enables efficient optimization over it. We then empirically evaluate the extent to which our instance-optimal approach outperforms other rules.

Further related work. Our model diverges from two major strands of prior work in computational social choice: (i) classical frameworks with minimal utility assumptions, and (ii) metric distortion approaches that adopt latent representations, but differ fundamentally in objective and structure.

In classical frameworks, various assumptions are imposed directly on the utility values assigned by voters to candidates—such as unit-sum (each voter's utilities sum to one), unit-range (utilities span a fixed range, e.g., $[0, 1]$), and approval (where each utility is either 0 or 1). Our approach strictly generalizes this first model by imposing an ℓ^1 normalization constraint on voter vectors and treating candidates as standard basis vectors, i.e., with $m = d$ (ℓ^∞ normalization would similarly generalize the second and third).

Metric distortion approaches are more closely aligned with our setting in that they also assume latent (metric) structure. However, their motivation typically centers on distance-based objectives in utility space—such as minimizing disutility in facility location problems—rather than maximizing social welfare. Although the relation $\text{disutil}_v(c) := 1 - u_v(c) = 1 - v^T c$ satisfies the triangle inequality and thus fits into the metric distortion framework, minimizing social cost in this sense is not equivalent to maximizing social welfare—our primary objective. As a result, our approach departs from metric distortion in both its goal and its fundamental structure.

Finally, we remark that while our work builds on the linear social choice framework introduced by Ge et al. [12], we impose additional structure on the voter and candidate vectors for analysis purpose. In addition, although our work and that of Gölz et al. [13] share an interest in alignment, their model of the *distortion of alignment* is substantively different from ours. First, they bound utilities within $[0, 1]$ but otherwise make no assumptions about latent preference structure. Second, they assume distributions over both voters and candidates, with pairwise comparisons drawn i.i.d. from these distributions; distortion is then evaluated in expectation over this distribution. In contrast, we make no distributional assumptions and evaluate distortion in the worst case over all possible voter-candidate profiles. Finally, we consider truthful ordinal rankings, while Gölz et al. [13] assume a Bradley–Terry noise model for reported preferences, which enables some access to voter preference intensities. This reflects different modeling goals: we aim for robustness under accurate ordinal preference reporting, while they seek insights under probabilistic reporting with minimal utility structure assumptions.

2 MODEL AND PRELIMINARIES

We consider a setting with n voters V and m candidates C, where each voter $v \in V$ and each candidate $c \in C$ corresponds to a vector in $\mathbb{R}^d_{\geq 0}$. The d dimensions correspond to positive attributes that determine both voters' preferences and candidates' characteristics. We use superscripts to denote coordinates of a vector.

We assume that the utility of a voter v for a candidate c is given by the inner product $u_v(c) = v^\top c$. This *utility function* $u_v : C \to \mathbb{R}_{\geq 0}$, in turn, induces a *preference ranking* σ_v where $c \succ_v c'$ if $u_v(c) \geq u_v(c')$, with ties broken arbitrarily. Below we will introduce some constraints for v and c, and use $\mathcal{U}$ to denote the set of all feasible utility functions induced by such v and c.

Let $\vec{u} = \{u_v\}_{v \in V}$ be the *utility profile* of all voters, and let $\vec{\sigma} = \{\sigma_v\}_{v \in V}$ be the *preference profile* of all voters. We use the notation $\vec{u} \triangleright \vec{\sigma}$ to indicate that u_v is consistent with σ_v for each voter $v \in V$. Voting rules have access to the rankings $\vec{\sigma}$, but *not* to the utilities $\vec{u}$.

If we know the utility functions u_v of all voters, a natural candidate selection criterion is *utilitarian welfare*, which is the sum of voter utilities: $\mathrm{UW}(\vec{u}, c) := \sum_{v \in V} u_v(c)$. In this case the winning candidate $c^*(\vec{u})$ is the one with highest welfare, i.e., $c^*(\vec{u}) \in \arg\max_{c \in C} \mathrm{UW}(\vec{u}, c)$.

Let f be a (possibly randomized) voting rule that maps a preference profile $\vec{\sigma}$ to a winner $c \in C$. For a fixed profile $\vec{\sigma}$, the *instance distortion* of f on $\vec{\sigma}$ is the worst-case ratio between the optimal utilitarian welfare and the utilitarian welfare of f; that is,

$$\mathrm{D}(f, \vec{\sigma}) := \max_{\vec{u} \triangleright \vec{\sigma}} \frac{\mathrm{UW}(\vec{u}, c^*(\vec{u}))}{\mathbb{E}_{c \sim f(\vec{\sigma})}[\mathrm{UW}(\vec{u}, c)]}.$$

For theoretical results, we are primarily interested in the overall *distortion* of f, which is the worst case over profiles:

$$\mathrm{D}(f) := \max_{\vec{u} \triangleright \vec{\sigma}, \vec{u} \in \mathcal{U}^n} \mathrm{D}(f, \vec{\sigma}).$$

We impose the following structural assumptions.

- **Non-negativity.** All voter and candidate vectors lie in the positive orthant, i.e., $v, c \in \mathbb{R}^d_{\geq 0}$. This ensures all utilities are non-negative, and avoids a mixed-sign objective.
- **Normalization.** We will assume $\|v\|_p = 1$ and $\|c\|_p = 1$ for all v, c. For $v \in V$ this can be viewed as constraining all voters to have equal influence on the mechanism (c.f. Aziz [2]), and for both candidates and voters as identifying the *relative* magnitude of embedding components. We focus on ℓ^p norms for $p = 1$ in the main body,[1] but also provide results for ℓ^2 normalization in the supplemental material.
- **Expressiveness.** For theoretical analysis, we further assume $V \subset \mathrm{Cone}(C)$, meaning every voter can be expressed as a non-negative linear combination of candidates. Intuitively, this can be understood as requiring that the candidate set is rich enough to describe voter preferences. This excludes the case where voters have 0 utility for all alternatives, ensuring each voter has meaningful preferences over the candidate set.

Normalization of both v and c is necessary to keep the model well-posed and the distortion finite, as the following example illustrates:

Example 1. *Consider $n-1$ voters who prefer c_1 to c_2 and one voter who strongly prefers c_2 due to a large utility spike in one coordinate. If candidate c_2 has an unbounded entry, it can yield unbounded utilitarian welfare even though $n-1$ voters rank c_1 above c_2.*

We study several established voting rules. We define them here for convenience.

Definition 1. *(Randomized Scoring Rules (RSRs)) Let $\vec{s} = (s^1, \ldots, s^m)$ be a scoring vector with $s^1 \geq s^2 \geq \cdots \geq s^m \geq 0$. For a candidate $c \in C$, let* $\mathrm{rank}_v(c)$ *denote the position of c in voter v's ranking (with $\mathrm{rank}_v(c) = 1$ meaning a is ranked first). Then the score assigned to c by agent v is:*

$$\mathrm{score}_v(c, \vec{s}) := s^{\mathrm{rank}_v(c)}.$$

The total score of c across all voters is:

$$\mathrm{score}_V(c, \vec{s}) := \sum_{i \in n} \mathrm{score}_v(c, \vec{s}).$$

The randomized scoring rule $f^{rand}_{\vec{s}}$ *selects each alternative $c \in C$ with probability proportional to its total score:*

$$\Pr[f^{rand}_{\vec{s}}(\vec{\sigma}) = c] := \frac{\mathrm{score}_V(c, \vec{s})}{n \cdot \|\vec{s}\|_1}.$$

Notable examples include *Random Dictatorship*, which corresponds to the plurality scoring rule $\vec{s} = (1, 0, \ldots, 0)$, and the *Randomized Harmonic* rule [5], which uses $\vec{s} = (1 + \frac{H_m}{m}, \frac{1}{2} + + \frac{H_m}{m}, \ldots, \frac{1}{m} + + \frac{H_m}{m})$, where H_m is the m^{th} harmonic number.

We will also make use of stable lotteries, which were shown to exist for any preference profile by Cheng et al. [7, Lemma 4].

Definition 2 (Stable Lotteries)**.** *Given a preference profile $\vec{\sigma}$, a committee W, and a candidate c, let $S_c(W) := \{v \in V : c \succ_v W\}$ be the set of voters who prefer c to all alternatives in W. We say a distribution $\mathcal{W}$ over committees of size k is a* stable lottery *if for all $c \in C$,*

$$\mathbb{E}_{W \sim \mathcal{W}}[|S_c(W)|] \leq \frac{n}{k}.$$

This has seen much recent use in computational social choice; our direct inspiration is its application by Ebadian et al. [10] to the design of distortion-optimal randomized rules in the unit-sum and related settings.

Special case: unit-sum utilities. The special case in which each candidate is a standard basis vector, i.e. $C = \{e_1, \ldots, e_d\}$, recovers the well-studied unit-sum model, in which distortion was first introduced [21], and for which the worst-case distortion-optimal randomized rule is has distortion $\Theta(\sqrt{m})$ [5, 10], while the worst-case optimal deterministic rule has distortion $\Theta(m^2)$ [6]. By generalizing this setting, we inherit its lower bounds for both deterministic and randomized rules (Theorems 2 and 4).

3 DETERMINISTIC RULES

What utility guarantees can deterministic rules provide in this setting? How well do prominent deterministic rules fare? We begin with a useful lower bound on every individual voter's utility for their favorite candidate. Throughout, we use $\mathrm{CH}(C)$ to denote the *convex hull* of a set C.

Lemma 1. *For any v and any candidates $C = \{c_j\}_{j \in [m]}$, the maximum utility of v is at least* $\max_{c \in C} u_v(c) \geq \frac{1}{d}$.

Proof. Let $\tilde{c} := \arg\max_{c \in C} v^T c$. For any $\alpha \in \Delta_m$, $v^T \tilde{c} \geq v^T(\sum_{i \in [m]} \alpha_i c_i) = \sum_{i \in [m]} \alpha_i (v^T c_i)$ since the maximum upper bounds any convex combination. Moreover, for ℓ^1 normalization, $v \in \mathrm{Cone}(C)$ implies $v \in \mathrm{CH}(C)$, i.e. $v = \sum_{i \in [m]} \beta_i c_i$ for some $\beta \in \Delta_m$. Substituting α with this specific β, we have $v^T \tilde{c} \geq v^T v \geq \frac{1}{d}$ by Cauchy-Schwarz (in particular, since $\|v\|_2 \cdot d \geq \|v\|_1 = 1$). □

An analogous claim (Lemma 2 in Appendix C) holds for ℓ^2 normalization.

[1] Note that this together with non-negativity is equivalent to assuming $v, c \in \Delta_d$, where Δ_k denotes the k-dimensional simplex.

A natural deterministic rule is plurality (f_{Plur}). In the unit-sum setting, f_{Plur} is known to have distortion $\Theta(\min(n, m) \cdot m)$, which is asymptotically optimal [6]. We find that f_{Plur} attains worst-case distortion $\Theta(\min(n, m) \cdot d)$ for linear utilities. This matches the unit-sum guarantee when $d = m$, but scales poorly for $m, n \gg d$. The proof is deferred to Section A.

Theorem 1. $\mathrm{D}(f_{\text{Plur}}) = \Theta(\min(m, n) \cdot d)$.

Is this dependence on m and n unavoidable for all deterministic rules, or can it be circumvented? For deterministic rules, we inherit lower bounds from the work of Caragiannis et al. [6] in the unit-sum setting.

Theorem 2. *[Theorem 1 of Caragiannis et al. [6]] Suppose $m, n \geq d$. Then for any deterministic voting rule f, we have $\mathrm{D}(f) = \Omega(d^2)$.*

This lower bound depends only on d, and raises the question of whether it is possible to improve upon plurality by avoiding a dependence on n and m in the distortion bound. Indeed it is: we introduce a rule we dub *maximum coordinate plurality (MCP)* and denote by f_{MCP}.

Definition 3 (Maximum Coordinate Plurality). *Let*

$$\hat{C} := \left\{ c_i \in C \mid c_i = \arg\max_{c \in C} c^i \text{ for each } i \in [d] \right\}$$

be a set of at most d candidates such that for each coordinate i, a candidate maximal in that coordinate is included. The maximum coordinate plurality rule f_{MCP} *then restricts the profile to $\hat{C}$ and selects the plurality winner.*

Theorem 3. $\mathrm{D}(f_{\text{MCP}}) = O(d^3)$.

Proof. For any fixed voter v, it has its maximum coordinate at least $\frac{1}{d}$; call this coordinate i. Since $v \in \mathrm{CH}(C)$, the candidate $c_i \in \hat{C}$ therefore has i-th coordinate at least $\frac{1}{d}$, and so the welfare conferred to v by their favorite choice in $\hat{C}$ is at least $\max_{c \in \hat{C}} v^T c \geq \frac{1}{d^2}$. Since $|\hat{C}| \leq d$, this plurality winner $\hat{c} \in \hat{C}$ receives at least $\frac{n}{d}$ votes, and so $\mathrm{UW}(\hat{c}) \geq \frac{n}{d} \frac{1}{d^2}$. As $\max_{c \in C} \mathrm{UW}(c) \leq n$, the claim follows. □

Though this guarantee still exceeds the lower bound by a factor of d, the $\Omega(d^2)$ lower bound is already quite large. Can this be improved by randomizing over candidates?

4 RANDOMIZED RULES

As previously mentioned, we also inherit a distortion lower bound for *any* randomized rule from the unit-sum setting.

Theorem 4. *[5] Suppose $m \geq d$ and $n \geq \sqrt{d}$. Then there exists a preference profile $\vec{\sigma}$ such that for any randomized rule f, we have $\mathrm{D}(f, \vec{\sigma}) = \Omega(\sqrt{d})$.*

We include a proof for completeness in Appendix A.

4.1 Known Candidate Embeddings

Consider the center of the simplex Δ_d, denoted by $\mu := (\frac{1}{d}, \ldots, \frac{1}{d})$. Observe that $\mathrm{UW}(\mu) = \frac{n}{d}$, which implies a distortion at most d is possible when $\mu \in C$. However in general, $\mu \notin \mathrm{CH}(C)$. This motivates the goal of *approximating* the uniform candidate μ by a point within the convex hull of candidates $\mathrm{CH}(C)$, assuming the candidate embeddings are known. We therefore introduce a randomized rule whose output distribution matches the *reverse information projection* of μ onto $\mathrm{CH}(C)$.

Definition 4 (Uniform Projection Rule (f_{UProj})). *Given candidate locations $C \subset \mathbb{R}^d_{\geq 0}$, the* uniform projection rule f_{UProj} *defines a distribution $\{p_c\}_{c \in C}$ over candidates such that the expected candidate vector $\hat{c} := \sum_{c \in C} p_c \cdot c$ minimizes the Kullback–Leibler (KL) divergence from the uniform vector, i.e., $\hat{c} := \arg\min_{x \in \mathrm{CH}(C)} \mathrm{KL}(\mu \| x)$.*

Theorem 5. *The expected welfare of f_{UProj} is at least $\mathrm{UW}(f_{\text{UProj}}) \geq \frac{n}{d}$. As a consequence, $\mathrm{D}(f_{\text{UProj}}) = O(d)$.*

Proof of Theorem 5. First observe that

$$\mathrm{KL}(\mu \| x) = \sum_{i=1}^{d} \mu^i \ln \frac{\mu^i}{x^i} = \sum_i \mu^i \ln \mu^i \; - \; \sum_{i=1}^{d} \mu^i \ln x^i,$$

so minimizing $\mathrm{KL}(\mu \| x)$ over the convex set $\mathrm{CH}(C)$ is equivalent to minimizing the smooth, convex function $f(x) = -\sum_{i=1}^d \mu^i \ln x^i$ subject to $x \in \mathrm{CH}(C)$. The first-order optimality condition gives $\nabla f(x^*)^\top (v - x^*) \;\geq\; 0$ for every voter $v \in \mathrm{CH}(C)$. Since $\frac{\partial f}{\partial x^i}(x) = -\frac{\mu^i}{x^i}$, we have

$$\nabla f(x^*)^\top (v - x^*) = -\sum_{i=1}^{d} \frac{\mu^i}{x^{*i}} (v^i - x^{*i}) \;\geq\; 0$$

$$\implies \; \sum_{i=1}^{d} \frac{\mu^i \, v^i}{x^{*i}} \;\leq\; \sum_{i=1}^{d} \mu^i = 1.$$

Since $\mu^i = \frac{1}{d}$, this yields $\sum_{i=1}^d \frac{v^i}{x^{*i}} \;\leq\; d$. Now define the auxiliary sequences $a_i := \sqrt{\frac{v^i}{x^{*i}}}$ and $b_i := \sqrt{v^i \, x^{*i}}$. Then $\sum_{i=1}^d a_i \, b_i = \sum_{i=1}^d v^i = 1$ and, by, Cauchy–Schwarz,

$$\begin{aligned} 1 = \Big(\sum_i a_i b_i\Big)^2 &\leq \Big(\sum_i a_i^2\Big)\Big(\sum_i b_i^2\Big) \\ &= \Big(\sum_i \tfrac{v^i}{x^{*i}}\Big)\Big(\sum_i v^i \, x^{*i}\Big). \end{aligned}$$

Combining with $\sum_i v^i / x^{*i} \leq d$ gives $\sum_{i=1}^d x^{*i} \, v^i \geq 1/d$. Now as we have n voters and the above inequality holds for arbitrary v, the total utilitarian welfare is at least $\frac{n}{d}$.

The bound on the distortion of f_{UProj} then follows because the maximum utility for any voter is at most 1. □

This randomized $O(d)$-distortion rule can be seen as generalizing uniform candidate selection in the unit-sum setting. A natural candidate for improving upon this is the harmonic rule f_{HR}, which obtains near-optimal distortion $\Theta(\sqrt{m \log m})$ [4, 5] for unit-sum utilities. However this performance does not generalize to linear utilities, wherein the introduction of duplicate candidates does not constrain the underlying utility profile. Instead, it turns out that here f_{HR} has distortion unbounded in d; we discuss this and other randomized positional scoring rules in Section 4.2.

Fortunately, Theorem 5 does indirectly lead to distortion sublinear in d. Using it, we may adapt the stable lottery rule of [10] to the linear utilities setting to achieve a better distortion bound.

Definition 5 (Linear Stable Lottery Rule (f_{LSLR})). *Given a stable lottery $\mathcal{W}$ over committees of size $k = \sqrt{d}$, the* linear stable lottery rule f_{LSLR} *on profile $\vec{\sigma}$ chooses each $c \in C$ with probability* $\frac{1}{2\sqrt{d}} \Pr_{W \sim \mathcal{W}(\vec{\sigma})}[c \in W] + \frac{1}{2} \Pr_{c' \sim f_{\text{UProj}}(\vec{\sigma})}[c' = c]$.

Here f_{UProj} takes the place of uniform selection.

Theorem 6. $\mathrm{D}(f_{\text{LSLR}}) = O(\sqrt{d})$.

Proof. We follow the proof of Ebadian et al. [10] closely, albeit in our notation. The principal difference is our use of f_{UProj} instead of uniform selection over candidates.

To that end, we combine Definition 2 and Theorem 5 in order to relate the expected welfare conferred from each part of f_{LSLR}. For any committee W of size k and any $a \in W$, let $S_a(W) \subseteq V$ denote the voters for which $a \succ_v W$, and $\overline{S}_a(W)$ its complement. Then

$$
\begin{aligned}
\sum_{v \in V} \sum_{c \in W} u_v(c) &= \sum_{v \in S_a(W)} \sum_{c \in W} u_v(c) + \sum_{v \in \overline{S}_a(W)} \sum_{c \in W} u_v(c) \\
&\geq \sum_{v \in S_a(W)} \sum_{c \in W} u_v(c) + \sum_{v \in \overline{S}_a(W)} u_v(c^*) \\
&\geq \sum_{v \in S_a(W)} (u_v(c^*) - 1) + \sum_{v \in \overline{S}_a(W)} u_v(c^*) \\
&= \sum_{v \in V} u_v(c^*) - |\{S_a(W)\}|. \qquad (1)
\end{aligned}
$$

Taking the expectation over a stable lottery $\mathcal{W}$ and applying Definition 2, we obtain

$$
\begin{aligned}
\mathbb{E}_{W \sim \mathcal{W}}\left[\sum_{c \in W} \mathrm{UW}(c)\right] &\geq \mathrm{UW}(c^*) - \mathbb{E}_{W \sim \mathcal{W}}\left[|\{S_a(W)\}|\right] \\
&\geq \mathrm{UW}(c^*) - \frac{n}{k} \\
&\geq \mathrm{UW}(c^*) - \frac{d}{k} \mathbb{E}_{c \sim f_{\text{UProj}}}\left[\mathrm{UW}(c)\right], \qquad (2)
\end{aligned}
$$

where the last step follows from Theorem 5.

We now let $x = \frac{1}{2}x_1 + \frac{1}{2}x_2$ denote the distribution over C of f_{LSLR}, where x_1 is the stable lottery part and x_2 is the f_{UProj} part. Then applying (2) and letting $k = \sqrt{d}$,

$$
\begin{aligned}
k \cdot \mathrm{UW}(x_1) &= \mathbb{E}_{W \sim \mathcal{W}}\left[\sum_{c \in W} \mathrm{UW}(c)\right] \\
&\geq \mathrm{UW}(c^*) - \frac{d}{k} \mathbb{E}_{c \sim f_{\text{UProj}}}\left[\mathrm{UW}(c)\right], \\
k \cdot \mathrm{UW}(x_1) + \frac{d}{k}\mathrm{UW}(x_2) &\geq \mathrm{UW}(c^*) \\
\mathbb{E}_{c \sim f_{\text{LSLR}}}\left[\mathrm{UW}(c)\right] &\geq \frac{1}{2\sqrt{d}} \mathrm{UW}(c^*). \\
\frac{\mathrm{UW}(c^*)}{\mathbb{E}_{c \sim f_{\text{LSLR}}}\left[\mathrm{UW}(c)\right]} &\leq 2\sqrt{d}
\end{aligned}
$$

This proves the claim. □

Computing a stable lottery $\mathcal{W}$ requires only ordinal information; however in order to identify the distribution from which f_{UProj} samples, the embedding of C into $\mathbb{R}^d_{\geq 0}$ must be known by the rule. This raises the question of what can be done with ordinal preferences when both voter *and* candidate embeddings are unknown.

4.2 Unknown Candidate Embeddings

Even when the locations of the candidate vectors $c \in \mathbb{R}^d_{\geq 0}$ are unknown, some established rules exhibit distortion that is bounded as a function of d. Our primary example is random dictatorship, for which proof is deferred to Appendix A.

Theorem 7. $\mathrm{D}(f_{\text{RD}}) = \Omega(d)$, *and also* $\mathrm{D}(f_{\text{RD}}) = O(d^3)$.

Proof of Theorem 7. The $\Omega(d)$ lower bound for the distortion of f_{RD} in the unit-sum setting is straightforward; we reproduce it here for completeness. Consider $m = d$, with each candidate embedded to a unit basis vector; i.e. $C = [d]$ and $c_i = e_i$ for each i. We will consider $d-1$ equal-size groups of voters V_i for $i \in [d-1]$. The voters in V_i have embedding $v = (e_i + e_d)/2$ (or $v = e_i(1/2 + \epsilon) + e_d(1/2 - \epsilon)$ for some small $\epsilon > 0$, if ties must be avoided), and so their rankings will be $c_i \succ_v c_d \succ_v \ldots$. On such an instance, f_{RD} chooses uniformly from $[d-1]$ and $\mathrm{UW}(f_{\text{RD}}) = \frac{n}{2(d-1)}$, while the welfare optimizer is c_d with $\mathrm{UW}(c_d) = \frac{n}{2}$. Therefore on this profile $\vec{\sigma}$ we have

$$\mathrm{D}(f_{\text{RD}}, \vec{\sigma}) = \frac{1}{d-1} = \Omega(d).$$

We now turn to upper bounding the distortion of f_{RD}. For each $v \in V$, let c_v denote the top-ranked candidate for v. By Lemma 1 we have that $u_v(c_v) = v^T c_v \geq 1/d$. So by an averaging argument over the d dimensions, there is some $\ell \in [d]$ such that $v^\ell c_v^\ell \geq 1/d^2$. Since $v, c_v \in \Delta_d$, this implies that $v^\ell \geq 1/d^2$ and $c_v^\ell \geq 1/d^2$.

Fix one such coordinate ℓ_v for each v and partition the voters V into groups $V_1, \ldots, V_d$ according to their ℓ_v. Let $p_\ell := |V_\ell|/n$ encode the proportion of voters in each group, and observe that $p \in \Delta_d$. We can now lower bound the expected welfare from f_{RD} by observing that if the dictator is chosen from group V_i, all other voters in V_i will also obtain some welfare from their choice.

$$
\begin{aligned}
E_{c \sim RD}\left[\mathrm{UW}(c)\right] &= \frac{1}{n} \sum_{v \in V} \mathrm{UW}(c_v) \\
&\geq \frac{1}{n} \sum_{v \in V} \sum_{v' \in V_{\ell_v}} v'^T c_v \\
&\geq \frac{1}{n} \sum_{v \in V} \sum_{v' \in V_{\ell_v}} (v')_{\ell_v} (c_v)_{\ell_v}.
\end{aligned}
$$

From here we will rearrange the sum to range over the groups v^ℓ. Observe that $v^\ell c_v^\ell \geq 1/d^2$ implies that $c_v^\ell \geq \frac{1}{v^\ell d^2}$. Then continuing,

$$
\begin{aligned}
&= \frac{1}{n} \sum_{\ell} \sum_{v, v' \in V_\ell} v^\ell c_\ell^{v'} \\
&\geq \frac{1}{n} \sum_{\ell} \sum_{v, v' \in V_\ell} v^\ell \frac{1}{d^2 v'_\ell} \\
&\geq \frac{1}{n} \sum_{\ell} \sum_{v, v' \in V_\ell : v^\ell \geq v'_\ell} v^\ell \frac{1}{d^2 v'_\ell} \\
&\geq \frac{1}{n} \sum_{\ell} \sum_{v, v' \in V_\ell : v^\ell \geq v'_\ell} v^\ell \frac{1}{d^2 v^\ell} \\
&\geq \frac{1}{nd^2} \sum_{\ell} \frac{1}{2} |v^\ell|^2 = \frac{n}{2d^2} p^T p \geq \frac{n}{2d^3}.
\end{aligned}
$$

The claim follows from the fact that welfare is at most n. □

We conclude by remarking that while the analysis of f_{RD} in this proof in some sense proceeds on a dimension-by-dimension basis, the rule is itself agnostic to (and indeed has no knowledge of) these coordinates or the embeddings of V and C. This is in contrast to f_{MCP}, where both the rule and its analysis proceed along these lines.

One feature of this model is that the worst-case distortion of many randomized positional scoring rules (Appendix 1) is quite similar. In particular, consider the RSRs where, for given m, the score vector is given by $\vec{s} = \frac{1}{S_m}(s_1, s_2, \ldots, s_m)$ for some fixed sequence $s_1, s_2, \ldots$, and where $S_i = \sum_{j \le i} s_j$. (Note that this contains f_{RD} and f_{HR} and the uniform distribution over candidates, but not Borda.) Then there are two cases. If $S_m \to S$ converges, then the distortion of f_s on *any* instance is within S/s_1 of f_{RD}, and we can make the behavior of f_s approach that of f_{RD} by cloning each candidate sufficiently many times. And if S_m diverges, then cloning bad candidates can lead to poor performance on instances which are easy for f_{RD}. We illustrate this by relating the performance of f_{HR} to that of f_{RD}, the proof of which also appears in Appendix A.

Theorem 8. *The distortion of f_{HR} is unbounded as a function of d. However, for any profile $\vec{\sigma}$,*

$$\mathrm{D}(f_{\text{HR}}, \vec{\sigma}) \le \mathrm{D}(f_{\text{RD}}, \vec{\sigma}) \cdot (\log m + 1).$$

Proof. We start with the first claim. Consider the instance for which $d = 2$, all voters are $v = (1, 0)$ and perfectly aligned with $c = (1, 0)$, and all other candidates are $c' = (0, 1)$. Then $\mathrm{UW}(f_{\text{HR}}) = \frac{n}{2H_m}$, where $H_m = \log m + \Theta(1)$ is the m^{th} harmonic number. On the other hand $\mathrm{UW}(c) = n$, so for this instance family $\{\vec{u}_m\}$ we have $\mathrm{D}(f_{\text{HR}}, \vec{u}_m) = \Theta(\log^{-1} m) \to 0$ as m grows large.

We now address the second. Observe that on any profile $\vec{\sigma}$ induced by a utility profile instance $\vec{u}$, and for every alternative a, $\Pr_{c\sim f_{\text{HR}}(\vec{\sigma})}[c = a] \ge \Pr_{c\sim f_{\text{RD}}(\vec{\sigma})}[c = a]/H_m$. Therefore

$$\begin{aligned}
\mathrm{UW}(f_{\text{HR}}(\vec{\sigma})) &= \sum_a \Pr_{c\sim f_{\text{HR}}(\vec{\sigma})}[c = a]\,\mathrm{UW}(a) \\
&\ge \frac{1}{H_m} \sum_a \Pr_{c\sim f_{\text{RD}}(\vec{\sigma})}[c = a]\,\mathrm{UW}(a) \\
&= \frac{\mathrm{UW}(f_{\text{RD}}(\vec{\sigma}))}{H_m}.
\end{aligned}$$

The claim follows from $H_m \le \log m + 1$. □

Thus, the distortion of f_{HR} is never much better than that of f_{RD}, and if—as we expect—the worst-case profiles for f_{RD} have $\mathrm{poly}(d)$ candidates, then its worst-case distortion cannot be much better than that of f_{RD}, even for small m.

Can we improve upon this bound when candidate embeddings are unknown? The role that f_{UProj} plays in the design and analysis of f_{LSLR}—and the role uniform selection plays in the stable lotteries for Ebadian et al. [10]—is in some sense both an absolute lower bound on welfare when the maximum candidate welfare is not large, and sample access to it. We might then let f_{RD} take the role of f_{UProj} in f_{LSLR} to attain distortion $O(d^{3/2})$ for unknown embeddings.

However even in the absence of a better rule, it turns out stable lotteries can still be used by furnishing a direct lower bound on the maximum welfare. Consider the following:

Definition 6 (Pure Stable Lottery Rule (f_{PSLR})). *Given a stable lottery $\mathcal{W}$ over committees of size $2d$, the* pure stable lottery rule *f_{PSLR} chooses $c \in C$ w.p. $\frac{1}{2d}\Pr_{W\sim\mathcal{W}}[c \in W]$.*

Using larger committees and that $\max_{c\in C}\mathrm{UW}(c) \ge \frac{n}{d}$, we avoid the need for a second sampling component.

Theorem 9. $\mathrm{D}(f_{\text{PSLR}}) = O(d)$.

Proof. As in the proof of Theorem 6, for any committee W of size k and any $a \in W$, let $S_a(W) \subseteq V$ denote the voters for which $a \succ_v W$, and $\overline{S}_a(W)$ its complement. Then via identical derivation we obtain

$$\sum_{v\in V}\sum_{c\in W} u_v(c) \ge \sum_{v\in V} u_v(c^*) - |\{S_a(W)\}|. \tag{3}$$

Taking the expectation over a stable lottery $\mathcal{W}$ and applying Definition 2 with $k = 2d$, we obtain

$$\begin{aligned}
\mathbb{E}_{W\sim\mathcal{W}}\left[\sum_{c\in W}\mathrm{UW}(c)\right] &\ge \mathrm{UW}(c^*) - \mathbb{E}_{W\sim\mathcal{W}}\left[|\{S_a(W)\}|\right] \\
&\ge \mathrm{UW}(c^*) - \frac{n}{2d} \\
&\ge \frac{1}{2}\cdot\mathrm{UW}(c^*),
\end{aligned} \tag{4}$$

where the last step follows from the fact that $\mathrm{UW}(c^*) \ge n/d$ by Theorem 5. Applying (4) and recalling our choice of $k = 2d$, we have

$$\begin{aligned}
k \cdot \mathbb{E}_{c\sim f_{\text{PSLR}}}[\mathrm{UW}(c)] &= \mathbb{E}_{W\sim\mathcal{W}}\left[\sum_{c\in W}\mathrm{UW}(c)\right] \\
&\ge \frac{1}{2}\cdot\mathrm{UW}(c^*),
\end{aligned}$$

and so

$$\mathbb{E}_{c\sim f_{\text{LSLR}}}[\mathrm{UW}(c)] \ge \frac{1}{4d}\cdot\mathrm{UW}(c^*).$$

This proves the claim. □

5 OPTIMIZING DISTORTION

In the preceding sections we analyzed the asymptotic behavior of voting rules, but the lower bounds are often driven by pathological cases. In practice, we are more interested in this welfare approximation guarantee for a given profile. That is: *Given a preference profile, can we compute the best possible distortion—and design a rule that achieves it?*

The computational tractability of the distortion-instance-optimal rule in the unit-sum setting was established by Boutilier et al. [5] and clarified by Ebadian et al. [9]. For linear utilities we also answer both questions affirmatively, showing that instance-optimal rules—both deterministic and randomized—can be computed efficiently.

Definition 7 (Feasible Region for $\bar{v}$). *Let $\mathcal{F}_j = \{v \in \Delta_d : \langle c_a - c_b, v\rangle \ge 0$ whenever $c_a \succ_j c_b\}$ be the set of voter vectors consistent with σ_{v_j}. Then the feasible region for the average voter $\bar{v} = \frac{1}{n}\sum_{j=1}^{n} v_j$ is $\mathcal{F} = \frac{1}{n}\sum_{j=1}^{n}\mathcal{F}_j$.*

Theorem 10 (Deterministic Instance-Optimal Rule). *Given a preference profile $\vec{\sigma}$ and any $\varepsilon > 0$, one can compute a deterministic voting rule f such that $\mathrm{D}(f, \vec{\sigma}) \le \min_{c \in C} \mathrm{D}(c, \vec{\sigma}) + \varepsilon$ in time* $\mathrm{poly}(n, m, \log \frac{1}{\varepsilon})$.

Proof. Let $\mathcal{F}$ be the feasible region for the average voter vector $\bar{v}$ induced by $\vec{\sigma}$ (Definition 7). For each pair $c_1, c_2 \in C$, we aim to compute the best possible upper bound on $\frac{\mathrm{UW}(c_1)}{\mathrm{UW}(c_2)} = \frac{\bar{v}^\top c_1}{\bar{v}^\top c_2}$ over all $\bar{v} \in \mathcal{F}$. Since we cannot compute this ratio exactly without knowing $\bar{v}$, we upper-bound it by the largest value $\frac{1}{\beta}$ such that $\bar{v}^\top c_1 \le \frac{1}{\beta} \bar{v}^\top c_2 \; \forall \, \bar{v} \in \mathcal{F}$.

Equivalently, for fixed c_1, c_2, we find the largest $\beta_{c_1,c_2} \in [0, 1]$ such that the following LP has a non-negative optimal value: $\min_{\bar{v} \in \mathcal{F}} \langle \beta c_1 - c_2, \bar{v} \rangle$. This is doable via binary search over β to precision ε, requiring $O(\log \frac{1}{\varepsilon})$ iterations per pair.

Once the distortion bounds are computed for all $O(m^2)$ candidate pairs, we select the candidate $\hat{c} \in C$ minimizing the worst-case distortion $\hat{c} := \arg\min_{c \in C} \max_{c' \in C} \beta_{c',c}$. Returning this candidate defines the deterministic rule f, and by construction we have $\mathrm{D}(f, \vec{\sigma}) \le \min_{c \in C} \mathrm{D}(c, \vec{\sigma}) + \varepsilon$. Since each linear program is solvable in polynomial time in n, m, d, and the number of binary search iterations is $O(\log \frac{1}{\varepsilon})$, the total runtime is $\mathrm{poly}(n, m, d, \log \frac{1}{\varepsilon})$. □

Note that the procedure described above can be used to compute the distortion of any rule on a given preference profile. We also apply it in our experiments.

Theorem 11 (Randomized Instance-Optimal Rule). *Given a preference profile $\vec{\sigma}$ and any $\varepsilon > 0$, a randomized voting rule f with $\mathrm{D}(f, \vec{\sigma}) \le \min_{f' \in \Delta(C)} \mathrm{D}(f', \vec{\sigma}) + \varepsilon$ can be computed in time $\mathrm{poly}(n, m, \frac{1}{\varepsilon})$, where the minimum is taken over all distributions over candidates.*

Proof. Let $p \in \Delta_C$ be a distribution over candidates, and let $\hat{c} := \sum_{i=1}^m p_i c_i$ be the expected candidate vector under p. We aim to find a distribution p that minimizes the distortion:

$$\mathrm{D}(p, \vec{\sigma}) = \sup_{\bar{v} \in \mathcal{F}} \frac{\max_{c \in C} \langle c, \bar{v} \rangle}{\langle \hat{c}, \bar{v} \rangle}.$$

Since this expression may be unbounded, we instead maximize its reciprocal $\beta \in [0, 1]$, subject to the constraint:

$$\langle \hat{c}, \bar{v} \rangle \ge \beta \cdot \max_{c \in C} \langle c, \bar{v} \rangle \quad \forall \, \bar{v} \in \mathcal{F}.$$

This yields the following optimization problem:

$$\max_{\substack{p \in \Delta(C) \\ \beta \in [0,1]}} \left\{ \beta : \sum_{i=1}^m p_i \langle c_i, \bar{v} \rangle \ge \beta \cdot \max_{c \in C} \langle c, \bar{v} \rangle \quad \forall \, \bar{v} \in \mathcal{F} \right\}.$$

Although $\mathcal{F}$ contains infinitely many constraints, we can apply the ellipsoid method using a separation oracle. For a candidate solution $(\hat{\beta}, \hat{p})$, feasibility reduces to verifying:

$$\sum_{i=1}^m \hat{p}_i \langle c_i, \bar{v} \rangle \ge \hat{\beta} \cdot \langle c, \bar{v} \rangle \qquad \forall \, c \in C, \; \bar{v} \in \mathcal{F}.$$

For each $c \in C$, this can be checked by solving an LP:

$$\min_{\bar{v} \in \mathcal{F}} \quad \left\langle \sum_{i=1}^m \hat{p}_i c_i - \hat{\beta} \cdot c, \; \bar{v} \right\rangle.$$

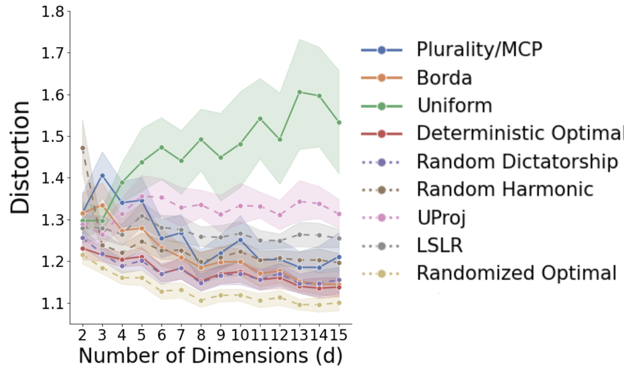

Figure 1: Instance distortion $\mathrm{D}(f, \vec{\sigma})$ for MovieLens

If the minimum is nonnegative for all $c \in C$, the solution is feasible; otherwise, the auxiliary LP provides a separating hyperplane. Since each $\bar{v} \in \mathcal{F}$ is defined via $O(m^2 n)$ linear constraints, each LP is of polynomial size. Running the ellipsoid method with this oracle to precision ε yields a solution in time $\mathrm{poly}(n, m, d, \frac{1}{\varepsilon})$. □

Practical implementation. Although the ellipsoid method is polynomial-time, it is often impractical. We instead use the more efficient *column generation* [8] (see Appendix B). As the algorithm uses only pairwise comparisons, it also supports partial rankings.

6 EXPERIMENTS

We evaluate our instance-optimal rules alongside f_{UProj}, f_{MCP}, f_{LSLR}, and several other common rules, on two real-world datasets. Specifically, we measure both the *instance distortion* (relative to the underlying utility profile $\vec{u}$) and the *worst-case distortion* (over all utility profiles consistent with the observed preferences $\vec{\sigma}$). Due to space constraints, we report results for the latter, and only as a function of the dimension d in the main text. Results varying n, m, running times, and instance distortion, are deferred to Appendix D.

Datasets. We consider the MovieLens 100K [14] and abortion opinion survey [11] datasets. MovieLens contains 100K movie ratings from 1000 users (voters) over 1700 movies. We translate it into our setting via approximate matrix decomposition for each dimension d, adding the nonnegativity and normalization constraints. We then randomly subsample 20 movie preference votes for each dimension d with $n = 100$ and $m = 25$. The abortion opinions dataset includes ratings from 100 individuals on 5 different opinion statements. For this setting, we embed the candidates using Matryoshka embeddings [16], followed by dimensionality reduction via principal component analysis (PCA). The embedding choice for either dataset is not essential to our method and is intended only as a representative example.

Results. Figure 2 presents instance-based distortion bounds for MovieLens (left) and abortion opinion survey (right). Our first observation is that, as we expect, the instance-optimal approaches we propose outperform all alternatives, with improvement appearing to increase with d. This is true for both deterministic and randomized rules. Second, and somewhat surprisingly, we find that worst-case distortion bounds *decrease with dimension d*, despite

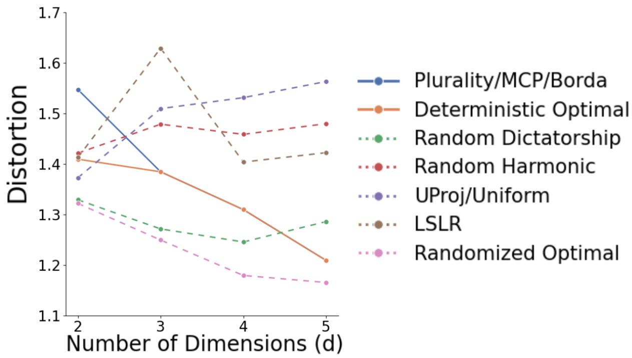

Figure 2: Instance distortion $\mathrm{D}(f, \vec{\sigma})$ for Abortion Survey

the fact that distortion lower bounds increase in d. Thus, as the dimension of candidate representations increases, using rankings in place of utilities becomes empirically less consequential.

7 CONCLUSION

We initiate the study of distortion in the setting of linear social choice, which we anticipate will play a role in the integration of social choice into neural and representation-based models. For deterministic rules we introduce the maximum coordinate plurality (MCP) rule, which we prove obtains worst-case distortion $O(d^3)$ and is within $\Theta(d)$ of optimal. Since MCP requires access to candidate positions, we leave open whether d-dependent distortion is even *possible* for deterministic rules without such access.

For randomized rules, we construct the *linear stable lotteries rule*, and show it attains asymptotically optimal worst-case distortion $\Theta(\sqrt{d})$ when candidate locations are known. We show stable lotteries yield $O(d)$ distortion even when candidate locations are unknown, leaving unresolved the optimal distortion in this setting. We also establish preliminary results for ℓ^2 normalization.

Beyond the utilitarian welfare objective, future work might also explore *Nash welfare* or *proportional fairness*, as studied by Ebadian et al. [10], or worst-case *regret* minimization [6, 15]. Finally, given the connection between linear social choice and reward learning, we plan to investigate the robustness of various rules under both learning errors and strategic voters.

ACKNOWLEDGMENTS

We thank Dominik Peters for helpful suggestions regarding the implementation of stable lotteries. This work was supported in part by the NSF (IIS-2214141, CCF-2403758), ARO (W911NF-25-1-0059), ONR (N00014-24-1-2663), the Foresight Institute, and Amazon.

REFERENCES

[1] Elliot Anshelevich, Aris Filos-Ratsikas, Nisarg Shah, and Alexandros A Voudouris. 2021. Distortion in Social Choice Problems: The First 15 Years and Beyond. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21)*.

[2] Haris Aziz. 2020. Justifications of Welfare Guarantees Under Normalized Utilities. *ACM SIGecom Exchanges* (2020).

[3] Ratip Emin Berker, Sílvia Casacuberta, Isaac Robinson, Christopher Ong, Vincent Conitzer, and Edith Elkind. 2025. From Independence of Clones to Composition Consistency: A Hierarchy of Barriers to Strategic Nomination. In *Proceedings of the 26th ACM Conference on Economics and Computation*.

[4] Umang Bhaskar, Varsha Dani, and Abheek Ghosh. 2018. Truthful and Near-Optimal Mechanisms for Welfare Maximization in Multi-winner Elections. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

[5] Craig Boutilier, Ioannis Caragiannis, Simi Haber, Tyler Lu, Ariel D Procaccia, and Or Sheffet. 2015. Optimal Social Choice Functions: A Utilitarian View. *Artificial Intelligence* (2015).

[6] Ioannis Caragiannis, Swaprava Nath, Ariel D Procaccia, and Nisarg Shah. 2017. Subset Selection via Implicit Utilitarian Voting. *Journal of Artificial Intelligence Research* (2017).

[7] Yu Cheng, Zhihao Jiang, Kamesh Munagala, and Kangning Wang. 2020. Group Fairness in Committee Selection. *ACM Transactions on Economics and Computation (TEAC)* (2020).

[8] George B Dantzig and Philip Wolfe. 1960. Decomposition Principle for Linear Programs. *Operations Research* (1960).

[9] Soroush Ebadian, Aris Filos-Ratsikas, Mohamad Latifian, and Nisarg Shah. 2024. Computational Aspects of Distortion. In *The 23rd International Conference on Autonomous Agents and Multiagent Systems*.

[10] Soroush Ebadian, Anson Kahng, Dominik Peters, and Nisarg Shah. 2024. Optimized Distortion and Proportional Fairness in Voting. *ACM Transactions on Economics and Computation* (2024).

[11] Sara Fish, Paul Gölz, David C. Parkes, Ariel D. Procaccia, Gili Rusak, Itai Shapira, and Manuel Wüthrich. 2024. Generative Social Choice. In *Proceedings of the 25th ACM Conference on Economics and Computation*.

[12] Luise Ge, Daniel Halpern, Evi Micha, Ariel D Procaccia, Itai Shapira, Yevgeniy Vorobeychik, and Junlin Wu. 2024. Axioms for AI Alignment from Human Feedback. In *Advances in Neural Information Processing Systems*.

[13] Paul Gölz, Nika Haghtalab, and Kunhe Yang. 2025. Distortion of AI Alignment: Does Preference Optimization Optimize for Preferences?. In *Advances in Neural Information Processing Systems*.

[14] F Maxwell Harper and Joseph A Konstan. 2015. The MovieLens Datasets: History and Context. *ACM Transactions on Interactive Intelligent Systems (TiiS)* (2015).

[15] Anson Kahng and Gregory Kehne. 2022. Worst-Case Voting When the Stakes Are High. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

[16] Aditya Kusupati, Gantavya Bhatt, Aniket Rege, Matthew Wallingford, Aditya Sinha, Vivek Ramanujan, William Howard-Snyder, Kaifeng Chen, Sham Kakade, Prateek Jain, and Others. 2022. Matryoshka Representation Learning. In *Advances in Neural Information Processing Systems*.

[17] Annie Liang. 2025. Artificial Intelligence Clones. In *Proceedings of the 26th ACM Conference on Economics and Computation, EC 2025*.

[18] Belle Lin. 2024. Grindr Aims to Build the Dating World's First AI 'Wingman'. *The Wall Street Journal* (2024). October 5.

[19] Patrick Ngatchou, Anahita Zarei, and A El-Sharkawi. 2005. Pareto Multi Objective Optimization. In *Proceedings of the 13th International Conference on Intelligent Systems Application to Power Systems*. IEEE.

[20] David M Pennock, Eric Horvitz, C Lee Giles, and Others. 2000. Social Choice Theory and Recommender Systems: Analysis of the Axiomatic Foundations of Collaborative Filtering. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

[21] Ariel D Procaccia and Jeffrey S Rosenschein. 2006. The Distortion of Cardinal Preferences in Voting. In *Proceedings of the International Workshop on Cooperative Information Agents*. Springer.

[22] Ariel D Procaccia, Benjamin Schiffer, and Shirley Zhang. 2025. Clone-Robust AI Alignment. In *Proceedings of the International Conference on Machine Learning*. PMLR.

[23] Eli Tan. 2024. Dating Apps Suck. A.I. Clones Are Making Them Even Weirder. *The New York Times* (2024). December 11.

[24] Banghua Zhu, Michael Jordan, and Jiantao Jiao. 2023. Principled Reinforcement Learning with Human Feedback from Pairwise or K-wise Comparisons. In *Proceedings of the International Conference on Machine Learning*. PMLR.

A OMITTED PROOFS

Theorem 1. $\mathrm{D}(f_{\mathrm{Plur}}) = \Theta(\min(m, n) \cdot d)$.

Proof of Theorem 1. We begin with the upper bound. If $m \leq n$, by Lemma 1, we know the plurality winner must have social welfare at least $\frac{n}{m} \cdot \frac{1}{d}$. Otherwise, the plurality winner still must have social welfare at least $\frac{1}{d}$. And the best possible candidate's social welfare is at most n. Hence, $\mathrm{D}(f_{\mathrm{Plur}}) = O(\min(m, n) \cdot d)$.

We now demonstrate an instance for which this is tight. Let $\mu := (\frac{1}{d}, \ldots, \frac{1}{d})$ and consider a candidate set where $c_1 = \mu$ and $c_2 = \cdots = c_m = (1, 0, \ldots, 0)$. First, assume $n \geq m$. Construct a profile where two voters vote for c_1 and each of the remaining $n - 2$ voters votes for a distinct candidate among $c_2, \ldots, c_m$. Let $v_1 = v_2 = \mu$ and $v_3, \ldots, v_n = (1, 0, \ldots, 0)$. The social welfare of c_1 is $2/d$, while any other candidate has welfare 1 from each of the other $n - 2$ voters and $1/d$ from the first two, giving at least $n - 2 + 2/d$. Thus, the distortion is $\mathrm{D}(f_{\mathrm{Plur}}, \vec{\sigma}) = \frac{n-2+2/d}{2/d} \in \Omega(n \cdot d)$.

Now assume $n < m$ and suppose m is divisible by n. Let exactly $\frac{n}{m} + 1$ voters rank c_1 first (denote this group as $\mathcal{G}_1$), and assign the remaining voters evenly among the other candidates. Let $v_j = \mu$ for $j \in \mathcal{G}_1$, and $v_j = (1, 0, \ldots, 0)$ for all other voters. Then the social welfare of c_1 is at most $(n/m + 1)/d$, while the best alternative has social welfare at least $n - n/m - 1 + (n/m + 1)/d$. The distortion becomes

$$\mathrm{D}(f_{\mathrm{Plur}}, \vec{\sigma}) = \frac{n - \frac{n}{m} - 1 + \frac{1}{d}(\frac{n}{m} + 1)}{\frac{1}{d} \cdot (\frac{n}{m} + 1)} = \Omega(m \cdot d).$$

In both cases, the lower bound is $\Omega(\min(m, n) \cdot d)$. □

Theorem 4. *[5] Suppose $m \geq d$ and $n \geq \sqrt{d}$. Then there exists a preference profile $\vec{\sigma}$ such that for any randomized rule f, we have $\mathrm{D}(f, \vec{\sigma}) = \Omega(\sqrt{d})$.*

Proof of Theorem 4. This is essentially a presentation of the proof of Boutilier et al. [5].

Without loss of generality, assume that $\sqrt{d}$ divides n. Partition the set of agents into $\sqrt{d}$ equal-size groups $N_1, \ldots, N_{\sqrt{d}}$. Consider the following preference profile: for each $k = 1, \ldots, \sqrt{d}$, all agents in N_k rank candidate c_k first (the remainder of their rankings can be arbitrary).

By the pigeonhole principle, for any randomized social choice function f, there exists a $k^* \in \{1, \ldots, \sqrt{d}\}$ such that $\Pr[f(\vec{\sigma}) = c_{k^*}] \leq \frac{1}{\sqrt{d}}$.

Now, construct the following instance (which is *unknown* to the social choice rule): Let $c_{k^*} = c_1$, and for $k \in [\sqrt{d}]$, let $c_k = e_k$ (the k-th standard basis vector). For all other candidates $j > \sqrt{d}$, let $c_j = e_2$. Assign all voters in N_{k^*} the vector $v_1 = e_1$, and all voters in other groups the vector μ. It is easy to check that these voters and candidates satisfy all constraints imposed in the model section.

For this configuration, the average social welfare is maximized by selecting c_{k^*}, yielding

$$\mathrm{UW}_1 = \frac{1}{\sqrt{d}} + \frac{1}{d}\left(1 - \frac{1}{\sqrt{d}}\right).$$

For any other candidate, the average social welfare is

$$\mathrm{UW}_2 < \frac{1}{d}.$$

Thus, the welfare-maximizing candidate is c_{k^*}. However, since $\Pr[f(\vec{\sigma}) = c_{k^*}] \leq \frac{1}{\sqrt{d}}$, the expected social welfare under f is at most

$$\mathbb{E}[\mathrm{UW}] < \mathrm{UW}_1 \cdot \frac{1}{\sqrt{d}} + \mathrm{UW}_2 \cdot \left(1 - \frac{1}{\sqrt{d}}\right).$$

Therefore, the distortion is at least

$$\mathrm{D}(f) > \frac{\mathrm{UW}_1}{\mathrm{UW}_1 \cdot \frac{1}{\sqrt{d}} + \mathrm{UW}_2 \cdot \left(1 - \frac{1}{\sqrt{d}}\right)}.$$

Plugging in $\mathrm{UW}_1 = \frac{1}{\sqrt{d}} + \frac{1}{d}\left(1 - \frac{1}{\sqrt{d}}\right)$ and $\mathrm{UW}_2 = \frac{1}{d}$, we find

$$\mathrm{D}(f) \in \Omega(\sqrt{d}).$$

Thus, the distortion is $\Omega(\sqrt{d})$, as claimed. □

B PSEUDOCODE FOR IMPLEMENTATIONS OF INSTANCE-OPTIMAL RULES

Algorithm 1 Deterministic Distortion-Optimal Rule (Column Generation)

INPUT: Candidate set C, preference profile $\vec{\sigma}$, tolerance ε
for each candidate $c_k \in C$ **do**
  Initialize $\mathcal{V} \leftarrow$ small subset of feasible average voters in $\mathcal{F}$
  repeat
    Solve the following LP to obtain $\hat{\beta}_k$:

$$\max_{\beta} \quad \beta \qquad \text{s.t.} \quad \langle c_k, \bar{v} \rangle \geq \beta \cdot \max_{c \in C} \langle c, \bar{v} \rangle \quad \forall \bar{v} \in \mathcal{V}$$

    Call SeparationOracle$(c_k, \hat{\beta}_k, \vec{\sigma})$ to find a violated $\bar{v} \in \mathcal{F}$
    if a violated $\bar{v}$ is found **then**
      ADD $\bar{v}$ to $\mathcal{V}$
    else
      BREAK
    end if
  until convergence
  Store worst-case distortion bound $\hat{\beta}_k$
end for
return Candidate $c_k \in C$ with largest $\hat{\beta}_k$

Algorithm 2 SeparationOracle$(\hat{c}, \hat{\beta}, \vec{\sigma})$

for each candidate $c \in C$ **do**
  Let $d_c \leftarrow \hat{c} - \hat{\beta} \cdot c$
  Solve LP:

$$\min_{\bar{v} \in \mathcal{F}} \quad \langle d_c, \bar{v} \rangle$$

  if optimal value < 0 **then**
    Return violating $\bar{v}$
  end if
end for
Return None

Algorithm 3 Column Generation for Randomized Distortion-Optimal Rule

INPUT: Candidate set C, preference profile $\vec{\sigma}$, tolerance ε
INITIALIZE: Constraint set $\mathcal{V} \leftarrow$ small subset of $\mathcal{F}$ (e.g., one feasible $\bar{v}$)
repeat
  Solve the following LP to obtain $(\hat{\beta}, \hat{p})$:

$$\begin{aligned} \max_{\beta, p} \quad & \beta \\ \text{s.t.} \quad & \sum_{i=1}^{m} p_i \langle c_i, \bar{v} \rangle \geq \beta \cdot \max_{c \in C} \langle c, \bar{v} \rangle \quad \forall \bar{v} \in \mathcal{V} \\ & \sum_{i=1}^{m} p_i = 1, \quad p_i \geq 0 \quad \forall i \end{aligned}$$

  CALL SeparationOracle$(\hat{p}^T C, \hat{\beta}, \vec{\sigma})$ to find a violated constraint $\bar{v} \in \mathcal{F}$
  if a violated $\bar{v}$ is returned **then**
    ADD $\bar{v}$ to $\mathcal{V}$
  else
    BREAK
  end if
until convergence
RETURN: Randomized rule defined by $\hat{p}$

Instead of performing binary search independently for each pair of candidates in the deterministic instance-optimal rule, we can adopt a structure similar to that used in the randomized rule. This still requires $O(m^2)$ different programs, but now, for each candidate, we compute its maximum distortion against all others directly. In this formulation, the separation oracle is invoked repeatedly with a fixed feasible region but varying objective functions. This allows us to use warm-starting to significantly accelerate computation.

C RESULTS UNDER ℓ^2 NORMALIZATION

When voter and candidate vectors are ℓ^2-normalized, meaning that $\|v\|_2 = \|c\|_2 = 1$ utilities are given by $u_v(c) = v^T c = \cos(\theta_{vc})$, where θ_{vc} is the angle between v and c.

A critical difference between these models is a voter's favorite candidate: under ℓ^1 normalization, any voter v has a favorite candidate in $\{e_1, \ldots, e_d\}$, where e_i is the standard basis vector for coordinate $i \in [d]$. By contrast, under ℓ^2 normalization the favorite candidate of any voter v is itself.

C Known?	Randomized		Deterministic	
Known	$O(d^{1/2})$	(Thm. 17)	$O(d^2)$	(Thm. 14)
	$\Omega(d^{1/4})$	(Thm. 13)	$\Omega(d^{3/2})$	(Thm. 12)
Unknown	$O(d)$	(Thm. 18)	-	

Table 2: Preliminary distortion bounds in the ℓ^2-normalized linear social choice setting. Again the 'C unknown' setting inherits lower bounds from the 'C known' setting.

C.1 Impossibilities

Throughout let $\mu_2 = (\frac{1}{\sqrt{d}}, \ldots, \frac{1}{\sqrt{d})})$ denote the uniform ℓ^2-normalized vector in $\mathbb{R}^d_{\geq 0}$.

Theorem 12. *Suppose $m, n \geq d$. Then there is a profile $\vec{\sigma}$ on which every deterministic rule f has distortion $\Omega(d^{3/2})$.*

Proof. We follow the helpful presentation by Anshelevich et al. [1] of the $\Omega(m^2)$ lower bound of Caragiannis et al. [6] on the distortion of any deterministic rule in the unit-sum setting.

Voters are organized into two equal-size groups V_a and V_b, where each voter has a unique preferred candidate c_v, all voters $v \in V_a$ rank a second, and all voters $v \in V_b$ rank b second. Suppose that $d = m$ and (as usual) that the embeddings of the candidates form the standard basis for $\mathbb{R}^d$. In this case, ranking suffixes may be completed arbitrarily; this defines our profile $\vec{\sigma}$.

What can our deterministic rule do? If it outputs a candidate in $\{a, b\}$ then suppose without loss of generality that $f(\vec{\sigma}) = b$ and a is the welfare maximizer. If all $v \in V_b$ are embedded as $v = c_v$, then $\mathrm{UW}(f(\vec{\sigma})) = 0$, while the maximizer can be anything supported by group V_a, leading to unbounded distortion.

On the other hand, suppose that $f(\vec{\sigma}) = c_v$, some voter's favorite candidate. Suppose without loss of generality $v \in V_b$ and that a is the welfare maximizer, and all $v' \in A$ are embedded as $v' = (a + c_{v'})/2$, so that $\mathrm{UW}(a) = \Omega(n)$. If v is essentially indifferent and embedded at μ_2, but no other voters place any value on c_v, then $\mathrm{UW}(c_v) = v^T c_v = d^{-1/2}$. In this case, we have

$$\mathrm{D}(f(\vec{\sigma})) = \Omega(d^{3/2}).$$

Since in all other cases f suffers infinite distortion, this is the best it can attain on such a profile $\vec{\sigma}$. □

In the same spirit as Theorem 4, we have a lower bound for any randomized rule that is agnostic to candidates' positions.

Theorem 13. *Suppose $m \geq d$ and $n \geq d^{1/4}$. Then there exists a preference profile p such that for any randomized social choice function f, we have $\mathrm{D}(p, f(p)) = \Omega(d^{1/4})$.*

Proof. Without loss of generality, assume that $d^{1/4}$ divides n. Partition the set of agents into $d^{1/4}$ equally sized groups, $N_1, \ldots, N_{d^{1/4}}$. Consider the following preference profile: for each $k = 1, \ldots, d^{1/4}$, all agents in N_k rank candidate c_k first (the remainder of their rankings can be arbitrary).

By the pigeonhole principle, for any randomized social choice function f, there exists $k^* \in \{1, \ldots, d^{1/4}\}$ such that $\Pr[f(p) = c_{k^*}] \leq \frac{1}{d^{1/4}}$.

Now, construct the following instance (which is *unknown* to the social choice rule): Let $c_{k^*} = c_1$, and for $k \in [d^{1/4}]$, let $c_k = e_k$ (the k-th standard basis vector). For all other candidates $j > d^{1/4}$, let $c_j = e_2$. Assign all voters in N_{k^*} the vector $v_1 = e_1$, and all voters in other groups the vector μ_2. It is easy to check that these voters and candidates satisfy all constraints imposed in the model section.

For this configuration, the average social welfare is maximized by selecting $c_{k^*} = e_1$, yielding

$$\mathrm{UW}_1 = \frac{1}{d^{1/4}} + \frac{1}{\sqrt{d}} \left(1 - \frac{1}{d^{1/4}}\right).$$

For any other candidate, the average social welfare is

$$\mathrm{UW}_2 < \frac{1}{\sqrt{d}}.$$

Thus, the welfare-maximizing candidate is c_{k^*}. However, since $\Pr[f(p) = c_{k^*}] \leq \frac{1}{d^{1/4}}$, the expected social welfare under f is at most

$$\mathbb{E}[\mathrm{UW}] < \mathrm{UW}_1 \cdot \frac{1}{d^{1/4}} + \mathrm{UW}_2 \cdot \left(1 - \frac{1}{d^{1/4}}\right).$$

Therefore, the distortion is at least

$$\mathrm{D}(p, f(p)) > \frac{\mathrm{UW}_1}{\mathrm{UW}_1 \cdot \frac{1}{d^{1/4}} + \mathrm{UW}_2 \cdot \left(1 - \frac{1}{d^{1/4}}\right)}.$$

Plugging in $\mathrm{UW}_1 = \frac{1}{d^{1/4}} + \frac{1}{\sqrt{d}}\left(1 - \frac{1}{d^{1/4}}\right)$ and $\mathrm{UW}_2 = \frac{1}{\sqrt{d}}$, we find

$$\mathrm{D}(p, f(p)) \in \Omega(d^{1/4}).$$

Thus, the distortion is $\Omega(d^{1/4})$, as claimed. □

Observation 1. *Random dictatorship has worst-case distortion $\Omega(d)$ under ℓ^2 normalization.*

Proof. The instance is the same as in the proof of Theorem 7, up to constant factors due to the renormalization of the sparse utilities. Consider $C = \{e_1, \ldots, e_d\}$ and $d-1$ equal-size groups of voters. Again voters in group i are nearly indifferent between c_i and c_d, but rank c_i above c_d, and garner 0 utility from all other candidates. Then the welfare-maximizing candidate is c_d with welfare $\mathrm{UW}(c^*) = \Omega(n)$, while $\mathrm{UW}(f_{\mathrm{RD}}) = O(\frac{n}{d})$. □

C.2 Positive Results

We start with an analog of Lemma 1.

Lemma 2. *For any set of candidates $C = \{c_j\}_{j \in [m]}$ and any voter $v \in \mathrm{Cone}(C)$,*

$$\max_{c \in C} u_v(c) \geq d^{-1/2}.$$

Proof of Lemma 2. We will denote the maximum utility for v by $\hat{u}_v := \max_{c \in C} u_v(c)$; our goal will then be to provide a lower bound on $\hat{u}_v$. Since v and $c \in C$ are ℓ^2-normalized, this coincides with lower bounding the cosine similarity between v and the most similar $c \in C$.

To begin, since $v \in \mathrm{Cone}(C)$ we may write $v = \sum_c \alpha_c c$ for some collection of $\alpha_c \geq 0$. Then to begin we have

$$1 = v^T v = \sum_c \alpha_c v^\top c \leq \hat{u}_v \sum_c \alpha_c,$$

and so $\hat{u}_v \geq (\sum_c \alpha_c)^{-1}$. Next, we use the standard ℓ^1-ℓ^2 norm inequalities, which state that for all $x \in \mathbb{R}^d$, $\|x\|_2 \leq \|x\|_1 \leq \|x\|_2 \sqrt{d}$. Therefore $1 = \|c\|_2 \leq \|c\|_1$, and so

$$\sum_c \alpha_c = \sum_c \alpha_c \|c\|_2 \leq \sum_c \alpha_c \|c\|_1 = \|v\|_1 \leq \sqrt{d}.$$

Combining gives us $\hat{u}_v \geq d^{-1/2}$, as claimed. □

We conclude by observing that Lemma 2 is tight when $v = \mu_2$ and C is the standard basis of $\mathbb{R}^d$.

As a consequence, we obtain an improved distortion bound for f_{MCP} under ℓ^2 normalization.

Theorem 14. *The distortion of Max Coordinate Plurality is $O(d^2)$ under ℓ^2 normalization of voters and candidates.*

Proof. Following the argument of Theorem 3, know that each $v \in V$ has a coordinate with value at least $d^{-1/2}$. Therefore there is some $c \in \hat{C}$ with value at least $d^{-1/2}$ for this same coordinate by the cone assumption, and so $\max_{c' \in \hat{C}} v^T c' \geq v^T c \geq 1/d$.

Since the plurality winner in $\hat{C}$ is supported by at least n/d voters, the winner $c^* \in \hat{C}$ chosen by MCP confers welfare

$$\mathrm{UW}(c^*) \geq \frac{n}{d} \cdot \frac{1}{d},$$

and the claim follows. □

Theorem 15. *Random Dictatorship has distortion at most $O(d^{5/2})$ under ℓ^2 normalization of voters and candidates.*

Proof. We follow the proof of Theorem 7.

For each agent $v \in V$, let c_v denote the top-ranked alternative for v. By Lemma 2 we have that $u_v(c_v) = v^T c_v \geq d^{-1/2}$. So by averaging, there is some $\ell \in [d]$ such that $v^\ell c_v^\ell \geq d^{-3/2}$. Since $v, c_v \in S_+^d$, this implies $v^\ell \geq d^{-3/2}$ and $c_v^\ell \geq d^{-3/2}$. Fix one such coordinate ℓ_v for each v and partition V into $V_1, \ldots, V_d$ according to ℓ_v. Let $p_\ell = |v^\ell|/n$, and observe that $p \in \Delta_d$. Let $c^* = \arg\max_{c \in C} \sum_{v \in V} u_v(c)$ be the welfare-maximizing candidate. Clearly,

$$\mathrm{UW}(c^*) \leq n.$$

On the other hand,

$$\begin{aligned}
E_{c \sim RD}\left[\mathrm{UW}(c)\right] &= \frac{1}{n} \sum_{v \in V} \mathrm{UW}(c_v) \\
&= \frac{1}{n} \sum_{v \in V} \sum_{v' \in V} u_{v'}(c_v) \\
&\geq \frac{1}{n} \sum_{v \in V} \sum_{v' \in V_{\ell_v}} v'^T c_v \\
&\geq \frac{1}{n} \sum_{v \in V} \sum_{v' \in V_{\ell_v}} (v')_{\ell_v} (c_v)_{\ell_v}.
\end{aligned}$$

From here we will rearrange the sum to range over the groups v^ℓ. Observe that $v^\ell c_v^\ell \geq d^{-3/2}$ implies that $c_v^\ell \geq \frac{1}{v^\ell d^{3/2}}$. Then continuing,

$$\begin{aligned}
&= \frac{1}{n} \sum_\ell \sum_{v, v' \in V_\ell} v^\ell c_\ell^{v'} \\
&\geq \frac{1}{n} \sum_\ell \sum_{v, v' \in V_\ell} v^\ell \frac{1}{d^{3/2} v'_\ell} \\
&\geq \frac{1}{n} \sum_\ell \sum_{v, v' \in V_\ell : v^\ell \geq v'_\ell} v^\ell \frac{1}{d^{3/2} v'_\ell} \\
&\geq \frac{1}{n} \sum_\ell \sum_{v, v' \in V_\ell : v^\ell \geq v'_\ell} v^\ell \frac{1}{d^{3/2} v^\ell} \\
&\geq \frac{1}{n \cdot d^{3/2}} \sum_\ell \frac{1}{2} |v^\ell|^2 \\
&= \frac{n}{2d^{3/2}} p^T p \\
&\geq \frac{n}{2d^{5/2}}.
\end{aligned}$$

The claim follows. □

The insights about the uniform projection rule and the stable lottery rule can also be adapted to the the setting of ℓ^2-normalized voter and candidate embeddings.

Definition 8 (Uniform Projection Rule (f_{UProj2}) for ℓ^2 Normalization)**.** *Given candidate locations* $C \subset \mathbb{R}^d_{\geq 0}$*, the* uniform projection rule f_{UProj2} *defines a distribution* $\{p_c\}_{c\in C}$ *over candidates such that the expected pseudo-candidate vector* $\hat{c} := \sum_{c\in C} p_c \cdot c$ *minimizes the Kullback–Leibler (KL) divergence from the uniform vector, i.e.,* $\hat{c} = \arg\min_{x\in \mathrm{CH}(C)} \mathrm{KL}\,(\mu_2 \| x)$.

Note that $\hat{c}$ in general will not satisfy $\|\hat{c}\|_2 = 1$, since the set of (even positive orthant) vectors with ℓ^2-norm 1 is not convex.

Theorem 16. *The expected welfare of* f_{UProj2} *is at least* $\mathrm{UW}(f_{\mathrm{UProj2}}) \geq \frac{n}{d}$*. As a consequence,* $\mathrm{D}(f_{\mathrm{UProj2}}) = O(d)$.

Proof of Theorem 16. First observe that

$$\mathrm{KL}\,(\mu_2\|x) = \sum_{i=1}^{d} \mu_2^i \ln \frac{\mu_2^i}{x^i} = \sum_i \mu_2^i \ln \mu_2^i \; - \; \sum_{i=1}^{d} \mu_2^i \ln x^i,$$

so minimizing $\mathrm{KL}\,(\mu_2\|x)$ over the convex set $\mathrm{CH}(C)$ is equivalent to minimizing the smooth, convex function $f(x) = -\sum_{i=1}^d \mu_2^i \ln x^i$ subject to $x \in \mathrm{CH}(C)$. The first-order optimality condition gives $\nabla f(x^*)^\top (v - x^*) \;\geq\; 0$ for every voter $v \in \mathrm{CH}(C)$. Since $\frac{\partial f}{\partial x^i}(x) = -\frac{\mu_2^i}{x^i}$, we have

$$\nabla f(x^*)^\top (v - x^*) = -\sum_{i=1}^{d} \frac{\mu_2^i}{x^{*i}} (v^i - x^{*i}) \;\geq\; 0$$

$$\implies \sum_{i=1}^{d} \frac{\mu_2^i \, v^i}{x^{*i}} \;\leq\; \sum_{i=1}^{d} \mu_2^i.$$

Since $\mu_2^i = 1/\sqrt{d}$, this yields $\sum_{i=1}^d \frac{v^i}{x^{*i}} \;\leq\; d$. Now define the auxiliary sequences $a_i := \sqrt{\frac{v^i}{x^{*i}}}$ and $b_i := \sqrt{v^i\, x^{*i}}$. Then $\sum_{i=1}^d a_i\, b_i = \sum_{i=1}^d v^i \in [1, \sqrt{d}]$ since $\|v\|_2 = 1$. And, by Cauchy–Schwarz,

$$\begin{aligned} 1 \leq \Big(\sum_i a_i b_i\Big)^2 &\leq \Big(\sum_i a_i^2\Big)\Big(\sum_i b_i^2\Big) \\ &= \Big(\sum_i \tfrac{v^i}{x^{*i}}\Big)\Big(\sum_i v^i\, x^{*i}\Big). \end{aligned}$$

Combining with $\sum_i v^i / x^{*i} \leq d$ gives $\sum_{i=1}^d x^{*i}\, v^i \geq 1/d$. Now as we have n voters and the above inequality holds for arbitrary v, the total utilitarian welfare is at least $\frac{n}{d}$.

The bound on the distortion of f_{UProj2} then follows because the maximum utility for any voter is again at most 1. □

As the original proof for linear stable lottery rule's distortion bound only relies on the bound of the uniform projection rule and the definition of stable lotteries, we immediately get the same result.

Definition 9 (Linear Stable Lottery Rule (f_{LSLR2}))**.** *Given a stable lottery* $\mathcal{W}$ *over committees of size* $k = \sqrt{d}$*, the* linear stable lottery rule f_{LSLR2} *on profile* $\vec{\sigma}$ *chooses each* $c \in C$ *with probability* $\frac{1}{2\sqrt{d}} \Pr_{W\sim \mathcal{W}(\vec{\sigma})}[c \in W] + \frac{1}{2} \Pr_{c' \sim f_{\mathrm{UProj2}}(\vec{\sigma})}[c' = c]$.

Theorem 17. $\mathrm{D}(f_{\mathrm{LSLR2}}) = O(\sqrt{d})$.

Since Theorem 16 implies that $\mathrm{UW}(c^*) \geq \frac{n}{d}$ for ℓ_2 normalization, the proof of Theorem 9 also proceeds identically. We hence also have:

Theorem 18. $\mathrm{D}(f_{\mathrm{PSLR}}) = O(d)$ *for* ℓ_2*-normalized candidates and voters.*

C.3 Optimizing distortion for ℓ^2-normalization

When switching to ℓ^2 normalization, the key change is that the feasible region $\mathcal{F}$ becomes *nonconvex*, due to the unit-norm constraint $\|v_j\|_2 = 1$ for each voter $j = 1, \ldots, n$. As a result, our instance-optimal problem likely cannot be solved exactly in polynomial time.

Nonetheless, several methods have been proposed for approximating such problems, including augmented Lagrangian techniques and projected gradient descent, which can be employed in practice to compute near-optimal solutions.

D ADDITIONAL EXPERIMENTS

All experiments were run on a single-socket Intel® Xeon® Gold 6130 CPU @ 2.10 GHz (16 physical cores, 32 threads), arranged in two NUMA nodes (cores 0–15 and 16–31).

In addition to evaluating the *instance distortion* as in Section 6, here we also measure the *empirical distortion*, defined as follows:

Definition 10 (Empirical distortion)**.** *Given a fixed utility profile* $\vec{u}$*, the empirical distortion of a mechanism* f *is the ratio of the optimal utilitarian social welfare to the expected welfare achieved by* f *on the preference profile* $\vec{\sigma}$ *induced by* $\vec{u}$*. Formally,*

$$\mathrm{D}(f, \vec{u}) := \frac{\mathrm{UW}(\vec{u}, c^*(\vec{u}))}{\mathbb{E}_{c\sim f(\vec{\sigma}(\vec{u}))}[\mathrm{UW}(\vec{u}, c)]}.$$

While the distortion results presented in the main body establish the optimal guarantee of our proposed instance-optimal rules, the empirical findings offer further insight into the structure of real-world instances. A particularly notable observation is that even a simple rule such as plurality has strong performance in practice.

The general trends observed for empirical distortion closely mirror those of instance distortion across variations in n, m, and d: both measures remain largely stable. In terms of computational cost, the deterministic optimal rule exhibits the most significant increase in running time as parameters grow, followed by the randomized optimal rule. In contrast, the computation times for the remaining mechanisms are negligible in comparison. Notably, the number of alternatives m is the most influential factor affecting runtime.

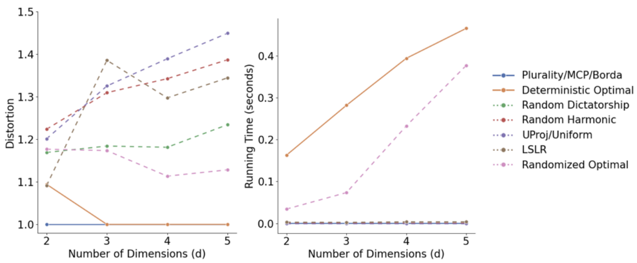

Figure 7: Abortion Opinion Survey. Left: Empirical Distortion; Right: Running Time.