Linear Social Choice with Few Queries: A Moment-Based Approach

Luise Ge* Daniel Halpern† Gregory Kehne‡ Yevgeniy Vorobeychik§

Abstract

Most social choice rules assume access to full rankings, while current alignment practice—despite aiming for diversity—typically treats voters as anonymous and comparisons as independent, effectively extracting only about one bit per voter. Motivated by this gap, we study social choice under an extreme communication budget in the linear social choice model, where each voter's utility is the inner product between a latent voter type and the embedding of the context and candidate. The candidate and voter spaces may be very large or even infinite. Our core idea is to model the electorate as an unknown distribution over voter types and to recover its moments as informative summary statistics for candidate selection. We show that one pairwise comparison per voter already suffices to select a candidate that maximizes social welfare, but this elicitation cannot identify the second moment and therefore cannot support objectives that account for inequality. We prove that two pairwise comparisons per voter, or alternatively a single graded comparison, identify the second moment; moreover, these richer queries suffice to identify all moments, and hence the entire voter-type distribution. These results enable principled solutions to a range of social choice objectives including inequality-aware welfare criteria such as taking into account the spread of voter utilities and choosing a representative subset.

1 Introduction

A fundamental problem in social choice is to map potentially diverse preferences of a collection of voters over a set of candidates to a subset of *winners*. The classical model assumes that voter preferences are elicited as full rankings. In many real settings, however, the voter population can be massive and the candidate space is enormous or unbounded—for example, voters may be users of an AI system (such as an LLM) and candidates may be possible outputs like images, music, recipes, or answers to complex prompts. Moreover, preferences over candidates may depend a great deal on context, which itself can defy enumeration: for example, preferences over responses clearly depend on the question. This situation has become particularly salient in the context of AI *value alignment*, such as training large language models (LLMs) to align with (i.e., behave according to) humans' subjective values such as helpfulness and harmlessness [Bai et al., 2022, Ji et al., 2023, Ouyang et al., 2022, 2025]. Since the space of candidates cannot be explicitly enumerated, a typical paradigm would present specific pairs of candidates to human annotators, who would select which of these they prefer, or indicate indifference. Such datasets are typically collected without retaining annotator identifiers; hence, the feedback per voter can be as sparse as one bit. Approaches such as reinforcement learning from human feedback (RLHF) then use these pairwise comparisons to first train a parametric reward (utility) function. This reward model is subsequently plugged into a conventional RL framework such as PPO [Schulman et al., 2017] to achieve context-dependent behavioral alignment [Christiano et al., 2017, Stiennon et al., 2020].

While there has been an increasing interest in incomplete vote elicitation [Halpern et al., 2023, 2024], the practice of AI alignment has pushed the feedback constraint to an extreme. On the one hand, if we wish to achieve useful guarantees about selecting good candidates in such settings (e.g., in terms of social

*Washington University in St. Louis. g.luise@wustl.edu
†Google Research. dhalpern@google.com
‡Washington University in St. Louis. kehne@wustl.edu
§Washington University in St. Louis. yvorobeychik@wustl.edu

welfare if voter preferences reflect latent utility functions), the situation seems hopeless. On the other hand, it is typical in settings such as value alignment that the space of candidates and voters is *structured*. In particular, generative AI methods rely on embedding digital objects such as music, images, or text, as vectors. It is then natural to posit that human preferences, too, have a structured representation as parametric utility functions over these vectors, an assumption that is exploited in the reward model learning step of RLHF. This motivates our central research question:

> *Is it possible to leverage vector representations of candidates and parametric representations of voter utilities to obtain sufficiently reliable information from few per-voter pairwise comparison queries to select good winning candidates, when voter and candidate spaces are both large?*

To study this question, we assume that each voter has a utility function over the vector space of candidates which is linear in the candidate embedding. Motivated by the linear representation hypothesis [Park et al., 2024], this effectively assumes that LLM text embeddings are rich enough that user preferences can be represented as linear functions over them. Yet even in this structured setting, the initial results have been bleak. Even with access to full rankings, many aggregation rules including the standard RLHF procedure fail the most basic social choice properties like Pareto Optimality [Ge et al., 2024a]. Moreover, it is information-theoretically impossible to identify a candidate to achieve social welfare within a constant factor of optimal [Ge et al., 2025].

Crucially, these negative conclusions are obtained in a finite-electorate setting with a fixed dataset of comparisons. Our focus is instead on preference elicitation under limited communication. We assume access to a large population that can be sampled repeatedly, and we can choose the comparisons we ask, but each voter provides only a small number of binary responses. Still, suppose that we aim to choose a single candidate to (approximately) maximize social welfare. How many pairwise rankings do we need to elicit from each voter? Moreover, assuming we cannot ask all possible voters (e.g., the entire population of a country), how many voters suffice, if we treat each as a sample from an unknown voter distribution? Furthermore, how does this picture change as we ask more complex questions, such as (a) maximizing welfare while accounting for inequality aversion, or (b) selecting a committee rather than a single candidate?

1.1 Our Contributions

We cast our setting as social choice under sparse elicitation: each voter has a latent preference vector $\theta \in \mathbb{S}^{d-1}$, but the mechanism can ask only a small number of comparison-style queries per voter. This raises three basic questions: (i) which social objectives are meaningful in this limited-information regime, (ii) what information should be elicited to evaluate those objectives, and (iii) how many samples (voters) are required.

Our answers are organized around a single principle: moments are appropriate summaries of the preference distribution. We show how different query families identify different moments, and we give finite-sample guarantees that translate moment estimation into guarantees for downstream social-choice objectives.

Concretely, first consider selecting a single candidate. If the goal is to maximize social welfare, it turns out only a single such query suffices. That is, we show that we can both *identify* (Section 3) and *effectively estimate* (Section 4) the first moment of the voter distribution *with only a single pairwise comparison query per voter*:

Theorem 1.1 (informal)**.** *The first moment of the voter distribution is identifiable using one pairwise comparison query per voter. Moreover, we can estimate it to within ε with sample complexity polynomial in $1/\varepsilon$ and d.*

As a direct consequence of this result, *we can find an approximately welfare-optimal candidate by asking each voter only one pairwise comparison query* (Section 5).

While maximizing welfare is a natural goal in social choice, it has a significant limitation: the result can be extremely inequitable, for example, with some voters having very high, while many others very low, utility over the final candidate. It is often desirable to moderate this criterion by using welfare objectives that also account for inequality [Atkinson et al., 1970]. fOne measure of inequality is utility variance, which we can leverage to construct welfare functions that combine average utility with variance. In order to identify a candidate that maximizes a variance-adjusted welfare function, a crucial subproblem is to estimate *the second moment of the voter distribution.* Is a single pairwise comparison query to each voter sufficient for this? We show that it is not (Section 3):

Theorem 1.2 (informal)**.** *The second moment of the voter distribution is not identifiable from one pairwise comparison query per voter.*

Thus, we need more than a single pairwise comparison per voter to obtain variance information. Do two such comparisons suffice? We show that they do—in fact, we show that this generalizes directly for any k (Section 3 shows identifiability while Section 4 considers estimation):

Theorem 1.3 (informal)**.** *The first k moments of the voter distribution are identifiable using k pairwise comparison queries per voter, and can be estimated from such data to within ε with sample complexity polynomial in $1/\varepsilon$ and d.*

Intuitively, it seems that k pairwise queries are also *necessary* to identify k moments of a distribution. Remarkably, this is not true: only 2 per voter queries suffice to estimate *any properties of the distribution* (including arbitrary moments):

Theorem 1.4 (informal)**.** *The voter distribution is identifiable using 2 pairwise comparison queries per voter, and the kth moment can be estimated from such data to within ε with sample complexity polynomial in $1/\varepsilon$ and d.*

We can also significantly generalize the above results to stochastic response models; details are provided in Appendix D. Furthermore, if voters only respond if they *strongly* prefer one response to another (formally defined in Section 2.5), then only one query per voter is necessary to identify the distribution.

Our moment-based approach is also useful in selecting candidates under other objectives. For example, we show how moments can be used to approximate Nash welfare. We also apply this to selecting committees of candidates in multi-winner elections. In this context, we need to extend the voter utility functions to *sets* of candidates. For example, we can assume that each voter's utility of a set is the maximum utility from any candidate in the set. By approximating this using a k-degree polynomial and maximizing the resulting function, we can obtain an approximately social-welfare-optimal committee.

1.2 Related Work

Our work operates within the *Linear Social Choice* framework, a new paradigm of social choice in which voter utilities are linear functions of context and candidate embeddings. Prior results emphasize worst-case aspects from axiom violations [Ge et al., 2024a] to distortion bounds [Ge et al., 2025]. In contrast, we take a bottom-up perspective and ask information-theoretic questions. There is a rich literature on dealing with incomplete information in social choice [Brandt et al., 2016, Chapter 10]. Our work is most closely related to recent frameworks that combine elicitation with distributional assumptions, where only minimal information is elicited from each voter [Halpern et al., 2023, 2024]. These works, however, focus on different feedback models (e.g., approval) and different questions (e.g., computing voting rules), and do not operate in the linear-utility setting.

Our contributions also intersect with the extensive preference learning literature whose roots lie in *learning to rank* [Cohen et al., 1997, Burges et al., 2005], and which has been adopted for *virtual democracy* [Noothigattu et al., 2018, Kahng et al., 2019]. More recently, learning a parametric utility model has been analyzed theoretically [Ge et al., 2024b] and used in practice as part of RLHF-style training [Christiano

et al., 2017] to enable generalization over large candidate sets. Notably, these works deal with a single preference, and observed disagreement is treated as noise.

In contrast, pluralistic alignment and the increasing use of LLM-based auto-raters [Li et al., 2025] shift the focus toward learning the voter population itself. Efforts in this direction are so far predominantly empirical [Chakraborty et al., 2024, Melnyk et al., 2024, Siththaranjan et al., 2024, Kim et al., 2025]. The more theoretical works to date [Chidambaram et al., 2025, Cherapanamjeri et al., 2025, Shirali et al., 2025] adopt different models but, broadly, resonate with our findings: the standard alignment data pipeline is insufficient for learning heterogeneity in the underlying population.

Our estimation strategy relies on the *Generalized Method of Moments* (GMM) [Hansen, 1982, Pearson, 1936]. While GMM has been applied to estimate parameters of probabilistic ranking models like Plackett-Luce [Azari Soufiani et al., 2013], we adapt the principle to the linear setting. At a technical level, our problem is also related to *1-Bit Compressed Sensing* in signal processing [Boufounos and Baraniuk, 2008, Plan and Vershynin, 2013], which recovers signals from the signs of random linear measurements. Our setting shares this structure of receiving single bits of information per sample (comparison queries), though we focus on recovering distributional moments rather than sparse vectors, and allowing multiple comparison queries at once. At a more fundamental level, our identifiability results connect to *geometric tomography* and the *Cramér-Wold theorem* [Cramér and Wold, 1936], which states that a high-dimensional distribution is determined by its lower-dimensional projections. We extend these insights to show how "projections" obtained via pairwise queries are sufficient to recover properties of the voter distribution.

1.3 Organization

The rest of the paper is organized as follows. First, we provide formal preliminaries in Section 2. Our main results regarding identifiability of moments of the voter distribution from pairwise comparison queries then follow in Section 3. Next, we build on the positive identifiability results to characterize how to estimate moments in Section 4. Finally, we illustrate our results on moment estimation in the context of social choice applications in both single-winner and committee selection in Section 5.

2 Preliminaries

2.1 Voters and Utilities

We assume that voters $v \in V$ are distributed according to a distribution $\mathcal{V}$ which represents our underlying voter population. Each voter is characterized by a type vector θ_v lying on the unit sphere $\mathbb{S}^{d-1} \coloneqq \{\theta \in \mathbb{R}^d : \|\theta\|_2 = 1\}$ to control for utility scale invariance of preference rankings. This mapping induces a voter type distribution Θ over $\mathbb{S}^{d-1}$. For ease of exposition, we assume that Θ is absolutely continuous with respect to the Lebesgue measure on the sphere.[1]

We consider a space of contexts (prompts) $\mathcal{X}$ and a space of candidates (responses) $\mathcal{Y}$. An embedding function $\Phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^d$ maps each context-candidate pair to a real vector. We assume that voter utilities are linear in these embeddings; specifically, the utility of a voter with type θ for a pair (x, y) is given by the inner product $u_\theta(x, y) = \theta \cdot \Phi(x, y)$. Since utilities are fully determined by voter types and candidate embeddings, we will henceforth identify voters directly with their types θ and refer to Θ as the voter distribution. Similarly, we will often refer to a context-candidate pair (x, y) by its embedding $\phi = \Phi(x, y)$, writing the utility function simply as $u_\theta(\phi) = \theta \cdot \phi$. Finally, we will write $\mathcal{U}_\phi$ for the distribution of utilities induced by ϕ, i.e., $\mathcal{U}_\phi$ is the distribution induced by $u_\theta(\phi)$ over the randomness of θ. A reference table is provided in Appendix A.

[1] Informally, this implies that the probability mass of the voter distribution is not concentrated on lower-dimensional subsets. However, with additional technical care, our results can be extended to hold without this assumption.

2.2 Welfare Objectives

The first objective to consider is *social welfare*, which is the expected utility w.r.t. the voter distribution $\mathbb{E}_{\theta\sim\Theta}[u_\theta(\phi)]$. However, simply maximizing welfare can yield considerable inequality in realized utility across voters. A number of alternative welfare notions therefore aim to adjust for potential inequality. A natural way to do this is maximizing *risk-adjusted* welfare (raw), maximize the expected welfare but penalize a candidate by α times the standard deviation:

Definition 2.1 (α-risk-adjusted welfare)**.** For $\alpha \geq 0$, the *α-risk-adjusted welfare* of ϕ is given by

$$\mathrm{raw}_\alpha(\phi) \coloneqq \mathbb{E}_\Theta[u_\theta(\phi)] - \alpha \cdot \sqrt{\mathbb{E}_\Theta\left[(u_\theta(\phi) - \mathbb{E}_\Theta[u_\theta(\phi)])^2\right]}.$$

Another common objective that has the effect of being more equitable than social welfare is *Nash welfare*, which in our setting is defined as $\mathbb{E}_{\theta\sim\Theta}[\log u_\theta(\phi)]$. Of course, for this to be well-defined and meaningful, voter utilities $u_\theta(\phi)$ must be strictly positive.

Alternatively, it is often possible and desirably to compute a set of winning candidates (that is, a *committee*) W rather than a single winner. In a welfarist context, we need to extend a voter's utility over individual candidates to a utility function over sets. One natural extension is that the voter's utility for W stems from their most preferred candidate in W, i.e., $u_\theta(W) = \max_{\phi\in W} u_\theta(\phi)$. This yields a welfare objective that we refer to as *top-choice welfare*:

Definition 2.2 (Top-choice welfare)**.** For a user distribution Θ and a set of candidate responses Φ, the *top-choice welfare* of a set of candidates $W \subseteq \Phi$ is given by

$$\mathrm{tcw}_\Theta(W) \coloneqq \mathbb{E}_\Theta\left[\max_{\phi\in W} u_\theta(\phi)\right].$$

We refer to the problem of optimizing tcw subject to $|W| \leq \ell$ for a given ℓ as ℓ-tcw *maximization*.

2.3 Preference Elicitation

As is typical in the value alignment literature, voters do not report their utilities directly (it is typically too much to ask). Instead, we elicit preference information via pairwise comparison queries. A single query consists of a context x and a pair of candidate responses y_1, y_2. When presented with such a query, a voter θ indicates which response yields higher utility, returning 1 if they prefer y_1 and 0 if they prefer y_2.[2] We encode this response as $\mathrm{resp}_\theta((x, y_1, y_2)) = \mathbb{I}\{u_\theta(x, y_1) \geq u_\theta(x, y_2)\}$.[3]

In our linear utility model, the condition $u_\theta(x, y_1) \geq u_\theta(x, y_2)$ is equivalent to $\theta\cdot(\Phi(x, y_1) - \Phi(x, y_2)) \geq 0$. Since the response depends solely on the *difference* between the embeddings, we define the query vector $q \coloneqq \Phi(x, y_1) - \Phi(x, y_2)$ and allow the response function to operate directly on these differences, i.e., $\mathrm{resp}_\theta(q) \coloneqq \mathbb{1}\{\theta^\top q \geq 0\}$.

Our results rely on a geometric assumption that for any direction, we can identify a context and pair of candidates that (approximately) induce this vector by the difference in the associated embeddings. In effect, this means that the space of contexts and candidates must be sufficiently rich, at least with respect to the induced embedding space. This assumption allows us to establish the fundamental limits of this form of preference elicitation. Impossibility results in our setting imply impossibility under any weaker query model. Conversely, positive identifiability results establish a theoretical ceiling, reducing the alignment problem to the engineering challenge of generation.

[2] Our framework can be naturally extended to incorporate stochastic responses; see Appendix D.

[3] This formulation effectively breaks ties in favor of y_1, but the specific tie-breaking mechanism does not impact our results. By the absolute continuity of Θ, the probability that a random voter assigns exactly equal utility to any two distinct candidate embeddings is zero.

Assumption 2.3. The embedding space is sufficiently expressive such that for any direction $q \in \mathbb{S}^{d-1}$, we can generate (x, y_1, y_2) such that $\Phi(x, y_1) - \Phi(x, y_2) \propto q$.

Under Assumption 2.3, the problem of selecting a comparison (x, y_1, y_2) reduces to directly choosing a direction $q \in \mathbb{S}^{d-1}$. We refer to such a vector q as a *query* and treat these queries as the primary decision variables for the data collector.

2.4 Multi-Query Responses

Given a sequence of t queries $\mathbf{q} = (q_1, \ldots, q_t)$, a random voter $\theta \sim \Theta$ arrives and provides a response vector $(\mathrm{resp}_\theta(q_1), \ldots, \mathrm{resp}_\theta(q_1)) \in \{0,1\}^t$ indicating their preferences. Consequently, a fixed query sequence $\mathbf{q}$ induces a distribution over binary response vectors. We define $Q_t(\mathbf{q}, \mathbf{b}; \Theta)$ as the probability that a random voter drawn from Θ produces the response vector $\mathbf{b} \in \{0,1\}^t$ when presented with queries $\mathbf{q}$. When the underlying distribution Θ is clear from context, we will omit it from the notation. Thus, for any fixed $\mathbf{q}$, the function $Q_t(\mathbf{q}, \cdot)$ represents a probability distribution over $\{0,1\}^t$.

Throughout, we will use the shorthand $Q_t(\boldsymbol{q})$ to denote the probability of the all-positive response vector, $Q_t(\boldsymbol{q}, \mathbf{1})$, where $\mathbf{1} = (1, \ldots, 1)$. Formally, we have

$$Q_t(q_1, \ldots, q_t) \coloneqq \Pr_{\theta \sim \Theta} \left[[\mathrm{resp}_\theta(q_1) = 1] \wedge \ldots \wedge [\mathrm{resp}_\theta(q_t) = 1] \right].$$

It is worth noting that the probability of any arbitrary response pattern $\mathbf{b} \in \{0,1\}^t$ can be recovered solely from the values of $Q_t(\mathbf{q})$ (for instance, by negating specific query vectors q_i to target zeros, since $\mathrm{resp}_\theta(-q) = 1 - \mathrm{resp}_\theta(q)$ almost surely).

Finally, we remark that we assume deterministic voter responses in the main paper for clarity of presentation. Nevertheless, our results for multi-query elicitation extend directly to stochastic response models (e.g., Bradley–Terry); see Appendix D.

2.5 Graded-Query Responses

Beyond ordinal comparisons, recent work shows that even a few bits of cardinal utility can improve distortion in single-winner elections [Amanatidis et al., 2021, Ebadian and Shah, 2025]. But reporting cardinal values remains cognitively challenging. By contrast, reporting preference intensity (e.g., "strong" versus "weak") is often less demanding and is already collected in many preference datasets. We therefore follow the line of work that incorporates preference intensity [Kahng et al., 2023]. Concretely, we model a graded preference query as whether the utility margin exceeds a threshold $\tau \in (0,1)$: $\mathrm{grad}_\theta(q) \coloneqq \mathbb{1}\{\theta^\top q \geq \tau\}$. Then for a distribution Θ, $G_\tau(q) = \Pr_{\theta \sim \Theta}[\mathrm{grad}_\theta(q)]$ returns the fraction of voters having a τ-strong preference.

Since the inner product $\theta^\top q$ depends on the norm of q, for results on graded queries, we need to further assume that the query space is rich enough to produce $q \in \mathbb{S}^{d-1}$.

Assumption 2.4. The embedding space is sufficiently expressive such that for any direction $q \in \mathbb{S}^{d-1}$, we can generate (x, y_1, y_2) such that $\Phi(x, y_1) - \Phi(x, y_2) = q$.

2.6 Distributions and their Moments

We adopt the standard measure-theoretic formalism of probability, in which probability distributions are defined over a sample space Ω with an associated σ-algebra $\mathcal{R}$ of measurable events. Probability measures are then measures μ on $(\Omega, \mathcal{R})$ that are normalized such that $\mu(\Omega) = 1$. We will use $\bar{\sigma}$ to describe the uniform probability measure over the sphere $\mathbb{S}^{d-1}$, to differentiate it from σ which is the *unnormalized* surface area measure over $\mathbb{S}^{d-1}$.

Any measure μ over $\mathbb{R}$ has an associated sequence of moments $\boldsymbol{m} = (m_0, m_1, m_2, \ldots)$ given by $m_n \coloneqq \int_{\mathbb{R}} x^n \, d\mu(x)$. Under certain assumptions, as in the Hausdorff moment problem, where μ is restricted to $[0, 1]$, they are sufficient to uniquely specify μ (up to sets of measure 0).

We use tensors to represent the higher-order moments. Formally, a kth order tensor $T \in (\mathbb{R}^d)^{\otimes k}$ is a multidimensional array indexed by a tuple $(i_1, \ldots, i_k)$ where each $i_j \in [d]$. For vectors $v_1, \ldots, v_k \in \mathbb{R}^d$, the *outer product* $v_1 \otimes \cdots \otimes v_k$ is a tensor with entries given by the product of the coordinates: $(v_1 \otimes \cdots \otimes v_k)_{i_1,\ldots,i_k} = (v_1)_{i_1} \cdot (v_2)_{i_2} \cdots (v_k)_{i_k}$. When $k = 1$, this is equivalent to the vector itself; when $k = 2$, this corresponds to the matrix outer product $v_1 v_2^\top$. For two tensors $A, B \in (\mathbb{R}^d)^{\otimes k}$, their inner product is the sum of the products of their corresponding entries: $\langle A, B \rangle \coloneqq \sum_{i_1,\ldots,i_k=1}^{d} A_{i_1,\ldots,i_k} B_{i_1,\ldots,i_k}$. We sometimes reshape tensors into matrices. For a partition of the modes $\{1, \ldots, k\}$ into two sets I and J, the *matricization* $\mathrm{mat}_{I|J}(T)$ flattens the tensor into a matrix where the rows are indexed by I and the columns by J. And sometimes we symmetrize a tensor of rank k as $\mathrm{Sym}(T) \coloneqq \frac{1}{k!} \sum_{\pi \in S_k} T^\pi$, with S_k the symmetric group and T^π the tensor re-indexed according to the permutation π.

For a vector-valued distribution Θ over $\mathbb{R}^d$, the moments $\boldsymbol{M} = (M_0, M_1, M_2, \ldots)$ can then be expressed using k-tensors:

$$M_k \coloneqq \mathbb{E}_{\theta \sim \Theta}\left[\theta^{\otimes k}\right] = \int_{\mathbb{R}^d} \theta^{\otimes k} \, d\mu(\theta).$$

Intuitively, this is a k-dimensional array with entries indexed by sequences of k indices, $i_1, \ldots, i_k \in [d]$, where entries are the scalars $(M_k)_{i_1,\ldots,i_k} = \mathbb{E}_{\theta \sim \Theta}[\theta_{i_1} \cdots \theta_{i_k}]$. We will write $\boldsymbol{M}(\Theta)$ and $M_k(\Theta)$ when the distribution (measure) is not clear from context.

We can now relate the moments of the distribution over utilities $\mathcal{U}_\phi$ for a candidate ϕ to the moments of θ. In particular, let $m_k = \mathbb{E}_{\theta \sim \Theta}\left[u_\theta(\phi)^k\right]$ be the kth moment. Then by tensor arithmetic and linearity of $u_\theta(\phi)$, $m_k = \langle M_k, \phi^{\otimes k} \rangle$.

2.7 Moment Identifiability and Estimation

Our first concern is information-theoretic: what properties $\mathcal{P}$ of the voter distribution Θ (e.g., its kth moment tensor $M_k(\Theta)$) can be determined by the responses? We say that $\mathcal{P}$ is *identifiable* from Q_t if, for any two distributions Θ, Θ' that yield identical query responses (e.g., $Q_t(\Theta) = Q_t(\Theta')$), it holds that $\mathcal{P}(\Theta) = \mathcal{P}(\Theta')$. For example, the kth moment is identifiable with t-sized queries if, for any two distributions with distinct kth moments, there exist a $\mathbf{q} = (q_1, \ldots, q_t)$ and $\mathbf{b} = (b_1, \ldots, b_t)$ such that $Q_t(\mathbf{q}, \mathbf{b}; \Theta_1) \neq Q_t(\mathbf{q}, \mathbf{b}; \Theta_2)$.

In applications, we must *estimate* moments from finitely many voters. Since we are ultimately interested in using moments to estimate the utility distribution of given ϕ, we measure estimation error using the *spectral norm*, which for a kth order tensor $T \in (\mathbb{R}^d)^{\otimes k}$ is defined as

$$\|T\| \coloneqq \sup_{u_1,\ldots,u_k \in \mathbb{S}^{d-1}} \left|\langle T, u_1 \otimes \cdots \otimes u_k \rangle\right|.$$

For $k = 1$, the spectral norm corresponds to the standard L_2 norm. Control over the spectral norm guarantees uniform accuracy in estimating the moments of the utility distribution. Specifically, if $\widehat{M}_k$ satisfies $\|\widehat{M}_k - M_k\| \leq \varepsilon$, then for any $\phi \in \mathbb{R}^d$, the estimated kth moment of the utility distribution satisfies: $\left|\left\langle \widehat{M}_k, \phi^{\otimes k} \right\rangle - \mathbb{E}_{\theta \sim \Theta}\left[\langle \theta, \phi \rangle^k\right]\right| \leq \varepsilon \|\phi\|^k$.

3 Identifiability of Moments from Queries

A key first step toward estimating moments from pairwise comparisons is understanding the *minimum* number of queries per voter needed for identifiability. In this section, we provide such results in the form of identifiability. We start by considering the first moment M_1, which in our setting translates to effective social welfare maximization, showing that only a single per-voter query suffices. Next, we show several

more general results: 1) we can identify the first k moments using k per-voter queries, and 2) only two queries or only a single graded query per voter are sufficient to identify all moments (and hence, the distribution).

3.1 Identifying the Average Voter

This first moment M_1 is the average voter, i.e., $M_1 \coloneqq \bar{\theta} = \mathbb{E}_{\theta\sim\Theta}[\theta]$.

Observation 3.1. Knowing $\bar{\theta} = M_1$ suffices to maximize welfare of Θ; in particular, for any candidate embedding ϕ, we have $\mathbb{E}_{\theta\sim\Theta}[u_\theta(\phi)] = \mathbb{E}_{\theta\sim\Theta}\left[\theta^\top\phi\right] = \bar{\theta}^\top\phi$.

We now show that M_1 can be identified from 1-sized queries alone. Recall that Q_1 denotes pairwise comparison queries. Consider the contribution of a voter type θ to Q_1 when averaged over all query directions q, which we denote by

$$I(\theta) \coloneqq \int_{\mathbb{S}^{d-1}} \mathrm{resp}_\theta(q)\cdot q\, d\bar{\sigma}(q).$$

As query directions q are uniform, symmetry implies that this is itself in the direction of θ.

Lemma 3.2. *For any $\theta \in \mathbb{S}^{d-1}$ it holds that $I(\theta) = c_d\cdot\theta$, where*

$$c_d = \frac{\Gamma(d/2)}{2\sqrt{\pi}\Gamma(\frac{d+1}{2})} = \Theta(d^{-1/2}). \tag{1}$$

In particular, $c_d \geq \frac{1}{\sqrt{2\pi}}\cdot d^{-1/2}$.

Proof of Lemma 3.2. Let R be any rotation in $\mathbb{R}^d$. Because $\langle R\theta, q\rangle = \langle\theta, R^{-1}q\rangle$ and $d\bar{\sigma}$ is rotation invariant,

$$\begin{aligned} I(R\theta) &= \int \mathbb{1}\{\langle R\theta, q\rangle \geq 0\}\cdot q\, d\bar{\sigma}(q) = \int \mathbb{1}\{\langle\theta, R^{-1}q\rangle \geq 0\}\cdot q\, d\bar{\sigma}(q) \\ &= \int \mathbb{1}\{\langle\theta, q'\rangle \geq 0\}\, Rq'\, d\bar{\sigma}(q') = R\, I(\theta). \end{aligned}$$

Let $\mathcal{R}_\theta$ be the subgroup of rotations of $\mathbb{R}^d$ that fix θ. For any $R\in\mathcal{R}_\theta$, we have $R\theta = \theta$, and therefore by the equivariance established above, $R\, I(\theta) = I(R\theta) = I(\theta)$.

Therefore $I(\theta)$ is the fixed point of the linear action of the group $\mathcal{R}_\theta$. However, as $\mathcal{R}_\theta$ is acting as the full rotation group on the orthogonal complement of θ, the only fixed point in $\theta^\perp$ is the zero vector, and the only dimension left is for the span of the vector itself; hence $I(\theta) = c_d\cdot\theta + 0$ for some scalar c_d. To calculate this constant, we set $\theta = e_1$ to obtain

$$\begin{aligned} c_d &\coloneqq \int_{\mathbb{S}^{d-1}} \mathbb{1}\{q_1\geq 0\} q_1\, d\bar{\sigma}(q) = \frac{1}{2}\frac{\Gamma(d/2)}{\sqrt{\pi}\Gamma((d-1)/2)}\int_0^1 t(1-t^2)^{\frac{d-3}{2}}dt \\ &= \frac{1}{d-1}\frac{\Gamma(d/2)}{\sqrt{\pi}\Gamma(\frac{(d-1)}{2})} = \frac{\Gamma(d/2)}{2\sqrt{\pi}\Gamma(\frac{d+1}{2})} \geq \frac{1}{\sqrt{2\pi}}\cdot d^{-1/2}, \end{aligned}$$

where the second to last step follows from the Gamma recurrence $\Gamma(x+1) = x\Gamma(x)$, and the last using Wendel's Inequality [Wendel, 1948, Luo and Qi, 2012]; plugging in $x = d/2$ and $s = 1/2$ yields $\frac{\Gamma(d/2)}{\Gamma(d/2+1/2)} > \frac{\sqrt{2}}{d^{1/2}}$. □

This identity allows us to recover the first moment.

Lemma 3.3. *The first moment M_1 is identifiable with access to Q_1.*

Proof. By applying first Lemma 3.2 and then Fubini's theorem, we can write M_1 as

$$\mathbb{E}_{\theta\sim\Theta}[\theta] = \mathbb{E}_{\theta\sim\Theta}\left[\frac{1}{c_d}\cdot\int \mathbb{1}\{\langle\theta,q\rangle\geq 0\}\cdot q\, d\bar{\sigma}(q)\right] = \frac{1}{c_d}\cdot\int\int \mathbb{1}\{\langle\theta,q\rangle\geq 0\}\cdot q\, d\bar{\sigma}(q)\, d\Theta(\theta)$$
$$= \frac{1}{c_d}\cdot\int\int \mathbb{1}\{\langle\theta,q\rangle\geq 0\}\cdot q\, d\Theta(\theta)\, d\bar{\sigma}(q) = \frac{1}{c_d}\cdot\int Q_1(q)\cdot q\, d\bar{\sigma}(q). \qquad \square$$

Combining Lemma 3.3 with Observation 3.1, we have our first major result.

Theorem 3.4. *Welfare-maximizing candidates are identifiable from Q_1.*

3.2 Pairwise Queries and Higher Moments

We may naturally wonder whether Q_1 suffices to identify higher moments. As we show in the following example, it does not even for the second moment.

Example 1. Consider the voter type distribution $\Theta_{\pm\theta}$ over $\mathbb{S}^{d-1}$ given by placing half of the probability mass in an ε-neighborhood around the vector θ and the other half (antipodally) symmetrically around $-\theta$. By symmetry, $Q_1(q) = 1/2$ for all queries q regardless of θ. Furthermore, all candidates $\phi\in\mathbb{S}^{d-1}$ confer expected welfare 0. But candidates orthogonal to θ have variance ≈ 0, while candidates $c \coloneqq \beta\theta$ for a fixed β have variance β^2.

This example illustrates that some distributions are indistinguishable from one another via Q_1 and therefore from their first moments M_1, but that optimizing well-motivated nonlinear objectives requires distinguishing them. Naturally, this trend continues: knowing the moment tensors $M_1,\ldots,M_k$ does not determine subsequent moments, or the distribution overall, even when its support is constrained to $\mathbb{S}^{d-1}$.

Observation 3.5. For $d\geq 2$ and let $k\geq 1$. There exist two probability measures μ_+ and μ_- on the sphere $\mathbb{S}^{d-1}$ such that they have the same first k moments but have different $(k+1)$-st moments.

We defer details to Appendix C. In particular, this means that we should not hope to derive M_2 from M_1. Can we derive M_2 from Q_2? We now show that we can; in fact, we show that we can derive M_k from Q_k, for any k.

Theorem 3.6. *The kth moment tensor M_k is identifiable from*

$$M_k = \frac{1}{c_d^k}\cdot\mathbb{E}_{q_1,\ldots,q_k\sim\bar{\sigma}^k}\left[Q_k(q_1,\ldots,q_k)\cdot q_1\otimes\cdots\otimes q_k\right],$$

where c_d is defined as in Lemma 3.2.

Proof. In Lemma 3.2, we showed that for any fixed θ and for uniform $q\sim\bar{\sigma}$, $\mathbb{E}_q[\mathbb{1}\{\langle\theta,q\rangle\geq 0\}\cdot q] = c_d\cdot\theta$. We will now generalize this to the expectation over k independently chosen pairwise comparison directions. Consider sampling k uniformly independent comparison directions $\boldsymbol{q}\sim\bar{\sigma}^k$, where $\boldsymbol{q} = (q_1,\ldots,q_k)$. For fixed θ, due to independence,

$$\mathbb{E}_{\boldsymbol{q}}[\mathbb{1}\{\langle\theta,q_1\rangle\geq 0,\ldots,\langle\theta,q_k\rangle\geq 0\}\cdot q_1\otimes\cdots\otimes q_k] = \bigotimes_{j=1}^{k}\mathbb{E}_{q_j}[\mathbb{1}\{\langle\theta,q_j\rangle\geq 0\}\cdot q_j] = c_d^k\cdot\theta^{\otimes k},$$

where last step follows from Lemma 3.2. Taking the expectation over Θ and applying Fubini yields

$$\mathbb{E}_{\theta\sim\Theta}[\mathbb{E}_{\boldsymbol{q}}[\mathbb{1}\{\langle\theta,q_1\rangle\geq 0,\ldots,\langle\theta,q_k\rangle\geq 0\}\cdot q_1\otimes\cdots\otimes q_k]] = \mathbb{E}_{\boldsymbol{q}}[Q_k(\boldsymbol{q})\cdot q_1\otimes\cdots\otimes q_k]$$
$$= c_d^k\cdot\mathbb{E}_{\theta\sim\Theta}\left[\theta^{\otimes k}\right].$$

Rearranging gives the stated claim. $\square$

3.3 Identifying the Voter Distribution via Size-2 Queries

Having established that kth moments of Θ are identifiable from k pairwise comparison queries, we now show an even stronger result: *the full voter distribution is identifiable from two pairwise queries.* Although in Section 4, we observe that the associated sample complexity is higher, it is quite remarkable that this is even possible.

First, we note that the set of all moments $\boldsymbol{M}$ uniquely identify Θ. This is a well-known fact in probability and real analysis, even for the more general setting where the support of the distribution is a compact subset of $\mathbb{R}^d$ (e.g. [Schmüdgen, 2017, Corollary 14.9]).

Lemma 3.7. *If voter distributions Θ and Θ' have equal moments $\boldsymbol{M}(\Theta) = \boldsymbol{M}(\Theta')$, then they are in fact equal: for all measurable sets $T \subseteq \mathbb{S}^{d-1}$ it holds that $\Theta(T) = \Theta'(T)$.*

We outline the argument for completeness in Appendix C. Combined with Theorem 3.6, this says that if we can ask each voter infinitely many queries, we can distinguish any two distributions Θ and Θ'. The problem is, of course, that asking a single voter an unbounded number of queries is infeasible. We now demonstrate that this is unnecessary.

Theorem 3.8. *Distributions are identifiable from 2-sized queries.*

The high-level proof relies on the spectral decomposition of functions on the sphere into *spherical harmonics*. Recall that a single query effectively measures the fraction of voters residing in a specific hemisphere. In the spherical harmonic literature, this mapping is known as the *hemispherical transform*. This transform is invertible on the subspace of odd-degree spherical harmonics, allowing us to identify $\mathbb{E}_\Theta[p(\theta)]$ for any odd-degree harmonic p. Using linearity of expectation, this allows us to reconstruct all odd-degree moments.

To identify even moments, we observe that any even-degree polynomial can be decomposed into a sum of a product of two odd-degree harmonics. This reduces the problem to identifying values of the form $\mathbb{E}_\Theta[p(\theta)p'(\theta)]$. By asking two queries to each arriving voter, we can estimate the joint expectation of the responses, which identifies this product and allows us to recover all even moments.

Proof of Theorem 3.8. We will show that all moments are identifiable. Lemma 3.7 then implies that distributions are identifiable.

Our analysis relies on the spectral theory of functions on the sphere [Groemer, 1996, Chapter 3]. For each $j \geq 0$, let $\mathcal{H}_j$ denote the finite-dimensional space of *spherical harmonics* of degree j on $\mathbb{S}^{d-1}$. An important fact is that any homogeneous polynomial can be decomposed into a sum of spherical harmonics [Groemer, 1996, Lemma 3.2.5]. Specifically, if p is a homogeneous polynomial of degree k, we can write it as a sum of spherical harmonics, one for each degree $j \leq k$ such that $k - j$ is even. We begin with odd k, and show how to extend our analysis to even k later. In particular, if k is odd, there exists $f_j \in \mathcal{H}_j$ for each odd $j \leq k$ such that for all $\theta \in \mathbb{S}^{d-1}$,

$$p(\theta) = \sum_{j \leq k \ j \text{ is odd}} f_j(\theta)\,.^4$$

For our purposes, p will be a k-degree monomial (e.g., $p(\theta) = \prod_{j=1}^k \theta_{i_j}$) for some indices $i_1, \ldots, i_k$). Our goal is equivalent to recover $\mathbb{E}_{\theta \sim \Theta}[p(\theta)]$ for all monomial p.

We now use a second property of spherical harmonics, i.e., Funk-Hecke Theorem (see Thm. B.1). It suggests that for any fixed spherical harmonic $f \in \mathcal{H}_j$ with j being odd, we have

$$f(\theta) = \frac{1}{\mu_j} \int_{S^{d-1}} \mathbb{1}\{\langle \theta, q \rangle \geq 0\} f(q)\, d\bar{\sigma}(q),$$

[4] Technically, Groemer [1996] proves this for polynomials *not* restricted to the sphere for all $x \in \mathbb{R}^d$, and it is of the form $p(x) = \sum_{j \leq k : j \text{ is odd}} \|x\|^{k-j} f_j(x)$. However, restricted to the sphere, $\|x\| = 1$.

for a nonzero constant μ_j.

Taking the expectation over Θ,

$$\begin{aligned}\mathbb{E}_{\theta\sim\Theta}\left[f(\theta)\right] &= \frac{1}{\mu_j}\iint \mathbb{1}\{\langle\theta,q\rangle\geq 0\}f(q)\,d\bar{\sigma}(q)d\Theta(\theta) = \frac{1}{\mu_j}\iint \mathbb{1}\{\langle\theta,q\rangle\geq 0\}f(q)\,d\Theta(\theta)d\bar{\sigma}(q)\\ &= \frac{1}{\mu_j}\int f(q)\int \mathbb{1}\{\langle\theta,q\rangle\geq 0\}\,d\Theta(\theta)d\bar{\sigma}(q).\end{aligned}$$

The exchange of integrals is justified by Fubini's theorem, as the integrand is bounded ($\mathbb{1}\{\langle\theta,q\rangle\geq 0\}$ is trivially bounded, and spherical harmonics are continuous functions on the compact sphere), and the measures are finite. The inner integral corresponds to the expected query response:

$$\int \mathbb{1}\{\langle\theta,q\rangle\geq 0\}\,d\Theta(\theta) = \Pr_{\Theta}[\langle q,\theta\rangle\geq 0] = Q_1(q).$$

Thus, the expected value of the harmonic is fully determined by Q_1:

$$\mathbb{E}_{\theta\sim\Theta}\left[f(\theta)\right] = \int f(\theta)\,d\Theta(\theta) = \frac{1}{\mu_j}\int f(q)Q_1(q,1)\,d\bar{\sigma}(q).$$

Next, consider even k. We can decompose any monomial p of degree k into a product of two monomials of odd degrees p_1, p_2 of degrees k_1 and k_2 (e.g., by partitioning the index set such that $k_1+k_2=k$). As we have just seen, both p_1 and p_2 have decompositions into odd-degree spherical harmonics $f_j\in\mathcal{H}_j$ for odd $j\leq k_1$ and $f'_{j'}\in\mathcal{H}_{j'}$ for $j'\leq k_2$ such that,

$$p_1 = \sum_{j\leq k_1 : j \text{ is odd}} f_j \qquad \text{and} \qquad p_2 = \sum_{j'\leq k_2 : j' \text{ is odd}} f'_{j'}.$$

Thus, we can write

$$p(x) = \sum_{j\leq k_1, j'\leq k_2, j, j' \text{ are odd}} f_j(x)\cdot f'_{j'}(x). \tag{2}$$

By (2), to determine $\mathbb{E}_{\theta\sim\Theta}\left[p(\theta)\right]$ for even-degree p it therefore suffices to deduce the expectation of the product of any two odd-degree spherical harmonics f and f'; that is, $\mathbb{E}_{\theta\sim\Theta}\left[f(\theta)f'(\theta)\right]$.

Fix two such spherical harmonics f and f' of odd degrees j and j'. As above, we can write

$$f(\theta) = \frac{1}{\mu_j}\int \mathbb{1}\{\langle\theta,q\rangle\geq 0\}f(q)\,d\bar{\sigma}(q) \qquad \text{and} \qquad f'(\theta) = \frac{1}{\mu_{j'}}\int \mathbb{1}\{\langle\theta,q\rangle\geq 0\}f'(q)\,d\bar{\sigma}(q).$$

By the linearity of the integral,

$$f(\theta)\cdot f'(\theta) = \frac{1}{\mu_j\mu_{j'}}\iint \mathbb{1}\{\langle\theta,q\rangle\geq 0\}\,\mathbb{1}\{\langle\theta,q'\rangle\geq 0\}f(q)f'(q')\,d\bar{\sigma}(q)\,d\bar{\sigma}(q').$$

Taking the expectation with respect to Θ, we have

$$\begin{aligned}\mathbb{E}_{\theta\sim\Theta}\left[f(\theta)\cdot f'(\theta)\right] &= \frac{1}{\mu_j\mu_{j'}}\iint \mathbb{1}\{\langle\theta,q\rangle\geq 0\}\,\mathbb{1}\{\langle\theta,q'\rangle\geq 0\}f(q)f'(q')\,d\bar{\sigma}(q)\,d\bar{\sigma}(q')\\ &= \frac{1}{\mu_j\mu_{j'}}\iint f(q)f'(q')\left[\int \mathbb{1}\{\langle\theta,q\rangle\geq 0\}\,\mathbb{1}\{\langle\theta,q'\rangle\geq 0\}\,d\Theta(\theta)\right]d\bar{\sigma}(q)\,d\bar{\sigma}(q'),\end{aligned}$$

where we again use Fubini's theorem for the integral swap. Finally, letting $\boldsymbol{q}=(q,q')$ we have

$$\int \mathbb{1}\{\langle\theta,q\rangle\geq 0\}\,\mathbb{1}\{\langle\theta,q'\rangle\geq 0\}\,d\Theta(\theta) = \Pr_{\Theta}[\mathbb{1}\{\langle q,\theta\rangle\rangle\geq 0\}\wedge\mathbb{1}\{\langle q',\theta\rangle\rangle\geq 0\}] = Q_2(\boldsymbol{q}).$$

Hence,

$$\mathbb{E}_{\theta\sim\Theta}\left[f(\theta)\cdot f'(\theta)\right] = \frac{1}{\mu_j\mu_{j'}}\iint f(q)f'(q')Q_2(\boldsymbol{q})\,d\bar{\sigma}(q)\,d\bar{\sigma}(q'),$$

implying identifiability. □

3.4 Identifying the Voter Distribution via Graded Queries

Next, we consider graded queries. While these require the stronger Assumption 2.4, they yield a powerful return: the entire distribution is identifiable using only a single query per voter.

Theorem 3.9. *For almost all $\tau \in (0,1)$, distributions are identifiable with access to G_τ.*

The proof is deferred to Appendix C. The high level approach is to show that each moment k is identifiable by (strong) induction on k. Suppose we know the first $k-1$ moments. We consider the correlation between the graded response probability $G_\tau(q)$ and the k-th tensor power of the query vector q, integrated over the uniform distribution of queries. Using the rotational symmetry of the sphere, this integral simplifies into a linear equation involving the unknown k-th moment of the voter distribution scaled by a specific scalar coefficient, plus terms composed entirely of lower-order moments (which are known by the inductive hypothesis). Crucially, this scaling coefficient depends on the threshold τ and behaves like a polynomial with a finite number of roots. Therefore, for almost all choices of τ (i.e., any τ not in this finite set of roots), the coefficient is non-zero, allowing us to invert the equation and uniquely recover the k-th moment. By excluding the countable union of such roots across all k, we can identify all moments, and hence the entire distribution, for almost every τ. Notably, if we only care about the first two moments, any $\tau \in (0,1)$ suffices.

4 Moment Estimation

In order to make our identifiability results useful, we must contend with the fact that we do not have perfect knowledge of these distributions, but instead can only obtain a finite set of samples of such comparisons. How many voters are sufficient in order to approximately learn the moments of Θ? In general, we could have an algorithm that chooses which query $\mathbf{q}$ to make based on all previous responses. For all of our positive results, it will suffice to have $\mathbf{q}$ be selected uniformly at random. Our goal in this section is to bound the *sample complexity* of moment estimation.

Definition 4.1 (Sample Complexity)**.** Estimating the kth moment from comparison queries has *sample complexity* T if there is an estimator $\widehat{M}_k$ taking in T t-sized regular query-response pairs $\boldsymbol{r} = \{\boldsymbol{q}_i, (\mathrm{resp}_{\theta_i}(\boldsymbol{q}_i)\}_{i=1}^T$ of T i.i.d. voters $\theta_i \sim \Theta$ to T uniformly random t-sized queries $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_T \sim \bar{\sigma}^t$, such that for all unknown Θ,

$$\Pr_{\boldsymbol{r}}\left[\left\|\widehat{M}_k(\boldsymbol{r}) - M_k(\Theta)\right\| \le \varepsilon\right] \ge 1-\delta.$$

4.1 Warmup: Welfare Maximization from Queries

Our first result shows that we can successfully estimate the first moment of the voter distribution by requesting a response to a single pairwise query from a polynomial number of voters.

Theorem 4.2. *The sample complexity of estimating the first moment from single queries is $O(\frac{d}{\varepsilon^2}\log\frac{1}{\delta})$.*

Proof. Consider the double integral over the probability measure for both the voter distribution Θ and the query distribution $\mathrm{Unif}\left(\mathbb{S}^{d-1}\right)$, which we denote by $d\Theta$ and $d\bar{\sigma}(q)$ respectively:

$$\iint_{\mathbb{S}^{d-1}\times\mathbb{S}^{d-1}} \mathrm{resp}_\theta(q)\cdot q\, d\bar{\sigma}(q)\, d\Theta(\theta) = \int_{\mathbb{S}^{d-1}} c_d\cdot\theta\, d\Theta(\theta) = c_d\cdot\bar{\theta}.$$

The corresponding estimator from T samples is $\widehat{\bar{\theta}} = \frac{1}{c_d T}\cdot\sum_{i\in[T]}\mathbb{1}\left\{\langle q_i,\theta_i\rangle \ge 0\right\}\cdot q_i$.

We would like to bound the rate at which this approaches $\bar{\theta}$ in spectral norm, which is just L_2 norm. That is, we would like to upper bound the T required to satisfy

$$\Pr\left[\left\|\bar{\theta} - \widehat{\bar{\theta}}\right\|_2 > \varepsilon\right] \le \delta.$$

We take a standard approach and appeal to McDiarmid's inequality [McDiarmid et al., 1989]. To this end, let $\hat{\theta}_i \coloneqq \frac{1}{c_d} \cdot \mathrm{resp}_{\theta_i}(q_i) \cdot q_i$ define the auxiliary random variable $y_i \coloneqq \bar{\theta} - \hat{\theta}_i$, and let $f(y_1, \ldots, y_T) \coloneqq \|\sum_i y_i\|_2 = \left\|\bar{\theta} - \widehat{\bar{\theta}}\right\|_2$. Our goal is then equivalent to upper bounding $\Pr[f(y_1, \ldots, y_T) \geq \varepsilon \cdot T]$. By the triangle inequality, this f satisfies the bounded difference property that for all y_i and y_i' it holds that $|f(y_1, \ldots, y_i, \ldots, y_T) - f(y_1, \ldots, y_i', \ldots, y_T)| \leq \left\|\hat{\theta}_i - \hat{\theta}_i'\right\|_2 = O(c_d^{-1})$, since our sampled q_i have unit norm. Therefore by McDiarmid's inequality,

$$\Pr\left[\left\|\bar{\theta} - \widehat{\bar{\theta}}\right\|_2 > \varepsilon\right] = \Pr[f(y_1, \ldots, y_T) \geq \varepsilon \cdot T] \leq \exp\left(\frac{-2\varepsilon^2 T}{C \cdot c_d^{-2}}\right)$$

for some constant C. Setting this upper bound on the failure probability to δ, solving for T, and recalling that $c_d = \Theta(d^{-1/2})$ finally yields $T = O(\frac{d}{\varepsilon^2} \log \frac{1}{\delta})$, as claimed. □

4.2 Moments from k-Wise Queries

Our next result significantly generalizes our observation above that the first moment can be efficiently estimated, showing that this, in fact, holds true more generally.

Theorem 4.3. *The sample complexity of estimating the kth moment from k-sized queries is*

$$O\left(\frac{(2\pi)^k \cdot k \cdot d^{\lceil k/2 \rceil}}{\varepsilon^2} \log\left(\frac{d}{\delta}\right)\right).$$

A technical challenge is that tensors are not simply matrices with more indices; even extending familiar matrix notions—such as computing the norm of higher-order tensors is nontrivial [Hillar and Lim, 2013]. And there is no comparably sharp, general-purpose concentration theory for tensors. We therefore work with matricizations of the relevant moment tensors, which allows us to leverage the matrix Bernstein inequality [Tropp, 2012, Theorem 1.6.2].

Theorem 4.4 (Matrix Bernstein Inequality for real-valued matrices)**.** *Let $S_1, \ldots, S_n$ be independent, centered real random matrices with common dimension $d_1 \times d_2$, and assume that each one is uniformly bounded, i.e., $\mathbb{E}[S_k] = 0$ and $\|S_k\| \leq L$ for each $k = 1, \ldots, n$. Let $Z = \sum_{k=1}^n S_k$, and let $\nu(Z)$ denote the matrix variance statistic of the sum:*

$$\nu(Z) = \max\left\{\|\mathbb{E}(ZZ^T)\|, \|\mathbb{E}(Z^TZ)\|\right\} = \max\left\{\left\|\sum_{k=1}^n \mathbb{E}(S_k S_k^T)\right\|, \left\|\sum_{k=1}^n \mathbb{E}(S_k^T S_k)\right\|\right\}.$$

Then, for all $t \geq 0$,

$$\mathbb{P}\{\|Z\| \geq t\} \leq (d_1 + d_2) \cdot \exp\left(\frac{-t^2/2}{\nu(Z) + Lt/3}\right).$$

The following standard result shows that it is sufficient to bound the norm of the matricization.

Lemma 4.5 (Proposition 4.1, [Wang et al., 2017])**.** *Let $T \in (\mathbb{R}^d)^{\otimes k}$ and let $I \sqcup J = \{1, \ldots, k\}$. Then $\|T\| \leq \|\mathrm{Mat}_{I|J}(T)\|$.*

Note that had we not matricized, typical vector-based concentration inequalities would end up with a dominant d^k factor. The matricization and more sophisticated matrix Bernstein inequality allow us to cut $\approx d^{\lfloor k/2 \rfloor}$ off of the sample complexity.

Proof of Theorem 4.3. We will make use of Theorem 3.6. Recall that $Q_k(q_1, \ldots, q_k)$ is the probability that a random voter θ has $\mathrm{resp}_\theta(q_i) = 1$ for all i. Thus, we can rewrite the statement as

$$M_k = \frac{1}{c_d^k} \cdot \mathbb{E}_{q_1,\ldots,q_k \sim \bar{\sigma}^k, \theta \sim \Theta} \left[\mathbb{1}\{\mathrm{resp}_\theta(q_i) = 1 \, \forall i\} \cdot q_1 \otimes \cdots \otimes q_k \right].$$

Given T sampled queries and responses $\boldsymbol{r} = \{\boldsymbol{q}_i, (\mathrm{resp}_{\theta_i}(\boldsymbol{q}_i)\}_{i=1}^T$, we let $\chi_i \coloneqq \mathbb{1}\{\mathrm{resp}_{\theta_i}(q_{i,j}) = 1 \, \forall j\}$ be the indicator that the i-th voter agreed with all k queries. We define the empirical estimator for the kth moment as:

$$\widehat{M}_k(\boldsymbol{r}) \coloneqq \frac{1}{T} \sum_{i=1}^{T} \frac{\chi_i}{c_d^k} (q_{i,1} \otimes \cdots \otimes q_{i,k}),$$

which, by above, is an unbiased estimate of M_k.

To bound the estimation error $\left\| \widehat{M}_k - M_k \right\|$, we fix an arbitrary partition of the modes $I \sqcup J = \{1, \ldots, k\}$ with $|I| = s$ and $|J| = k - s$. Let $(d_1, d_2) \coloneqq (d^s, d^{k-s})$. By Lemma 4.5, it suffices to bound the spectral norm of the matricized difference as $\left\| \mathrm{Mat}_{I|J}(\widehat{M}_k - M_k) \right\|$. We define independent random matrices $A_i \in \mathbb{R}^{d_1 \times d_2}$ such that:

$$A_i \coloneqq \mathrm{Mat}_{I|J} \left(\frac{1}{c_d^k} \cdot \chi_i \cdot q_{i,1} \otimes \cdots \otimes q_{i,k} \right).$$

Notice that $\mathbb{E}[A_i] = \mathrm{Mat}_{I|J}(M_k)$. Then, $\mathrm{Mat}_{I|J}(\widehat{M}_k - M_k) = \frac{1}{T} \sum_{i=1}^T S_i$, where $S_i \coloneqq A_i - \mathbb{E}[A_i]$ are independent, zero-mean random matrices. To apply the Matrix Bernstein inequality, we must compute a uniform bound L on the spectral norm of the summands and a bound ν on the total variance statistic.

Towards establishing this uniform bound, for each i we can write the random matrix as an outer product $A_i = \frac{\chi_i}{c_d^k} \cdot u_i v_i^\top$, where $u_i \coloneqq \bigotimes_{j \in I} q_{i,j} \in \mathbb{R}^{d_1}$ and $v_i \coloneqq \bigotimes_{j \in J} q_{i,j} \in \mathbb{R}^{d_2}$. Using the identity that $\|x \otimes y\| = \|x\| \cdot \|y\|$, since the queries are all unit norm, $\|u_i\|_2 = \|v_i\|_2 = 1$. Because $\chi_i \in \{0, 1\}$, we have: $\|A_i\| = \frac{\chi_i}{c_d^k} \cdot \|u_i\|_2 \|v_i\|_2 \le \frac{1}{c_d^k}$. Since all norms are convex, by Jensen's inequality $\|\mathbb{E}[A_i]\| \le \mathbb{E}[\|A_i\|] \le \max_i \|A_i\| \le c_d^{-k}$. Then by the triangle inequality, we have

$$\|S_i\| \le \|A_i\| + \|\mathbb{E}[A_i]\| \le \frac{2}{c_d^k} =: L.$$

We now bound the variance statistic $\nu \coloneqq \max\left\{ \left\| \sum_{i=1}^T \mathbb{E}\left[S_i S_i^\top\right] \right\|, \left\| \sum_{i=1}^T \mathbb{E}\left[S_i^\top S_i\right] \right\| \right\}$. Since the T voters are independent, the total variance is bounded by:

$$\nu \le T \cdot \max_i \max \left\{ \mathbb{E}\left[\left\| S_i S_i^\top \right\| \right], \left\| \mathbb{E}\left[S_i^\top S_i \right] \right\| \right\}. \tag{3}$$

Next note that $\mathbb{E}\left[S_i S_i^\top\right] = \mathbb{E}\left[A_i A_i^\top\right] - \mathbb{E}[A_i]\,\mathbb{E}[A_i]^\top$. We show that this implies

$$\left\| \mathbb{E}\left[S_i S_i^\top\right] \right\| \le \left\| \mathbb{E}\left[A_i A_i^\top\right] \right\|. \tag{4}$$

Indeed, for any unit vector x,

$$x^\top \mathbb{E}\left[S_i S_i^\top\right] x = x^\top \left(\mathbb{E}[A_i A_i^\top] - \mathbb{E}[A_i]\mathbb{E}[A_i]^\top \right) x = x^\top \mathbb{E}[A_i A_i^\top] x - x^\top \mathbb{E}[A_i]\mathbb{E}[A_i]^\top x.$$

Letting $y = \mathbb{E}[A_i]^\top x$, this is equal to $x^\top \mathbb{E}[A_i A_i^\top] x - y^\top y$. Since $y^\top y \ge 0$, this implies $x^\top \mathbb{E}[A_i A_i^\top] x \ge x^\top \mathbb{E}\left[S_i S_i^\top\right] x$ for all x. Hence $\left\| \mathbb{E}\left[S_i S_i^\top\right] \right\| \le \left\| \mathbb{E}[A_i A_i^\top] \right\|$.

With this established, we turn to upper bounding $\left\|\mathbb{E}\left[A_i A_i^\top\right]\right\|$. Utilizing the structure of A_i, the fact that v_i is a unit vector (and hence $v_i^\top v_i = 1$) and $\chi_i^2 = \chi_i \leq 1$, we have:

$$\left\|A_i A_i^\top\right\| = \left\|\frac{\chi_i^2}{c_d^{2k}} \cdot u_i (v_i^\top v_i) u_i^\top\right\| \leq \left\|\frac{1}{c_d^{2k}} \cdot u_i u_i^\top\right\|.$$

Since the query vectors $q_{i,j}$ are drawn independently and uniformly from $\mathbb{S}^{d-1}$, $\mathbb{E}\left[q_{i,j} q_{i,j}^\top\right] = \frac{1}{d} I_d$. By independence across the distinct queries:

$$\mathbb{E}\left[u_i u_i^\top\right] = \left(\mathbb{E}\left[q_{i,j} q_{i,j}^\top\right]\right)^{\otimes s} = \left(\frac{1}{d} I_d\right)^{\otimes s} = \frac{1}{d^s} I_{d_1}.$$

Furthermore, $\left\|\frac{1}{d^s} I_{d_1}\right\| = \frac{1}{d^s}$, so we get that $\left\|\mathbb{E}\left[A_i A_i^\top\right]\right\| \leq \frac{1}{c_d^{2k} d^s}$. Applying the analogous argument to $\left\|\mathbb{E}\left[A_i^\top A_i\right]\right\|$, we get $\left\|\mathbb{E}\left[A_i^\top A_i\right]\right\| \leq \frac{1}{c_d^{2k} d^{k-s}}$. From (3) and (4), this implies $\nu \leq \frac{T}{c_d^{2k}} \max\left(d^{-s},\, d^{s-k}\right) = \frac{T}{c_d^{2k}} d^{-\min(s,k-s)}$. Applying the Matrix Bernstein inequality then yields:

$$\Pr\left(\left\|\frac{1}{T}\sum_{i=1}^{T} S_i\right\| \geq \varepsilon\right) \leq (d_1 + d_2) \exp\left(-\frac{T^2\varepsilon^2/2}{\nu + LT\varepsilon/3}\right).$$

To ensure the error satisfies $\left\|\widehat{M}_k - M_k\right\| \leq \varepsilon$ with probability at least $1 - \delta$, we enforce the failure probability bound $\leq \delta$. Substituting the derived bounds for ν and L, a sufficient condition for the sample complexity is:

$$T \geq \frac{2}{\varepsilon^2}\left(\frac{1}{c_d^{2k} d^{\min(s,k-s)}} + \frac{2\varepsilon}{3c_d^k}\right) \log\left(\frac{d^s + d^{k-s}}{\delta}\right).$$

We set $s = \lfloor k/2 \rfloor$. Furthermore, recall from Lemma 3.2 that $c_d > \frac{1}{\sqrt{2\pi}} d^{-1/2}$. Furthermore, it is without loss of generality to assume $\varepsilon \leq 1$, as the 0-tensor is a trivial 1-approximation as $\|M_k\| \leq 1$. Thus,

$$T = O\left(\frac{1}{\varepsilon^2}\left(\frac{d^{-\lfloor k/2 \rfloor}}{c_d^{2k}} + \frac{1}{c_d^k}\right) \log\left(\frac{d^{\lceil k/2 \rceil}}{\delta}\right)\right) = O\left(\frac{1}{\varepsilon^2}\left(d^{k-\lfloor k/2 \rfloor} \cdot (2\pi)^k + d^{k/2} \cdot (2\pi)^{k/2}\right) \log\left(\frac{d^{\lceil k/2 \rceil}}{\delta}\right)\right)$$
$$= O\left(\frac{(2\pi)^k \cdot k \cdot d^{\lceil k/2 \rceil}}{\varepsilon^2} \log\left(\frac{d}{\delta}\right)\right). \qquad \square$$

4.3 Estimating All Moments from Few Queries

Our general result about moment estimation above requires k queries per voter. In Section 3, on the other hand, we showed that we can *identify* all moments with only two regular queries. How does this translate into sample efficiency?

The following result shows that it is still possible to estimate, although our upper bound is weaker, requiring on the order of d^{3k+2} rather than $d^{\lceil k/2 \rceil}$.

Theorem 4.6. *The sample complexity of estimating the kth moment from size-2 queries is*

$$O\left(\frac{k(d+1)^{3k+2}}{\varepsilon^2} \log\left(\frac{d}{\delta}\right)\right).$$

The proof is deferred to Appendix C. Our high-level approach mirrors the strategy used for k-wise queries in Theorem 4.3: we construct an unbiased estimator for the k-th moment tensor and bound its convergence using the Matrix Bernstein inequality. However, the construction of this estimator is

considerably more involved than in the k-wise case, analogous to how Theorem 3.8 was more involved than Theorem 3.6. Recall that the constructive proof of Theorem 3.8 establishes that the expectation of any degree-k monomial can be recovered by integrating the product of two specific spherical harmonic functions against the query distribution Q_2. Consequently, we can define our estimator via the empirical mean of these harmonic functions evaluated on sampled query pairs. The central technical challenge lies in bounding the magnitude of this estimator; for this, we must turn to more explicit constructions of spherical harmonics. These magnitude bounds lead to less efficient reconstruction: the sample complexity scales as $O(d^{3k+2})$ compared to the $O(d^{\lceil k/2 \rceil})$ achieved with k-wise queries.

Regarding graded queries, we leave the sample complexity bound to future work. While Theorem 3.9 establishes identifiability for almost all thresholds τ and provides a corresponding unbiased estimator, the reconstruction relies on inverting a specific coefficient that depends on τ. This coefficient vanishes at certain "blind spots," and without a quantitative bound on how close an arbitrary τ is to these problematic points, we cannot control the magnitude or variance of the resulting estimator.

5 Social Choice with Sparse Query Responses

Classical social choice rules take full rankings as input. Here we show it is possible to design a complementary class of rules that only require extremely sparse query responses from voters. All missing proofs in this section are deferred to Appendix C.

5.1 Moment-based Objectives

We begin by considering our moment-based-objectives. Leveraging our sample complexity results for estimating moments, we can provide end-to-end guarantees.

Maximizing Social Welfare. In a utilitarian setting, a natural and common objective is to maximize social welfare. Our results imply that this can be done with polynomial sample complexity:

Proposition 5.1. *By asking 1 query per voter, with at least $T \in O(\frac{d}{\varepsilon^2} \log \frac{1}{\delta})$ arriving voters, with probability $1-\delta$ over the arriving voters, for any candidate ϕ with $\|\phi\| \leq B$, we can estimate the social welfare $\mathbb{E}_\theta [u_\theta(\phi)]$ up to additive error εB. In particular, given a context x and ℓ candidates $y_1, \ldots, y_\ell$ such that $\|\Phi(x, y_i)\| \leq B$, we can select $\hat{y}_i$ within $2\varepsilon B$ of the optimal welfare.*

Maximizing Risk-Adjusted Welfare. A significant limitation of social welfare maximization is that it can yield extremely inequitable outcomes when voter preferences are highly diverse. Risk-adjusted welfare addresses this by explicitly accounting for inequality aversion. Armed with the first two moments, we can approximately maximize such welfare objectives.

Proposition 5.2. *By asking 2 queries per voter, for all $\varepsilon \leq 1$, with at least $T \in O(\frac{d}{\varepsilon^2} \log \frac{d}{\delta})$ arriving voters, with probability $1-\delta$ over the arriving voters, for any candidate ϕ with $\|\phi\| \leq B$, we can estimate $\mathrm{raw}_\alpha(\phi)$ up to $\sqrt{3} \cdot (\alpha+1) B \sqrt{\varepsilon}$. In particular, given a prompt x and ℓ possible responses $y_1, \ldots, y_\ell$ such that $\|\Phi(x, y_i)\| \leq B$, we will be able to select one within $2\sqrt{3} \cdot (\alpha+1) B \sqrt{\varepsilon}$ of the optimal risk-adjusted welfare.*

5.2 Beyond Moment-based Objectives

Next, we show how moments can be used to approximate other, more general objectives.

Maximizing Nash Welfare. Like risk-adjusted welfare, Nash welfare, aims to balance social welfare and distributed fairness. Defined as $\mathrm{Nash}(\phi) = \mathbb{E}_\Theta [\log(u_\theta(\phi))]$, this objective corresponds to maximizing the geometric mean of utilities (or the product in the finite case). Naturally, this is only well-defined if voter utilities are strictly positive. While Nash welfare cannot be computed exactly from a finite set of moments, we show that if the candidate's induced utilities lie within a known bounded interval, we can efficiently approximate the objective using moments.

Theorem 5.3. *Let ϕ be a candidate and let $[a, b] \subset (0, \infty)$ be a known interval such that voter utilities satisfy $u_\theta(\phi) \in [a, b]$ almost surely. Let $r \coloneqq a/b$. The Nash welfare can be estimated using the first k moment tensors $M_1, \ldots, M_k$ up to an additive error of:*

$$|Est_k - Nash(\phi)| \leq \frac{\sqrt{r}-1}{k+1}\left(1 - \frac{2}{\sqrt{r}+1}\right)^k.$$

Thus, we can achieve an approximation that improves exponentially quickly once $k \in \Omega(\sqrt{r})$. The proof uses standard techniques for approximating logarithm using kth-degree polynomials, specifically, using *Chebyshev polynomials* [Mason and Handscomb, 2002]. The expectation of any kth-degree polynomial can then be computed using the first k moments.

Proof of Theorem 5.3. The proof proceeds in two steps: first, we construct a kth-degree polynomial P_k such that

$$|P_k(x) - \log(x)| \leq \frac{\sqrt{r}-1}{k+1}\left(1 - \frac{2}{\sqrt{r}+1}\right)^k$$

for all $x \in [a, b]$. Then, we will show how to use P_k to estimate Nash welfare.

We will use *Chebyshev polynomials* to approximate log on the interval. See Mason and Handscomb [2002] for an overview. Each $T_k(x)$ is a kth-degree polynomial, defined on $[-1, 1]$, such that $|T_k(x)| \leq 1$. They are defined such that $T_k(\cos(\theta)) = \cos(k\theta)$ (which, perhaps surprisingly, defines a kth-degree polynomial mapping $[-1, 1]$ to $[-1, 1]$).

We would like to approximate $\log(x)$ on $[a, b]$. We first reduce this to approximating on $[-1, 1]$. We map the interval $[-1, 1]$ to $[a, b]$ using the affine transformation $g(t) \coloneqq \frac{b+a}{2} + \frac{b-a}{2}t$, and we let $\delta \coloneqq \frac{b-a}{b+a} = \frac{1-r}{1+r}$ where $r = a/b$. Substituting this into the logarithm function, we have $\log(g(t)) = \log(1+\delta t) + \log((b+a)/2)$. Now, if we have a kth-degree polynomial p' that is uniformly within ε of $\log \circ g$ for $t \in [-1, 1]$, then $p'(g^{-1}(\cdot))$ is a kth-degree polynomial that is within ε of $\log(x)$ on $[a, b]$. Furthermore, it suffices to find a kth-degree polynomial that approximates $\log(1 + \delta t)$, as $\log(g(t))$ differs from this function only by the constant $\log(\frac{b+a}{2})$.

To get this in a form using Chebyshev polynomials, we will make use of the identity [Gradshteyn and Ryzhik, 2014, Equation 1.514]

$$\log(1 - 2\alpha\cos(\varphi) + \alpha^2) = -2\sum_{j=1}^{\infty}\frac{\cos(j\varphi)}{j}\cdot\alpha^j,$$

for $\alpha^2 \leq 1$ and $\alpha\cos(\varphi) \neq 1$. In Appendix C, we show how to use this to derive that

$$\log(1 + t\delta) = -\log(1 + \alpha^2) - 2\sum_{j=1}^{\infty}\frac{\alpha^j}{j}\cdot T_j(t), \tag{5}$$

for $\alpha = -\frac{\sqrt{r}-1}{\sqrt{r}+1}$. We can choose the kth-degree polynomial $R_k(t) \coloneqq -\log(1 + \alpha^2) - 2\sum_{j=1}^{k}\frac{\alpha^j}{j}\cdot T_j(t)$ to approximate $\log(1 + \delta t)$. The error is

$$|\log(1 + t\delta) - R_k(t)| = \left|2\sum_{j=k+1}^{\infty}\frac{\alpha^j}{j}\cdot T_j(t)\right|. \tag{6}$$

Using that $|T_j(t)| \leq 1$ for $t \in [-1, 1]$, we can upper bound this error by

$$\begin{aligned}\sum_{j=k+1}^{\infty}\frac{2|\alpha|^j}{j} &\leq \frac{2}{k+1}\sum_{j=k+1}^{\infty}|\alpha|^j = \frac{2|\alpha|^{k+1}}{(k+1)(1-|\alpha|)} = \frac{\sqrt{r}+1}{k+1}\left(\frac{\sqrt{r}-1}{\sqrt{r}+1}\right)^{k+1} \\ &= \frac{\sqrt{r}-1}{k+1}\left(\frac{\sqrt{r}-1}{\sqrt{r}+1}\right)^k = \frac{\sqrt{r}-1}{k+1}\left(1 - \frac{2}{\sqrt{r}+1}\right)^k.\end{aligned}$$

We have therefore constructed a kth-degree polynomial $P_k(x) \coloneqq R_k \circ g^{-1}$, which can be written as $P_k(x) = \sum_{j=0}^{k} c_j x^j$ for some coefficients c_j. Our proposed estimator is the expected value of this polynomial:

$$\widehat{\text{Nash}} \coloneqq \mathbb{E}_{\theta\sim\Theta}[P_k(u_\theta(\phi))] = \sum_{j=0}^{k} c_j \mathbb{E}_{\theta\sim\Theta}[(u_\theta(\phi))^j] = \sum_{j=0}^{k} c_j \cdot \langle M_j \cdot \phi^{\otimes j} \rangle.$$

Since $|P_{k+1}(u_\theta(\phi)) - \log(u_\theta(\phi))|$ is uniformly bounded for $u_\theta(\phi) \in [a, b]$ as in (6) by the arguments above, this yields the same bound on the Nash approximation. □

Maximizing Welfare over Candidate Sets. In the paradigm of pluralistic alignment, a natural objective is to identify a *set* of candidates that represents most "reasonable" viewpoints [Sorensen et al., 2024].[5] One way to operationalize this is to define the utility of each voter for a set of candidates as their *maximum* utility over the candidates in this set. In Section 2, we defined the associated welfare notion as *top-choice welfare*. Next, we show that we can obtain an approximate welfare maximizer in this setting by considering k moments of the voter distribution.

Theorem 5.4 (Approximability of ℓ-tcw maximization from moments)**.** *Fix $\varepsilon > 0$, ℓ, and a set of candidate responses Φ. Let $\mathcal{W}_\ell \coloneqq \binom{\Phi}{\ell}$ be the collection of all sets of ℓ candidate. It is possible to identify a $\widehat{W} \in \mathcal{W}_\ell$ for which*

$$\text{tcw}_\Theta(\widehat{W}) \geq \max_{W\in\mathcal{W}_\ell} \text{tcw}_\Theta(W) - \varepsilon$$

from only the moments $M_1, \ldots, M_k$ of Θ for $k = 2Bd/\varepsilon$, where B is an upper bound on all $u_\theta(\phi)$.

The proof relies on establishing that any two distributions sharing the first k moments are close in the 1-Wasserstein distance, which in turn bounds the estimation error for any Lipschitz-continuous welfare function.

6 Discussion

We study social choice problems when each voter provides only minimal preference feedback. In the linear social choice model—where candidates have vector embeddings and voter utilities are linear—such inference is possible. We analyze pairwise queries with and without intensity, formalizing the latter as a thresholded additive difference.

We find a sharp divide between one comparison per voter and slightly richer elicitation. A single pairwise comparison suffices to identify and efficiently estimate the first moment (the "average voter"), enabling approximately welfare-optimal selection. However, the second moment is not identifiable from one comparison, so objectives that depend on dispersion or inequality, such as risk-adjusted and Nash welfare, cannot generally be supported by asking only a single query per voter. On the other hand, with more comparisons per voter (or appropriately designed graded comparisons), higher-order structure becomes recoverable: with k comparisons we can identify the first k moments with polynomial sample complexity. Remarkably, two comparisons or one graded comparison per voter suffice to identify the full distribution, though with substantially weaker estimation efficiency. These moment-recovery results enable principled social choice methods beyond social welfare maximization. For single-winner selection, estimating the first two moments supports risk-adjusted and Nash welfare objectives, while for multi-winner selection, moments can be used to approximately optimize coverage-style committee objectives.

Several challenges remain. Our analysis relies on geometric expressivity assumptions, making learnability under constrained query spaces an important open problem. Moreover, existing approximation bounds suggest that large k may be required; developing objectives and algorithms that perform well with low-order moments and realistic sample sizes is key. Additionally, while using one size-1 graded query

[5] This goes by the name of *Overton* Pluralism.

per vote to identify the distribution has theoretical guarantees, how to properly elicit and use preference intensity queries in practice is nontrivial. Finally, extending moment-based approaches such as tensor decomposition [Anandkumar et al., 2014] and sum-of-squares methods [Laurent, 2008] for more complex tasks is an interesting direction.

More broadly, our work highlights a design principle for preference collection in social choice and beyond: small increases in per-user feedback richness can qualitatively expand what society-level properties are learnable, enabling alignment systems that capture not only average preferences but also disagreement and representation.

References

Georgios Amanatidis, Georgios Birmpas, Aris Filos-Ratsikas, and Alexandros A Voudouris. Peeking behind the ordinal curtain: Improving distortion via cardinal queries. *Artificial Intelligence*, 296: 103488, 2021.

Animashree Anandkumar, Rong Ge, Daniel J Hsu, Sham M Kakade, Matus Telgarsky, et al. Tensor decompositions for learning latent variable models. *J. Mach. Learn. Res.*, 15(1):2773–2832, 2014.

Anthony B Atkinson et al. On the measurement of inequality. *Journal of economic theory*, 2(3):244–263, 1970.

Kendall Atkinson and Weimin Han. *Spherical harmonics and approximations on the unit sphere: an introduction*. Springer Science & Business Media, 2012.

Sheldon Axler, Paul Bourdon, and Ramey Wade. *Harmonic function theory*, volume 137. Springer Science & Business Media, 2001.

Hossein Azari, David Parks, and Lirong Xia. Random utility theory for social choice. *Advances in Neural Information Processing Systems*, 25, 2012.

Hossein Azari Soufiani, William Chen, David C Parkes, and Lirong Xia. Generalized method-of-moments for rank aggregation. *Advances in Neural Information Processing Systems*, 26, 2013.

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.

Petros T Boufounos and Richard G Baraniuk. 1-bit compressive sensing. In *2008 42nd Annual Conference on Information Sciences and Systems*, pages 16–21. IEEE, 2008.

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

Chris Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Greg Hullender. Learning to rank using gradient descent. In *Proceedings of the 22nd international conference on Machine learning*, pages 89–96, 2005.

Souradip Chakraborty, Jiahao Qiu, Hui Yuan, Alec Koppel, Furong Huang, Dinesh Manocha, Amrit Bedi, and Mengdi Wang. Maxmin-rlhf: Towards equitable alignment of large language models with diverse human preferences. In *ICML 2024 Workshop on Models of Human Feedback for AI Alignment*, 2024.

Sitan Chen, Vasilis Kontonis, and Kulin Shah. Learning general gaussian mixtures with efficient score matching. In Nika Haghtalab and Ankur Moitra, editors, *The Thirty Eighth Annual Conference on Learning Theory, 30-4 July 2025, Lyon, France*, volume 291 of *Proceedings of Machine Learning Research*, pages 1029–1090. PMLR, 2025. URL `https://proceedings.mlr.press/v291/chen25e.html`.

Yeshwanth Cherapanamjeri, Constantinos Daskalakis, Gabriele Farina, and Sobhan Mohammadpour. Learning correlated reward models: Statistical barriers and opportunities. *arXiv preprint arXiv:2510.15839*, 2025.

Keertana Chidambaram, Karthik Vinary Seetharaman, and Vasilis Syrgkanis. Direct preference optimization with unobserved preference heterogeneity: The necessity of ternary preferences. *arXiv preprint arXiv:2510.15716*, 2025.

Paul Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

William W Cohen, Robert E Schapire, and Yoram Singer. Learning to order things. *Advances in neural information processing systems*, 10, 1997.

Harald Cramér and Herman Wold. Some theorems on distribution functions. *Journal of the London Mathematical Society*, 1(4):290–294, 1936.

Feng Dai and Yuan Xu. *Approximation theory and harmonic analysis on spheres and balls*. Springer, 2013.

Soroush Ebadian and Nisarg Shah. Every bit helps: Achieving the optimal distortion with a few queries. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 13788–13795, 2025.

Paul Funk. Beiträge zur theorie der kugelfunktionen. *Mathematische Annalen*, 77(1):136–152, 1915.

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

Luise Ge, Brendan Juba, and Yevgeniy Vorobeychik. Learning linear utility functions from pairwise comparison queries. *arXiv preprint arXiv:2405.02612*, 2024b. doi: 10.48550/arXiv.2405.02612.

Luise Ge, Gregory Kehne, and Yevgeniy Vorobeychik. Optimized distortion in linear social choice. In *AAAI Conference on Artificial Intelligence*, 2025.

Izrail Solomonovich Gradshteyn and Iosif Moiseevich Ryzhik. *Table of integrals, series, and products*. Academic press, 2014.

Helmut Groemer. *Geometric applications of Fourier series and spherical harmonics*, volume 61. Cambridge University Press, 1996.

Daniel Halpern, Gregory Kehne, Ariel D Procaccia, Jamie Tucker-Foltz, and Manuel Wüthrich. Representation with incomplete votes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 5657–5664, 2023.

Daniel Halpern, Safwan Hossain, and Jamie Tucker-Foltz. Computing voting rules with elicited incomplete votes. In *Proceedings of the 25th ACM Conference on Economics and Computation*, pages 941–963, 2024.

Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the econometric society*, pages 1029–1054, 1982.

Erich Hecke. Über orthogonal-invariante integralgleichungen. *Mathematische Annalen*, 78(1):398–404, 1917.

Christopher J Hillar and Lek-Heng Lim. Most tensor problems are np-hard. *Journal of the ACM (JACM)*, 60(6):1–39, 2013.

Jiaming Ji, Mickel Liu, Josef Dai, Xuehai Pan, Chi Zhang, Ce Bian, Boyuan Chen, Ruiyang Sun, Yizhou Wang, and Yaodong Yang. Beavertails: Towards improved safety alignment of llm via a human-preference dataset. *Advances in Neural Information Processing Systems*, 36:24678–24704, 2023.

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

Anson Kahng, Mohamad Latifian, and Nisarg Shah. Voting with preference intensities. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 5697–5704, 2023.

Kihyun Kim, Jiawei Zhang, Asuman Ozdaglar, and Pablo A Parrilo. Population-proportional preference learning from human feedback: An axiomatic approach. *arXiv preprint arXiv:2506.05619*, 2025.

Monique Laurent. Sums of squares, moment matrices and optimization over polynomials. In *Emerging applications of algebraic geometry*, pages 157–270. Springer, 2008.

Zhuohang Li, Xiaowei Li, Chengyu Huang, Guowang Li, Katayoon Goshvadi, Bo Dai, Dale Schuurmans, Paul Zhou, Hamid Palangi, Yiwen Song, and Bradley A. Malin. Judging with confidence: Calibrating autoraters to preference distributions. *arXiv preprint arXiv:2510.00263v1*, 2025. URL `https://arxiv.org/html/2510.00263v1`.

Qiu-Ming Luo and Feng Qi. Bounds for the ratio of two gamma functions—from wendel's and related inequalities to logarithmically completely monotonic functions. *Banach Journal of Mathematical Analysis*, 6(2):132–158, 2012.

John C Mason and David C Handscomb. *Chebyshev polynomials*. Chapman and Hall/CRC, 2002.

Colin McDiarmid et al. On the method of bounded differences. *Surveys in combinatorics*, 141(1):148–188, 1989.

Igor Melnyk, Youssef Mroueh, Brian Belgodere, Mattia Rigotti, Apoorva Nitsure, Mikhail Yurochkin, Kristjan Greenewald, Jiri Navratil, and Jarret Ross. Distributional preference alignment of llms via optimal transport. *Advances in Neural Information Processing Systems*, 37:104412–104442, 2024.

Donald J. Newman and Harold S. Shapiro. Jackson's theorem in higher dimensions. In *Proceedings of Conference in Oberwolfach*, volume 5 of *International Series of Numerical Mathematics*, pages 208–219, Basel-Stuttgart, 1964. Birkhäuser. ISNM 5.

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

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.

Sheng Ouyang, Yulan Hu, Ge Chen, Qingyang Li, Fuzheng Zhang, and Yong Liu. Towards reward fairness in rlhf: From a resource allocation perspective. *arXiv preprint*, 2025.

Kiho Park, Yo Joong Choe, and Victor Veitch. The linear representation hypothesis and the geometry of large language models. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

Karl Pearson. Method of moments and method of maximum likelihood. *Biometrika*, 28(1/2):34–59, 1936.

Yaniv Plan and Roman Vershynin. One-bit compressed sensing by linear programming. *Communications on pure and Applied Mathematics*, 66(8):1275–1297, 2013.

Konrad Schmüdgen. *The moment problem*, volume 9. Springer, 2017.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Ali Shirali, Arash Nasr-Esfahany, Abdullah Alomar, Parsa Mirtaheri, Rediet Abebe, and Ariel Procaccia. Direct alignment with heterogeneous preferences. *arXiv preprint arXiv:2502.16320*, 2025.

Anand Siththaranjan, Cassidy Laidlaw, and Dylan Hadfield-Menell. Distributional preference learning: Understanding and accounting for hidden context in rlhf. *Neurips*, 2024.

Taylor Sorensen, Jared Moore, Jillian Fisher, Mitchell Gordon, Niloofar Mireshghallah, Christopher Michael Rytting, Andre Ye, Liwei Jiang, Ximing Lu, Nouha Dziri, et al. A roadmap to pluralistic alignment. *arXiv preprint arXiv:2402.05070*, 2024.

Nisan Stiennon, Long Ouyang, Jeff Wu, Daniel M. Ziegler, Chelsea Voss Ryan Lowe, Alec Radford, Dario Amodei, and Paul Christiano. Learning to summarize with human feedback. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012. doi: 10.1007/s10208-011-9099-z. URL `https://doi.org/10.1007/s10208-011-9099-z`.

Miaoyan Wang, Khanh Dao Duc, Jonathan Fischer, and Yun S Song. Operator norm inequalities between tensor unfoldings on the partition lattice. *Linear algebra and its applications*, 520:44–66, 2017.

James G Wendel. Note on the gamma function. *The American Mathematical Monthly*, 55(9):563, 1948.

A Notation Reference

Symbol	Description
$\mathbb{S}^{d-1}$	The unit sphere in $\mathbb{R}^d$
$\bar{\sigma}$	Uniform probability measure (normalized surface area measure) on $\mathbb{S}^{d-1}$
Θ	The underlying population distribution of voter types over $\mathbb{S}^{d-1}$
$x \in \mathcal{X}, y \in \mathcal{Y}$	Prompts and Responses
$\Phi(x, y)$ or ϕ	Embedding vector of a candidate (or prompt-response pair)
$u_\theta(\phi)$	Utility of voter θ for candidate ϕ, given by $\theta \cdot \phi$
q	Pairwise comparison query vector (difference of embeddings $q \coloneqq \phi_1 - \phi_2$)
$\mathrm{resp}_\theta(q)$	Binary response of voter θ to query q (e.g., $\mathbb{1}\{\theta \cdot q \geq 0\}$)
$\mathrm{grad}_\theta(q)$	Binary response of voter θ to query q (e.g., $\mathbb{1}\{\theta \cdot q \geq \tau\}$)
$Q_k(\boldsymbol{q}; \Theta)$ or $Q_k(\boldsymbol{q})$	Probability that a random voter $\theta \sim \Theta$ responds positively to all queries $\boldsymbol{q}$
$G_\tau(\boldsymbol{q}; \Theta)$ or $G_\tau(\boldsymbol{q})$	Probability that a random voter $\theta \sim \Theta$ has strong preference to $\boldsymbol{q}$
c_d	The constant from the first moment identity (see Equation (1))
Z_k	The set of $\tau \in (0, 1)$ for which G_τ fails to identify M_k

Table 1: Notation used throughout this work.

B Spherical Harmonics Basics

A spherical harmonic of degree j is the restriction to $\mathbb{S}^{d-1}$ of a degree-j homogeneous polynomial whose Laplacian $\Delta \coloneqq \sum_{i=1}^{d} \partial_i^2$ vanishes. Spherical harmonics arise naturally in our analysis because the distributions we study are supported on the sphere, and moment functionals correspond to homogeneous polynomials. There are many excellent references on spherical harmonics, including the classical text [Groemer, 1996] and the more recent [Dai and Xu, 2013]; we refer interested readers to these sources for further background.

Spherical harmonics enjoy many useful properties, including orthogonality and an L^2-decomposition into harmonic subspaces. The main tool deferred from the main text is the classic Funk–Hecke formula, which dates back to the work of Funk [1915] and Hecke [1917].

Theorem B.1 (Funk-Hecke Theorem)**.** *Suppose $d \geq 2$ and let $Y \in \mathcal{H}_j$, $x \in \mathbb{S}^{d-1}$, K is bounded. Then, for all $y \in \mathbb{S}^{d-1}$,*

$$\int_{S^{d-1}} K(x \cdot y) Y(x)\, d\sigma(x) = \mu_j \cdot Y(y)$$

where $\mu_j = \frac{\Gamma(\frac{d}{2})}{\sqrt{\pi}\Gamma(\frac{d-1}{2})} \int_{-1}^{1} P_j(t) K(t) (1 - t^2)^{\frac{d-3}{2}} dt$, where P_j being the j-th Gegenbauer polynomial with parameter $\frac{d-2}{2}$.

Gegenbauer polynomials form a classical family of orthogonal polynomials on $[-1, 1]$ and are closely connected to spherical harmonics. In the main text, we use the basic parity property: the polynomial is even when j is even and odd when j is odd. Further background and properties can be found in Appendix B of [Dai and Xu, 2013].

C Omitted Results and Proofs

C.1 Proof of Observation 3.5

Proof. Having the same k-th moment tensor is the same as having same expected values for each monomials of degree k, and if the two distribution have the same first k moments, for any polynomial with

degree less or equal to k, their expected values over the distributions will also be the same by the linearity of the expected values.

Let σ denote the uniform probability measure on $\mathbb{S}^{d-1}$. By [Axler et al., 2001][Proposition 5.9] for any polynomial p and any homogeneous harmonic polynomial q on $\mathbb{R}^d$ satisfying $\deg(q) > \deg(p)$,

$$\int_{\mathbb{S}^{d-1}} p(x)\, q(x)\, d\sigma(x) = 0. \tag{1}$$

Choose any nonzero homogeneous harmonic polynomial of degree $k+1$; for instance

$$h(x) = \Re\big((x_1 + ix_2)^{k+1}\big),$$

where x_1, x_2 are the first and second coordinates of the vector x and $\Re$ represents the real part of a complex number.

By the mean value property of a harmonic function, $\int h(x)d\sigma(x) = h(0) = 0$. Thus we are able to define

$$d\mu_{\pm}(x) = (1 \pm \varepsilon h(x))\, d\sigma(x)$$

for some sufficiently $\varepsilon > 0$ such that

$$1 \pm \varepsilon h(x) \geq 0 \qquad \text{for all } x \in \mathbb{S}^{d-1}.$$

Let x^α be a monomial with $|\alpha| \leq k$. Applying Proposition 5.9 with $p(x) = x^\alpha$ and $q(x) = h(x)$ gives

$$\int_{S^{d-1}} x^\alpha(x)\, h(x)\, d\sigma(x) = 0,$$

since $\deg(h) = k+1 > |\alpha| = \deg(p)$. Hence

$$\int x^\alpha\, d\mu_+ - \int x^\alpha\, d\mu_- = 2\varepsilon \int_{S^{d-1}} x^\alpha(x)\, h(x)\, d\sigma(x) = 0,$$

so all moments up to order k agree.

Because h is not identically zero,

$$\int_{S^{d-1}} h(x)^2\, d\sigma(x) > 0.$$

Thus

$$\int h\, d\mu_+ - \int h\, d\mu_- = 2\varepsilon \int_{S^{d-1}} h(x)^2\, d\sigma(x) \neq 0.$$

Since h is a homogeneous polynomial of degree $k+1$, this implies that the $(k+1)$-st moments of μ_+ and μ_- differ. □

C.2 Proof of Lemma 3.7

Sketch. Proofs of this sort proceed generally in three steps. The first is to argue that the moment tensor M_k for a distribution μ suffices to compute the expectation $\mathbb{E}_\mu[p(x)]$ of any homogeneous degree-k polynomial p on $\mathbb{S}^{d-1}$.

The second step is invoke the Stone-Weierstrass theorem for $\mathbb{S}^{d-1}$, which says that the set of polynomials is dense in the larger space of continuous functions $C(\mathbb{S}^{d-1})$. Here density means uniform convergence; for all continuous $f \in C(\mathbb{S}^{d-1})$ there is a sequence of polynomials which converges to it in $\|f - p\|_\infty = \max_{x \in \mathbb{S}^{d-1}} |f(x) - p(x)|$. This is quite strong, and it is possible because the domain is compact.

The last step is to invoke the Riesz-Markov-Kakutani representation theorem, which states that any two measures that agree on all continuous functions $\mathbb{E}[f(x)]$ are in fact the same. Since the polynomials are dense in $C(\mathbb{S}^{d-1})$, if μ and μ' agree on their moments (and therefore on $\mathbb{E}[p(x)]$ for all polynomials), then $\mathbb{E}_\mu[f(x)] = \mathbb{E}_{\mu'}[f(x)]$ for all continuous f and so this representation theorem applies. □

C.3 Proof of Theorem 3.9

To prove Theorem 3.9, we need the following lemma. For a fixed voter direction $\theta \in \mathbb{S}^{d-1}$, consider its contribution

$$I_{k,\tau}(\theta) = \int_{\mathbb{S}^{d-1}} \mathbb{1}\{\theta^\top q \geq \tau\}\, q^{\otimes k}\, d\bar{\sigma}(q).$$

Lemma C.1. *Let $d \geq 2$, $k \geq 1$, for any $\theta \in \mathbb{S}^{d-1}$, it holds that*

$$I_{k,\tau}(\theta) = \sum_{j=0}^{\lfloor k/2 \rfloor} \lambda_{k,j}\, \mathrm{Sym}\big(\theta^{\otimes(k-2j)} \otimes I^{\otimes j}\big),$$

where the constants $\lambda_{k,0}, \ldots, \lambda_{k,\lfloor k/2 \rfloor}$ depend on τ, d, k, I is the rank-2 identity tensor (i.e. matrix) on $\mathbb{R}^d$, and Sym *denotes full symmetrization over all indices.*

Proof. Let $x \in \mathbb{R}^d$ be an arbitrary vector. We define the scalar polynomial $p(x)$ by contracting $I_{k,\tau}(\theta)$ with k copies of x:

$$p(x) \coloneqq \langle I_k(\theta), x^{\otimes k} \rangle = \left\langle \int_{\mathbb{S}^{d-1}} \mathbb{1}\{\theta^\top q \geq \tau\}\, q^{\otimes k}\, d\sigma(q),\ x^{\otimes k} \right\rangle.$$

As $\langle q^{\otimes k}, x^{\otimes k} \rangle = (\langle q, x \rangle)^k$, we have:

$$p(x) = \int_{\mathbb{S}^{d-1}} \mathbb{1}\{\theta^\top q \geq \tau\}\ (q^\top x)^k\, d\sigma(q).$$

Let $O(d)$ denote the orthogonal group acting on $\mathbb{R}^d$. Consider the stabilizer subgroup of θ, denoted $\mathcal{R}_\theta$. For $R \in \mathcal{R}_\theta$:

$$p(Rx) = \int_{\mathbb{S}^{d-1}} \mathbb{1}\{\theta^\top q \geq \tau\}\ (q^\top Rx)^k\, d\sigma(q).$$

Now let $u = R^\top q$. Since R is orthogonal, the measure is invariant, so $d\sigma(q) = d\sigma(u)$. Furthermore, $\theta^\top q = \theta^\top (Ru) = (R^\top \theta)^\top u = \theta^\top u$. Substituting these back into the integral:

$$p(Rx) = \int_{\mathbb{S}^{d-1}} \mathbb{1}\{\theta^\top u \geq \tau\}\ (u^\top Rx)^k\, d\sigma(u) = p(x).$$

Thus, $P(x)$ is a polynomial in x that is invariant under all rotations about the axis θ.

Therefore, $p(x)$ must be of the form: $p(x) = F(\langle x, \theta \rangle, |x|^2)$. Since $p(x)$ is defined by an integral of $(q^\top x)^k$, $p(x)$ is a homogeneous polynomial of degree k in x. Consequently, F must be a linear combination of terms of the form $(\langle x, \theta \rangle)^a (|x|^2)^b$ such that the total degree matches k: $a + 2b = k$. Since a, b must be non-negative integers, let $b = j$. Then $a = k - 2j$. The possible values for j are integers satisfying $0 \leq 2j \leq k$, i.e., $0 \leq j \leq \lfloor k/2 \rfloor$. Thus, there exist scalar constants $\lambda_{k,j}$ such that:

$$p(x) = \sum_{j=0}^{\lfloor k/2 \rfloor} \lambda_{k,j}\, (\langle x, \theta \rangle)^{k-2j} (|x|^2)^j. \tag{7}$$

We now map the scalar terms back to their tensor equivalents.

- The term $(\langle x, \theta \rangle)^{k-2j}$ corresponds to the contraction of the rank-$(k-2j)$ tensor $\theta^{\otimes(k-2j)}$ with $x^{\otimes(k-2j)}$.
- The term $|x|^{2j} = (\langle x, Ix \rangle)^j$ corresponds to the contraction of j copies of the identity matrix, $I^{\otimes j}$, with $x^{\otimes 2j}$.

The product corresponds to the tensor product $\theta^{\otimes(k-2j)} \otimes I^{\otimes j}$ contracted with $x^{\otimes k}$. However, the original tensor $I_{k,\tau}(\theta)$ is fully symmetric, while the tensor product $\theta^{\otimes(k-2j)} \otimes I^{\otimes j}$ is not. Since the equality holds for all x, the symmetric tensors associated with the polynomials must be equal. Finally, we apply the symmetrization operator Sym to the basis tensors: $I_{k,\tau}(\theta) = \sum_{j=0}^{\lfloor k/2 \rfloor} \lambda_{k,j} \operatorname{Sym}\left(\theta^{\otimes(k-2j)} \otimes I^{\otimes j}\right)$. □

Proof of Theorem 3.9. We first show that all moments are identifiable through induction on k. Let $k \geq 1$ be fixed and assume that all moments $\mathbb{E}_{\theta\sim\Theta}[\theta^{\otimes \ell}]$ for $\ell < k$ are known. Now by Lemma C.1 and Fubini's swap,

$$\begin{aligned}
T_{k,\tau} &\coloneqq \int_{\mathbb{S}^{d-1}} G_\tau(q)\, q^{\otimes k}\, d\bar{\sigma}(q) \\
&= \int_{\mathbb{S}^{d-1}} \left(\int_{\mathbb{S}^{d-1}} \mathbf{1}(\theta^\top q \geq \tau)\, q^{\otimes k}\, d\bar{\sigma}(q) \right) d\Theta(\theta) \\
&= \mathbb{E}_{\theta\sim\Theta}[I_{k,\tau}(\theta)] \\
&= \lambda_{k,0}(\tau)\, \mathbb{E}[\theta^{\otimes k}] + \sum_{j=1}^{\lfloor k/2 \rfloor} \lambda_{k,j}(\tau) \operatorname{Sym}\Big(\mathbb{E}_{\theta\sim\Theta}[\theta^{\otimes(k-2j)}] \otimes I^{\otimes j} \Big),
\end{aligned}$$

The summation term involves only moments of order $k-2, k-4, \ldots$, which are known by the inductive hypothesis. Therefore, the kth moment tensor $\mathbb{E}[\theta^{\otimes k}]$ is uniquely determined if and only if the coefficient $\lambda_{k,0}(\tau)$ is non-zero.

To rigorously determine $\lambda_{k,0}(\tau)$, we utilize spherical harmonics again. First, for a fixed x, the homogeneous polynomial $f(q) = (q^T x)^k$ can be decomposed as a sum of spherical harmonics of degree $k, k-2, \ldots$: $f(q) = \sum_{j=0}^{\lfloor k/2 \rfloor} |x|^{2j} Y_{k-2j}(q)$. Now applying the linearity of the integral and the Funk-Hecke Theorem (see Theorem B.1 in Appendix C), we have that

$$p(x) = \sum_{j=0}^{\lfloor k/2 \rfloor} \mu_{k-2j}\, |x|^{2j} Y_{k-2j}(\theta).$$

Comparing this with Equation (7), we see that $\mu_k = \lambda_{k,0}$. Thus

$$\lambda_{k,0}(\tau) = \frac{\Gamma(\frac{d}{2})}{\sqrt{\pi}\Gamma(\frac{d-1}{2})} \int_\tau^1 P_k(t)(1-t^2)^{\frac{d-3}{2}} dt, \tag{8}$$

where P_k is the Gegenbauer polynomial with parameter $\frac{d-2}{2}$. Note that the derivative of $\lambda_{k,0}(\tau)$ (specifically, $P_k(t)$) has k distinct roots on $(-1,1)$, and $\lfloor k/2 \rfloor$ on $(0,1)$ by symmetry. So $\lambda_{k,0}(\tau)$ has $\lfloor k/2 \rfloor$ stationary points. Now consider the boundary points for $\lambda_{k,0}(\tau)$. Clearly, $\lambda_{k,0}(1) = 0$ .When $\tau = 0$ and k is even, both $P_{k,d}(t)$ and $(1-t^2)^{(d-3)/2}$ are even and due to the orthogonality of Gegenbauer polynomials with parameter $\frac{d-2}{2}$, $\int_0^1 P_k(t)(1-t^2)^{(d-3)/2} = \frac{1}{2}\int_{-1}^1 P_k(t) \cdot 1 \cdot (1-t^2)^{(d-3)/2} = 0$, so $\lambda_{k,0}(\tau) = 0$ and the function has one less interior zero. Hence the number of roots of the function $\lambda_{k,0}(\tau)$ is $\lfloor \frac{k-1}{2} \rfloor$. And $Z_k \coloneqq \{\tau \in (0,1) : \lambda_{k,0}(\tau) = 0\}$ must be discrete. Thus, for any $\tau \notin Z_k$, we can solve for the kth moment:

$$\mathbb{E}[\theta^{\otimes k}] = \frac{1}{\lambda_{k,0}(\tau)} \left(T_{k,\tau} - \sum_{j=1}^{\lfloor k/2 \rfloor} \lambda_{k,j}(\tau) \operatorname{Sym}\big(\mathbb{E}[\theta^{\otimes(k-2j)}] \otimes I^{\otimes j} \big) \right)$$

Now since the set $\bigcup_{k\geq 1} Z_k$ is countable, and has measure zero on $(0,1)$, for almost every $\tau \in (0,1)$, all moments are identifiable with G_τ. Lemma 3.7 then implies that distributions are identifiable. □

C.4 Proof of Theorem 4.6

Proof. Fix odd ℓ and let $i_1 \ldots, i_\ell$ be a sequence of indices. Let $p(\theta) = \theta_{i_1} \cdot \cdots \cdot \theta_{i_\ell}$. Recall, from the proof of Theorem 3.8, there are spherical harmonics f_j for odd $j \le \ell$ such that $p(\theta) = \sum_{j \le \ell : j \text{ is odd}} f_j(\theta)$. However, we did not give an explicit construction of them. It turns out, that finding them is possible. The primary property we need is that in our case, $|f_j(\theta)| \le d^{j/2}$ for all $\theta \in \mathbb{S}^{d-1}$. We encapsulate this into the following lemma, whose proof we differ until later:

Lemma C.2. *Let $p(x) = x_{i_1} x_{i_2} \ldots x_{i_\ell}$ be a monomial of degree ℓ in $\mathbb{R}^d$. Let $f_j \in \mathbb{Y}_j^d$ denote the j-th spherical harmonic component of the restriction of $p(x)$ to the unit sphere $\mathbb{S}^{d-1}$. Then,*

$$|f_j(x)| \le d^{j/2}$$

for all $x \in \mathbb{S}^{d-1}$.

Proof of Lemma C.2. The key tool we will need is the *projection operator* as a way to determine the j'th spherical harmonic in the decomposition. Atkinson and Han [2012, Definition 2.11] call this function $\mathcal{P}_{j,d} f$).

The only property we will need is Equation (2.49) which shows that

$$\|\mathcal{P}_{j,d} f\|_{C(\mathbb{S}^{d-1})} \le N_{j,d}^{1/2} \|f\|_{C(\mathbb{S}^{d-1})}.$$

where, by Equation (2.10), $N_{j,d} = \binom{j+d-1}{j} - \binom{j+d-3}{j-2}$ and, $\|g\|_{C(\mathbb{S}^{d-1})} = \sup\{|g(\xi)| : \xi \in \mathbb{S}^{d-1}\}$ (see Section 1.3). Note that trivially

$$N_{j,d} \le \binom{j+d-1}{j} = \frac{j+d-1}{j} \cdot \cdots \cdot \frac{d+1-1}{1} \le d^j.$$

Plugging in our monomial, note that on $\mathbb{S}^{d-1}$, each coordinate $|\theta_i| \le 1$, therefore, $|p(\theta)| \le 1$. Combining these yields the lemma statement. □

Next, we also need an explicit construction of λ_j. From Groemer [1996, Lemma 3.4.6], using our normalized measure $d\bar{\sigma}$, this is given by

$$\lambda_j = (-1)^{(d-1)/2} \frac{1}{d} \cdot \frac{1 \cdot 3 \cdot \cdots \cdot (j-2)}{(d+1)(d+3)\cdots(d+j-2)}.$$

For our purposes, we will lower bound $|\lambda_j| \ge (d+1)^{-(j+1)/2}$.

Combining these, we see that

$$\left| \sum_{j \le \ell : j \text{ is odd}} \frac{1}{\lambda_j} f_j(\theta) \right| \le \sum_{j \le \ell : j \text{ is odd}} (d+1)^{j+1/2} \le 2(d+1)^{\ell+1/2}. \tag{9}$$

Now we will apply this to our concentration inequalities.

We start with odd k. We will work with the matricized form where the entire $[k]$ is on one side, so this is essentially a $d^k \times 1$ sized vector. Define $\Psi^k(q)$ as the matricized vector where for a each index sequence $\alpha = i_1 \ldots i_k$, $\Psi^k(q,b)_\alpha = b \cdot \sum_{j \le k : j \text{ is odd}} \frac{1}{\lambda_j} f_j^\alpha(q)$ where f_j^α is the spherical harmonic of degree j in the decomposition of the monomial corresponding to α. As we showed, $\mathbb{E}_{q \sim \bar{\sigma}, \theta \sim \Theta} \left[\Psi^k(q, \text{resp}_\theta(q))_\alpha\right]$ is the α entry of M_k. Thus, we will set our estimator $\widehat{M_k}$ to be the unmatricized version of $\frac{1}{T} \sum_{i=1}^T \Psi^k(q^i, \text{resp}_{\theta_i}(q^i))$ where q^i is the i'th sampled query and $\text{resp}_{\theta_i}(q^i)$ is the response of the sampled voter θ_i.

We would like to apply Matrix Bernstein to this. By Inequality (9), each entry of $\Psi^k(q_i, \text{resp}_{\theta_i}(q_i))$ is bounded by $2(d+1)^{k+1/2}$. The expected value must also lie in $[-2(d+1)^{k+1/2}, 2(d+1)^{k+1/2}]$, and thus the

maximum distance from the expected value is at most $4(d+1)^{k+1/2}$. As there are d^k entries, the L_2 norm (equivalent to the operator norm for vectors) is upper bounded by $d^{k/2} \cdot 4(d+1)^{k+1/2} \leq 4(d+1)^{(3k+1)/2}$. As for the variance bound, note that for a vector v, both $\|v^\top v\|$ and $\|vv^\top\|$ are bounded by $\|v\|_2^2$. Thus, we can upperbound this by $16(d+1)^{3k+1}$. Plugging this into matrix Bernstein, we get

$$T \geq \frac{2}{\varepsilon^2}\left(16(d+1)^{3k+1} + \frac{4(d+1)^{(3k+1)/2}\varepsilon}{3}\right)\log\left(\frac{d^k+1}{\delta}\right) = O\left(\frac{(d+1)^{3k+1}}{\varepsilon^2}\log\left(\frac{d^k}{\delta}\right)\right).$$

Next, consider even k. Let $s = k/2 - 1$. We will matricize such that $|I| = s$ and $|J| = k - s$. In particular, we will set our estimator $\widehat{M}_k$ to be the unmatricized version of

$$\frac{1}{T}\sum_{i=1}^{T} \Psi^s(q_1^i, \text{resp}_{\theta_i}(q_1^i))^\top \Psi^{k-s}(q_2^i, \text{resp}_{\theta_i}(q_2^i).$$

Let α be a sequence of indices and let α^s and α^{k-s} be its first s and last $k-s$ indices collectively. The (α^s, α^{k-s}) entry of this outer product (which corresponds to α) will have expectation (by linearity) exactly the moment corresponding to α.

We now apply matrix Berstein on these. Note that the individual Φ^s and Φ^{k-s} have terms bounded by $(d+1)^{s+1/2}$ and $(d+1)^{(k-s)+1/2}$ respectively. Thus, the distance each term is from the expectation is at most $2(d+1)^{s+1/2}$ and $2(d+1)^{(k-s)+1/2}$, respectively. Hence, their L_2 norms are at most $2(d+1)^{2s+1/2}$ and $2(d+1)^{2(k-s)+1/2}$, respectively. Therefore, the spectral norm of their outer product is at most $2(d+1)^{2k+1}$.

Next, we consider the variance. In general, we have two vectors u and v of dimensions n and n' and entries bounded by L and L', respectively, then $\|(u^\top v)^\top (u^\top v)\|$ and $\|(u^\top v)(u^\top v)^\top\|$ are bounded by $n \cdot n' \cdot L^2 \cdot (L')^2$. Plugging this in for use, we have a bound of $d^s \cdot d^{k-s} \cdot (2(d+1)^{s+1/2})^2 (2(d+1)^{k-s+1/2})^2 \leq 16(d+1)^{3k+2}$. Plugging this into matrix Berstein, we get

$$T \geq \frac{2}{\varepsilon^2}\left(16(d+1)^{3k+2} + \frac{4(d+1)^{(3k+2)/2}\varepsilon}{3}\right)\log\left(\frac{d^{k/2+1}+d^{k/2-1}}{\delta}\right) = O\left(\frac{(d+1)^{3k+2}}{\varepsilon^2}\log\left(\frac{d^k}{\delta}\right)\right). \quad \square$$

C.5 Proof of Proposition 5.1

Proof. By Theorem 4.2, T samples suffice to estimate the first moment M_1 such that $\|\hat{M}_1 - M_1\|_{op} \leq \varepsilon$. The estimated welfare for any candidate ϕ is $\langle \hat{M}_1, \phi\rangle$. The estimation error is $|\langle \hat{M}_1 - M_1, \phi\rangle| \leq \|\hat{M}_1 - M_1\|_{op}\|\phi\| \leq \varepsilon B$. If we select the candidate maximizing the estimated welfare, the true welfare of the selected candidate is at most $2\varepsilon B$ suboptimal. $\square$

C.6 Proof of Proposition 5.2

Proof. It is without loss of generality to assume $\varepsilon \leq 1$, as the 0-tensor is a trivial 1-approximation since $\|M_k\| \leq 1$. Using Theorem 4.3 with $k = 2$, we obtain estimates $\hat{M}_1$ and $\hat{M}_2$ such that $\|\hat{M}_1 - M_1\| \leq \varepsilon$ and $\|\hat{M}_2 - M_2\| \leq \varepsilon$. Recall that $\text{raw}_\alpha(\phi) = M_1^\top\phi - \alpha\sqrt{\phi^\top M_2\phi - (M_1^\top\phi)^2}$. Let $\mu = M_1^\top\phi$ and $\hat{\mu} = \hat{M}_1^\top\phi$. Similarly, let $\sigma^2 = \phi^\top M_2\phi - \mu^2$ and $\hat{\sigma}^2 = \phi^\top\hat{M}_2\phi - \hat{\mu}^2$. We have $|\hat{\mu} - \mu| \leq \varepsilon B$ and $|\phi^\top(\hat{M}_2 - M_2)\phi| \leq \varepsilon B^2$. The error in the variance term is bounded by:

$$|\hat{\sigma}^2 - \sigma^2| \leq \varepsilon B^2 + |\hat{\mu}^2 - \mu^2| \leq \varepsilon B^2 + 2B(\varepsilon B) \leq 3\varepsilon B^2.$$

Using the fact that $|\sqrt{x} - \sqrt{y}| \leq \sqrt{|x-y|}$, the error in the standard deviation is bounded by $\sqrt{3\varepsilon}B$. Thus, the total estimation error is bounded by $\varepsilon B + \alpha\sqrt{3\varepsilon}B \leq \sqrt{3}\cdot(\alpha+1)B\sqrt{\varepsilon}$ because $\varepsilon \leq 1$. If we select the candidate maximizing estimated risk-adjusted welfare, the true risk-adjusted welfare of the selected candidate is at most double. $\square$

C.7 Missing Portion of Proof of Theorem 5.3

Here we derive (5). Recall that we will make use of the identity [Gradshteyn and Ryzhik, 2014, Equation 1.514]

$$\log(1 - 2\alpha\cos(\varphi) + \alpha^2) = -2\sum_{j=1}^{\infty} \frac{\cos(j\varphi)}{j} \cdot \alpha^j,$$

for $\alpha^2 \leq 1$ and $\alpha\cos(\varphi) \neq 1$. Our goal is to show that

$$\log(1 + \delta t) = -2\sum_{j=1}^{\infty} \frac{\alpha^j}{j} \cdot T_j(t),$$

where $\delta = \frac{r-1}{r+1}$.

As long as $|\alpha| < 1$, for real φ, $\alpha\cos(\varphi) \neq 1$. By our definition of Chebyshev polynomials, this implies that for $t \in [-1, 1]$,

$$\log(1 - 2\alpha t + \alpha^2) = -2\sum_{j=1}^{\infty} \frac{\alpha^j}{j} \cdot T_j(t).$$

To get it in our form, we can normalize by $1 + \alpha^2$ to get

$$\log\left(1 + \frac{-2\alpha}{1+\alpha^2} t\right) = -\log(1 + \alpha^2) - 2\sum_{j=1}^{\infty} \frac{\alpha^j}{j} \cdot T_j(t).$$

We will set $\alpha = -\frac{\sqrt{r}-1}{\sqrt{r}+1}$, which clearly has $|\alpha| < 1$ and yields

$$\frac{-2\alpha}{1+\alpha^2} = \frac{2 \cdot \frac{\sqrt{r}-1}{\sqrt{r}+1}}{1 + \left(\frac{\sqrt{r}-1}{\sqrt{r}+1}\right)^2} = \frac{2 \cdot \frac{\sqrt{r}-1}{\sqrt{r}+1}}{\frac{2(r+1)}{(\sqrt{r}+1)^2}} = \frac{(\sqrt{r}-1)(\sqrt{r}+1)}{r+1} = \frac{r-1}{r+1} = \delta.$$

First, observe that $\alpha^2 < 1$. This also implies that $\alpha\cos(\varphi) \neq 1$ for real φ. Thus, for $t \in [-1, 1]$,

$$\log(1 + \delta) = \log(1 - 2\alpha x + \alpha^2) = -2\sum_{j=1}^{\infty} \frac{\alpha^j}{j} \cdot T_j(x). \quad \square$$

C.8 Proof of Theorem 5.4

We now prove a key general result which shows that any (welfare) function that is Lipschitz continuous can be approximated using the first k moments. This result relies on the following lemma, which can be understood as a counterpart to Observation 3.5. It quantifies the extent to which two measures that agree on their first k moments can substantively differ. Here the difference between distributions μ and ν is measured according to the *1-Wasserstein distance*, or earth-mover's distance, given by

$$W_1(\mu, \nu) \coloneqq \inf_{\gamma \in \Gamma} \iint_{\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}} \left\|\theta - \theta'\right\|_2 \, d\mu(\theta)\, d\nu(\theta'),$$

where $\Gamma = \Gamma(\mu, \nu)$ is the set of all statistical couplings of μ and ν.

Lemma C.3 (Wasserstein Bound via Moment Matching)**.** *Let μ and ν be two probability measures on $\mathbb{S}^{d-1} \subset \mathbb{R}^d$. Suppose the first k moments of μ and ν are equal. Then the 1-Wasserstein distance satisfies*

$$W_1(\mu, \nu) \leq C \cdot \frac{d}{k},$$

where $C > 0$ is an absolute constant independent of the dimension d, the degree k, and the measures.

Proof. We will use the fact that the 1-Wasserstein distance admits the following alternative definition via Kantorovich-Rubinstein duality:

$$W_1(\mu,\nu) = \sup_{f\in \mathrm{Lip}(\mathbb{S}^{d-1})} \left| \int_{\mathbb{S}^{d-1}} f\, d\mu - \int_{\mathbb{S}^{d-1}} f\, d\nu \right|, \tag{10}$$

where $\mathrm{Lip}(\mathbb{S}^{d-1})$ is the set of functions on $\mathbb{S}^{d-1}$ that are 1-Lipschitz with respect to L_2. For the purposes of this proof it will be more convenient to work with the *geodesic distance*, given by $d_g(\theta,\theta') = \arccos(\langle\theta,\theta'\rangle)$ between $\theta,\theta' \in \mathbb{S}^{d-1}$. They are equivalent for our purposes because $\|\theta,\theta'\|_2 \le d_g(\theta,\theta') \le \frac{\pi}{2}\|\theta,\theta'\|_2$ on the unit sphere.

To begin, fix $f \in \mathrm{Lip}(\mathbb{S}^{d-1})$. Our goal is to approximate f using a spherical polynomial $P_k \in \Pi_k^d$, where Π_k^d denotes the space of all spherical polynomials of degree at most k. By the triangle inequality,

$$\begin{aligned} |\mathbb{E}_\mu[f] - \mathbb{E}_\nu[f]| &= |\mathbb{E}_\mu[f - P_k + P_k] - \mathbb{E}_\nu[f - P_k + P_k]| \\ &\le |\mathbb{E}_\mu[f - P_k]| + |\mathbb{E}_\nu[P_k] - \mathbb{E}_\mu[P_k]| + |\mathbb{E}_\nu[f - P_k]| \\ &\le |\mathbb{E}_\mu[f - P_k]| + |\mathbb{E}_\nu[f - P_k]|. \end{aligned} \tag{11}$$

This last step follows since $\mathbb{E}[P_k]$ can be written as a function of the at-most-k'th moments, which are equal for μ and ν by assumption.

We will bound the remaining two terms by the uniform norm over $\mathbb{S}^{d-1}$. To this end, let $E_k(f) \coloneqq \min_{P\in\Pi_k^d} \|f - P\|_\infty$ be the best possible uniform approximation to f achievable using degree-k polynomials. Choosing P_k to be the polynomial approximation of f witnessing this value, we have

$$|\mathbb{E}_\mu[f - P_k]| \le \int_{\mathbb{S}^{d-1}} \|f - P_k\|_\infty \, d\mu = E_k(f), \tag{12}$$

and likewise for ν. Therefore from (11) we have

$$|\mathbb{E}_\mu[f] - \mathbb{E}_\nu[f]| \le 2 \cdot E_k(f).$$

To bound $E_k(f)$ we invoke the so-called Jackson theorem for the sphere, which relates the approximability of a function by polynomials to its smoothness. As established by Newman and Shapiro [1964] (see also [Chen et al., 2025, Lemma 41]), for any continuous function f, the approximation error is bounded by the *modulus of continuity*, which on $\mathbb{S}^{d-1}$ is given by $\omega(f,\delta)_\infty \coloneqq \sup_{x,y:\, d_g(x,y)\le\delta} |f(x) - f(y)|$. In particular, it holds that

$$E_k(f) \le C_{NS} \cdot \omega\left(f, \frac{d}{k}\right)_\infty \tag{13}$$

for some absolute constant C_{NS} independent of the dimension d. Since f is 1-Lipschitz, its modulus of continuity satisfies $\omega(f,t)_\infty \le t$. Applying this to (13) and substituting it into (12) yields

$$|\mathbb{E}_\mu[f - P_k]| \le 2 \cdot C_{NS} \cdot \frac{d}{k}.$$

Taking the supremum over all 1-Lipschitz f, from (10) we finally have

$$W_1(\mu,\nu) \le 2 \cdot C_{NS} \cdot \frac{d}{k}. \qquad \square$$

Now, we are ready to present the general result. We do not claim that the resulting dependence of the required number of moments on ε is optimal, opting instead for a readily generalizable approach.

Lemma C.4. *Let μ and ν be probability measures on the unit sphere $\mathbb{S}^{d-1} \subset \mathbb{R}^d$ such that their first k moments match. For any L-Lipschitz function $f : \mathbb{S}^{d-1} \to \mathbb{R}$, the integration error is bounded by*

$$\left| \int_{\mathbb{S}^{d-1}} f\, d\mu - \int_{\mathbb{S}^{d-1}} f\, d\nu \right| \le C \cdot L \cdot \frac{d}{k},$$

where $C > 0$ is an absolute constant independent of the dimension d.

Proof. This follows directly from Lemma C.3. Given some L-Lipschitz f, observe that $g \coloneqq f/L$ is 1-Lipschitz. By the dual description of $W_1(\mu,\nu)$ (10), we therefore have

$$|\mathbb{E}_\mu[g] - \mathbb{E}_\nu[g]| \le W_1(\mu,\nu) \le 2C \cdot \frac{d}{k}.$$

Multiplying both sides by L, by linearity of expectation we obtain

$$|\mathbb{E}_\mu[f] - \mathbb{E}_\nu[f]| \le 2C \cdot L \cdot \frac{d}{k}. \qquad \square$$

We now use these lemmas to prove our main result on top-choice welfare.

Proof of Theorem 5.4. This proof proceeds from Lemma C.4. We will apply it to the functions $\mathrm{tc}_W(\theta) \coloneqq \max_{\phi \in W} u_\theta(\phi)$. Observe that in our model all such $\mathrm{tc}_W(\theta)$ are B-Lipschitz in θ, where B is an upper bound on $\|\phi\|$ for all $\phi \in W$.

Given the moments $M_1, \ldots, M_k$ of Θ, construct $\widehat{\Theta}$ to be an arbitrary distribution over $\mathbb{S}^{d-1}$ consistent with these moments and choose $\widehat{W} \coloneqq \arg\max_{W \in \mathcal{W}_\ell} \mathrm{tcw}_{\widehat{\Theta}}(W)$. Let W^* be the (unknown) ℓ-tcw optimum for Θ.

Then letting $\varepsilon' = C \cdot B \cdot \frac{d}{k}$, by the optimality of $\widehat{\Theta}$ and applying Lemma C.4 twice we have

$$\begin{aligned}
\mathrm{tcw}_\Theta(\widehat{W}) &\ge \mathrm{tcw}_{\widehat{\Theta}}(\widehat{W}) - \varepsilon' \\
&\ge \mathrm{tcw}_{\widehat{\Theta}}(W^*) - \varepsilon' \\
&\ge \mathrm{tcw}_\Theta(W^*) - 2\varepsilon'
\end{aligned}$$

Setting $\varepsilon = 2\varepsilon'$ and solving for k, we have $k = 2Bd/\varepsilon$, as claimed. $\square$

D Extending to Stochastic Responses

Instead of assuming voters respond deterministically with $\mathrm{resp}_\theta(q) = \mathbb{1}\{\theta \cdot q \ge 0\}$, we now allow responses to be stochastic.

We consider the random utility model (RUM) [Azari et al., 2012], which is a standard model in both social choice and alignment. Concretely, we can capture the stochasticity by introducing a generic function relating each single voter's utility difference to their probability of response

$$\psi : \mathbb{R} \to [0,1] \quad \text{with} \quad \psi(t) + \psi(-t) = 1 \quad \text{for all } t \in \mathbb{R}.$$

Then, when presented with a comparison (x, y_1, y_2), a voter responds 1 with probability $\psi(u_\theta(x,y_1) - u_\theta(x,y_2)) = \psi(\theta \cdot q)$ for $q = \Psi(x,y_1) - \Psi(x,y_2)$. More formally, given a voter type θ,

$$\Pr[\mathrm{resp}_\theta(q) = 1 |\ \theta] = \psi(\theta^\top q).$$

A common choice of the model is Bradley-Terry: $\psi_{\mathrm{BT}}(t) = \frac{1}{1+\exp(-t)}$.

In general, we can now generalize multi-query responses for independent $q_1, \ldots, q_t$ as

$$\tilde{Q}_t(\boldsymbol{q}) = \Pr_{\theta \sim \Theta}[[\mathrm{resp}_\theta(q_1) = 1] \wedge \ldots \wedge [\mathrm{resp}_\theta(q_t) = 1]] = \mathbb{E}_{\theta \sim \Theta}\left[\prod_{i=1}^{t} \Pr[\mathrm{resp}_\theta(q_i) = 1 |\ \theta]\right]$$

First, consider the problem of selecting a social-welfare–maximizing candidate. The first step of Lemma 3.2 requires rotational equivariance of inner-product–based responses. This argument continues to hold for the stochastic model, since for any rotation R we have $\psi((R\theta)^\top q) = \psi(\theta^\top (R^{-1} q))$.

Lemma D.1. *Let $\psi : [-1, 1] \to [0, 1]$, and $\theta \in \mathbb{S}^{d-1}$ be fixed. Then*

$$I_\psi(\theta) \coloneqq \int_{\mathbb{S}^{d-1}} \psi(\theta^\top q)\, q\, d\bar{\sigma}(q) = c_d(\psi)\ \theta,$$

where the scalar $c_d(\psi)$ is given by

$$c_d(\psi) = \frac{\Gamma\left(\frac{d}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{d-1}{2}\right)} \int_{-1}^{1} \psi(t)\, t\, (1 - t^2)^{\frac{d-3}{2}}\, dt$$

Proof. Let R be any rotation in $\mathbb{R}^d$. Because $\langle R\theta, Rq\rangle = \langle \theta, q\rangle$ and $d\bar{\sigma}$ is rotation equivariant,

$$I_\psi(R\theta) = \int \psi(\langle R\theta, q\rangle)\, q\, d\bar{\sigma}(q) = \int \psi(\langle \theta, R^{-1}q\rangle)\, q\, d\bar{\sigma}(q) = R \int \psi(\langle \theta, u\rangle)\, u\, d\bar{\sigma}(u) = R\, I_\psi(\theta).$$

Let $\mathcal{R}_\theta$ be the group of rotations that keep θ fixed. For any $R \in \mathcal{R}_\theta$, we have $R\theta = \theta$, and therefore by the equivariance established above,

$$R\, I_\psi(\theta) = I\psi(R\theta) = I_\psi(\theta).$$

Hence $I_\psi(\theta)$ is the fixed point of the linear action of the group $\mathcal{R}_\theta$. However, as $\mathcal{R}_\theta$ is acting as the full rotation group on the orthogonal complement of θ, the only fixed point in $\theta^\perp$ is the zero vector, and the only dimension left is for the span of the vector itself:

$$I_\psi(\theta) = c_d(\psi)\, \theta + 0, \text{ for some scalar } c_d(\psi).$$

Now to calculate the constant, we set $\theta = e_d$. Then

$$\begin{aligned} c_d(\psi) &= \int_{\mathbb{S}^{d-1}} \psi(q_d)\, q_d\, d\bar{\sigma}(q) \\ &= \frac{\Gamma\left(\frac{d}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{d-1}{2}\right)} \int_{-1}^{1} \psi(t)\, t\, (1 - t^2)^{\frac{d-3}{2}}\, dt. \end{aligned}$$

□

Observation D.2. To improve the signal-to-noise ratio in our moment estimators, we would like $c_d(\psi)$ to be as big as possible. Given that $\psi(x) + \psi(-x) = 1$,

$$\begin{aligned} &\int_{-1}^{1} \psi(t)\, t\, (1 - t^2)^{\frac{d-3}{2}}\, dt \\ &= \int_{0}^{1} \psi(t)\, t\, (1 - t^2)^{\frac{d-3}{2}}\, dt + \int_{-1}^{0} \psi(t)\, t\, (1 - t^2)^{\frac{d-3}{2}}\, dt \\ &= \int_{0}^{1} \psi(t)\, t\, (1 - t^2)^{\frac{d-3}{2}}\, dt + \int_{-1}^{0} (1 - \psi(-t))\, t\, (1 - t^2)^{\frac{d-3}{2}}\, dt \\ &= \int_{0}^{1} [2\psi(t) - 1]\, t\, (1 - t^2)^{\frac{d-3}{2}} dt \end{aligned}$$

Since $t\,(1 - t^2)^{\frac{d-3}{2}}$ is always nonnegative on $[0, 1]$, maximizing c_d is the same as maximizing $\psi(t)$ point-wise on $t \in (0, 1]$, which indicates that the more deterministic the preference is, the stronger is the signal.

The remainders of the proofs for Lemma 3.3 and Theorem 3.4 carry over without modification.

Theorem D.3. *The welfare-maximizing candidate is identifiable from* $\tilde{Q}_1$.

And since we can draw i.i.d. voters from the distribution Θ, $\tilde{Q}_1$ can be estimated as before.
Next, consider the task of estimating higher moments. Theorem 3.6 can be extended due to Lemma D.1.

Theorem D.4. *The* k*th moment* M_k *is identifiable with* $\tilde{Q}_k$.

And due to the properties of spherical harmonics including Funk-Hecke (Theorem B.1),

Theorem D.5. *Distributions are identifiable with* $\tilde{Q}_2$.