Optimal Communication-Distortion Tradeoff in Voting*

Debmalya Mandal
Columbia University
dm3557@columbia.edu

Nisarg Shah
University of Toronto
nisarg@cs.toronto.edu

David P. Woodruff
Carnegie Mellon University
dwoodruf@cs.cmu.edu

Abstract

In recent work, Mandal et al. [33] study a novel framework for the winner selection problem in voting, in which a voting rule is seen as a combination of an elicitation rule and an aggregation rule. The elicitation rule asks voters to respond to a query based on their preferences over a set of alternatives, and the aggregation rule aggregates voter responses to return a winning alternative. They study the tradeoff between the communication complexity of a voting rule, which measures the number of bits of information each voter must send in response to its query, and its distortion, which measures the quality of the winning alternative in terms of utilitarian social welfare. They prove upper and lower bounds on the communication complexity required to achieve a desired level of distortion, but their bounds are not tight. Importantly, they also leave open the question whether the best randomized rule can significantly outperform the best deterministic rule.

We settle this question in the affirmative. For a winner selection rule to achieve distortion d with m alternatives, we show that the communication complexity required is $\widetilde{\Theta}\left(\frac{m}{d}\right)$ when using deterministic elicitation, and $\widetilde{\Theta}\left(\frac{m}{d^3}\right)$ when using randomized elicitation; both bounds are tight up to logarithmic factors. Our upper bound leverages recent advances in streaming algorithms. To establish our lower bound, we derive a new lower bound on a multi-party communication complexity problem.

We then study the k-selection problem in voting, where the goal is to select a set of k alternatives. For a k-selection rule that achieves distortion d with m alternatives, we show that the best communication complexity is $\widetilde{\Theta}\left(\frac{m}{kd}\right)$ when the rule uses deterministic elicitation and $\widetilde{\Theta}\left(\frac{m}{kd^3}\right)$ when the rule uses randomized elicitation. Our optimal bounds yield the non-trivial implication that the k-selection problem becomes strictly easier as k increases.

1 Introduction

Making collective decisions through voting has been a subject of interest (at least) since the rise of democracy in ancient Athens. However, formal study of voting theory has more recent roots in the work of de Condorcet [21] in the late 18th Century. In the centuries subsequent to his work, social choice theorists pondered about the following canonical voting problem: If n voters express ranked preferences over a set of m alternatives, how should their preferences be aggregated to find the most socially desirable alternative? However, the lack of an objective notion of what is "socially desirable" led to a plethora of voting rules being proposed, with no clear consensus, even among experts, as to which voting rule is the best [15].

*A preliminary version of this paper was published in the proceedings of the 21st ACM Conference on Economics and Computation, 2020.

In the recent decades, the marriage between social choice theory and computer science has given rise to the field of computational social choice [16], and one of the key influences of computer science has been to view voting as an optimization problem. Specifically, Procaccia and Rosenschein [38] proposed the framework of *implicit utilitarian voting*, whereby voters' expressed ranked preferences over alternatives are seen as proxy for their underlying numerical utility functions. The overall framework consists of two steps. First, we must set an objective that one would want to optimize if voters' numerical utility functions were known. For example, one goal could be to simply choose an alternative maximizing the sum of voters' utilities (a.k.a. the *utilitarian social welfare*); this objective function has firm foundations in economic theory [11, 27, 29]. Given that it is impossible to perfectly optimize such an objective given the lack of complete information about voters' utility functions, the next step is to seek the best *worst-case approximation* of the objective function that can be achieved given available information. This worst-case approximation is referred to as *distortion* in this framework. Arguably, this notion of a worst-case approximation is another key contribution of computer science to economic theory, which has led to successful paradigms such as algorithmic mechanism design [37, 28, 39] and the price of anarchy [32].

One benefit of this distortion framework is that it yields an *optimal* voting rule to aggregate voters' expressed ranked preferences. Caragiannis et al. [17] and Boutilier et al. [14] identify the optimal deterministic and randomized aggregation rules, and show that their distortion is $\Theta(m^2)$ and $\widetilde{\Theta}(\sqrt{m})$, respectively, where m is the number of alternatives and $\widetilde{\Theta}$ hides logarithmic factors.

Later, Benadè et al. [9] observe that the implicit utilitarian voting framework has another benefit: not only can it be used to derive the optimal method to aggregate ranked votes, or k-approval votes, or votes expressed in any other *input format* for that matter, it can also be used to compare the efficacy of different input formats. Pushing this idea to the next level, Mandal et al. [33] propose optimizing both the input format and the vote aggregation method simultaneously. Specifically, they view a *winner selection* rule (i.e. a rule which returns a single winning alternative) as a combination of an elicitation rule, which specifies how voters should submit their votes in a certain format given their numerical utility functions, and an aggregation rule, which specifies how voters' votes should be aggregated to find a single winning alternative. Then, they formally study the elicitation-distortion tradeoff: to achieve a desired distortion of d, what is the minimum number of bits of information that must be elicited from each voter about her utility function? For deterministic elicitation rules, they give an upper bound of $\widetilde{O}(m/d)$ and a lower bound of $\Omega(m/d^2)$. For randomized elicitation rules, they give a lower bound of $\Omega(m/d^3)$. We omit a discussion of the deterministic or randomized nature of the aggregation rule for now; a detailed discussion is presented in Section 7. Their work leaves open two key questions:

- What is the optimal communication complexity required to achieve distortion d in the winner selection problem using deterministic and randomized elicitation?
- Can the optimal winner selection rule with randomized elicitation significantly (i.e. beyond logarithmic factors) outperform that with deterministic elicitation?

In this paper, we answer both these questions by identifying the optimal elicitation-distortion tradeoff in the winner selection problem in voting.

We also examine the *k-selection* problem, where the goal is to return a set of k alternatives. This is a widely studied problem in voting, often known as *committee selection* or *multiwinner voting* [23, 22, 17]. For the classical setting where the elicitation rule is fixed to be the one which elicits ranked preferences, Caragiannis et al. [17] study the optimal distortion which can be achieved using deterministic and randomized aggregation rules. But no prior work considers optimizing the elicitation rule, along with the aggregation rule, for this problem. We provide optimal bounds in this case as well.

1.1 Our Results

Let us briefly introduce our problem a bit more formally (a detailed model is presented in Section 2. For $k \in \mathbb{N}$, we define $[k] \triangleq \{1, \ldots, k\}$. There is a set of voters $N = [n]$ and a set of m alternatives A. Each voter i has a valuation function $v_i : A \to \mathbb{R}_{\geq 0}$. Following a standard assumption in voting theory [14, 17, 33], we assume normalized valuations, where $\sum_{a \in A} v_i(a) = 1$ for each voter i. Given the vector of valuations $\vec{v} = (v_1, \ldots, v_n)$, we want to maximize a certain objective function. However, eliciting real-valued v_i precisely requires asking voters to communicate potentially infinitely many bits of information. We are interested in examining how well we can approximate a given objective function in the worst case given a bound on the number of bits of information we are allowed to elicit from each voter. This worst-case approximation ratio is termed *distortion* in the literature, and the (expected) number of bits elicited from each voter is termed the *communication complexity*. We note that following traditional modeling, we assume that the rule asks the same "query" (i.e. how to map valuation function to a discrete response) to all voters. The model is more formally laid out in Section 2.

1.1.1 Winner Selection

In the winner selection (i.e. 1-selection) problem, we are interested in finding a single alternative with the goal of maximizing the social welfare: $\mathrm{sw}(a, \vec{v}) = \sum_{i \in N} v_i(a)$. Specifically, we are interested in the communication complexity required to achieve distortion at most d, for a given d. For this problem, the results of Mandal et al. [33] Pareto dominate prior results in the literature. Hence, we only present a comparison of our results to theirs.

With deterministic elicitation, Mandal et al. propose a voting rule — PrefThreshold — which achieves an upper bound of $\widetilde{O}(m/d)$ communication complexity, and establish a weaker $\Omega(m/d^2)$ lower bound. We show that their upper bound is tight up to logarithmic factors by proving that every winner selection rule with deterministic elicitation requires $\Omega(m/d)$ communication complexity. We note that the PrefThreshold voting rule of Mandal et al. uses deterministic aggregation, whereas our lower bound holds even for randomized aggregation, thus establishing that with optimal deterministic elicitation, randomized aggregation does not provide significant benefit over deterministic aggregation for the winner selection problem.

With randomized elicitation, the story is inverted. Mandal et al. do not offer any better upper bounds than the $\widetilde{O}(m/d)$ achieved with deterministic elicitation (except in very restricted cases), but do offer a lower bound of $\Omega(m/d^3)$. In this case, we show that their lower bound is tight (up to logarithmic factors) by proposing a new winner selection rule with randomized elicitation which uses $\widetilde{O}(m/d^3)$ communication complexity.

Our optimal results imply that for the winner selection problem, randomized elicitation indeed offers a significant benefit (i.e. beyond logarithmic factors) over deterministic elicitation.

1.1.2 k-Selection

In the k-selection problem, the goal is to select a set of k alternatives, where k is given. Trivially, $k = 1$ is exactly the winner selection problem mentioned above. Hence, we focus on the case of

$k > 1$. To fully specify this problem, we need to define the objective function we want to maximize. Following Caragiannis et al. [17], we say that the social welfare of a set S of k alternatives is $\text{sw}(S, \vec{v}) = \sum_{i \in N} \max_{a \in S} v_i(a)$. In other words, each voter i derives value from only her most favorite alternative in S. This formulation is applicable in contexts where the k alternatives act as substitutes. For example, in political elections for choosing a committee of k representatives, it is often assumed that a voter is "represented" by her most favorite candidate who is elected in the committee [35, 19]. We follow this framework, albeit note that there are other equally interesting formulations of social welfare of a set of alternatives, which could lead to interesting future work (see Section 7).

For the k-selection problem, the only known bounds on distortion are those established by Caragiannis et al. [17]. They consider the specific elicitation rule which asks each voter to provide a ranking of the alternatives by value; this requires $O(m \log m)$ bits of elicitation. As noted by Mandal et al. [33], rankings are not a very efficient form of elicitation in the winner selection problem: they allow achieving only $\widetilde{\Theta}(\sqrt{m})$ distortion with $\Theta(m \log m)$ bits of communication (even with randomized aggregation), whereas their PrefThreshold method achieves $O(1)$ distortion with just $O(m \log \log m)$ bits of elicitation (and deterministic aggregation!). Our results imply that the same holds for the k-selection problem. Since our bounds significantly outperform those of Caragiannis et al., we omit a detailed presentation of their (complicated) bounds, and directly present our results.

With deterministic elicitation, we show that the optimal communication complexity required to achieve distortion d is $\widetilde{\Theta}(\frac{m}{kd})$. Note that this bound decreases linearly as k increases. The same holds for randomized elicitation, which leads to optimal communication complexity of $\widetilde{\Theta}\left(\frac{m}{kd^3}\right)$.

We remark that apriori it is not even clear that the problem becomes easier as k increases. Increasing the value of k allows returning larger sets that achieve higher social welfare, but it also raises the *optimal* social welfare against which a voting rule needs to compete. For example, the bounds established by Caragiannis et al. [17] for ranked elicitation do not monotonically decrease with k (but they are also loose enough that they do not rule out the possibility of the optimal bounds for ranked elicitation decreasing monotonically with k). A more detailed discussion is presented in Section 6.3.

1.2 Related Work

We begin by describing the results of Mandal et al. [33] in detail, as their work is the most relevant to ours. For deterministic elicitation, they construct an intuitive voting rule PrefThreshold, in which each voter is asked to report her approximate utility (with granularity parametrized by ℓ) for her t most preferred alternatives. Using a simple deterministic aggregation rule and by setting appropriate values of t and ℓ, they show that distortion d can be achieved with $\widetilde{O}(m/d)$ bits of communication. Our result establishes this rule as asymptotically optimal (up to logarithmic factors) for deterministic elicitation. For randomized elicitation, they construct a voting rule RandSubset, which outperforms PrefThreshold by logarithmic factors, but leave open the possibility of a rule that significantly outperforms PrefThreshold, only establishing a $\Omega(m/d^3)$ lower bound. We show that their lower bound is tight (again, up to logarithmic factors) by constructing a voting rule that achieves it. We note that for certain cases, Mandal et al. provide exactly tight bounds; for example, they show that the optimal distortion with $\log m$ bits of communication is $\Theta(m^2)$, achieved by the plurality voting rule. We are only concerned with optimality up to logarithmic factors.

More recently, Amanatidis et al. [2] also consider the elicitation-distortion tradeoff in voting. However, they take a *query complexity* approach to measuring communication. Specifically, they consider a setting where the preference rankings of the agents are known to the center, and the center can perform two types of queries: a *value query* asks a voter to report her precise utility for an alternative, and a *comparison query* asks a voter whether her utility for one alternative is at least x times her utility for another alternative. Our upper bound results are incomparable to theirs because they work with cardinal queries on top of known ordinal information. We do not assume any upfront knowledge about voters' preferences, and the set of possible queries are restricted to eliciting finitely many bits. At first glance, it may seem that our lower bounds carry over to their framework. However, they allow adaptively asking different queries to different voters, whereas our lower bounds apply when the queries are common across voters.

When asking different questions to different voters is allowed, Caragiannis and Procaccia [18] showed that one can achieve significantly lower distortion using simple techniques (e.g. $O(1)$ distortion with only $\log m$ bits per voter). They show that this is possible via specific limited communication schemes, which they call embeddings. Recently [12] consider a setting where the center can ask a (randomized) threshold query to different voters with possibly different thresholds for different voters, and then the agents approve all the alternatives that they rank higher than this threshold. For this case, Bhaskar et al. [12] show that achieving constant distortion is possible even with *vanishing* number of bits per voter (specifically, with total number of bits independent of the number of voters). We argue that having a common ballot that all voters respond to is a natural assumption and is the most common practice for conducting voting in the real world.

Broadly, our work sits within the framework of implicit utilitarian voting in which no assumptions are made on voters' underlying numerical utility functions [38, 14, 10, 17, 12]. In certain contexts (especially for political elections), it is also common to assume that voters and alternatives lie in an underlying metric space, and voters' utilities (or costs) for alternatives respect the triangle inequality [4, 5, 25, 13, 36, 1, 24].

We also note that our use of sketching voter utility functions closely resembles the line of work on sketching combinatorial valuation functions [8, 7]. Their goal is to compress exponentially many numbers into polynomially many bits, whereas in our case, there are only polynomially many numbers (but infinitely many bits in exact representation) which need to be compressed.

To the best of our knowledge, ours is the first work to use sketching to design optimal voting rules. In particular, we use L_p samplers [34, 31, 30] which given a sequence of updates to an underlying vector, process the stream and finally output a coordinate proportional to it's p-th power. Space-efficient L_p-samplers can be used for various applications, like moment estimation, or finding heavy hitters. The reader is referred to the work of Monemizadeh and Woodruff [34] and Andoni et al. [3] for further applications.

2 Model

For $k \in \mathbb{N}$, define $[k] = \{1, \ldots, k\}$. Let $x \sim D$ denote that random variable x has distribution D. Let log denote logarithm to base 2, ln denote logarithm to base e, and med denote the median.

There is a set of alternatives A with $|A| = m$, and a set of voters $N = [n]$. Each voter $i \in N$ is endowed with a valuation $v_i : A \to \mathbb{R}_{\geq 0}$, where $v_i(a) \geq 0$ represents the value of voter i for alternative a. Equivalently, we view $v_i \in \mathbb{R}^m_{\geq 0}$ as a vector which contains the voter's value for each alternative. Collectively, voter valuations are denoted by valuation profile $\vec{v} = (v_1, \dots, v_n)$. Given a valuation profile $\vec{v}$, the (utilitarian) social welfare of an alternative a is $\text{sw}(a, \vec{v}) = \sum_{i \in N} v_i(a)$. In the k-selection problem, we are interested in social welfare of a set of k alternatives. For a set $S \subseteq A$, define $v_i(S) = \max_{a \in S} v_i(a)$ and $\text{sw}(S, \vec{v}) = \sum_{i \in N} v_i(S)$. Our goal is to elicit information about voter valuations, and use it to find an alternative (in the winner selection problem) or a set of k alternatives (in the k-selection problem) with high social welfare.

Valuations: We adopt the standard normalization assumption [6] that $\sum_{a \in A} v_i(a) = 1$ for each $i \in N$. This is akin to a "one voter, one vote" principle. Alternatively, one can think of $v_i(a)$ as the intensity of voter i's relative preference for a as compared to other alternatives. Let Δ^m denote the m-simplex, i.e., the set of all vectors in $\mathbb{R}^m_{\geq 0}$ whose coordinates sum to 1. Hence, we have that $v_i \in \Delta^m$ for each $i \in N$.

Query space: The literature on voting considers different types of responses, e.g., plurality votes, k-approval votes (which ask voters to report the set of their k favorite alternatives), threshold approval votes (which ask voters to approve all alternatives for which their value is at least a given threshold), and ranked votes. Mandal et al. [33] unify these through the framework of communication complexity.

Consider an interaction with voter i which elicits finitely many bits of information and in which the voter responds deterministically. In this interaction, the voter must provide one of finitely many (say k) possible responses. Following Mandal et al. [33], we say that this interaction elicits $\log k$ bits of information. It effectively partitions Δ^m into k *compartments*, where the compartment corresponding to each response is the set of all valuations which would result in the voter choosing that response. In other words, any interaction which elicits $\log k$ bits of information is equivalent to a *query* which partitions Δ^m into k compartments and asks the voter to pick the compartment in which her valuation belongs. Let $\mathcal{Q}$ denote the set of all queries which partition Δ^m into finitely many compartments. For a query $q \in \mathcal{Q}$, let $k(q)$ denote the number of compartments created by q; the number of bits elicited is $\log k(q)$.

Voting Rule: A voting rule consists of two parts: an *elicitation rule* Π_f and an *aggregation rule* Γ_f. The (randomized) elicitation rule Π_f is a distribution over $\mathcal{Q}$, according to which a query q is sampled. Each voter i provides a response ρ_i to this query, depending on her valuation v_i. We say that the elicitation rule is *deterministic* if it has singleton support (i.e., it chooses a query deterministically). The (randomized) aggregation rule Γ_f, takes voter responses $\vec{\rho} = (\rho_1, \dots, \rho_n)$ as input, and returns an output. For the 1-selection problem, this is a distribution over the alternatives, and for the k-selection problem, this is a distribution over sets of k alternatives. We say that the aggregation rule is *deterministic* if it always returns a distribution with singleton support. Slightly abusing notation, we denote by $f(\vec{v})$ the (randomized) alternative / subset returned by f when voter valuations are $\vec{v} = (v_1, \dots, v_n)$. We measure the performance of f via two metrics.

1. The *communication complexity* of f for m alternatives, denoted $\mathrm{C}^m(f) = \mathbb{E}_{q \sim \Pi_f}[\log k(q)]$, is the expected number of bits of information elicited by f from each voter. We drop m from the superscript when its value is clear from the context.

2. The *distortion* of f for m alternatives, denoted $\text{dist}^m(f)$, is the worst-case ratio of the optimal social welfare to the expected social welfare achieved by f. Again, we drop m from the

superscript when its value is clear from the context. Formally,

$$\mathrm{dist}(f) = \sup_{\vec{v}\in(\Delta^m)^n} \frac{\max_{a\in A} \mathrm{sw}(a,\vec{v})}{\mathbb{E}_{\widehat{a}\sim f(\vec{v})} \mathrm{sw}(\widehat{a},\vec{v})}, \qquad \text{[1-selection]}$$

$$\mathrm{dist}(f) = \sup_{\vec{v}\in(\Delta^m)^n} \frac{\max_{S\subseteq A:|S|=k} \mathrm{sw}(S,\vec{v})}{\mathbb{E}_{\widehat{S}\sim f(\vec{v})} \mathrm{sw}(\widehat{S},\vec{v})}. \qquad [k\text{-selection}]$$

3 Winner Selection: Randomized Elicitation Upper Bound

While it is desirable for a voting rule to have low communication complexity and low distortion, typically eliciting more information from voters enables achieving low distortion. Mandal et al. [33] study this trade-off for the winner selection problem. They propose a voting rule with deterministic elicitation and aggregation which achieves $O(d)$ distortion with $\widetilde{O}(m/d)$ communication complexity, but their lower bound on communication complexity for achieving $O(d)$ distortion with deterministic elicitation (and possibly randomized aggregation) is only $\Omega(m/d^2)$. For randomized elicitation, they do not propose any voting rule which improves upon their deterministic elicitation bound for general d, and present a weaker $\Omega(m/d^3)$ lower bound.

In this and next section, we fill the gaps for both deterministic and randomized elicitation. We start by presenting our new voting rule with randomized elicitation and deterministic aggregation, which achieves distortion d with communication complexity $\widetilde{O}(m/d^3)$. This matches, up to logarithmic factors, the lower bound established by Mandal et al. [33] for randomized elicitation. Note that their lower bound holds even for randomized aggregation, whereas our rule achieves it with deterministic aggregation. The main tool we use in our improved algorithm is the notion of an L_p-sampler introduced by Monemizadeh and Woodruff [34]. An L_p-sampler processes a sequence of updates to an underlying vector $x \in \mathbb{R}^m$ (with less memory than what is required to simply store x). After processing the entire stream of updates, its goal is to output a random coordinate of x such that the i-th coordinate is sampled with probability approximately proportional to $|x(i)|^p$.

Definition 1. Let $x \in \mathbb{R}^m$ and $\delta > 0$ be a small constant. An L_2-sampler with relative error ν is an algorithm that, with probability at least $1-\delta$, outputs a coordinate $\widehat{j}$ such that for any $j \in [m]$,

$$\Pr[\widehat{j} = j] \in \left[(1-\nu)\frac{|x(j)|^2}{\|x\|_2^2}, (1+\nu)\frac{|x(j)|^2}{\|x\|_2^2}\right] \pm O(m^{-c}),$$

where $c \geq 1$ is an arbitrary constant. With the remaining probability (at most δ), the algorithm can output FAIL. When $\nu = 0$, this is known as a perfect L_2-sampler.

Our goal is to use such a sampler to sample according to the social welfare vector $\mathrm{sw} = \sum_{i\in N} v_i$. However, each v_i is held privately by voter i. Hence, we need a multi-agent version of the L_2-sampler, which can obtain the required information from each agent i about her vector v_i, and then perform L_2-sampling on $\mathrm{sw} = \sum_{i\in N} v_i$. This is where we crucially use the fact that the L_2-sampler of Jayaram and Woodruff [30] uses a linear "sketch" A. Therefore, we can obtain the linear sketch $A(v_i)$ from each voter i, and combine these to compute $A(\mathrm{sw}) = \sum_{i\in N} A(v_i)$. Let us describe the high-level template of the L_2-sampler of Jayaram and Woodruff [30]:

(a) Duplicate the input $x \in \mathbb{R}^m$ by copying each coordinate m^{c-1} times to obtain $X \in \mathbb{R}^{m^c}$, and then scale each coordinate by an i.i.d. random variable to get a vector $\zeta \in \mathbb{R}^{m^c}$.

(b) Run the duplicated input X and the scaled input ζ through the count-sketch algorithm [20] to get a sketch $A(x)$, which consists of four components.

(c) Select an index $\widehat{j}$ using a statistic of $A(x)$.

(d) Use a statistical test to determine whether to output $\widehat{j}$ or to output FAIL.

ALGORITHM 1: L_2-Sampler

Input: $\{x_i\}_{i=1}^n$, δ.

1. Run COMMUNICATE $(\{x_i\}_{i=1}^n, \delta)$ to get $A(x_i) = (A_{i,1}, A_{i,2}, A_{i,3}, A_{i,4})$ from each agent i.
2. Compute $A_r = \sum_{i=1}^n A_{i,r}$ for $r \in [4]$.
3. Run SAMPLE $(\{A_r\}_{r\in[4]})$ to get (STATUS,$(j^*, \widehat{x}(j^*))$).
4. IF: STATUS == FAIL:
   - RETURN FAIL.
5. ELSE:
   - RETURN $(j^*, \widehat{x}(j^*))$.

In our multi-agent setup, we assume that there is a vector x_i held by each agent i, where $x_i(j) \in \{j/\Delta : j \in [\Delta] \cup \{0\}\}$ for some fixed $\Delta \in \mathbb{N}$. We do not allow arbitrary real numbers because we are interested in analyzing the exact number of bits that each agent will need to communicate. We want to perform perfect L_2-sampling on the vector $x = \sum_i x_i$. To do so, we essentially require that each agent i run steps (a) and (b) to compute $A(x_i)$ and communicate it to the center. Since these sketches are linear, the center computes $A(x) = \sum_i A(x_i)$, and uses it to select a random index $\widehat{j}$ or output FAIL. The algorithm is presented as Algorithm 1. It first runs COMMUNICATE, which takes discretized agent vectors (and a failure probability δ) as input, performs Steps (a) and (b) of the L_2-sampler on these valuations, and communicates the obtained sketches $A(x_i)$ back to the center. Next, it sums the sketches from the individual agents. And finally, it runs SAMPLE, which takes the combined sketches as input, and performs Steps (c) and (d) of the L_2-sampler to decide whether to return FAIL or an index j^* along with an estimate of its value x_{j^*}. The details of the procedures COMMUNICATE and SAMPLE are provided in the appendix; for further explanation of these procedures, we refer the reader to the work of Jayaram and Woodruff [30]. The purpose of the next theorem is to re-cast the guarantees provided by Jayaram and Woodruff [30] for their perfect L_2-sampler into a guarantee of our multi-agent variant, Algorithm 1.[1]

Theorem 1. *Let $\Delta \in \mathbb{N}$, c be a sufficiently large constant, and $\delta \geq 1/poly(m)$. Suppose each agent i holds a vector x_i such that $x_i(j) \in \{0, 1/\Delta, 2/\Delta, \ldots, 1\}$ for each $j \in [m]$. Let $x = \sum_i x_i$. Then, Algorithm 1 outputs* FAIL *with probability at most δ, and with the remaining probability, its output $(j^*, \widehat{x}(j^*))$ satisfies the following two conditions.*

[1] All the missing proofs are provided in the appendix.

- *For each $j \in [m]$,*

$$\Pr[j^* = j] = \frac{x(j)^2}{\|x\|_2^2} \pm O(m^{-c}).$$

- *Conditioned on the event that $j^* = j$,*

$$\widehat{x}(j^*) \in [1/2 \cdot x(j), 2 \cdot x(j)] \, .$$

Moreover, under this algorithm, each agent communicates $O(\log^2 m \log(\Delta \log m))$ bits.

Next, we want to use this algorithm to design our voting rule. The elicitation rule of our voting rule is simple: each voter i approximates her valuation function v_i to v_i^Δ by rounding each $v_i(a)$ to the nearest multiple of $1/\Delta$, and then sends the sketch $A(v_i^\Delta)$ as specified by Algorithm 1. We show that for an appropriately chosen Δ, the error introduced in this step does not significantly affect the final result. This is one of the contributions of the next result.

Another contribution is that instead of requesting just one sketch from each voter, we generate t independent sketch functions (for an appropriately chosen t), and request the corresponding sketches from each voter. Recall that our randomized voting rule is allowed to select a random query q which maps each voter's valuation to a response and ask voters to respond to q. Communicating query q to the voters is free of cost. Equivalently, one can imagine that there is a public tape, and the voting rule can write any information required to represent query q on this tape, free of cost. Hence, to request these sketches, the voting rule generates four random hash functions as well as m^c i.i.d. exponential random variables for each of t sketches, and writes them onto the public tape, so voters know exactly which sketches they are supposed to compute.

In the aggregation rule, we run the t samplers on the combined sketches obtained from the voters. If any of those samplers fail (we show this happens with a low probability), then we simply return an arbitrary alternative. Otherwise, we return the alternative for which the corresponding estimated count returned by our sampler is the highest. The voting rule is formally presented as Algorithm 2. A key contribution of the next result is to show that using L_2-samplers to sample alternatives according to the square of their social welfare helps achieve the optimal distortion. To the best of our knowledge, this is the first result using L_2-samplers to design a voting rule. Although the elicitation rule closely follows the L_2-sampler designed by [30], the novelty lies in designing the aggregation rule and analyzing the distortion achieved.

Theorem 2. *For any d,* MAX-L_2-SAMPLER *achieves dist*(MAX-L_2-SAMPLER) $= O(d)$ *with*

$$C(\text{MAX-}L_2\text{-SAMPLER}) = O\left(\frac{m}{d^3} \log^3 m\right).$$

Proof. Recall that our parameter choices are $\Delta = 128m^3$, and $t = 4m/d^3$. Let E denote the event that none of t independent L_2-samplers fail. Since each L_2-sampler is run with failure probability $\delta = 1/(4t) \geq 1/\text{poly}(m)$, the probability that each L_2-sampler fails is at most $1/(4t)$. Hence, by the union bound, the probability of the event E is at least $1 - t \times 1/(4t) = 3/4$. Since the expected social welfare achieved by the voting rule is at least $3/4$ times the expected social welfare achieved conditioned on E, we condition on E being true for the rest of the proof. The final distortion can only increase by a factor of at most $4/3$.

ALGORITHM 2: MAX-L_2-SAMPLER

Elicitation Rule:

- Set $t = 4m/d^3$ and $\Delta = 128m^3$.
- Each voter i computes an approximate valuation v_i^Δ, where, for each alternative $a \in A$, $v_i^\Delta(a)$ is $v_i(a)$ rounded to the nearest multiple of $1/\Delta$.
- Run t independent copies of the procedure $\mathsf{COMMUNICATE}\left(\{v_i^\Delta\}_{i=1}^n, \frac{1}{4t}\right)$.
- Obtain responses $A_r^k(v_i^\Delta)$ for each $k \in [t]$, $r \in [4]$, and $i \in [n]$.

Aggregation Rule:

- For each $k \in [t]$ and $r \in [4]$, compute $A_r^k = \sum_{i=1}^n A_r^k(v_i^\Delta)$.
- For each $k \in [t]$, run procedure $\mathsf{SAMPLE}\left(\{A_r^k\}_{r\in[4]}\right)$ to obtain either FAIL or a pair $(a^k, \widehat{\mathrm{sw}}(a^k))$.
- If at least one algorithm returns FAIL, then output an arbitrary alternative.
- Otherwise, return $\widehat{a} = a^{k^*}$, where $k^* \in \arg\max_{k\in[t]} \widehat{\mathrm{sw}}(a^k)$.

Communication Complexity:

$$\mathrm{C}(\text{MAX-}L_2\text{-SAMPLER}) = O\left(\frac{m}{d^3}\log^3 m\right).$$

Distortion:

$$\mathrm{dist}(\text{MAX-}L_2\text{-SAMPLER}) = O(d).$$

Since our algorithm just calls t L_2-samplers in parallel, using Theorem 1, the communication complexity of the voting rule is clearly bounded by $O(t\log^2 m \log(m^3 \log m)) = O\left(\frac{m}{d^3}\log^3 m\right)$. It remains to show that its distortion is $O(d)$.

Let $(a^k, \widehat{\mathrm{sw}}(a^k))$ denote the output of the k-th L_2 sampler. Throughout the proof, we will use three notions of welfare: (1) the *true welfare* $\mathrm{sw}(a) = \sum_{i=1}^n v_i(a)$, (2) the *rounded welfare* $\mathrm{sw}^\Delta(a) = \sum_{i=1}^n v_i^\Delta(a)$ where $v_i^\Delta(a)$ is $v_i(a)$ rounded to the nearest multiple of $1/\Delta$, and (3) the *estimated welfare* $\widehat{\mathrm{sw}}(a^k)$ returned by the k-th L_2-sampler, which is an estimate of $\mathrm{sw}^\Delta(a^k)$. We will write sw and sw^Δ to denote the vectors of true and rounded welfare.

First we notice the following obvious relationship between $\mathrm{sw}(a)$ and $\mathrm{sw}^\Delta(a)$.

$$\forall a \in A : \left|\mathrm{sw}^\Delta(a) - \mathrm{sw}(a)\right| = \left|\sum_{i=1}^n v_i^\Delta(a) - \sum_{i=1}^n v_i(a)\right| \le \sum_{i=1}^n \left|v_i^\Delta(a) - v_i(a)\right| \le \frac{n}{\Delta} \tag{1}$$

Let $a^* \in \arg\max_{a\in A} \mathrm{sw}(a)$ be an alternative with the highest social welfare. Next, we show a lower bound on the expected social welfare of the alternative $\widehat{a}$ returned by our voting rule. The proof of the next lemma is given in the appendix.

Lemma 1. $\mathbb{E}[\mathrm{sw}(\widehat{a})] \ge \frac{1}{256}\frac{n}{m^3}$.

Finally, to derive an upper bound on the distortion (i.e. to derive an upper bound on $\frac{\mathrm{sw}(a^*)}{\mathbb{E}[\mathrm{sw}(\widehat{a})]}$), we consider two cases.

Case 1: Suppose $\mathrm{sw}^\Delta(a^*) \ge \frac{\|\mathrm{sw}^\Delta\|_2}{\sqrt{td/3}}$. In this case, $\mathrm{sw}^\Delta(a^*)^2 \ge \frac{3\|\mathrm{sw}^\Delta\|_2^2}{td}$. Since each a^k is generated by a perfect L_2 sampler, we have

$$\Pr[a^k = a^*] \geq \frac{\mathrm{sw}^\Delta(a^*)^2}{\|\mathrm{sw}^\Delta\|_2^2} - O(m^{-c}) \geq \frac{3}{td} - O(m^{-c}).$$

Note that $td = 4m/d^2 = O(m)$. Hence, for a sufficiently large c, we can ensure that this probability is at least $2/td$. Therefore, the probability that none of a^k, $k \in [t]$, are equal to a^* is at most $\left(1 - \frac{2}{td}\right)^t \leq 1 - \frac{1}{d}$. This implies that with probability at least $1/d$, a^* appears as a^k for at least one $k \in [t]$. Since we select the final alternative $\widehat{a}$ as a^k with the highest estimated welfare $\widehat{\mathrm{sw}}$, we have

$$\begin{aligned}\mathbb{E}[\mathrm{sw}(\widehat{a})] &\geq \mathbb{E}[\mathrm{sw}^\Delta(\widehat{a})] - \frac{n}{\Delta} \geq \frac{1}{2} \cdot \mathbb{E}[\widehat{\mathrm{sw}}(\widehat{a})] - \frac{n}{\Delta} \geq \frac{1}{2d} \cdot \mathbb{E}[\widehat{\mathrm{sw}}(a^*)] - \frac{n}{\Delta} \\ &\geq \frac{1}{4d} \cdot \mathrm{sw}^\Delta(a^*) - \frac{n}{\Delta} \geq \frac{1}{4d} \cdot \mathrm{sw}(a^*) - \frac{n}{\Delta}\left(1 + \frac{1}{4d}\right).\end{aligned}$$

The first inequality follows from Equation (1). The second inequality follows from the guarantee in Theorem 1. The third inequality follows because $\widehat{\mathrm{sw}}(\widehat{a}) \geq \widehat{\mathrm{sw}}(a^k)$ for each $k \in [t]$, and $a^* \in \{a^k : k \in [t]\}$ with probability at least $1/d$. The fourth inequality again uses Theorem 1. The final inequality again uses Equation (1). Rearranging, we get the following bound on the distortion.

$$\frac{\mathrm{sw}(a^*)}{\mathbb{E}[\mathrm{sw}(\widehat{a})]} \leq 4d + \frac{\frac{n}{\Delta}\left(1 + \frac{1}{4d}\right)}{\mathbb{E}[\mathrm{sw}(\widehat{a})]} \leq 4d + 2\left(1 + \frac{1}{4d}\right) \leq 8d,$$

where the second inequality follows from substituting $\Delta = 128m^3$ and using Lemma 1.

Case 2: Suppose $\mathrm{sw}^\Delta(a^*) < \frac{\|\mathrm{sw}^\Delta\|_2}{\sqrt{td/3}}$. Fix any $k \in [t]$. We claim that with probability at least $1/2$, we have $\left(\mathrm{sw}^\Delta(a^k)\right)^2 / \|\mathrm{sw}^\Delta\|_2^2 \geq \frac{1}{2m}$. This is because every alternative a with $\left(\mathrm{sw}^\Delta(a)\right)^2 / \|\mathrm{sw}^\Delta\|_2^2 \leq \frac{1}{2m}$ is picked with probability at most $\frac{1}{2m}$ and there are at most m such alternatives.

Thus, for each $k \in [t]$, the following holds.

$$\begin{aligned}\mathbb{E}[\mathrm{sw}(a^k)] &\geq \mathbb{E}[\mathrm{sw}^\Delta(a^k)] - \frac{n}{\Delta} \geq \frac{1}{2} \cdot \frac{\|\mathrm{sw}^\Delta\|_2}{\sqrt{2m}} - \frac{n}{\Delta} \\ &\geq \frac{1}{9} \cdot \sqrt{\frac{td}{m}} \cdot \mathrm{sw}^\Delta(a^*) - \frac{n}{\Delta} \geq \frac{1}{9} \cdot \sqrt{\frac{td}{m}} \cdot \mathrm{sw}(a^*) - \frac{n}{\Delta}\left(1 + \frac{1}{9}\sqrt{\frac{td}{m}}\right).\end{aligned}$$

The first and the final inequalities use Equation (1), and the third inequality uses the fact that we are in the case of $\|\mathrm{sw}^\Delta\|_2 > \mathrm{sw}^\Delta(a^*) \cdot \sqrt{td/3}$.

Since the final alternative $\widehat{a} = a^k$ for some $k \in [t]$, we get the following bound on the distortion.

$$\frac{\mathrm{sw}(a^*)}{\mathbb{E}[\mathrm{sw}(\widehat{a})]} \leq 9\sqrt{\frac{m}{td}} + \frac{\frac{n}{\Delta}\left(1 + \frac{1}{9}\sqrt{\frac{td}{m}}\right)}{\mathbb{E}[\mathrm{sw}(\widehat{a})]} \leq \frac{9d}{2} + 2\left(1 + \frac{2}{9d}\right) \leq 9d,$$

where the second inequality follows from substituting $\Delta = 128m^3$ and $t = 4m/d^3$, and using the lower bound on $\mathbb{E}[\mathrm{sw}(\widehat{a})]$ from Lemma 1. $\square$

4 Winner Selection: Deterministic Elicitation Lower Bound

In this section, we derive an $\Omega(m/d)$ lower bound on the communication complexity of voting rules which achieve distortion at most d using deterministic elicitation (and possibly randomized aggregation). Mandal et al. [33] provide an upper bound of $\widetilde{\Theta}(m/d)$, establishing that our lower bound is optimal up to logarithmic factors.

To prove our lower bound, we use a reduction to the multi-party set disjointness problem. To keep the paper self-contained, we begin with a brief background of the multi-party communication complexity literature.

4.1 Setup of Multi-Party Communication Complexity

In multi-party communication complexity, there are t players, denoted 1 through t. Each player i holds a private input $X_i \in \mathcal{X}_i$. We refer to $(X_1, \ldots, X_t)$ as the *input profile*. The players are assumed to be computationally omnipotent and limited only in their communication capabilities. The goal is to compute the output of a function $f : \mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_t \to \{0, 1\}$ on an input profile. To compute this function, players communicate messages about their private input. We assume a *blackboard* model, in which each player writes her messages on the blackboard, and they are visible to all other players for free.

In this framework, a *deterministic protocol* Γ specifies how players should write messages on the blackboard given their private input and any previous messages they see on the blackboard. We use $\Gamma(X_1, \ldots, X_t)$ to denote the transcript generated on the blackboard from messages written by all players given input profile $(X_1, \ldots, X_t)$.

Definition 2 (Deterministic Communication Cost)**.** The *deterministic communication cost* of protocol Γ, denoted by $D(\Gamma)$, is the maximum length of the transcript $\Gamma(X_1, \ldots, X_t)$, where the maximum is taken over all input profiles $(X_1, \ldots, X_t)$.

We say that Γ is a protocol for f if there exists a function Π_{out} mapping the set of possible transcripts to $\{0, 1\}$ such that $\Pi_{out}(\Pi(X_1, \ldots, X_t)) = f(X_1, \ldots, X_t)$ for every input profile $(X_1, \ldots, X_t)$.

Definition 3 (Deterministic Communication Complexity)**.** The *deterministic communication complexity* of f, denoted $D(f)$, is the deterministic communication cost of the best deterministic protocol for f, i.e. $D(f) = \min_{\Gamma\,:\,\Gamma \text{ is a protocol for } f} D(\Gamma)$.

We now briefly review randomized protocols and randomized communication complexity. We assume *public randomness*. That is, players have free access to a shared stream of random coin tosses. A *randomized protocol* Π specifies how players should write messages on the blackboard given their private input, random coin tosses, and any previous messages they see on the blackboard. Let $\Pi(X_1, \ldots, X_t)$ be the random variable that denotes the transcript generated when all the players follow the protocol given input profile $(X_1, \ldots, X_t)$; here the randomness is due to the public coin tosses.

Definition 4 (Randomized Communication Cost)**.** The *randomized communication cost* of protocol Π, denoted by $R(\Pi)$, is the maximum length of the transcript $\Pi(X_1, \ldots, X_t)$, where the maximum is taken over all input profiles $(X_1, \ldots, X_t)$, and all public coin tosses.

Given $\delta \in [0,1]$, we say that Π is a *δ-error protocol* for f if there exists a function Π_{out} mapping the set of possible transcripts to $\{0,1\}$ such that $\Pr\left[\Pi_{\text{out}}(\Pi(X_1,\ldots,X_t)) = f(X_1,\ldots,X_t)\right] \geq 1-\delta$ for every input profile $(X_1,\ldots,X_t)$; again, the randomness here is over the public coin tosses.

Definition 5 (Randomized Communication Complexity)**.** The *δ-error randomized communication complexity* of f, denoted $R_\delta(f)$, is the randomized communication cost of the best δ-error randomized protocol for f, i.e. $R_\delta(f) = \min_{\Pi\,:\,\Pi \text{ is a } \delta\text{-error protocol for } f} R(\Pi)$.

4.2 Multi-Party Fixed-Size Set Disjointness

Our main tool is to analyze the multi-party fixed-size set disjointness problem [33], which is a refinement of the classic multi-party set disjointness problem.

Definition 6 (Multi-Party (Fixed-Size) Set Disjointness)**.** In the classic *multi-party set disjointness* problem, denoted $\text{DISJ}_{m,t}$, there are t players and a universe of m elements. Each player i holds a subset S_i of the universe. The goal is to determine whether all sets are pairwise disjoint (i.e. $S_i \cap S_j = \emptyset$ for all distinct $i,j \in [t]$). If this is the case in an input, it is referred to as a *NO* instance; otherwise, it is referred to as a *YES* instance.

In multi-party fixed-size set disjointness, denoted $\text{FDISJ}_{m,s,t}$, there is an additional parameter $s \in [m]$ such that the set held by each player has size exactly s (i.e. $|S_i| = s$ for each $i \in [t]$).

Often, set disjointness is studied under a promise, which allows assuming additional structure of YES instances (equivalently, the protocol is free to return any answer on YES instances without this structure).

Definition 7 (Unique Intersection Promise)**.** The classical promise in the literature is the *unique intersection* promise, which guarantees that in every YES instance, there exists an element x such that $x \in S_i$ for each $i \in [t]$ and $(S_i \setminus \{x\}) \cap (S_j \setminus \{x\}) = \emptyset$ for all distinct $i,j \in [t]$.

We refer to such an element x as a *common element.* Mandal et al. [33] show that $R_\delta(\text{FDISJ}_{m,m/t,t}) = \Omega(m/t)$ under the unique intersection promise, and then they give a reduction from multi-party fixed-size set disjointness to voting. Specifically, they use a voting rule f with deterministic elicitation to construct a 0-error protocol for $\text{FDISJ}_{m,\Theta(m/d),\Theta(d)}$ that uses roughly $\Theta(d)\cdot \text{C}(f)$ bits of total elicitation. Then, using $R_\delta(\text{FDISJ}_{m,\Theta(m/d),\Theta(d)}) = \Omega(m/d)$, they derive $\text{C}(f) = \Omega(m/d^2)$.[2] In order to improve the lower bound on the communication complexity of a voting rule, we first show that to $D(\text{FDISJ}_{m,m/t,t}) = \Omega(m)$. The change from randomized communication complexity to its deterministic counterpart is not an issue because the reduction of Mandal et al. [33] eventually constructs a deterministic protocol anyway. The improvement from $\Omega(m/t)$ to $\Omega(m)$ helps shave off a factor of $t = \Theta(d)$. However, this requires moving from the strong unique intersection promise to the weaker *substantial intersection promise*, which we define next.

Definition 8 (Substantial Intersection Promise)**.** We introduce a weaker promise, which we refer to as the *substantial intersection* promise, which guarantees that for a given constant $\gamma > 0$, in every YES instance, there exists at least one element x and a subset of players $P \subseteq [t]$ with $|P| \geq \gamma \cdot t$ such that $x \in S_i$ for each $i \in P$.

[2] The additional factor of d appears because $R_\delta(\text{FDISJ}_{m,s,t})$ measures the *total* communication from all players, whereas $\text{C}(f)$ measures the communication from each player. So a lower bound of $\Omega(m/d)$ on the *total* communication translates to a lower bound of $\Omega(m/d^2)$ on the communication from each of $t = \Theta(d)$ players in their reduction.

Note that unique intersection promise provides a stronger guarantee than substantial intersection promise; hence, any communication complexity lower bound under the former promise immediately implies the same lower bound under the latter promise.

4.3 Lower Bound on Communication Complexity of Voting Rules

We begin by establishing an improved lower bound on the deterministic communication complexity of FDISJ under the weaker substantial intersection promise.

Theorem 3. *Under the substantial intersection promise with $\gamma \le 1/76$ and $t \le (m/2)\cdot(1-1/e)$, we have $D(\mathrm{FDISJ}_{m,m/t,t}) \ge m$.*

Proof. Suppose for contradiction that there exists a deterministic protocol Γ for $\mathrm{FDISJ}_{m,m/t,t}$ with $D(\Gamma) = r < m$. For $i \in [t]$, let $\mathcal{X}_i = \{S_i : S_i \subseteq [m] \wedge |S_i| = m/t\}$ be the collection of possible sets held by player i, and let $\mathcal{X} = \{(S_1,\dots,S_t) : S_i \subseteq [m] \wedge |S_i| = m/t, \forall i \in [t]\}$ be the collection of possible input profiles.

A subset $\mathcal{S} \subseteq \mathcal{X}$ of input profiles is called a rectangle if $\mathcal{S} = \prod_{i\in[t]} \mathcal{S}_i$ where $\mathcal{S}_i \subseteq \mathcal{X}_i$ for all $i \in [t]$. A rectangle $\mathcal{S}$ is monochromatic with respect to deterministic protocol Γ if Γ generates identical transcripts on all input profiles in $\mathcal{S}$. Since we are working with the *blackboard* model and at most r bits are communicated under Γ in the worst case, it is easy to observe that Γ partitions $\mathcal{X}$ into at most 2^r monochromatic rectangles.[3]

Under a NO instance, each of t players hold a disjoint subset of size m/t. Hence, the total number of NO instances is $m!/((m/t)!)^t$. Thus, by the pigeonhole principle, at least one of the monochromatic rectangles (call it $\mathcal{S}^*$) must contain at least $m!/(2^r((m/t)!)^t)$ NO instances. Hence,

$$\begin{aligned}|\mathcal{S}^*| &\ge \frac{m!}{2^r\left((m/t)!\right)^t} > \frac{m^{m+1/2}e^{-m}}{2^m\left(e(m/t)^{m/t+1/2}e^{-m/t}\right)^t} = \frac{m^{m+1/2}e^{-m}}{2^m e^t (m/t)^{m+t/2}e^{-m}} = \frac{m^{1/2}\cdot t^m \cdot t^{t/2}}{2^m\cdot e^t\cdot m^{t/2}}\\ &\ge \frac{t^m}{2^m\cdot e^{m/2}} = \left(\frac{t}{2\sqrt{e}}\right)^m. \end{aligned} \tag{2}$$

Here, the first inequality holds due to Stirling's approximation[4] and the fact that $r < m$, and the second inequality holds because for $t \le (m/2)\cdot(1-1/e)$, we have

$$\frac{t^{t/2}}{e^t m^{t/2}} = \frac{1}{e^t(m/t)^{t/2}} \ge \frac{1}{e^{m/2\cdot(1-1/e)}e^{m/(2e)}} = \frac{1}{e^{m/2}}.$$

Here, we used the fact that $(m/t)^t$ achieves its maximum value at $t = m/e$, and hence is bounded from above by $e^{m/e}$.

Now, let us write the rectangle $\mathcal{S}^* = \prod_{i\in[t]} \mathcal{S}_i^*$, where $\mathcal{S}_i^* \subseteq \mathcal{X}_i$ for each $i \in [t]$. For each $i \in [t]$, let us also define $\mathrm{supp}(\mathcal{S}_i^*) = \bigcup_{S_i : S_i \in \mathcal{S}_i^*} S_i$ to be the set of all elements which appear in at least one set in $\mathcal{S}_i^*$. Then, we have the following upper bound on the number of possible instances in $\mathcal{S}^*$.

$$|\mathcal{S}^*| = \prod_{i\in[t]} \binom{|\mathcal{S}_i^*|}{m/t} \le \prod_{i\in[t]} \left(\frac{|\mathcal{S}_i^*|\, et}{m}\right)^{m/t}. \tag{3}$$

[3]This is a standard argument. Note that Γ can equivalently be described by a binary tree, where each internal node represents one of the players writing the next bit on the blackboard. Since at most r bits are written, the tree has at most 2^r leaves, and each leaf is obtained on a set of input profiles that form a monochromatic rectangle.

[4]For all $n \in \mathbb{N}$, $\frac{n!}{n^{n+1/2}e^{-n}} \in [1, e]$.

We now claim that at least $t/4$ players i must have $|\mathcal{S}_i^*| \geq m/19$. Suppose this is not true. That is, all but at most $t/4$ players i have $|\mathcal{S}_i^*| \leq m/19$, and the remaining at most $t/4$ players i have $|\mathcal{S}_i^*| \leq m$. Then, by Equation (3), we would have

$$|\mathcal{S}^*| \leq \left((et)^{m/t}\right)^{t/4} \cdot \left(\left(\frac{et}{19}\right)^{m/t}\right)^{3t/4} \leq \left(\frac{et}{19^{3/4}}\right)^m < \left(\frac{t}{2\sqrt{e}}\right)^m,$$

which would contradict Equation (2).

Hence, there exist at least $t/4$ players i with $|\text{supp}(\mathcal{S}_i^*)| \geq m/19$. Thus, by the pigeonhole principle, at least one element x^* must appear in $\text{supp}(\mathcal{S}_i^*)$ for at least $t/76 \geq \gamma t$ players i. Selecting a corresponding set from $\mathcal{S}_i^*$ containing x^* for each such player i and an arbitrary set from $\mathcal{S}_i^*$ for each remaining player i generates a YES instance under the substantial intersection promise, and this YES instance also belongs to $\mathcal{S}^*$ which also contains at least one NO instance. Hence, protocol Γ would not be able to distinguish these YES and NO instances, establishing the contradiction. This shows that $D(\text{FDISJ}_{m,m/t,t}) \geq m$. □

We now use Theorem 3 to derive a lower bound on the communication complexity of voting rules with deterministic elicitation.

Theorem 4. *For any d, if voting rule f uses deterministic elicitation and satisfies dist$(f) \leq d$, then $C(f) = \Omega(m/d)$.*

Note that our lower bound of $\Omega(m/d)$ from Theorem 4 applies even when the rule is allowed to use randomized aggregation, whereas Mandal et al. [33] achieve a matching upper bound of $\widetilde{O}(m/d)$ using deterministic aggregation. This establishes that, when using deterministic elicitation, there is no significant asymptotic benefit of using randomized aggregation. This matches our observation for the case of randomized elicitation.

However, such observation applies only when we are considering both the optimal elicitation and optimal aggregation rule. It need not be true when considering a fixed (possibly suboptimal) elicitation method. For example, when each voter sends a ranking of the alternatives by their value for her, it is known that randomized aggregation yields optimal distortion of $\widetilde{\Theta}(\sqrt{m})$ [14], which is significantly lower than the optimal distortion of $\Theta(m^2)$ with the deterministic aggregation [17].

5 k-Selection: Upper Bounds

We now turn to the k-selection problem, where the goal is to select a set $S \subseteq A$ of $|S| = k$ alternatives. Recall that the value that voter i derives from set S is defined as $v_i(S) = \max_{a \in S} v_i(a)$. In Section 5.1 below, we present a deterministic k-selection rule which achieves a distortion of at most d with communication complexity $\widetilde{O}(m/(kd))$. Then, in Section 5.2, we present a randomized k-selection rule with distortion d and communication complexity $\widetilde{O}(m/(kd^3))$. Later, in Section 6, we show that these bounds are almost tight. While these results for the k-selection problem subsume those for the winner selection problem ($k = 1$) from the previous sections, they require much more intricate algorithms and techniques, and have larger logarithmic factors hidden within the $\widetilde{O}$ notation.

In Section 3, we used L_2-sampling to design a new winner selection with randomized elicitation that achieves the optimal trade-off between communication complexity and distortion. Unfortunately, it seems that L_2-sampling (or more generally, L_p-sampling) is not well-suited for the k-selection problem. The difficulty is that in the k-selection problem, a voter's value for a set of alternatives is defined as her *maximum* value for any alternative in the set. This inevitably leads to the structure of a coverage problem, where, after choosing one alternative in the desired set, a good choice of the next alternative crucially depends on the alternative just chosen. It is unlikely to be able to obtain a set of alternatives that *collectively* provide "good coverage" by taking *independent* L_2-samples obtained from a sampler. Hence, we use different techniques to design k-selection rules.

5.1 Deterministic Elicitation

Let us first focus on designing a k-selection rule with deterministic elicitation. The detailed algorithm is provided in the appendix. Here, we discuss the main idea behind the algorithm. We partition the range of possible values that a voter might have for an alternative into $\log m$ exponentially spaced "value-buckets", and observe that there is one bucket such that if we zero-out valuations that are not in that bucket, the optimal set with respect to the modified valuations is still an $O(\log m)$ approximation to the original optimal set. Furthermore, for each voter, we can look at the number of alternatives for which the voter's value is in this bucket. This number can be further placed into $\log m$ exponentially spaced "quantity-buckets". We again show that there exists one quantity bucket such that only considering this bucket further loses only a factor of $O(\log m)$ social welfare. Once we are restricted to a fixed value-bucket and a fixed quantity-bucket, the problem is simple: if the size of the quantity-bucket is large ($\geq \frac{m}{kd}$), then we show that a uniformly random set of size k well-approximates the optimal solution. On the other hand, if the size of the quantity-bucket is small ($\leq \frac{m}{kd}$), then every voter can simply send all the alternatives that are in this bucket for her, and we can estimate the optimal solution. Of course, the algorithm is not aware of the *right* pair of value- and quantity-buckets, so it simply chooses one uniformly at random, further losing at most $O(\log^2 m)$ factor.

Theorem 5. *For $d \geq 144 \log^4 m$, there is a voting rule with deterministic elicitation for the k-selection rule with distortion $O(d)$ and communication complexity $O\left(\frac{m}{kd} \log^6 m\right)$.*

5.2 Randomized Elicitation

We now discuss our k-selection rule with randomized elicitation. Once again, the detailed algorithm is presented in the appendix, but we discuss the high-level idea below. This algorithm is developed in two stages: first, we design an algorithm that works when the number of voters n is polynomially bounded by m (say $n \leq m^4$). Later, we show how the general problem can be reduced to this case.

Suppose $n \leq m^4$. As we did in the case of deterministic elicitation, by losing only logarithmic factors, we can reduce to an instance where all voters have either 0 or a non-zero value for each alternative, the non-zero values are approximately equal (as they are in a single "value-bucket"), and the voters have non-zero value for approximately the same number of alternatives (as they are in a single "quantity-bucket"). If this number is large ($\geq \frac{m}{kd}$), then again a uniformly random set of size k provides a good solution. On the other hand, if this number is small ($\leq \frac{m}{kd^3}$), then voters can just send the alternatives for which they have non-zero values, and we are only using about $\frac{m}{kd^3}$ bits of communication. So the only interesting case is when this number is in the range $[\frac{m}{kd^3}, \frac{m}{kd}]$.

A first attempt would be the following. The voting rule creates a subset of size m/d, in which each of m alternatives is included with probability $1/d$. Then, every voter reports which of these alternatives she has a non-zero value for, and the voting rule finds the optimal set subject to this information. It can be shown that this leads to distortion $O(d)$. However, this protocol requires each voter to send $\widetilde{O}\left(\frac{m}{kd^2}\right)$ bits, as each voter may have non-zero value for up to $\frac{m}{kd}$ alternatives.

This problem can be easily resolved if we allow voters to further sample a smaller set of alternatives on which they place non-zero value, and then take the intersection with the common set, so each element is retained with probability only $\mathrm{poly}(\log m)/d$. However, recall that, while our model allows choosing a randomized query, it does not allow having voters respond randomly to a fixed query (e.g. two voters with the same valuation responding differently to a query). To circumvent this problem, the voting rule can create random seeds and send them to the voters as part of the query for subsampling. Although voters need to subsample independently, the seeds must be anonymous, as the query must be the same for all voters. To circumvent this issue, we "identify" each voter by the set of alternatives on which it places non-zero value. So the voting rule samples a large prime ϕ, and sends it to each voter. Now, voter i interprets her set (on which she places non-zero value) as a number, and computes her ID modulo the prime number ϕ. When the number of voters is small ($\mathrm{poly}(m)$), we show that we can choose a prime large enough so that voters with distinct sets get distinct IDs. Now, the voters use their IDs to read the random seeds provided by the rule, sample a set where each alternative is retained with probability $\mathrm{poly}(\log m)/d$, and send the intersection with the common set back along with the ID they computed for themselves. To keep the communication complexity low, we thus need the ID to be small, which is why we need the prime number ϕ, and in turn, the number of voters n to be $\mathrm{poly}(m)$. This rule has communication complexity $\widetilde{O}\left(\frac{m}{kd^3}\right)$. Our proof shows that its distortion is $O(d)$ by carefully arguing that some alternative from the optimal set survives with a non-negligible probability in the intersection, and will be present enough times in the sets sent by the voter, so we can identify it with high probability.

Theorem 6. *When $n \le m^4$ and $d = \Omega\left(\log^6 m\right)$, there is a voting rule with randomized elicitation for the k-selection rule with distortion $O(d)$ and communication complexity $O\left(\frac{m}{kd^3}\log^{21} m\right)$.*

For $n \ge m^4$, we reduce the problem to $n \le m^4$ by choosing a random subset of m^4 voters and only considering their responses. Using Chernoff bounds, we show that welfare of this subset of voters provides a good approximation of welfare of all voters, with high probability, and thus the optimal set computed from the responses from this subset of voters provides low distortion.

Theorem 7. *When $d = \Omega(\log^6 m)$, there is a voting rule with randomized elicitation for the k-selection rule with distortion $O(d)$ and communication complexity $O\left(\frac{m}{kd^3}\log^{21} m\right)$.*

6 k-Selection: Lower Bounds

In this section, we derive lower bounds on the communication complexity required to achieve a given distortion d in the k-selection problem. For voting rules with deterministic (resp. randomized) elicitation, we establish $\Omega(\frac{m}{kd})$ (resp. $\Omega\left(\frac{m}{kd^3}\right)$) lower bound on the communication complexity; note that both these bounds are tight up to logarithmic factors given our upper bounds from Section 5. To establish the lower bound for deterministic voting rules, we extend the technique introduced in Section 4 for 1-selection to k-selection. To establish the lower bound for randomized voting rules, we use a new embedding technique in Section 6.2.

Before we introduce our lower bounds, we remark that it is also possible to reduce the 1-selection problem to the k-selection problem for a fixed $k > 1$. That is, we show how to use a k-selection voting rule to solve the 1-selection problem, and use a lower bound on the latter to derive a lower bound on the former. Since $k > 1$ is fixed, we cannot trivially *set* $k = 1$ and then run the k-selection rule. Nonetheless, our reduction is simple. However, it produces lower bounds that are weaker by a factor of k compared to the optimal bounds we present below. We still present this reduction in the appendix due to its conceptual novelty; to the best of our knowledge, this is the first reduction from the 1-selection problem to the k-selection problem for a fixed $k > 1$.

6.1 Deterministic Elicitation

We first present a lower bound of $\Omega(\frac{m}{kd})$ on the communication complexity of k-selection rules which achieves distortion at most d using deterministic elicitation. As noted above, this is achieved by using our new lower bound on the total communication complexity of multi-party fixed-size set disjointness problem from Theorem 3, and using a technique similar to that from Theorem 4 to reduce this problem to the k-selection problem.

Theorem 8. *Let f be a k-selection voting rule which uses deterministic elicitation and achieves $dist(f) \leq d$. Then, $C(f) = \Omega(\frac{m}{kd})$.*

6.2 Randomized Elicitation

We now move to presenting a lower bound on the communication complexity $\mathrm{C}(f)$ of a voting rule f which achieves distortion at most d using randomized elicitation. For this, the approach of Theorem 8 unfortunately fails. Specifically, the argument goes through to show that, with probability at least k/t, the voting rule must return a set S containing an element appearing in more than one set. However, the randomness here is no longer only in aggregation, but in elicitation as well. Since our protocol is communication-restricted, we cannot assume direct access to all sets that are returned with probability at least k/t by the voting rule on this instance. This can be handled using a trick similar to what Mandal et al. [33] use in their lower bound proof for randomized elicitation with $k = 1$. We could run the voting rule $\frac{t}{k} \cdot \ln(1/\delta)$ times, and record the generated sets. This ensures that with probability at least $1 - \delta$, at least one of the sets returned contains at least one element appearing in more than one set. Finding such an element results in a δ-error protocol for $\mathrm{FDISJ}_{m,m/t,t}$, which must have communication cost $\Omega(m/t)$. However, the communication cost of this protocol is roughly $\frac{t}{k} \cdot t \cdot \mathrm{C}(f)$, which, using $t = \Theta(kd)$, gives $\mathrm{C}(f) = \Omega\left(\frac{m}{k^2 d^3}\right)$. This bound is a factor of k looser than the bound we want.

Hence, we take a very different approach here. We go back to using FDISJ under the unique intersection promise, for which Mandal et al. [33] show that $R_\delta(\mathrm{FDISJ}_{m,s,t}) = \Omega(s)$ whenever $m \geq \frac{3}{2}st$. However, instead of using a big instance with m elements and $t = \Theta(kd)$ players like they do, we use a small instance with m/k elements and $t = \Theta(d)$ players, and embed it among a set of k instances where the other $k - 1$ are randomly generated YES instances also containing m/k elements and t players. We use a symmetrization trick to ensure that the voting rule cannot distinguish the real instance from the generated ones, and must return its common element with a sufficiently high probability. We then repeat this setup to find the common element of the real instance with probability at least $1 - \delta$, and analyze the communication required. Interestingly, this gives us the tight bound we desire.

Theorem 9. *Let f be a k-selection voting rule which uses randomized elicitation and achieves $dist(f) \leq d$. Then, $C(f) = \Omega\left(\frac{m}{kd^3}\right)$.*

6.3 Discussion of Our Lower Bounds

Given that our matching upper and lower bounds decrease with k, this shows that the k-selection problem becomes strictly easier as k increases. This is not obvious apriori because, while higher values of k allow a voting rule to achieve higher expected social welfare by selecting more alternatives, they also raise the optimal social welfare against which the expected social welfare of the voting rule is compared.

It is worth noting that Caragiannis et al. [17] also study the k-selection problem, but under the specific elicitation rule where each voter provides a ranking of the alternatives. They examine the optimal distortion which can be achieved using any aggregation rule. For deterministic aggregation, their bounds imply that the optimal distortion is roughly inversely proportional to k. For randomized aggregation, their upper bound *increases* as $\sqrt{k}$ up to $k \approx \sqrt[3]{m}$, and then decreases as $1/k$, whereas their lower bound stays constant until $k \approx \sqrt{m}$, and then decreases roughly as $1/k$, leaving open the question of the optimal bound for small k. In contrast, when the communication complexity is kept fixed, Theorems 8 and 9 show that the optimal distortion for k-selection decreases as $1/k$ using deterministic elicitation and as $1/\sqrt[3]{k}$ using randomized elicitation.

7 Discussion & Future Work

Our work leaves open a number of directions for future research. On a technical level, our upper and lower bounds are tight, but only up to logarithmic factors. When the desired distortion is either very small ($d = \text{polylog}(m)$) or very large ($d = m/\text{polylog}(m)$), improving the logarithmic factors and determining the exact communication complexity becomes important. For example, our lower bounds show that to achieve $O(1)$ distortion, $\Omega(m/k)$ communication is necessary in the k-selection problem even with randomized elicitation and aggregation. Can this be achieved using only $O(m/k)$ communication, without any logarithmic factors? Does randomized elicitation offer any asymptotic advantage over deterministic elicitation in the $d = O(1)$ regime?

We also note that for the 1-selection problem with deterministic elicitation, the upper bound of $\widetilde{O}(m/d)$ achieved by Mandal et al. [33] uses deterministic aggregation whereas our matching lower bound of $\Omega(m/d)$ holds even for randomized aggregation, thus establishing that there is no significant advantage of using randomized aggregation over deterministic aggregation. However, for the 1-selection problem with randomized elicitation, and the k-selection problem with deterministic or randomized elicitation, our upper bounds use randomized aggregation. We believe that it should be possible to achieve the same bounds using deterministic aggregation (implying that deterministic aggregation is almost as powerful as randomized aggregation), but leave this for future work.

We note that our lower bounds use a reduction from the set-disjointness problem, where lower bounds are known in a very powerful setup which allows a protocol to use multiple rounds of adaptive elicitation. In contrast, our voting model is defined for voting rules which use a single round of uniform elicitation (where the same query is posed to all voters). Establishing lower bounds for more general forms of elicitation is an interesting direction for future work.

We remark that our results also have implications for other problems studied in voting. One interesting example is *participatory budgeting*, where each alternative has an associated cost, there is a total budget, and the goal is to return the optimal feasible set of alternatives (where feasibility means that the total cost should not exceed the budget). This problem models the real-world process of participatory budgeting, where residents of a city vote over which public projects should be funded. We view designing optimal voting rules for participatory budgeting as a key direction for future work.

Finally, we note that our use of communication complexity (i.e. the number of bits that each voter needs to transmit) can be viewed as an extremely crude measure of the cognitive burden that an elicitation rule imposes on the voters. Quite possibly, humans may find it easier to provide a certain response even if it technically requires transmitting a larger number of bits. The study of more realistic measures of cognitive burden in voting is a challenging direction for future work, which may require interdisciplinary ideas.

Acknowledgments

We thank Ariel D. Procaccia for useful discussions. Mandal was supported in part by the Post-Doctoral fellowship from the Columbia Data Science Institute. Shah was partially supported by the Natural Sciences and Engineering Research Council under a Discovery grant. Woodruff was partially supported by the National Science Foundation under grant CCF-1815840, and part of this work was done while he was visiting the Simons Institute for the Theory of Computing.

References

[1] B. Abramowitz, E. Anshelevich, and W. Zhu. 2019. Awareness of voter passion greatly improves the distortion of metric social choice. In *Proceedings of the 15th Conference on Web and Internet Economics (WINE)*. 3–16.

[2] G. Amanatidis, G. Birmpas, A. Filos-Ratsikas, and A. A. Voudouris. 2020. Peeking Behind the Ordinal Curtain: Improving Distortion via Cardinal Queries. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*. Forthcoming.

[3] A. Andoni, R. Krauthgamer, and K. Onak. 2011. Streaming algorithms via precision sampling. In *Proceedings of the 52nd IEEE Annual Symposium on Foundations of Computer Science*. 363–372.

[4] E. Anshelevich, O. Bhardwaj, E. Elkind, J. Postl, and P. Skowron. 2018. Approximating optimal social choice under metric preferences. *Artificial Intelligence* 264 (2018), 27–51.

[5] E. Anshelevich and J. Postl. 2017. Randomized social choice functions under metric preferences. *Journal of Artificial Intelligence Research* 58 (2017), 797–827.

[6] H. Aziz. 2020. Justifications of Welfare Guarantees under Normalized Utilities. *ACM SIGecom Exchanges* 17, 2 (2020), 71–75.

[7] A. Badanidiyuru, S. Dobzinski, H. Fu, R. Kleinberg, N. Nisan, and T. Roughgarden. 2012. Sketching valuation functions. In *Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 1025–1035.

[8] M.-F. Balcan and N. J. Harvey. 2018. Submodular Functions: Learnability, Structure, and Optimization. *SIAM Journal on Computing* 47, 3 (2018), 703–754.

[9] G. Benadè, S. Nath, A. D. Procaccia, and N. Shah. 2017. Preference Elicitation for Participatory Budgeting. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*. 376–382.

[10] G. Benadè, A. D. Procaccia, and M. Qiao. 2019. Low-Distortion Social Welfare Functions. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI)*. 1788–1795.

[11] J. Bentham. 1996. *The Collected Works of Jeremy Bentham: An Introduction to the Principles of Morals and Legislation*. Clarendon Press.

[12] U. Bhaskar, V. Dani, and A. Ghosh. 2018. Truthful and Near-Optimal Mechanisms for Welfare Maximization in Multi-Winner Elections. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*. 925–932.

[13] A. Borodin, O. Lev, N. Shah, and T. Strangway. 2019. Primarily about Primaries. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI)*. 1804–1811.

[14] C. Boutilier, I. Caragiannis, S. Haber, T. Lu, A. D. Procaccia, and O. Sheffet. 2015. Optimal Social Choice Functions: A Utilitarian View. *Artificial Intelligence* 227 (2015), 190–213.

[15] S. J. Brams and P. C. Fishburn. 2002. Voting Procedures. In *Handbook of Social Choice and Welfare*. 173–236.

[16] F. Brandt, V. Conitzer, U. Endress, J. Lang, and A. D. Procaccia (Eds.). 2016. *Handbook of Computational Social Choice*. Cambridge University Press.

[17] I. Caragiannis, S. Nath, A. D. Procaccia, and N. Shah. 2017. Subset Selection Via Implicit Utilitarian Voting. *Journal of Artificial Intelligence Research* 58 (2017), 123–152.

[18] I. Caragiannis and A. D. Procaccia. 2011. Voting Almost Maximizes Social Welfare Despite Limited Communication. *Artificial Intelligence* 175, 9–10 (2011), 1655–1671.

[19] J. R. Chamberlin and P. N. Courant. 1983. Representative Deliberations and Representative Decisions: Proportional Representation and the Borda Rule. *American Political Science Review* 77, 3 (1983), 718–733.

[20] M. Charikar, K. Chen, and M. Farach-Colton. 2002. Finding frequent items in data streams. In *Proceedings of the 29th International Colloquium on Automata, Languages and Programming (ICALP)*. 693–703.

[21] M. de Condorcet. 1785. Essai sur l'application de l'analyse à la probabilité de décisions rendues à la pluralité de voix. Imprimerie Royal. Facsimile published in 1972 by Chelsea Publishing Company, New York.

[22] E. Elkind, P. Faliszewski, P. Skowron, and A. Slinko. 2017. Properties of Multiwinner Voting Rules. *Social Choice and Welfare* 48, 3 (2017), 599–632.

[23] P. Faliszewski, P. Skowron, A. Slinko, and N. Talmon. 2017. Multiwinner Voting: A New Challenge for Social Choice Theory. *Trends in Computational Social Choice* 74 (2017), 27–47.

[24] V. Gkatzelis, D. Halpern, and N. Shah. 2020. Resolving the Optimal Metric Distortion Conjecture. arXiv:2004.07447.

[25] A. Goel, A. K. Krishnaswamy, and K. Munagala. 2017. Metric distortion of social choice rules: Lower bounds and fairness properties. In *Proceedings of the 18th ACM Conference on Economics and Computation (EC)*. 287–304.

[26] Godrey H Hardy. 1917. On the normal number of prime factors of a number n. *Quart. J. Math. Oxford Ser.* 48 (1917), 76–92.

[27] J. C. Harsanyi. 1955. Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility. *Journal of Political Economy* 63, 4 (1955), 309–321.

[28] J. D. Hartline. 2012. Approximation in Economic Design. Book in preparation.

[29] C. Hillinger. 2005. The Case for Utilitarian Voting. *Homo Oeconomicus* 22, 3 (2005).

[30] R. Jayaram and D. P. Woodruff. 2018. Perfect L_p Sampling in a Data Stream. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 544–555.

[31] H. Jowhari, M. Sağlam, and G. Tardos. 2011. Tight Bounds for Lp Samplers, Finding Duplicates in Streams, and Related Problems. In *Proceedings of the 30th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*. 49–58.

[32] E. Koutsoupias and C. Papadimitriou. 1999. Worst-Case Equilibria. In *Annual Symposium on Theoretical Aspects of Computer Science*. 404–413.

[33] D. Mandal, A. D. Procaccia, N. Shah, and D. Woodruff. 2019. Efficient and Thrifty Voting by Any Means Necessary. In *Proceedings of the 33rd Annual Conference on Neural Information Processing Systems (NeurIPS)*. 7178–7189.

[34] M. Monemizadeh and D. P. Woodruff. 2010. 1-Pass Relative-Error Lp-Sampling with Applications. In *Proceedings of the 21st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 1143–1160.

[35] B. L. Monroe. 1995. Fully Proportional Representation. *American Political Science Review* 89, 4 (1995), 925–940.

[36] K. Munagala and K. Wang. 2019. Improved Metric Distortion for Deterministic Social Choice Rules. In *Proceedings of the 20th ACM Conference on Economics and Computation (EC)*. 245–262.

[37] N. Nisan and A. Ronen. 2001. Algorithmic Mechanism Design. *Games and Economic Behavior* 35, 1–2 (2001), 166–196.

[38] A. D. Procaccia and J. S. Rosenschein. 2006. The Distortion of Cardinal Preferences in Voting. In *Proceedings of the 10th International Workshop on Cooperative Information Agents (CIA)*. 317–331.

[39] A. D. Procaccia and M. Tennenholtz. 2009. Approximate mechanism design without money. In *Proceedings of the 10th ACM Conference on Economics and Computation (EC)*. 177–186.

Appendix

A Missing Algorithms from Section 3

The multi-agent version of the L_2-sampler given in Algorithm 1 and the winner selection rule Algorithm 2 with randomized elicitation that we designed based on it use two procedures named COMMUNICATE and SAMPLE. Here, we provide a formal description of these procedures.

ALGORITHM 3: COMMUNICATE

Input: $\{x_i\}_{i=1}^n$, δ.

1. Set $d = \Theta(\log m)$, $\eta = 1/\sqrt{\log m}$, and $\mu \sim \text{Unif}[1/2, 3/2]$. Let c and c' be sufficiently large constants.
2. Generate four independent hash tables and m^c i.i.d. random variables as follows.
   a. Hash table A_1 of dimension $d \times 6/\eta^2$ used for selecting an alternative.
   b. Hash tables A_2 and A_3 of dimensions $c' \log(1/\delta) \times O(1)$ used for L_2-norm estimation.
   c. Hash table A_4 of dimension $c' \log(1/\delta) \times O(1)$ used for total frequency estimation.
   d. I.i.d. exponential random variables t_j for $j \in [m^c]$ used for scaling valuations.
3. Each agent i performs the following computation.
   a. Duplicate x_i to get $X_i \in \mathbb{R}^{m^c}$.
   b. Scale X_i as $\zeta_i(j) = X_i(j)/t_j$ for $j \in [m^c]$.
   c. Run count-sketches using hash tables A_1, A_2, A_3, and A_4 on duplicated vector X_i, and the scaled vector ζ_i to get $A_{i,1}$, $A_{i,2}$, $A_{i,3}$, and $A_{i,4}$.
4. Each agent i returns $A(x_i) = (A_{i,1}, A_{i,2}, A_{i,3}, A_{i,4})$ for $i \in [n]$ to the center.

B Algorithms for k-Selection

We now provide the aggregation and elicitation rules that achieve the optimal distortion for the k-selection problem.

B.1 Deterministic Elicitation

Algorithm 5 describes the voting rule for the k-selection problem with deterministic elicitation.

ALGORITHM 4: SAMPLE

Input: $\{A_r\}_{r=1}^4$.

1. Get an approximation y of $\zeta = \sum_i \zeta_i$, where $y_j = \text{med}_{r\in[d]}\, g_{1,r}(j) A_{1,r,h_r^1(j)}$ for each $j \in [m^c]$.
2. Get an approximation of $\|X\|_2$ (where $X = \sum_{i\in N} X_i$) by $R = \text{med}_{r\in[c'\log(1/\delta)]}\|A_2[r,\cdot]\|_2$.
3. Get an approximation of $\|\zeta\|_2$ by $R' = \text{med}_{r\in[c'\log(1/\delta)]}\|A_3[r,\cdot]\|_2$.
4. Let $y_{(k)}$ denote the k^{th} largest value in y.
5. IF: $y_{(1)} - y_{(2)} < 100\mu\eta R + \eta R'$ OR $y_{(2)} < 50\eta\mu R$:
   - RETURN (FAIL,$\emptyset$).
6. ELSE:
   - Let $\widehat{j} \in \arg\max_{j\in[m^c]} y_j$. Let j^* be the corresponding non-duplicated index.
   - Get an approximation $\widetilde{y}$ of ζ, where $\widetilde{y}_j = \text{med}_{r\in[d]}\, g_{4,r}(j) A_{4,r,h_r(j)}$ for each $j \in [m^c]$.
   - Compute an approximation of $x(j^*)$: $\widehat{x}(j^*) = t_{\widehat{j}} \times \widetilde{y}_{\widehat{j}}$.
   - RETURN (SUCCESS, $(j^*, \widehat{x}(j^*))$).

ALGORITHM 5: k-Selection (Deterministic Elicitation)

Elicitation Rule:

1. Set $t = d/(144\log^4 m)$.
2. Partition the interval $[0,1]$ into $2\log m + 1$ buckets: $B_0 = [0, \frac{1}{m^2})$, $B_j = [\frac{2^{j-1}}{m^2}, \frac{2^j}{m^2}]$ for $j = 1,\ldots,2\log m$.
3. Partition the set of integers $[m]\cup\{0\}$ into $\log m + 1$ buckets: $C_0 = \{0\}$, $C_j = \{2^{j-1},\ldots,2^j\}$ for $j = 1,\ldots,\log m$.
4. Let $q_s = \lfloor \log \frac{m}{tk} \rfloor$. Note that this is the largest q such that the upper end of C_q is at most $\frac{m}{tk}$.
5. For each $p \in \{0,1,\ldots,2\log m\}$ and $q \in \{0,1,\ldots,q_s\}$:
   (a) Voter i constructs valuation V_i^p such that for each $a \in A$, $V_i^p(a)$ is the upper endpoint of the interval B_p if $v_i(a) \in B_p$ and 0 otherwise.
   (b) Send $S_i^{pq} = \begin{cases} \{a : v_i(a) \in B_p\} & \text{if } |\{a : v_i(a) \in B_p\}| \in C_q \\ \emptyset & \text{o.w.} \end{cases}$

Aggregation Rule:

- For each $p \in \{0,1,\ldots,2\log m\}$ and $q \in \{0,1,\ldots,\log m\}$:
  - IF: $q > q_s$, then choose $\widehat{A}_{pq}$ to be a uniformly random subset of k alternatives, i.e., uniformly randomly from $\{S \subseteq [m] : |S| = k\}$.
  - ELSE: Obtain S_i^{pq} from each voter $i \in [n]$. Choose $\widehat{A}_{pq} \in \arg\max_{S\subseteq A:|S|=k} \sum_{i=1}^n \mathbb{1}\{S_i^{pq} \cap S \neq \emptyset\}$.
- Return one of the $(1+2\log m)\times(1+\log m)$ subsets $\{\widehat{A}_{pq}\}_{p\in\{0,1,\ldots,2\log m\},q\in\{0,1,\ldots,\log m\}}$ uniformly at random.

B.2 Randomized Elicitation

We now turn to the k-selection problem with randomized elicitation. First we consider the case when the number of agents $n \leq m^4$. Algorithms 6 and 7 respectively provide the elicitation and aggregation rule for the k-selection problem with randomized elicitation. Finally algorithm 8 provides the voting rule for general n by reducing the k-selection problem to the problem with $n \leq m^4$.

ALGORITHM 6: k-Selection Upper Bound (Elicitation Rule): For $n \leq m^4$

Elicitation Rule:

1. Set $t = \frac{d}{\log^6 m}$.
2. Partition the interval $[0, 1]$ into $2\log m + 1$ buckets: $B_0 = [0, \frac{1}{m^2})$, $B_j = [\frac{2^{j-1}}{m^2}, \frac{2^j}{m^2}]$ for $j = 1, \ldots, 2\log m$.
3. Partition the set of integers $[m] \cup \{0\}$ into $\log m + 1$ buckets: $C_0 = \{0\}$, $C_j = \{2^{j-1}, \ldots, 2^j\}$ for $j = 1, \ldots, \log m$.
4. Let $q_s = \lfloor \log \frac{m}{t^3 k} \rfloor$. Note that this is the largest q such that the upper end of C_q is at most $\frac{m}{t^3 k}$.
5. Let $q_r = \lfloor \log \frac{m}{tk} \rfloor$. Note that this is the largest q such that the upper end of C_q is at most $\frac{m}{tk}$.
6. For $p \in \{0, 1, \ldots, 2\log m\}$, and $q \in \{0, 1, \ldots, \log m\}$:
   (a) Voter i constructs valuation V_i^p such that for each $a \in A$, $V_i^p(a)$ is the upper endpoint of the interval B_p if $v_i(a) \in B_p$ and 0 otherwise.
   (b) Set $S_i^{pq} = \begin{cases} \{a : v_i(a) \in B_p\} & \text{if } |\{a : v_i(a) \in B_p\}| \in C_q \\ \emptyset & \text{o.w.} \end{cases}$
   (c) IF: $q \leq q_s$, send S_i^{pq} to the coordinator.
   (d) ELSE IF: $q \leq q_r$:
      i. Construct a random subset $S \subseteq [m]$ by selecting each element of $\{1, \ldots, m\}$ iid with probability $1/t$.
      ii. Let ϕ be a random prime in the range $[m^{20}]$.
      iii. For each value $r \in \{0, 1, \ldots, p - 1\}$
         - Let $f(r)$ be an independent collection of random bits that given a set of size at most $\frac{10m\log m}{kt^2}$ selects each of its elements uniformly at random with probability $\frac{\log^2 m}{t}$.
      iv. The elicitation rule copies ϕ, S, and $f(r)$ for $r \in \{0, 1, \ldots, \phi - 1\}$ on the "public tape" as part of the query.
      v. Voter i computes $x_i^{pq} = s_i^{pq} \bmod \phi$, where s_i^{pq} is the integer corresponding to the set S_i^{pq}.
      vi. Voter i computes $\widehat{S}_i^{pq} = S_i^{pq} \cap S$.
      vii. Voter i uses $f(x_i^{pq})$ to sample a subset $\widetilde{S}_i^{pq}$ which includes each element of $\widehat{S}_i^{pq}$ with probability $\frac{\log^2 m}{t}$.
      viii. Voter i sends back x_i^{pq} and $\widetilde{S}_i^{pq}$ to the coordinator.

ALGORITHM 7: k-Selection Upper Bound (Aggregation Rule): For $n \le m^4$

Aggregation Rule:

1. Set $t = \frac{d}{\log^6 m}$.
2. For each $p \in [1 + \log m]$ and $q \in [\log m]$
   (a) IF: $q \le q_s$,
      i. Obtain $\widetilde{S}_i^{pq}$, $i \in [n]$, which are sent by the voters.
      ii. Set $\widehat{A}_{pq} \in \arg\max_{S \subseteq A : |S| = k} \sum_{i=1}^{n} \mathbf{1}\{S_i^{pq} \cap S \neq \emptyset\}$.
   (b) ELSE IF: $q \le q_r$,
      i. Obtain $\widetilde{S}_i^{pq}, x_i^{pq}$, $i \in [n]$, which are sent by the voters.
      ii. Partition $[m^4]$ into $4 \log m$ bins $D_j = \{2^{j-1}, \ldots, 2^j\}$ for $j = 1, \ldots, 4 \log m$.
      iii. For every $j \in \{0, 1, \ldots, \phi - 1\}$ compute frequency $f_j = |\{i : x_i^{pq} = j\}|$.
      iv. For each $j \in \{0, 1, \ldots, \phi - 1\}$, round f_j to the smallest multiple of two that is at least f_j.
      v. For each frequency f in $\{1, 2, 2^2, \ldots, 2^{4 \log m}\}$:
         A. Let U_f be the set of distinct subsets with frequency f.
         B. Set $\widehat{A}_1^{fpq}$ to be set of size k that covers at least k subsets in U_f.
         C. Let B_f^{pq} be the set of alternatives in S with number of occurrences at least $\log^2 m$.
         D. Set $\widehat{A}_2^{fpq}$ to be the k most frequent elements from B_f^{pq}.
         E. $\widehat{A}^{fpq} = \begin{cases} \widehat{A}_1^{fpq} & \text{w.p. } 1/2 \\ \widehat{A}_2^{fpq} & \text{w.p. } 1/2 \end{cases}$
      vi. Set $\widehat{A}^{pq}$ to be one of the $4 \log m$ sets $\{\widehat{A}^{fpq}\}_{f \in [4 \log m]}$ uniformly at random.
   (c) ELSE: Set $\widehat{A}_{pq}$ to be a uniform random set of size k.
3. Return one of the $(1 + 2\log m) \times (1 + \log m)$ subsets $\{\widehat{A}_{pq}\}_{p \in \{0,1,\ldots,2\log m\}, q \in \{0,1,\ldots,\log m\}}$ uniformly at random.

C Missing Proofs

C.1 Proof of Theorem 1

Proof. We note that Algorithm 1 effectively runs the perfect L_2-sampler of Jayaram and Woodruff [30] because their sketches are linear, and therefore sketches of x_i from all agents i can be merged into sketches of x by a simple summation. The first condition is then a direct implication of Theorem 3 by Jayaram and Woodruff [30] (which essentially states that their L_2-sampler is perfect). The second condition is a direct implication of Theorem 4 by Jayaram and Woodruff [30], who show that their algorithm can output a $1 \pm \epsilon$ approximation of the actual frequency $x(j)$ when $j^* = j$. In our case, we do not need arbitrarily good approximations. Simply setting $\epsilon = 1/2$ suffices for our purpose.

For the communication, we note that each entry of the count-sketch table A_i^1 is computed by summing at most $6/\eta^2$ elements. Since each element is a multiple of $1/\Delta$, this guarantees that each element of A_i^1 can be represented using $\log(6/(\eta^2\Delta))$ bits. Therefore, the total number of bits sent by each agent for the first sketch is $O(d \times 6/\eta^2 \times \log(6/(\eta^2\Delta))) = O(\log^2 m \log(\Delta \log m))$. This step ignores the fact that the scaling random variables are exponential and can be any real number.

ALGORITHM 8: k-Selection Upper Bound: For $n \geq m^4$

Elicitation Rule:

- Run the elicitation rule given by Algorithm 6 for distortion d and get voter responses $\{e_i\}_{i=1}^n$.

Aggregation Rule:

- Sample a subset $T \subseteq [n]$ of size m^4 uniformly at random.
- Run aggregation rule given by Algorithm 7 on voter responses $\{e_i\}_{i \in T}$ with distortion parameter d and return the obtained subset of k alternatives.

However, this can be easily resolved by first rounding the exponential variables to the nearest power of $(1+\nu)$ and then scaling with them. This step introduces a relative error of $O(\nu)$ in the sampling. We then choose $\nu = O(m^{-c})$ so that the relative error is consumed by the additive error of $O(m^{-c})$. Additionally, this choice of ν does not asymptotically change the number of bits needed to send the sketch. By a similar argument, it follows that each agent i needs to send $O(\log^2 m \log(\Delta \log m))$ bits each for the remaining three sketches A_i^2, A_i^3, and A_i^4 as well. □

C.2 Proof of Lemma 1

Proof.

$$\begin{aligned}
\mathbb{E}[\mathrm{sw}(\widehat{a})] &\geq \mathbb{E}[\mathrm{sw}^{\Delta}(\widehat{a})] - \frac{n}{\Delta} \\
&\geq \frac{1}{2} \cdot \mathbb{E}[\widehat{\mathrm{sw}}(\widehat{a})] - \frac{n}{\Delta} \\
&= \frac{1}{2t} \cdot \sum_{k=1}^{t} \mathbb{E}[\widehat{\mathrm{sw}}(a^k)] - \frac{n}{\Delta} \\
&\geq \frac{1}{4t} \cdot \sum_{k=1}^{t} \mathbb{E}[\mathrm{sw}^{\Delta}(a^k)] - \frac{n}{\Delta} \\
&= \frac{1}{4t} \cdot \sum_{k=1}^{t} \sum_{a} \Pr[a^k = a] \cdot \mathrm{sw}^{\Delta}(a) - \frac{n}{\Delta} \\
&\geq \frac{1}{4t} \cdot \sum_{k=1}^{t} \Pr[a^k = a^*] \cdot \mathrm{sw}^{\Delta}(a^*) - \frac{n}{\Delta} \\
&\geq \frac{1}{4} \cdot \left(\frac{\mathrm{sw}^{\Delta}(a^*)^3}{\|\mathrm{sw}^{\Delta}\|_2^2} - O(m^{-c}) \right) - \frac{n}{\Delta}
\end{aligned}$$

Here, the third inequality follows because $\widehat{a}$ maximizes the estimated welfare among all a^k, and the last equality follows from the definition of the perfect L_2-sampler.

Next, note that $\mathrm{sw}(a^*) \geq n/m$ by the pigeonhole principle because the total of social welfare of all alternatives is n (recall that our valuations are normalized). Hence,

$$\mathrm{sw}^{\Delta}(a^*) \geq \mathrm{sw}(a^*) - \frac{n}{\Delta} \geq \frac{n}{m} - \frac{n}{\Delta} \geq \frac{n}{2m},$$

where the last inequality holds because $\Delta = 128m^3 \geq 2m$.

Further, we also have

$$
\begin{aligned}
\|\mathrm{sw}^{\Delta}\|_2^2 &= \sum_a \mathrm{sw}^{\Delta}(a)^2 \le \sum_a (\mathrm{sw}(a) + n/\Delta)^2 \\
&= \sum_a \left(\mathrm{sw}^2(a) + 2\mathrm{sw}(a)\cdot n/\Delta + n^2/\Delta^2\right) \\
&\le \sum_a v(a) + (2n/\Delta)\sum_a v(a) + n^2 m/\Delta^2 && (\because v(a) \le 1) \\
&= n + 2n^2/\Delta + n^2 m/\Delta^2 \\
&\le n + \frac{2n^2 m}{\Delta} && (\because m \ge 2 \text{ and } \Delta > 1) \\
&\le 2n^2 && (\because \Delta = 128m^3 \ge 2m).
\end{aligned}
$$

Plugging these into the lower bound for $\mathbb{E}[\mathrm{sw}(\widehat{a})]$, we get

$$
\mathbb{E}[\mathrm{sw}(\widehat{a})] \ge \frac{1}{4}\frac{(n/2m)^3}{2n^2} - O(m^{-c}) - \frac{n}{\Delta} = \frac{n}{64m^3} - O(m^{-c}) - \frac{n}{128m^3}.
$$

We can set a sufficiently large c to make $O(m^{-c})$ at most $1/(256m^3)$, which would give the desired lower bound. □

C.3 Proof of Theorem 4

Proof. Set $\gamma = 1/76$. If $d > (\gamma m/4)\cdot(1-1/e)$, then the lower bound trivially holds. Hence, assume that $d \le (\gamma m/4)\cdot(1-1/e)$.

Consider a voting rule with deterministic elicitation rule Π_f, possibly randomized aggregation rule Γ_f, and distortion $\mathrm{dist}(f) \le d$. We use f to construct a deterministic protocol for $\mathrm{FDISJ}_{m,s,t}$, where $t = 2d/\gamma$ and $s = m/t$, under the substantial intersection promise with parameter γ. Note that for these choices of t and s, Theorem 3 shows that $D(\mathrm{FDISJ}_{m,s,t}) \ge m$.

Consider an input profile $(S_1,\dots,S_t)$ of $\mathrm{FDISJ}_{m,s,t}$ with a universe U of size m and substantial intersection promise with parameter γ. Let us create an instance of the voting problem with a set of n voters N and a set of m alternatives A. Each alternative in A corresponds to a unique element of U. Partition the set of voters N into t equal-size buckets: for $i \in [t]$, bucket N_i consists of n/t voters corresponding to player i, each of whom has valuation v^{S_i} given by $v^{S_i}(a) = 1/s$ for each $a \in S_i$ and $v^{S_i}(a) = 0$ for each $a \notin S_i$. Let $\vec{v}$ denote the resulting profile of voter valuations.

Under these valuations, for each alternative a, we have $\mathrm{sw}(a,\vec{v}) = \frac{n}{ts}\sum_{i=1}^{t}\mathbb{1}[a \in S_i]$, where $\mathbb{1}$ is the indicator variable. Due to the substantial intersection promise, every YES instance admits at least one alternative a^* that appears in at least γt sets and has $\mathrm{sw}(a^*,\vec{v}) \ge \frac{n}{ts}\cdot\gamma t$. Hence, when f is run on this voting instance, to achieve distortion at most $d = \gamma t/2$, the voting rule must return a random alternative a with $\mathbb{E}[\mathrm{sw}(a,\vec{v})] \ge \frac{2n}{ts}$.

In contrast, note that an alternative a which appears in at most one set has social welfare $\mathrm{sw}(a,\vec{v}) \le \frac{n}{ts}$. Hence, the voting rule can only return alternatives appearing at most once with probability at most $1-1/t$. Observe that if this probability were more than $1-1/t$, then the expected social welfare of the rule would have been less than $(1/t)\cdot\frac{n}{s} + 1\cdot\frac{n}{ts} = \frac{2n}{ts}$, which would contradict the bound obtained above.

Thus, we have established that on every YES instance, f returns an alternative that appears in more than one set with probability at least $1/t$. We are now ready to construct a deterministic protocol for $\mathrm{FDISJ}_{m,m/t,t}$.

The protocol runs the voting rule f on the voting instance constructed above. That is, each player i responds to the query posed by the elicitation rule of f according to valuation v^{S_i}. Note that this requires a total of $t \cdot \mathrm{C}(f)$ bits of communication from the players.

Next, we take players' responses, create n/t copies of the response of each player, pass the resulting profile as input to the aggregation rule of f, and obtain the distribution over alternatives that it returns. Next, we let B be the set of all alternatives which are returned with probability at least $1/t$. Note that $|B| \leq t$.[5] From the argument above, we know that in every YES instance, there exists an alternative $x^* \in B$ that appears in at least two sets. In contrast, in every NO instance, every alternative in B appears in at most one set.

Next, recall that we are in the blackboard model. Hence, each player can see the full transcript and compute this set B without any communication from the center. Next, player 1 writes $S_1 \cap B$ on the blackboard. For $i = 2, \ldots, t$, player i first checks the set of alternatives written on the blackboard against her own set S_i. If any of those alternatives (which must be in the set of at least one other player) appear in her set, then she writes 'YES' on the blackboard, indicating that the instance is necessarily a YES instance. Otherwise, she adds $S_i \cap B$ to the blackboard.

Recall that in a YES instance, at least one alternative appears in more than one set. Hence, the second player (by the index) containing that alternative must see it written on the blackboard, and correctly write 'YES' on the blackboard. In a NO instance, no player would ever write 'YES' on the blackboard. This allows us to distinguish between YES and NO instances. Also, note that the total communication in this round is $O(t)$ because until a player writes 'YES', the alternatives written on the blackboard are distinct alternatives of B.

Thus, the total communication complexity of this protocol is at most $t \cdot \mathrm{C}(f) + O(t)$, which must be at least m by Theorem 3. Hence, we get that $\mathrm{C}(f) = \Omega(m/t) = \Omega(m/d)$. □

C.4 Proof of Theorem 5

Proof. Let $A^* \in \arg\max_{S \subseteq A : |S| = k} \mathrm{sw}(S, \vec{v})$ denote an optimal set of size k. For $p \in \{0, 1, \ldots, 2\log m\}$, recall that V_i^p is the valuation where $V_i^p(a)$ is the upper endpoint of bucket B_p if $v_i(a) \in B_p$ and 0 otherwise. Let $\vec{V}^p = \{V_i^p\}_{i \in N}$ denote the corresponding valuation profile. Let $A_p^* \in \arg\max_{S \subseteq A : |S| = k} \mathrm{sw}(S, \vec{V}^p)$ be the optimal subset of k alternatives with respect to the valuation profile $\vec{V}^p$. Then,

$$\begin{aligned}
\sum_{p=0}^{2\log m} \mathrm{sw}(A_p^*, \vec{V}^p) &\geq \sum_{p=0}^{2\log m} \mathrm{sw}(A^*, \vec{V}^p) \\
&= \sum_{i=1}^{n} \max_{a \in A^*} \frac{1}{m^2} \mathbb{1}\{v_i(a) \in B_0\} + \sum_{p=1}^{2\log m} \sum_{i=1}^{n} \max_{a \in A^*} \frac{2^p}{m^2} \mathbb{1}\{v_i(a) \in B_p\} \\
&= \sum_{i=1}^{n} \frac{1}{m^2} \max_{a \in A^*} \mathbb{1}\{v_i(a) \in B_0\} + \sum_{i=1}^{n} \sum_{p=1}^{2\log m} \frac{2^p}{m^2} \max_{a \in A^*} \mathbb{1}\{v_i(a) \in B_p\} \\
&\geq \sum_{i=1}^{n} \max_{a \in A^*} v_i(a) = \mathrm{sw}(A^*, \vec{v}).
\end{aligned}$$

[5] If we do not have access to the exact distribution returned by the aggregation rule, and only to a sampler that samples a random alternative from this distribution, we can take infinitely many samples and precisely estimate the distribution since we are not computationally bounded.

The first inequality follows because A_p^* is the optimal subset with respect to the profile $\vec{V}^p$. In order to see why the inequality on the last line is true, suppose for a voter i the maximum is attained at $a' \in A^*$ and $v_i(a') \in B_{p'}$. If $p = 0$, $v_i(a') \leq 1/m^2$, and if $p \geq 1$, $v_i(a') \leq 2^p/m^2$.

This implies that there exists $p \in \{0, 1, \dots, 2\log m\}$ for which

$$\text{sw}(A_p^*, \vec{V}^p) \geq \frac{1}{1 + 2\log m} \cdot \text{sw}(A^*, \vec{v}). \tag{4}$$

Fix this value of p. Next, for $q \in \{0, 1, \dots, \log m\}$, define the valuation profile $\vec{V}^{pq}$, where for each $i \in N$,

$$V_i^{pq} = \begin{cases} V_i^0 & \text{if } \left|\{a : V_i^0(a) = \frac{1}{m^2}\}\right| \in C_q \text{ and } p = 0 \\ V_i^p & \text{if } \left|\{a : V_i^p(a) = \frac{2^p}{m^2}\}\right| \in C_q \text{ and } p \geq 1 \\ 0 & \text{o.w.} \end{cases}$$

Let $A_{pq}^* \in \arg\max_{S \subseteq A : |S| = k} \text{sw}(S, \vec{V}^{pq})$ be an optimal subset of k alternatives with respect to $\vec{V}^{pq}$. Then,

$$\sum_{q=0}^{\log m} \text{sw}(A_{pq}^*, \vec{V}^{pq}) \geq \sum_{q=0}^{\log m} \text{sw}(A_p^*, \vec{V}^{pq}) = \sum_{i=1}^{n} \sum_{q=0}^{\log m} \max_{a \in A_p^*} V_i^{pq}(a) \geq \sum_{i=1}^{n} \max_{a \in A_p^*} V_i^p(a) = \text{sw}(A_p^*, \vec{V}^p).$$

The first inequality is true because A_{pq}^* is the optimal set with respect to $\vec{V}^{pq}$. In order to see why the final inequality is true, suppose $p \neq 0$ and let $C_{q(i)}$ be the bucket such that the number of alternatives with positive valuation i.e. $\left|\{a : V_i^p(a) = \frac{2^p}{m^2}\}\right|$ is in $C_{q(i)}$. Then $\sum_q \max_{a \in A_p^*} V_i^{pq}(a) \geq \max_{a \in A_p^*} V_i^{pq(i)}(a) = \max_{a \in A_p^*} V_i^p(a)$. A similar argument holds for $p = 0$.

This implies that there exists $q \in \{0, 1, \dots, \log m\}$ such that

$$\text{sw}(A_{pq}^*, \vec{V}^{pq}) \geq \frac{1}{1 + \log m} \cdot \text{sw}(A_p^*, \vec{V}^p). \tag{5}$$

Using Equations (4) and (5), we see that there exists a pair $(p, q) \in \{0, 1, \dots, 2\log m\} \times \{0, 1, \dots, \log m\}$ for which

$$\text{sw}(A_{pq}^*, \vec{V}^{pq}) \geq \frac{\text{sw}(A^*, \vec{v})}{(1 + \log m)(1 + 2\log m)} \geq \frac{\text{sw}(A^*, \vec{v})}{6\log^2 m}. \tag{6}$$

Our goal is to show that the set $\widehat{A}_{pq}$ selected by our rule well-approximates the optimal set A_{pq}^* with respect to valuation profile $\vec{V}^{pq}$. However, the actual welfare we are interested in is $\text{sw}(\widehat{A}_{pq}, \vec{v})$ with respect to valuation profile $\vec{v}$. To close the gap, we show that for any set S, its social welfare with respect to $\vec{v}$ is lower bounded by an expression involving its social welfare with respect to $\vec{V}^{pq}$. Fix any $S \subseteq A$ with $|S| = k$. Let B_p^u denote the upper endpoint of bucket B_p. For each voter $i \in N$, note that $V_i^{pq}(a)$ is either $V_i^p(a)$ or 0, and in turn, $V_i^p(a)$ is either B_p^u (when $v_i(a) \in B_p$) or 0. Further, when $v_i(a) \in B_p$, we have $v_i(a) \geq \frac{B_p^u}{2} - \frac{1}{2m^2}$ (note that this applies even for $p = 0$). Hence, we have that for every alternative $a \in A$, $v_i(a) \geq \frac{V_i^{pq}(a)}{2} - \frac{1}{2m^2}$. This gives us the following result.

$$\text{sw}(S, \vec{v}) = \sum_{i=1}^{n} \max_{a \in S} v_i(a) \geq \frac{1}{2} \sum_{i=1}^{n} \max_{a \in S} V_i^{pq}(a) - \frac{n}{2m^2} = \frac{1}{2}\text{sw}(S, \vec{V}^{pq}) - \frac{n}{2m^2} \tag{7}$$

Let us now consider the pair (p, q) for which Equation (6) holds. We consider the case of $p \neq 0$. The proof for $p = 0$ is analogous. Our goal is to show that $\widehat{A}_{pq}$ is close to A^*_{pq} in terms of social welfare with respect to $\vec{V}^{pq}$.

If $q \leq q_s$, notice that $\widehat{A}_{pq}$ computed by our rule is precisely A^*_{pq}. This is because each voter i has the same non-zero value for every alternative in set S_i^{pq} under valuation V_i^{pq} and zero value for every other alternative. Hence, maximizing the number of voters i for which $\widehat{A}_{pq}$ contains some alternative of S_i^{pq} precisely yields $\widehat{A}_{pq} = A^*_{pq}$.

If $q > q_s$, then $\widehat{A}_{pq}$ is a set of k alternatives chosen uniformly at random. Let $\widehat{S}$ denote a random subset of k alternatives. We want to show that $\mathbb{E}[\text{sw}(\widehat{S}, \vec{V}^{pq})] \geq \frac{1}{2t} \cdot \text{sw}(A^*_{pq}, \vec{V}^{pq})$. Note that under $\vec{V}^{pq}$, each voter has value either $2^p/m^2$ or 0 for the set A^*_{pq}. Hence, $\text{sw}(A^*_{pq}, \vec{V}^{pq}) \leq n \cdot 2^p/m^2$. Thus, it suffices to show that $\mathbb{E}[\text{sw}(\widehat{S}, \vec{V}^{pq})] \geq \frac{1}{2t} \cdot \frac{n 2^p}{m^2}$. However, note that $\text{sw}(\widehat{S}, \vec{V}^{pq}) = \frac{2^p}{m^2} \sum_{i \in N} \mathbb{1}\{\widehat{S} \cap S_i^{pq} \neq \emptyset\}$. Hence, it would suffice to show that for each voter $i \in N$, $\Pr[\widehat{S} \cap S_i^{pq} \neq \emptyset] \geq 1/(2t)$.

Fix a voter i. If $m - |S_i^{pq}| < k$ then this probability is 1. So assume $m - |S_i^{pq}| \geq k$. Now,

$$
\begin{aligned}
\Pr[\widehat{S} \cap S_i^{pq} \neq \emptyset] &= 1 - \Pr[\widehat{S} \cap S_i^{pq} = \emptyset] = 1 - \frac{\binom{m - |S_i^{pq}|}{k}}{\binom{m}{k}} \\
&= 1 - \frac{(m-k)(m-k-1)\dots(m-k-|S_i^{pq}|-1)}{m(m-1)\dots(m-|S_i^{pq}|-1)} = 1 - \prod_{j=0}^{|S_i^{pq}|-1} \left(1 - \frac{k}{m-j}\right) \\
&\geq 1 - \left(1 - \frac{k}{m}\right)^{|S_i^{pq}|} \geq 1 - e^{-\frac{k|S_i^{pq}|}{m}} \geq 1 - e^{-1/t} \geq \frac{1}{t} - \frac{1}{t^2} \geq \frac{1}{2t},
\end{aligned}
$$

as desired. Crucially, the third inequality holds because we are in the case of $q > q_s$, for which $|S_i^{pq}| \geq \frac{m}{tk}$ (this is why the voter did not send the set), and thus, $k\,|S_i^{pq}|\,/m \geq 1/t$. The last inequality holds because we can assume $t \geq 2$ (otherwise the theorem trivially holds).

Therefore, irrespective of the value of q, we have the following guarantee:

$$
\begin{aligned}
\mathbb{E}[\text{sw}(\widehat{A}_{pq}, \vec{v})] &\geq \frac{1}{2}\, \mathbb{E}[\text{sw}(\widehat{A}_{pq}, \vec{V}^{pq})] - \frac{n}{2m^2} \\
&\geq \frac{1}{4t} \cdot \text{sw}(A^*_{pq}, \vec{V}^{pq}) - \frac{n}{2m^2} \\
&\geq \frac{\text{sw}(A^*, \vec{v})}{24t \log^2 m} - \frac{n}{2m^2}
\end{aligned}
$$

Therefore, we have

$$
\frac{\mathbb{E}[\text{sw}(\widehat{A}_{pq}, \vec{v})]}{\text{sw}(A^*, \vec{v})} \geq \frac{1}{24t \log^2 m} - \frac{n}{2m^2 \text{sw}(A^*, \vec{v})} \geq \frac{1}{24t \log^2 m} - \frac{1}{2m} \geq \frac{1}{48t \log^2 m}
$$

The second inequality uses the fact that $\text{sw}(A^*, \vec{v}) \geq n/m$. To see this, recall that the total welfare of all alternatives is n (due to normalization of values). Hence, there exists an alternative a^* with welfare at least n/m. Hence, any set of size k containing a^* has welfare at least n/m, which implies that $\text{sw}(A^*, \vec{v}) \geq n/m$. The third inequality uses the fact that $t = d/(144 \log^4 m) \leq m/(24 \log^2 m)$.

Now, the subset $\widehat{A}_{pq}$ is returned with probability $((1 + 2\log m) \times (1 + \log m))^{-1}$, which is at least $1/(6 \log^2 m)$. Hence, the distortion of the algorithm is at most $144t \log^4 m = d$. For communication complexity, note that each voter i sends at most $6 \log^2 m$ sets of size at most $m/(tk) = 144m \log^4 m/(dk)$ each. Hence, the total communication from each voter is at most $O\left(\frac{m}{dk} \log^6 m\right)$. $\square$

C.5 Proof of Theorem 6

Proof. Let $A^* \in \arg\max_{A:|A|=k} \sum_{i=1}^n \max_{a \in A} v_i(a)$ and $\mathrm{sw}(A^*, \vec{v}) = \sum_{i=1}^n \max_{a \in A^*} v_i(a)$. We will write $\{V_i^{pq}\}_{i \in [n]}$ to denote the valuation profile restricted to bucket B_p of the interval $[0,1]$ and bucket C_q of the alternatives $[m]$. Let A^*_{pq} be the optimal subset of size k with respect to the valuation profile $\{V_i^{pq}\}_{i \in [n]}$. Then, as shown in the proof of theorem 5, there exist p and q such that the following holds.

$$\mathrm{sw}(A^*_{pq}, \vec{V}^{pq}) \geq \frac{1}{6 \log^2 m} \mathrm{sw}(A^*, \vec{v}). \tag{8}$$

We also recall the following statement from the proof of theorem 5.

$$\mathrm{sw}(S, \vec{v}) \geq \frac{1}{2} \mathrm{sw}(S, \vec{V}^{pq}) - \frac{n}{2m^2}$$

Consider such a pair (p, q). We first show that it is sufficient to consider the case $q_s < q \leq q_r$, i.e. when the upper end point of C_q is between $m/(kt^3)$ and $m/(kt)$.

1. If $q > q_s$, the center outputs a random subset of size k and by the proof before, such a subset will have distortion of at most $2t$ with respect to the valuation profile $\{V_i^{pq}\}_{i \in [n]}$. As this solution is returned with probability at least $1/(6 \log^2 m)$, by 8 with respect to the actual valuation of the voters, this gives a distortion of $O(t \log^4 m)$ with zero communication.

2. If $q \leq q_r$, the voters send their actual subsets restricted to partitions B_p and C_q, and the center solves the optimal problem restricted to the valuation profiles $\{V_i^{pq}\}_{i \in [n]}$. This gives a distortion of $O(1)$ with communication $O(m/(kt^3))$ with respect to the valuation profile $\{V_i^{pq}\}_{i \in [n]}$. As this solution is returned with probability at least $1/(6 \log^2 m)$, by 8, with respect to the actual valuation of the voters, this gives a distortion of at most $O(t \log^4 m)$ with communication $O(m/(kt^3))$.

So we assume that the upper end of C_q is between $m/(kt^3)$ and $m/(kt)$. This also implies that for each voter i, the size of the set S_i^{pq} is between $m/(2kt^3)$, and $m/(kt)$. We now establish the bound on communication complexity for this case.

- Let S be a random subset of $\{1, \ldots, m\}$ where each element is selected independently with probability $1/t$. Then by Chernoff bound 2, we have $\Pr(|S| \geq \frac{10 m \log m}{t}) \leq m^{-9}$.

- Let ϕ be a random prime in the range $[m^{20}]$ and x be an integer less than or equal to 2^m. By the Hardy-Ramanujan theorem [26], the number of distinct prime factors of x is at most $2 \log x / \log \log x \leq 2m$. And by the prime number theorem, the number of primes in the range $[m^{20}]$ is at most $m^{20}/\log m$. Therefore, the probability that ϕ divides x is at most $O\left(\frac{\log m}{m^{19}}\right)$. Therefore, if there are at most m^4 distinct numbers less than or equal to 2^m, then the probability that ϕ divides the difference of any two of them is at most $O\left(\frac{\binom{m^4}{2} \log m}{m^{19}}\right) = O(\log m / m^{11})$. Therefore, with probability at least $1 - O(\log m/m^{11})$, two voters i, i' with $s_i^{pq} \neq s_{i'}^{pq}$ are assigned to unique id $x_i^{pq} \neq x_{i'}^{pq}$. This implies that for a given subset S provided by the coordinator, the subsets $\widetilde{S}_i^{pq}$ and $\widetilde{S}_{i'}^{pq}$ are independent as long as voters i and i' have different valuations restricted to the partitions B_p and C_q.

- Since $m/(2kt^3) \leq |S_i^{pq}| \leq m/(kt)$, by Chernoff bound we have $\Pr(|S_i^{pq} \cap S| \geq 20 \log m\, |S_i^{pq}|/t) \leq \exp(-19 \log m\, |S_i^{pq}|/t)$. This implies that size of $\widehat{S}_i^{pq}$ is at most $20m \log m/(kt^2)$ with probability at least $1 - m^{-19}$, because conditioned on the event $\left|\widehat{S}_i^{pq}\right| \leq 20 \log m\, |S_i^{pq}|/t$, $\left|\widehat{S}_i^{pq}\right| \geq 20m \log m/(kt^2)$ iff $|S_i^{pq}| = \Theta(m/t)$. Finally, by a union bound over the m^4 voters, this result holds simultaneously for all voters with probability at least $1 - m^{-15}$.

- The sets $\widetilde{S}_i^{pq}$ are constructed by selecting each element of $\widehat{S}_i^{pq}$ independently with probability $\log m/t$. Therefore, following the same argument as in the previous step, we can guarantee that for all voters i, the size of $\widetilde{S}_i^{pq}$ is at most $400m \log^3 m/(kt^3)$ with probability at most $1 - m^{-15}$.

The above argument guarantees that for any pair of partitions p and q, the communication of each voter is at most $O\left(m \log^3 m/(kt^3)\right)$ with probability at least $1-m^{-8}$. Therefore the expected number of bits communicated by each voter is at most $O\left(m \log^3 m/(kt^3)\right) + m^{-8}m = O\left(m \log^3 m/(kt^3)\right)$. Substituting $t = d/\log^6 m$, we get that the expected communication is $O(m \log^{21} m/(kd^3)$.

We now establish the required bound on the distortion of the voting rule.

- Our voting rule first rounds the frequencies to the smallest multiple of two greater than or equal to each frequency. This increases the distortion by at most two. In order to see this, for every $j \in \{0, \ldots, \phi - 1\}$ we write f_j to denote the frequency of the j-th possible id of a voter. We write f_j^+ to denote the smallest power of two greater than or equal to f. Then $1/2 f_j^+ \leq f_j \leq f_j^+$. This implies that the social welfare of any subset A of size k with respect to the new frequencies is within a factor of two of the welfare of A with respect to the original frequencies.

- After duplicating the responses of the voters, suppose A^*_{pq} denote the optimal subset of size k with the corresponding optimal value $\mathrm{sw}(A^*_{pq})$. Then, we claim that there must exist a frequency f^* such that the optimal solution with respect to the valuations mapped to bucket f, provides a welfare of at least $\mathrm{sw}(A^*_{pq})/(8 \log m)$. Suppose this is not the case. Let $A^*_{fpq})$ be the optimal solution with respect to the valuation profiles mapped to frequency f, $\{V_i^{pq}\}_{i \in L_f}$. Here is the set of valuation profiles mapped to frequency f. Since the valuations mapped to the $8 \log m$ frequencies are disjoint, we have.

$$\mathrm{sw}(A^*_{pq}, \{V_i^{pq}\}_{i=1}^{m^4}) \leq \sum_f \mathrm{sw}(A^*_{fpq}, \{V_i^{pq}\}_{i \in L_f}) < 8 \log m \frac{\mathrm{sw}(A^*_{pq})}{8 \log m} = \mathrm{sw}(A^*_{pq}),$$

a contradiction. Therefore, there exists a frequency f such that

$$\begin{aligned} \mathrm{sw}(A^*_{fpq}, \{V_i^{pq}\}_{i \in L_f}) \geq \frac{1}{8 \log m} \mathrm{sw}(A^*_{pq}, \vec{V}^{pq}) &\geq \frac{1}{8 \log m (1 + \log m)^2} \mathrm{sw}(A^*, \vec{v}) \\ &\geq \frac{1}{32 \log^3 m} \mathrm{sw}(A^*, \vec{v}). \end{aligned} \tag{9}$$

- Recall that the valuation profile mapped to frequency $\{V_i^{pq}\}_{i \in L_f}$ may have several duplicates. In fact, each distinct type is repeated exactly f times. We now claim that it is sufficient to consider the case when the number of distinct voter types / subsets is at least tk. Suppose this is not the case. Then, the total number of subsets mapped to frequency f is at most ftk. On the other hand, the set $\widehat{A}_1^{fpq}$ covers k distinct voters and kf voters in total. Since $\widehat{A}^{fpq}$ is $\widehat{A}_1^{fpq}$ with probability $1/2$, the distortion is guaranteed to be at most $2(fkt)/(kf) = O(t)$.

Therefore, we have now reduced our problem of providing a distortion of $O(t)$ to the valuation profiles mapped to frequency f, $\{V_i^{pq}\}_{i \in L_f}$. As each valuation appears exactly f times in L_f, we will write $\{U_i^{pq}\}_{i \in \ell_f}$ to denote the set of unique valuations belonging to L_f. Additionally, we are guaranteed that there are at least kt distinct type of voters in L_f i.e. $|L_f| \geq kt$.

We now recall how different sets of a voter i were constructed. Restricted to a partition B_p of the interval $[0,1]$ and the elements $[m]$, voter i constructed S_i^{pq}. Then, voter i constructed $\widehat{S}_i^{pq}$, where each element of S_i^{pq} were retained with probability $1/t$. Finally, voter i constructed $\widetilde{S}_i^{pq}$ where each element of $\widehat{S}_i^{pq}$ was retained with probability $\log^2 m/t$. Now, let us write A_f^* to denote the optimal solution with respect to the sets S_i^{pq}, for $i \in \ell_f$. Formally,

$$A_f^* \in \underset{A:|A|=k}{\arg\max} \sum_{i \in \ell_f} \mathbf{1}\{A \cap S_i^{pq} \neq \emptyset\},$$

and let OPT_f be the number of distinct voters covered by A_f^*. Then there exists $k' \leq k$ such that there are k' alternatives in A_f^* each contributing OPT_f/k', i.e. covering at least OPT_f/k' distinct voters. Otherwise we get a contradiction as $\mathrm{OPT}_f \leq \sum_{i \in \ell_f} \sum_{a \in A_f^*} \mathbf{1}\{a \in S_i^{pq}\} < k\mathrm{OPT}_f/k = \mathrm{OPT}_f$. As argued before, there exist k' elements each covering at least $\mathrm{OPT}_f/k' \geq kt/k' \geq t$ distinct voters. The probability that one of such element, say a_j^* belongs to S is at least $1 - (1 - 1/t)^{k'}$. This probability is at least $k'/(2t)$ if $k' \leq t$. Otherwise, this probability is at least $1 - e^{-1}$. In either case, the probability that one of these k' elements is sampled in S is at least $k'/(2t)$.

Now we proceed conditioning on the event that such an element $a_j^* \in S$. Consider an element a that covers at least $n_a = t$ distinct voters. Since each distinct voter i independently samples a to be in $\widetilde{S}_i^{pq}$ with probability $2\log^2 m/t$, by Chernoff bound the number of times such an alternative a appears in the bucket f is at least $1 - m^{-11}$. Moreover, by a union bound, this simultaneously holds for any alternative that covers at least t distinct voters with probability at least $1 - m^{-10}$. By a similar argument any alternative that appears less than $t/4$ times, covers less than $2\log^2 m$ distinct voters and are rejected. Since $a_j^* \in S$, the set of elements with no of appearances at least $2\log^2 m$, is non-empty. If a_j^* is included in $\widehat{A}_2^{fpq}$, it covers OPT_f/k' distinct voters, which gives an expected distortion of $2t$ as a_j^* is included with probability $k'/(2t)$. If the element a_j^* is not included within A_2^{fpq}, then we select an element which covers even more distinct elements, giving improved distortion.

The above argument shows that the subset $\widehat{A}^{fpq}$ provides a $O(t)$ distortion with respect to the valuation profile $\{V_i^{pq}\}_{i \in L_f}$ assigned to the frequency f, bucket pair p and q. Therefore, by 9 the solution $\widehat{A}^{fpq}$ guarantees a $O(t\log^3 m)$ distortion with respect to the optimal solution A^*. Now the solution $\widehat{A}^{fpq}$ is output with probability at least $1/(32\log^3 m)$. This shows that the distortion of the voting rule is $O(t\log^6 m) = O(d)$. □

C.6 Proof of Theorem 7

Proof. We first recall the Hoeffding inequality.

Lemma 2. *(Hoeffding Inequality) Let $X = \sum_{i=1}^n X_i$, where each $X_i \in [0,1]$, and all X_i are independent. Let $\mu = \mathbb{E}[X]$. Then*

$$\Pr(|X-\mu| \geq t) \leq e^{-2t^2/n}$$

Let $S^* \in \arg\max_{S\subseteq A:|S|=k} \mathrm{sw}(S,\vec{v})$ be an optimal set of k alternatives with respect to the full valuation profile $\vec{v}$. As argued previously, $\mathrm{sw}(S^*,\vec{v}) \geq n/m$.

Consider any good set S i.e. a set S with $\mathrm{sw}(S,\vec{v}) \geq 2n/(m^{1.5})$. Then by Hoeffding inequality,

$$\Pr\left[\left|\mathrm{sw}(S,\vec{v}_T) - \frac{m^4}{n}\mathrm{sw}(S,\vec{v})\right| \geq \frac{m^4}{2n}\mathrm{sw}(S,\vec{v})\right] \leq e^{-2\left(\frac{m^4}{2n}\mathrm{sw}(S,\vec{v})\right)^2/m^4} \leq e^{-2m}. \tag{10}$$

The last inequality follows because $\mathrm{sw}(S,\vec{v}) \geq n/m^{1.5}$. By a union bound over the sets S with social welfare at least $n/m^{1.5}$, equation 10 holds with probability at least $1-e^{-m}$.

Now we consider any bad subset S i.e. a set S with $\mathrm{sw}(S,\vec{v}) \leq 2n/m^{1.5}$.

$$\Pr(\mathrm{sw}(S,\vec{v}_T) \geq 2m) \leq \min_{t>0} \Pr(e^{t\mathrm{sw}(S,\vec{v}_T)} \geq e^{2tm}) \leq \min_{t>0} e^{-2tm} \prod_{i\in T} E[e^{tv_i(S)}] \tag{11}$$

Now we bound $\mathbb{E}[e^{tv_i(S)}]$.

$$\begin{aligned}
\mathbb{E}[e^{tv_i(S)}] &= \frac{1}{n}\sum_{i=1}^n e^{tv_i(S)} = \frac{1}{n}\sum_{i=1}^n (1+tv_i(S)+\frac{t^2}{2}(v_i(S))^2+\ldots) \\
&= \frac{1}{n}(1+t\|v(S)\|_1+\frac{t^2}{2}\|v(S)\|_2^2+\ldots) \\
&\leq \frac{1}{n}(1+t\|v(S)\|_1+\frac{t^2}{2}\|v(S)\|_1^2+\ldots) \\
&\leq \frac{1}{n}e^{t\|v(S)\|_1} \leq \frac{1}{n}e^{2tn/m^{1.5}}
\end{aligned}$$

Substituting the above bound in equation 11, we get

$$\Pr(\mathrm{sw}(S,\vec{v}_T) \geq 2m) \leq \min_{t>0} e^{-2tm}\frac{e^{2tnm^{2.5}}}{n^{m^4}}.$$

We now substitute $t = 1/(nm^{2.5})$ and get

$$\Pr(\mathrm{sw}(S,\vec{v}_T) \geq 2m) \leq e^{-2/(nm^{1.5})}\frac{e^2}{n^{m^4}} \leq e^{-m^2}.$$

By taking a union bound over all the bad sets, we get that their welfare is at most $2m$ with probability at least $1-e^{-m}$. Since our algorithm returns a subset with at most d approx for the agents in S, the welfare of the returned subset is at least $m^4/(dm) >= m^2$. But the welfare of any bad subset is at most 2m. So with high probability it never returns a bad subset. Conditioned on the event that algorithm 8 returns a good subset, equation 10 gives us that the welfares of all the good subsets are preserved upto a factor of two with probability at least $1-e^{-m}$. Therefore, if S_T is the set returned by algorith 8, we have

$$\mathbb{E}[\mathrm{sw}(S_T,\vec{v})] \geq (1-2e^{-m})\frac{\mathrm{sw}(S^*,\vec{v})}{2d} \geq \frac{\mathrm{sw}(S^*,\vec{v})}{4d}.$$

□

C.7 Proof of Theorem 8

Proof. Consider a k-selection rule f with deterministic elicitation rule Π_f, possibly randomized aggregation rule Γ_f, and distortion $\text{dist}(f) \le d$. This proof is similar in structure to the proof of Theorem 4, where we use f to construct a deterministic protocol for $\text{FDISJ}_{m,m/t,t}$. However, in this reduction, instead of using $t = \Theta(d)$, we use $t = \Theta(dk)$.

Set $\gamma = 1/76$. Set $t = 2dk/\gamma$ and $s = m/t$. First, note that $d \ge 1 \ge \gamma/2$. Hence, $t \ge k$. Next, if $dk > m/480$, then our lower bound trivially holds. Hence, assume that $dk \le m/480$. This implies $t \le (m/2) \cdot (1 - 1/e)$. Hence, by Theorem 3, $D(\text{FDISJ}_{m,s,t}) \ge m$ for our choices of s and t. We use f to construct a deterministic protocol for $\text{FDISJ}_{m,s,t}$ under the substantial intersection promise with parameter γ.

Consider an input profile $(S_1, \ldots, S_t)$ of $\text{FDISJ}_{m,s,t}$ with a universe U of size m and substantial intersection promise with parameter γ. Let us create an instance of the k-selection problem like in the proof of Theorem 4. We have a set of n voters N and a set of m alternatives A. Each alternative in A corresponds to a unique element of U. Partition the set of voters N into t equal-size buckets: for $i \in [t]$, bucket N_i consists of n/t voters corresponding to player i, each of whom has valuation v^{S_i} given by $v^{S_i}(a) = 1/s$ for each $a \in S_i$ and $v^{S_i}(a) = 0$ for each $a \notin S_i$. Let $\vec{v}$ denote the resulting profile of voter valuations.

Under these valuations, for each alternative a, we have $\text{sw}(a, \vec{v}) = \frac{n}{ts} \sum_{i=1}^{t} \mathbb{1}[a \in S_i]$, where $\mathbb{1}$ is the indicator variable. Due to the substantial intersection promise, every YES instance admits at least one alternative a^* that appears in at least γt sets and has $\text{sw}(a^*, \vec{v}) \ge \frac{n}{ts} \cdot \gamma t = \frac{\gamma n}{s}$. Let OPT denote the optimal social welfare that can be achieved by any set of size k. Then, trivially, $\frac{\gamma n}{s} \le OPT \le \frac{n}{s}$.

Let S denote the (possibly random) set of alternatives returned by f on this instance. Then, to achieve distortion at most $d = \gamma t/(2k)$, we must have $\mathbb{E}[\text{sw}(S, \vec{v})] \ge \frac{\gamma n}{sd} = \frac{2nk}{ts}$.

Now, if a subset of alternatives T of size k only consists of elements which appear at most once, then $\text{sw}(T, \vec{v}) \le \frac{kn}{ts}$. Given that $OPT \le \frac{n}{s}$, with probability at least k/t, the voting rule must return a set that contains at least one element which appears in more than one sets.

We are now ready to construct a deterministic protocol for $\text{FDISJ}_{m,m/t,t}$. The protocol runs the voting rule f on the voting instance constructed above. That is, each player i responds to the query posed by the elicitation rule of f according to valuation v^{S_i}. Note that this requires a total of $t \cdot \text{C}(f)$ bits of communication from the players.

Next, we take players' responses, create n/t copies of the response of each player, pass the resulting profile as input to the aggregation rule of f, and obtain the distribution over alternatives that it returns. As before, because there is no computational restriction on our protocol, we can assume it has access to the exact distribution. Let B denote the set of all alternatives contained in any set of size k returned with probability at least k/t. Since there can be at most t/k such distinct sets, $|B| \le t$. From the argument above, we know that in every YES instance, there exists an alternative $x^* \in B$ that appears in at least two sets. In contrast, in every NO instance, every alternative in B appears in at most one set.

Next, we use the same argument as in the proof of Theorem 4 to determine if an element of B appears in more than one sets using $O(t)$ communication, whereby players successively write the intersection of B with their own set S_i, until one player spots an alternative written by another player which also appears in her set.

Once again, the total communication complexity of this protocol is at most $t \cdot \text{C}(f) + O(t)$, and this must be at least m by Theorem 3. Hence, we get that $\text{C}(f) = \Omega(\frac{m}{t}) = \Omega(\frac{m}{kd})$. $\square$

C.8 Proof of Theorem 9

Proof. Consider a k-selection rule f with randomized elicitation rule Π_f, possibly randomized aggregation rule Γ_f, and distortion $\mathrm{dist}(f) \leq d$.

Let $\delta > 0$ be a small constant. Set $t = 2d$ and $s = (2m)/(3kt)$. Note that we can assume $s \geq 1$, otherwise our lower bound is trivially true. We use f to construct a δ-error protocol for $\mathrm{FDISJ}_{m/k,s,t}$ under the unique intersection promise. Under our parameter choices, note that the result of Mandal et al. [33] applies, and we have $R_\delta(\mathrm{FDISJ}_{m/k,s,t}) = \Omega(s) = \Omega(m/(kt))$.

Consider an input profile $I = (S_1, \ldots, S_t)$ of $\mathrm{FDISJ}_{m/k,s,t}$ with a universe U of size m/k. Choose a uniformly random permutation σ of the m/k elements, and apply it to the sets to generate an instance $\widehat{I} = (\widehat{S}_1, \ldots, \widehat{S}_t)$. Note that its universe is still $\widehat{U} = U$. Now, in a YES instance of $\mathrm{FDISJ}_{m/k,s,t}$ under the unique intersection promise, one element appears in every set and the sets are otherwise pairwise disjoint. Thus, all YES instance look identical up to a permutation of the elements. More specifically, applying a uniformly random permutation of elements to a given YES instance of $\mathrm{FDISJ}_{m/k,s,t}$ generates a uniformly random YES instance of $\mathrm{FDISJ}_{m/k,s,t}$.

Let μ denote the uniform distribution over YES instances of $\mathrm{FDISJ}_{m/k,s,t}$. We sample $k-1$ random YES instances $I^1, \ldots, I^{k-1}$ from this distribution, but give each YES instance a new universe $U^1, \ldots, U^{k-1}$ with disjoint m/k elements. Hence, $U, U^1, \ldots, U^{k-1}$ each contain disjoint m/k elements, or m elements in total. Next, we take a random permutation of the k instances $(\widehat{I}, I^1, \ldots, I^{k-1})$ to obtain $(\widetilde{I}^1, \ldots, \widetilde{I}^k)$. Structurally, every instance in this vector is a random sample from μ (with element relabeling). As such, there is no information available to distinguish $\widehat{I}$ from the $k-1$ generated instances in this vector due to symmetry. For $i \in [k]$, let instance $\widetilde{I}^i$ have input profile $(\widetilde{S}^i_1, \ldots, \widetilde{S}^i_t)$ and universe $\widetilde{U}^i$. Let $\widetilde{U} = \cup_{i\in[k]} \widetilde{U}^i$.

Next, we construct a k-selection voting instance as follows. We have a set of n voters N and a set of m alternatives A. N and A are partitioned into k equal-sized buckets $N^1, \ldots, N^k$ and $A^1, \ldots, A^k$. For each $i \in [k]$, voters in N^i and alternatives in A^i are used to construct a sub-instance corresponding to $\widetilde{I}^i$. Each alternative in A^i corresponds to a unique alternative in $\widetilde{U}^i$. Voters in N^i are further partitioned into t equal-size buckets: for $j \in [t]$, bucket N^i_j consists of $n/(kt)$ voters corresponding to player j of instance $\widetilde{I}^i$, each of whom has valuation $v^{\widetilde{S}^i_j}$ given by $v^{\widetilde{S}^i_j}(a) = 1/s$ for each $a \in \widetilde{S}^i_j$ and $v^{S_i}(a) = 0$ for every other a. Note that voters in N^i have zero values for alternatives in $A^{i'}$ whenever $i \neq i'$. Let $\vec{v}$ denote the resulting profile of voter valuations.

Under these valuations, for each alternative a, we have $\mathrm{sw}(a, \vec{v}) = \frac{n}{kts} \sum_{i\in[k]} \sum_{j\in[t]} \mathbb{1}[a \in \widetilde{S}^i_j]$, where $\mathbb{1}$ is the indicator variable. Suppose the original instance I is a YES instance. Then, due to the unique intersection promise, all k instances $\widetilde{I}^i$, $i \in [k]$, contain a common element which appears in the set of every player in that instance. Thus, the set T^* composed of alternatives corresponding to these k common elements has welfare $\mathrm{sw}(T^*, \vec{v}) = n/s$.

Let T denote the random k-set returned by f on this instance. To get distortion at most d, we must have $\mathbb{E}[\mathrm{sw}(T, \vec{v})] \geq \frac{n}{ds} = \frac{2n}{ts}$. Let p be the number of common elements contained in T. Since each common element appears in the set of t players, while every other element appears in the set of one player, we have $\mathrm{sw}(T, \vec{v}) \leq p \cdot \frac{n}{ks} + (k-p) \cdot \frac{n}{kts}$. Thus, to get distortion at most d, we must have

$$\mathbb{E}\left[p \cdot \frac{n}{ks} + (k-p) \cdot \frac{n}{kts}\right] \geq \frac{2n}{ts} \Rightarrow \mathbb{E}[p] \geq \frac{k}{t}.$$

Hence, the expected number of common elements returned by f must be at least k/t. Now, since f cannot distinguish between the k instances, it must return the common element of the real instance I with probability at least $1/t$. Further, again due to symmetry, the probability that f returns more than $2t\ln(2/\delta)/\delta$ elements from the real instance I is at most $\delta/(2t\ln(2/\delta))$.

We are now ready to construct a δ-error protocol for $\mathrm{FDISJ}_{m/k,s,t}$. The protocol runs the voting rule f on the voting instance constructed above. Each player i of the real instance responds to the query of f by communicating $\mathrm{C}(f)$ bits, thus using a total of $t \cdot \mathrm{C}(f)$ bits of communication from all the players. We make $n/(kt)$ copies of each such message, and generate messages from voters corresponding to the players from the $k-1$ generated instances (without any communication from the real players) and feed to the aggregation rule of f. This process is repeated $t\ln(2/\delta)$ times, and the sets $Z^1, \ldots, Z^{t\ln(2/\delta)}$ returned by f are recorded. Let $Z = \cup_{r\in[t\ln(2/\delta)]} Z^r$. Note that this uses a total of $t^2\ln(2/\delta)\mathrm{C}(f)$ bits of communication.

When the original instance I is a YES instance with common element x^*, we have shown that $\Pr[x^* \in Z^r] \geq 1/t$ for each $r \in [t\ln(2/\delta)]$. Hence,

$$\Pr[x^* \in Z] \geq 1 - \left(1 - \frac{1}{t}\right)^{t\ln(2/\delta)} \geq 1 - \frac{2}{\delta}.$$

Also, we have shown that $\Pr[|Z^r \cap U| > 2t\ln(2/\delta)/\delta] \leq \delta/(2t\ln(2/\delta))$ for each $r \in [t\ln(2/\delta)]$. Hence, by the union bound, $\Pr[|Z \cap U| \leq 2t^2\ln(2/\delta)^2/\delta] \geq 1 - \delta/2$. Thus, with probability at least $1-\delta$, we have that $x^* \in Z$ and $|Z \cap U| = O(t^2)$. Assume that this happens, since our δ-error protocol is allowed to fail with probability at most δ.

Thus, the remaining task is to check if any of $O(t^2)$ elements in $Z \cap U$ appear in more than one player's set in the original instance I. Since we are in the unique intersection promise, this can be done by asking two arbitrary players of I to report all elements in $Z \cap U$ that appear in their sets, and taking the intersection. Note that we are in the shared blackboard model with public randomness, so all players can do the abovementioned computation and compute Z themselves.

Thus, we have constructed a δ-error protocol for $\mathrm{FDISJ}_{m/k,s,t}$ with total communication cost at most $t^2\ln(2/\delta)\mathrm{C}(f) + O(t^2)$. However, this must be $\Omega(\frac{m}{kt})$ [33]. Hence, we get $\mathrm{C}(f) = \Omega\left(\frac{m}{kt^3}\right) = \Omega\left(\frac{m}{kd^3}\right)$. □

D Reduction from 1-Selection to k-Selection

In this section, we present our reduction from 1-selection to k-selection. We note that to the best of our knowledge, this is the first such reduction, despite the fact that both problems are widely studied in the computational social choice literature [16, 23]. This is perhaps the case because the k-selection problem is often viewed as one where the voting rule may need to pick k winners for *any* given value of k; as such, it would be a strict generalization of the $k = 1$ case. In contrast, we compare the k-selection problem with a *fixed* $k > 1$ to the 1-selection problem.

Theorem 10. *Fix $k > 1$. Let f be any k-selection voting rule for m alternatives which uses deterministic (resp. randomized) elicitation and achieves d distortion with c bits of communication complexity. Then, there exists a 1-selection voting rule f' for m/k alternatives which also uses deterministic (resp. randomized) elicitation and achieves at most d distortion with at most $k \cdot c$ bits of communication complexity.*

Proof. Suppose we are given a 1-selection voting instance I' with a set of voters N' and a set of m/k alternatives A'. Each voter $i \in N'$ has valuation v'_i. On this instance, our voting rule f' would work as follows.

Construct a new k-selection voting instance I by creating k identical copies of instance I', each with a fresh set of voters and alternatives. That is, instance I contains a set of voters $N = \{(i, j) : i \in N', j \in [k]\}$ and a set of alternatives $A = \{(a, j) : a \in A', j \in [k]\}$. In each copy j, each voter has values for alternatives in the same copy j as per instance I' and zero value for the alternatives in the other copies. That is, for all $i \in N'$, $a \in A'$, and $j, j' \in [k]$, $v_{i,j}(a, j') = v'_i(a)$ if $j = j'$ and 0 otherwise.

Then, we use the elicitation rule of f. That is, each voter $i \in N'$ in the original instance I' provides k responses, where the j-th response is how voter (i, j) with valuation $v_{i,j}$ would have responded to the query posed by f. Hence, voting rule f' elicits at most $k \cdot c$ bits from each voter in instance I'.

Next, these responses are fed to the aggregation rule of f, which returns a set S of k alternatives (possibly selected in a randomized fashion). The aggregation rule of f' then returns one alternative from S, selected uniformly at random.

To get a bound on the distortion of f', let a^* denote an optimal alternative in instance I' and let $\mathrm{sw}(a^*)$ denote its social welfare in I'. Then, notice that by selecting the k copies of a^* in instance I, we can get social welfare that is $k \cdot \mathrm{sw}(a^*)$. Because the distortion of f is d, the expected social welfare of S in instance I must be at least $k \cdot \mathrm{sw}(a^*)/d$. The fact that one alternative selected uniformly at random from S generates expected social welfare at least $\mathrm{sw}(a^*)/d$ in instance I' follows from submodularity of valuations (recall that the value of each voter for a set of alternatives is the maximum of her value for any alternative in the set) and the fact that the k copies in our construction are completely independent. $\square$

Recall the optimal lower bounds for the 1-selection problem with m/k alternatives. Theorem 4 shows that using deterministic elicitation, $\Omega((m/k)/d)$ communication complexity is required to achieve distortion d, whereas Mandal et al. [33] show that using randomized elicitation, $\Omega((m/k)/d^3)$ communication complexity is required to achieve distortion d. Since Theorem 10 shows that k times the communication complexity of any k-selection voting rule with distortion d must be at least this much, we immediately get $\Omega\left(\frac{m}{k^2 d}\right)$ lower bound for deterministic elicitation and $\Omega\left(\frac{m}{k^2 d^3}\right)$ lower bound for randomized elicitation. Both are weaker by a factor of k than the optimal lower bounds we establish in Section 6.