Distortion in Voting with Top-t Preferences

Allan Borodin[1] , **Daniel Halpern**[2] , **Mohamad Latifian**[1]***,** and **Nisarg Shah**[1]

[1]University of Toronto
[2]Harvard University

bor@cs.toronto.edu, dhalpern@g.harvard.edu, {latifian, nisarg}@cs.toronto.edu

Abstract

A fundamental question in social choice and multi-agent systems is aggregating ordinal preferences expressed by agents into a measurably prudent collective choice. A promising line of recent work views ordinal preferences as a proxy for underlying cardinal preferences. It aims to optimize distortion, the worst-case approximation ratio of the (utilitarian) social welfare. When agents rank the set of alternatives, prior work identifies near-optimal voting rules for selecting one or more alternatives. However, ranking all the alternatives is prohibitive when there are many alternatives.

In this work, we consider the setting where each agent ranks only her t favorite alternatives and identify almost tight bounds on the best possible distortion when selecting a single alternative or a committee of alternatives of a given size k. Our results also extend to approximating higher moments of social welfare. Along the way, we close a gap left open in prior work by identifying asymptotically tight distortion bounds for committee selection given full rankings.

1 Introduction

A common task in multi-agent systems is to make collective decisions that serve multiple agents well in a measurable sense, and voting is a frequently-used tool for this purpose [Shoham and Leyton-Brown, 2008; Pitt *et al.*, 2006], in applications such as human computation [Procaccia *et al.*, 2012], distributed sensor networks [Lesser *et al.*, 2003], meeting scheduling [Haynes *et al.*, 1997], planning [Ephrati and Rosenschein, 1997], and rank aggregation for the web [Dwork *et al.*, 2001].

Voting has been studied for centuries in social choice theory, dating back to the early work by Condorcet [1785], in which voters rank candidates, and the goal is to select one or more candidates. But the prominent approach for evaluating the efficacy of voting rules has been the axiomatic approach, which is more qualitative in nature and has resulted in celebrated impossibility results [Arrow, 1951]. Arguably, this has

---
*Contact Author

led to a lack of consensus, even among social choice theorists, on which voting rules are the "best."

A recent wave of interest in voting from computer science has provided a fundamentally new perspective for quantitatively evaluating voting rules. Procaccia and Rosenschein [2006] propose to view the ranked preferences submitted by voters over candidates as proxies for their underlying numerical utility functions. This assumption allows one to focus on a canonical quantitative goal: maximizing the (utilitarian) social welfare [Bentham, 1780]. They propose to judge voting rules by their distortion, the worst-case approximation ratio between the maximum possible social welfare given complete utility functions and the (expected) social welfare achieved by the voting rule given only the partial preference information. Hence, distortion is the "price" of missing information and acts as a yardstick for answering the age-old question: *Which voting rules are the best?*

Boutilier *et al.* [2015] identify a near-optimal randomized voting rule for selecting a single candidate given ranked preferences of voters. Caragiannis *et al.* [2017] extend their analysis to deterministic and randomized rules for selecting a committee of candidates of a given size k. Since then, the distortion literature has proliferated, and the idea has been applied to settings even beyond voting; we refer the reader to the recent survey by Anshelevich *et al.* [2021] for a detailed overview of the results.

Of particular interest to us is the observation that once we surmise the existence of underlying utility functions, we do not need to stick with asking voters to rank candidates. As Benade *et al.* [2021] observe, distortion can be used to evaluate and compare different elicitation formats (i.e., ballot designs). Mandal *et al.* [2019] and Mandal *et al.* [2020] stretch this to the extreme, allowing arbitrary elicitation formats and studying the tradeoff between the number of bits they extract from each voter and the distortion they enable. However, this can lead to unintuitive elicitation formats, which may be difficult for humans to answer.

Another line of work focuses on intuitive elicitation formats that are either more expressive than ranked preferences [Amanatidis *et al.*, 2021] or less expressive [Gross *et al.*, 2017; Kempe, 2020b; Halpern and Shah, 2021]. A common less-expressive format is top-t preferences, where each voter ranks only her t most favorite candidates instead of ranking all candidates. This is particularly well-suited in ap-

Setting	Upper and Lower Bounds
$k = 1$	$\Theta\left(\max\left(\sqrt{m}, \frac{m}{t}\right)\right)$
$t = m$	$\Theta\left(\min\left(\frac{m}{k}, \sqrt{m}\right)\right)$
$k > 1,$	Lower Bound: $\Omega\left(\min\left(\frac{m}{k}, \max\left(\frac{m}{kt}, \sqrt{m}\right)\right)\right)$
$t < m$	Upper Bound: $O\left(\min\left(\frac{m}{k}, \max\left(\frac{m}{t}, \sqrt{m}\right)\right)\right)$

Table 1: A summary of our bounds for the first moment distortion. Although there is a gap in the general case, bounds are tight in two interesting special cases.

plications where we have far too many candidates to choose from [Procaccia *et al.*, 2012]. Kempe [2020b] studies the distortion under this type of preferences in the *metric framework*, where voters have costs rather than utilities. In this paper, our main goal is to study distortion with top-t preferences under the original utilitarian framework.

1.1 Our Results

We consider selecting a committee of a given size k given top-t preferences of the voters over the candidates, under the model in which a voter's utility for a committee is her maximum utility for any candidate in the committee. We consider approximating the social welfare, and also generalize our results to approximating the p-th power of the social welfare for $p \geqslant 1$, as advocated by Fain *et al.* [2020]. As described in Section 2, the results of Caragiannis *et al.* [2017] can be used to immediately settle this question for deterministic voting rules; hence, we focus exclusively on randomized rules.

Considering the standard first-moment distortion ($p = 1$) and single-winner selection ($k = 1$), we identify a tight bound of $\Theta(\max(\sqrt{m}, m/t))$ using a recently-introduced technique [Ebadian *et al.*, 2022]. Strikingly, the optimal distortion is already achieved asymptotically with $t = \sqrt{m}$. In practical terms, this is encouraging, as it may be much more feasible for voters to rank $\sqrt{m}$ rather than m alternatives.

Furthermore, for committee selection ($k \geqslant 1$), we are able to extend this to an upper bound of $O\left(\min\left(m/k, \max\left(\sqrt{m}, m/t\right)\right)\right)$. For the case of full rankings ($t = m$), this matches the lower bound of Caragiannis *et al.* [2017], closing a $m^{1/6}$ gap they left open. This resolves the optimal distortion for committee selection as $\Theta(\min(m/k, \sqrt{m}))$. The lower bound of Caragiannis *et al.* [2017] continues to match our upper bound when $k \geqslant \sqrt{m}$ (with arbitrary t) or $k \leqslant \sqrt{m} \leqslant t$. However, when $k, t \leqslant \sqrt{m}$, our lower bound of $\Omega(\max(\sqrt{m}, m/kt))$ is weaker than our upper bound of $O(m/\max(k,t))$. A visualization of our results is show in Table 1.

Finally, as a theoretical curiosity, we extend our bounds to higher moments ($p > 1$), where for single winner selection our bound is tight up to the $\log t$ factor, but in the case of committee selection, we leave open substantial gaps.

1.2 Related Work

To the best of our knowledge, the only work that comes close to studying distortion under top-t preferences in the utilitarian framework is that of Mandal *et al.* [2019]. They propose

a voting rule, PREFTHRESHOLD, which asks voters to report the *set* of their t most preferred candidates along with their approximate utilities for these candidates, by partitioning the utility space into discrete buckets and asking the voters to identify the appropriate buckets. Their distortion bound is only comparable to ours when their rule uses a single bucket, for which their bound is infinite.

In the metric framework, distortion under top-t preferences is better understood. For single winner selection ($k = 1$) with the first moment ($p = 1$), Kempe; Kempe [2020b; 2020a] proves a lower bound of $(2m - t)/t$ and an asymptotically matching upper bound of $12m/t$. Recently, Anagnostides *et al.* [2021] improve the upper bound to $6m/t$ and show that this can be further improved to $(4m - t)/t$ if a generalization of a combinatorial lemma due to Gkatzelis *et al.* [2020] holds.

While top-t preferences are relatively less explored in the distortion setting, they are very well studied more broadly in voting [Oren *et al.*, 2013; Lee *et al.*, 2014; Lu and Boutilier, 2011; Filmus and Oren, 2014] and in settings beyond voting [Drummond and Boutilier, 2013; Hosseini *et al.*, 2021].

Finally, following Caragiannis *et al.* [2017], we use the model where the utility of a voter for a committee is her maximum utility for any candidate in the committee. This is in the style of common voting rules such as the Chamberlin-Courant rule and the Monroe rule (see [Lang and Xia, 2016] for definitions), which aim to select committees in which every voter has a candidate representing her. An alternative model would be to model the utility of a voter for a committee as the *sum* of her utilities for the candidates in the committee, which has also been considered in the literature [Benade *et al.*, 2021].

2 Preliminaries

For $t \in \mathbb{N}$, define $[t] = \{1, \ldots, t\}$. Let $V = [n]$ be a set of n voters and C be a set of m candidates. We use indices i, j to denote voters and letters a, b, c to denote candidates. A committee is a subset of candidates. In this work, we consider selecting a committee of a given size $k \in [m]$. Let $\mathcal{P}_k(C)$ denote the set of all committees of size k. When $k = 1$, we refer to it as single winner selection.

Voter utilities. Each voter i has a utility $u_i(c) \in \mathbb{R}_{\geqslant 0}$ for every candidate c; we assume the standard normalization that $\sum_{c \in C} u_i(c) = 1$ for every i [Aziz, 2020]. We refer to $u = (u_1, \ldots, u_n)$ as the utility profile. For a committee $X \in \mathcal{P}_k(C)$, we define, with slight abuse of notation, the utility of voter i for X as $u_i(X) = \max_{c \in X} u_i(c)$. This is a standard extension studied in the literature [Caragiannis *et al.*, 2017], whereby a voter cares about having some representative in the committee that they like. The (utilitarian) social welfare of X is then given by $\mathsf{sw}(X, u) = \sum_{i=1}^{n} u_i(X)$; for $X = \{c\}$, we simply write $\mathsf{sw}(c, u)$. When the utilities are clear from context, we may drop them from the notation and simply write $\mathsf{sw}(X)$ or $\mathsf{sw}(c)$.

Preference profile. We do not directly observe voters' underlying utility functions. Instead, we ask each voter i to submit a ranking of her t most preferred candidates, denoted by a one-to-one function $\sigma_i : [t] \rightarrow C$ satisfying $u_i(\sigma_i(1)) \geqslant \ldots \geqslant u_i(\sigma_i(t)) \geqslant u_i(c)$ for all $c \in C \setminus \sigma_i([t])$. We allow the voter to break any ties arbitrarily. We refer to

$\sigma = (\sigma_1, \ldots, \sigma_n)$ as the preference profile. We use $u \triangleright \sigma$ to denote that preference profile σ is induced from the underlying utility profile u.

Voting rule. A (randomized) voting rule f takes as input a preference profile σ and outputs a distribution $f(\sigma)$ over committees of size k. We say that the voting rule is deterministic if it always returns a distribution with singleton support, in which case we use $f(\sigma)$ to denote the unique committee of size k in the support.

Distortion. Roughly speaking, the distortion captures how far the candidate elected by f can be from the optimal alternative. More formally, let I be an instance in this model which is given by the tuple (V, C, u). When evaluating distortion, we fix the number of candidates m. Let $\mathcal{I}$ denote the set of all instances with m candidates. Fix $p \in \mathbb{N}_{>0}$. Following Fain *et al.* [2017] and Fain *et al.* [2020], the p-th moment distortion of voting rule f on an instance $I = (V, C, u)$ is given by

$$\text{dist}^p(f, I) = \sup_{\sigma: u \triangleright \sigma} \frac{\max_{Y \in \mathcal{P}_k(C)} [\text{sw}(Y, u)]^p}{\mathbb{E}_{X \sim f(\sigma)}[(\text{sw}(X, u))^p]}.$$

The p-th moment distortion of f is obtained by taking the worst case over all instances: $\text{dist}^p(f) = \sup_{I \in \mathcal{I}} \text{dist}^p(f, I)$.

Note that for deterministic rules, since there is no expectation in the denominator, the choice of p does not affect the distortion as it cancels out; hence, analyzing $p = 1$ is sufficient. For $p = 1$, Caragiannis *et al.* [2017] prove that a deterministic rule achieves distortion $1 + m(m - k)/k$ even with $t = 1$, and no deterministic rule can be asymptotically better even when $t = m$. Hence, this provides asymptotically optimal distortion bounds for all values of t, k, and p. Consequently, in this work, we focus exclusively on randomized voting rules.

3 Single Winner Selection

Let us begin by analyzing the distortion for single winner selection given top-t preferences. Given only plurality votes ($t = 1$), it is known that the best possible distortion is m, which can be achieved by selecting a uniformly random candidate (see, e.g. [Mandal *et al.*, 2019, Proposition 1]). On the other hand, given ranked preferences ($t = m$), Boutilier *et al.* [2015] pinpoint the optimal distortion to be between $O(\sqrt{m} \cdot \log^* m)$ and $\Omega(\sqrt{m})$, and Ebadian *et al.* [2022] close this gap to establish a $\Theta(\sqrt{m})$ bound.

In this section, we fill the gap between these two extremes. We show that the optimal distortion for top-t preferences is $\Theta(\max(m/t, \sqrt{m}))$. Hence, it first decreases from m to $\Theta(\sqrt{m})$ as t increases from 1 to $\Theta(\sqrt{m})$, but then remains $\Theta(\sqrt{m})$ as t increases further. In a sense, this shows that after eliciting the top-$\Theta(\sqrt{m})$ preferences of the voters, eliciting the rest of their preference ranking does not significantly help. Our analysis extends to the p-th moment with a logarithmic gap when $p > 1$.

3.1 Upper Bound

For ranked preferences, Boutilier *et al.* [2015] show that a simple rule achieves $O(\sqrt{m \log m})$ distortion, which is only logarithmically worse than the optimal distortion. They

define the harmonic score of candidate a as $\text{hsc}(a) = \sum_i 1/\sigma_i^{-1}(a)$; that is, candidate a gets $1/r$ points whenever it appears in the r-th position in a voter's preference ranking. Then, their rule chooses each candidate a with probability $\frac{1}{2} \frac{\text{hsc}(a)}{\sum_b \text{hsc}(b)} + \frac{1}{2} \frac{1}{m}$.

We show that a natural extension of this rule to top-t preferences achieves near-optimal distortion simultaneously for all p. We define the *truncated harmonic score* of candidate a, whereby the candidate still gets $1/r$ points whenever it appears in the r-th position for $r \leqslant t$, but gets zero points if it does not appear in the top t positions; that is, $\text{hsc}_t(a) = \sum_{i: a \in \sigma_i([t])} 1/\sigma_i^{-1}(a)$. Then, our rule, f^h, chooses every candidate a with probability $\frac{1}{2} \frac{\text{hsc}_t(a)}{\sum_b \text{hsc}_t(b)} + \frac{1}{2} \frac{1}{m}$.

Theorem 1. *For all $p \geqslant 1$ and $t \in [m]$, we have that*

$$\text{dist}^p(f^h) \leqslant 2 \min \left(m, \frac{p}{W(p)} \max \left((H_t \cdot m)^{\frac{p}{p+1}}, \left(\frac{m}{t} \right)^p \right) \right).$$

Here, $H_t = \sum_{r=1}^t \frac{1}{r}$ is the t-th harmonic number and $W(p) = \Theta(\log p)$ is the solution of $W(p)e^{W(p)} = p$.[1] Asymptotically, $H_t \in \Theta(\log t)$ and $W(p) \in \Theta \log(p)$.

Proof. Fix an arbitrary instance $I = (V, C, u)$ with top-t preference profile $\sigma = (\sigma_1, \ldots, \sigma_n)$ induced by u. Fix an optimal candidate $a \in \arg \max_{c \in C} \text{sw}(c)$. Let q_c be the probability by which f^h chooses candidate c on this profile.

We will show two separate upper bounds on the welfare approximation ratio $\frac{\text{sw}(a)^p}{\sum_{c \in C} q_c \cdot \text{sw}(c)^p}$; then, taking the minimum of the two ratios yields the bound stated in the theorem.

First, an upper bound of $2m$ follows directly from the fact that $q_a \geqslant 1/(2m)$. Hence,

$$\frac{\text{sw}(a)^p}{\sum_{c \in C} q_c \cdot \text{sw}(c)^p} \leqslant \frac{\text{sw}(a)^p}{q_a \cdot \text{sw}(a)^p} = \frac{1}{q_a} \leqslant 2m.$$

For the second upper bound, we consider two cases depending on the truncated harmonic score $\text{hsc}_t(a)$ of the optimal candidate a. Fix $\tau = \frac{W(p)}{p} \cdot \left(\frac{H_t}{m^p} \right)^{1/(p+1)}$. We consider $\text{hsc}_t(a) \geqslant n\tau$ and $\text{hsc}_t(a) < n\tau$, and show that the desired upper bound holds in both cases.

Case 1: First, suppose $\text{hsc}_t(a) \geqslant n\tau$. We have that

$$q_a \geqslant \frac{1}{2} \cdot \frac{\text{hsc}_t(a)}{\sum_{c \in C} \text{hsc}_t(c)} = \frac{1}{2} \cdot \frac{\text{hsc}_t(a)}{nH_t} \geqslant \frac{\tau}{2H_t}.$$

Hence, by the same argument as above, the welfare approximation ratio is at most

$$\frac{1}{q_a} \leqslant \frac{2H_t}{\tau} = \frac{2p}{W(p)} \cdot (H_t \cdot m)^{\frac{p}{p+1}}$$

$$\leqslant \frac{2p}{W(p)} \cdot \max \left((H_t \cdot m)^{\frac{p}{p+1}}, \left(\frac{m}{t} \right)^p \right).$$

Case 2: Next, suppose $\text{hsc}_t(a) < n\tau$. Note that the utility of voter i for a is at most $1/r$ if the voter ranks a in the r-th position, for some $r \leqslant t$, and at most $1/t$ otherwise. Hence,

$$\text{sw}(a) \leqslant \text{hsc}_t(a) + n/t \leqslant n \cdot (\tau + 1/t).$$

---

[1]This is known as the Lambert W function.

The expected social welfare when picking a uniformly random candidate is n/m, which implies that, by Jensen's inequality, the expected p-th moment of social welfare is at least $(n/m)^p$. Since our rule f^h implements this with probability $1/2$, we have $\sum_{c \in C} q_c \cdot \mathsf{sw}(c)^p \geqslant (1/2) \cdot (n/m)^p$. Together, these imply that

$$
\begin{aligned}
\frac{\mathsf{sw}(a)^p}{\sum_{c \in C} q_c \cdot \mathsf{sw}(c)^p} &\leqslant \frac{n^p \cdot (\tau + 1/t)^p}{(1/2) \cdot (n/m)^p} \\
&= 2 \cdot (m\tau + m/t)^p \\
&= 2 \cdot \left((W(p)/p) \cdot (H_t \cdot m)^{\frac{1}{p+1}} + \frac{m}{t} \right)^p \\
&\leqslant 2 \cdot (1 + W(p)/p)^p \cdot \max\left((H_t \cdot m)^{\frac{1}{p+1}}, \frac{m}{t} \right)^p \\
&\leqslant 2 \cdot e^{W(p)} \cdot \max((H_t \cdot m)^{\frac{p}{p+1}}, (m/t)^p) \\
&= \frac{2p}{W(p)} \cdot \max((H_t \cdot m)^{\frac{p}{p+1}}, (m/t)^p).
\end{aligned}
$$

Combining this with Case 1 yields the desired bound. $\qquad\square$

For $p = 1$, this bound is $O(\max(\sqrt{m \log t}, m/t))$. In this special case, we can eliminate the $\sqrt{\log t}$ factor by extending a recent technique due to Ebadian *et al.* [2022]. This follows directly from Theorem 4 presented in Section 4.

Proposition 1. *For $t \in [m]$ and $p = 1$, there exists a randomized rule whose distortion is $O(\max(\sqrt{m}, m/t))$.*

3.2 Lower Bound

Next, we show that the bound achieved in the previous subsection is tight up to the $(\log t)^{p/(p+1)}$ factor. For $p = 1$, the following lower bound is $\Omega(\max(\sqrt{m}, m/t))$, precisely matching the upper bound from Proposition 1. The proof of the following result, along with the other missing proofs, can be found in the full version.[2]

Theorem 2. *Fix constant $p \geqslant 1$. Every randomized rule f for selecting a single winner given top-t preferences has*

$$
\mathrm{dist}^p(f) = \Omega\left(\min\left(m, \max\left(m^{\frac{p}{p+1}}, \left(\frac{m}{t}\right)^p \right) \right) \right).
$$

4 Committee Selection for the First Moment

We now turn our attention to selecting a committee of size $k \geqslant 1$ given top-t preferences. In this section, we focus on the first moment, for which we are able to derive tight distortion bounds. The next section focuses on committee selection with higher moments, for which our bounds are not tight.

4.1 Upper Bound

In order to derive the upper bound, we extend a recent approach introduced by Ebadian *et al.* [2022]. They use it to derive an optimal $\Theta(\sqrt{m})$ bound for single winner selection ($k = 1$) given full rankings. We extend this to all $k, t \in [m]$.

The approach relies on another recent result due to Cheng *et al.* [2020]. They consider randomized committee selection that satisfies a compelling stability/fairness property. For a pair of committees $S, S' \subseteq C$, we say that $S' \succ_i S$ if voter i ranks her most preferred candidate in S' above her most preferred candidate in S. Let $V(S, S') = \{i \in V : S' \succ_i S\}$.

---
[2]https://www.cs.toronto.edu/~nisarg/papers/distortion-top-t.pdf

Definition 1 (Stable Lotteries). Fix $t \in [m]$. A distribution $\mathcal{S}$ over committees of size t is said to be *stable* if, for every committee S' with $|S'| \leqslant t$, we have $E_{S \sim \mathcal{S}}[|V(S, S')|] \leqslant n \cdot |S'|/t$.

Note that when a committee S is sampled from a stable lottery $\mathcal{S}$, the fraction of voters preferring any other fixed committee S' over S is bounded, in expectation, by the ratio of the sizes of S' and S. In other words, a small committee cannot be preferred by many voters. It is worth noting that if the property is satisfied for all S' with $|S'| = 1$, then it is satisfied for all S' with $|S'| \leqslant t$ (see [Cheng *et al.*, 2020]).

Theorem 3 (Cheng *et al.* [2020]). *Given ranked preferences and $t \in [m]$, a stable lottery over committees of size t always exists.*

We note that Cheng *et al.* [2020] also provide a $poly(m^t, 1/\epsilon)$ time algorithm to compute an ϵ-approximately stable lottery. Using that in our analysis only affects the distortion bound by a factor of $1 + \epsilon$. For simplicity, we work with exactly stable lotteries.

Given ranked preferences, Ebadian *et al.* [2022] show that if $\mathcal{S}$ is a stable lottery over committees of size $t = \sqrt{m}$, then picking a candidate uniformly at random from a committee $S \sim \mathcal{S}$ with probability $1/2$ and picking a uniformly random candidate from C with probability $1/2$ yields distortion $O(\sqrt{m})$ for single winner selection with $p = 1$.

We want to extend this to select a committee of size k given only top-t preferences. Our rule, f^{mix}, is a combination of two rules.

- f^{unif} picks a uniformly random committee U of size k.
- f^{stable} arbitrarily completes the partial preference profile into a ranked preference profile, finds a stable lottery $\mathcal{S}$ over committees of size $k\sqrt{m}$, samples $S \sim \mathcal{S}$, and then picks a uniformly random subset $S' \subseteq S$ of size k.

If $k > \sqrt{m}$, f^{mix} applies f^{unif}. Otherwise, it applies f^{stable} with probability $1/2$ and f^{unif} with probability $1/2$.

Note that while Ebadian *et al.* [2022] use a stable lottery over committees of size $\sqrt{m}$ to pick a single candidate, f^{stable} uses a stable lottery over committees of size $k\sqrt{m}$ to pick a committee of size k. While this approach does not work when $k > \sqrt{m}$ (since then $k\sqrt{m} > m$), that case turns out to be rather easy to address. Finally, note that we are able to handle the partial top-t preferences by simply extending them arbitrarily to complete ranked preferences!

Theorem 4. *For all $k, t \in [m]$, we have that*

$$
\mathrm{dist}(f^{mix}) \leqslant \min\left(\frac{2m}{k}, 4\max\left(\frac{m}{t}, \sqrt{m} \right) \right).
$$

Proof. We prove two separate upper bounds of $2m/k$ and $4\max(m/t, \sqrt{m})$ on $\mathrm{dist}(f^{\mathrm{mix}})$. Fix an arbitrary instance (V, C, u) with top-t preference profile σ induced by u. Let $D = f^{\mathrm{mix}}(\sigma)$ be the distribution return by our rule, and $q_a = \Pr_{S \sim D}[a \in S]$ be the marginal probability of candidate a being included in the chosen committee. Fix an optimal committee $S^* \in \arg\max_{S \in \mathcal{P}_k(C)} \mathsf{sw}(S)$.

First bound: Since f^{mix} executes f^{unif} with probability at least $1/2$, we have that $q_a \geqslant k/(2m)$ for all $a \in C$. Hence,

we have $\mathbb{E}_{S \sim D}[\mathsf{sw}(S, u)] \geq (k/(2m)) \cdot \mathsf{sw}(S^*)$. Rearranging yields the desired distortion bound.

Second bound: Our desired bound is $4 \max(m/t, \sqrt{m})$. We assume $k \leq \sqrt{m}$, otherwise $2m/k$ is already a stronger bound. Let $\hat{\sigma}$ denote the arbitrarily completed ranked preference profile, and let $\mathcal{S}$ be the stable lottery computed in f^{stable} for $\hat{\sigma}$. Fix a committee S in the support of $\mathcal{S}$. Let us partition the set of voters V into three:

- $V(S, S^*)$ includes every voter i for whom $S^* \succ_i S$ under $\hat{\sigma}$. From Definition 1, $E_{S \sim \mathcal{S}}[|V(S, S^*)|] \leq n/\sqrt{m}$.

- $G(S^*, S)$ includes every voter i for whom $S \succ_i S^*$ and she ranks her favorite candidate from S in the first t positions. This guarantees $u_i(S) \geq u_i(S^*)$.

- $N(S^*, S)$ includes every voter i for whom $S \succ_i S^*$ but she ranks her favorite candidate from S after the first t positions. In this case, $u_i(S^*) \leq 1/t$.

Now, we have

$\mathsf{sw}(S^*, u)$

$$= \sum_{i \in V(S, S^*)} u_i(S^*) + \sum_{i \in N(S, S^*)} u_i(S^*) + \sum_{i \in G(S, S^*)} u_i(S^*)$$

$$\leq |V(S, S^*)| \cdot 1 + n \cdot (1/t) + \sum_{i \in G(S, S^*)} u_i(S)$$

$$\leq |V(S, S^*)| + n/t + \mathsf{sw}(S, u).$$

Next, we take the expectation over $S \sim \mathcal{S}$.

$$\mathsf{sw}(S^*, u) \leq \frac{n}{\sqrt{m}} + \frac{n}{t} + \mathbb{E}_{S \sim \mathcal{S}}[\mathsf{sw}(S, u)]$$

$$\leq \frac{2n}{\min(\sqrt{m}, t)} + \mathbb{E}_{S \sim \mathcal{S}}[\mathsf{sw}(S, u)]. \qquad (1)$$

Let W_1 be the expected social welfare under f^{unif} and W_2 be the expected social welfare under f^{stable}. The expected social welfare under f^{mix} is $(W_1 + W_2)/2$. We express the RHS in Equation (1) in terms of W_1 and W_2.

First, Caragiannis *et al.* [2017] argue that W_1 is at least n/m. Next, consider $S' \subseteq S$ of size $|S'| = k$ chosen uniformly at random. For each voter i, her most favorite candidate in S is included in S' with probability $|S'|/|S| = 1/\sqrt{m}$. Hence, $\mathbb{E}_{S'}[u_i(S')] \geq u_i(S)/\sqrt{m}$. Summing over all voters and taking the expectation over $S \sim \mathcal{S}$, $W_2 = \mathbb{E}_{S, S'}[\mathsf{sw}(S', u)] \geq \mathbb{E}_S[\mathsf{sw}(S, u)]/\sqrt{m}$. Hence, $\mathbb{E}_S[\mathsf{sw}(S, u)] \leq \sqrt{m} \cdot W_2$.

Plugging these into Equation (1), we get

$$\mathsf{sw}(S^*, u) \leq \frac{2m \cdot W_1}{\min(\sqrt{m}, t)} + \sqrt{m} \cdot W_2$$

$$\leq 2 \max(\sqrt{m}, m/t) \cdot (W_1 + W_2),$$

which yields the desired distortion bound of $4 \max(\sqrt{m}, m/t)$ upon rearranging. □

4.2 Lower Bound

We now turn our attention to lower bounds. Caragiannis *et al.* [2017] already prove a lower bound of $\Omega(\min(m/k, \sqrt{m}))$ that holds even with fully ranked preferences ($t = m$), which obviously holds for all $t \leq m$. This matches the upper bound from Theorem 4 when $k \geq \sqrt{m}$ or when $k \leq \sqrt{m} \leq t$. In the remaining region of $k, t \leq \sqrt{m}$, the upper bound from Theorem 4 is $O(\min(m/k, m/t))$; for this case, we are able to establish a weaker lower bound of $\Omega(m/(kt))$. We do not provide a separate proof of the $\Omega(m/(kt))$ lower bound because it is implied by Theorem 6 in the next section.

Proposition 2. *Every randomized rule f for selecting a committee of size k given top-t preferences has*

$$\text{dist}(f) = \Omega\left(\min\left(\frac{m}{k}, \max\left(\frac{m}{kt}, \sqrt{m}\right)\right)\right).$$

Crucially, note that there is no gap between our upper and lower bounds when $k = O(1)$, $t = O(1)$, $k \geq \sqrt{m}$, or $k \leq \sqrt{m} \leq t$. Particularly, for ranked preferences ($t = m$), we derive a tight distortion bound of $\Theta(\min(m/k, \sqrt{m}))$, which was posed as an open question by Caragiannis *et al.* [2017]. Their upper bound was loose by a factor of $O(m^{1/6})$. While our upper bound in Theorem 4 eliminates this completely using a technique very different from theirs, our upper bound in the next section would show that even their technique can be modified to eliminate this factor up to a logarithmic term.

5 Committee Selection for Higher Moments

Finally, we consider the p-th moment distortion, with $p > 1$, for selecting a committee of size k given top-t preferences. Unfortunately, the ingenious approach of Ebadian *et al.* [2022] to utilize stable lotteries to bound distortion seems to break down for higher moments. If we want a committee sampled from such a lottery to well-approximate the optimal p-th moment of the social welfare, it forces us to use a lottery over committees larger than kt; however, this reduces the performance of subsampling of a size-k committee, resulting in unappealing distortion bounds.

In contrast, we prove that the approach of Caragiannis *et al.* [2017], which extends the harmonic score based approach of Boutilier *et al.* [2015], continues to work reasonably well for higher moments. In doing so, we identify and improve upon a suboptimal step in their approach. For $p = 1$ and $t = m$, this is what reduces their $m^{1/6}$ gap to a logarithmic gap, as mentioned above.

5.1 Upper Bound

Let us define our harmonic score based rule f^{hc} for committee selection. The rule is independent of p. Given a top-t preference profile σ, the rule works as follows. With probability $1/2$, it picks a uniformly random committee of size k. With the remaining probability $1/2$, it does the following. First, it computes the truncated harmonic score $\mathsf{hsc}_t(c)$ of every candidate c, as defined in Section 3. At this point, Caragiannis *et al.* [2017], whose rule is only defined for complete rankings with $t = m$, hence using hsc instead of hsc_t, define a marginal probability $q_a = \alpha \cdot (k/m) + (1-\alpha) \cdot k \cdot \frac{\mathsf{hsc}(a)}{\sum_{c \in C} \mathsf{hsc}(c)}$, find α such that $q_a \leq 1$ for all a, and compute a distribution over committees matching these marginal probabilities (which can be done efficiently using an extension of the

Birkhoff-von Neumann theorem due to Budish *et al.* [2013]). Instead, we compute a distribution over committees such that the marginal probability of each candidate a being included is *at least* $q_a = \min(k \cdot \frac{\text{hsc}_t(a)}{\sum_{c \in C} \text{hsc}_t(c)}, 1)$. It can be shown that this is always feasible (indeed, $\sum_{a \in C} q_a \leqslant k$ and $q_a \in [0, 1]$ for all a) and efficiently computable. This change in the marginal probabilities allows us to improve upon their bounds.

Theorem 5. *For all $p \geqslant 1$ and $k, t \in [m]$, we have that*

$$\text{dist}^p(f^{hc}) \leqslant 2 \cdot \min \left(\binom{m}{k}, \, m \cdot k^{p-2}, \right.$$

$$\left. 4^p \cdot \max \left((H_t \cdot m \cdot k^{p-1})^{\frac{p}{p+1}}, \left(\frac{m}{t} \right)^p \right) \right).$$

5.2 Lower Bounds

Next, we establish two lower bounds via different proof methodologies. The first bound is achieved using a more straightforward analysis.

Theorem 6. *Fix constant $p \geqslant 1$. Every randomized rule f for selecting a committee of size k given top-t preferences has*

$$\text{dist}^p(f) = \Omega \left(\min \left(\frac{m}{k}, \left(\frac{m}{kt} \right)^p \right) \right).$$

The next bound requires a more intricate analysis; a proof sketch is presented below.

Theorem 7. *Fix constant $p \geqslant 1$. Every randomzied rule f for selecting a committee of size k given top-t preferences has*

$$\text{dist}^p(f) = \begin{cases} \Omega \left(k^{\Theta(p)} \right), & \text{if } k = O(\sqrt{m}), \\ \Omega \left(\left(\frac{m-k}{k} \right)^p \right), & \text{if } k = \Omega(\sqrt{m} \log m). \end{cases}$$

Proof sketch. At a high level, the proof works as follows. We construct a simple profile with $n = m$ voters, each ranking a different candidate first; the rest of the profile is arbitrary. Further, the utilities are always such that there is a unique optimal committee S^*. All voters that rank a candidate $a \in S^*$ first have utility 1 for a and 0 for the other candidates; all other voters have utility $1/m$ for all candidates. Note that under these utilities, S^* uniquely has the highest p-th moment social welfare of $(k + \frac{m-k}{m})^p$. We then use an averaging argument to claim that regardless of the distribution picked by the rule on this profile, we can find some S^* and its corresponding utility profile so that the rule only achieves expected p-th moment social welfare of at most $\frac{1}{\binom{m}{k}} \cdot \sum_{h=0}^{k} \binom{k}{h} \cdot \binom{m-k}{k-h} \cdot \left(h + \frac{m-k}{m} \right)^p$. The remainder of the proof is dedicated to bounding the ratio of these two values. Namely, we use the Chernoff bounds on binomial random variables to upper bound the binomial sum. This allows us to make the following claim: for all $c \geqslant 2 \cdot \frac{k^2}{m-k}$, the distortion is at least

$$\frac{1}{2} \cdot \min \left(\left(\frac{k+1}{c+1} \right)^p, 2^c \right).$$

The final part of the proof is optimizing the value of c to get the tightest bound. □

6 Discussion

Our work identifies exciting technical open questions. While we identify tight distortion bounds for single winner selection ($k = 1$), for committee selection ($k > 1$) with the first moment ($p = 1$), there is a gap between our upper bound of $O(m/kt)$ and our lower bound of $O(m/\max(k,t))$ in the case where $k, t \leqslant \sqrt{m}$. We remark that in practice, it is common for k and t to be very small, for which the gap between our upper and lower bounds is also small. However, it would be interesting to close these gaps.

In the full version, we also prove encouraging preliminary results for the metric distortion framework. We identify tight distortion bounds for single winner selection ($k = 1$) with an arbitrary moment p, but our bounds for committee selection are off by a polynomial factor. Closing this gap would also be of immediate interest.

More broadly, an interesting direction for future work is to study distortion under other realistic settings and thrifty elicitation methods. For example, what is the best possible distortion if we have access to the ranked (or top-t) preferences of only a subset of randomly sampled voters? What if these voters are not sampled randomly? What if the preferences of the voters are not worst case, but instead stochastic (and possibly correlated)?

Casting an even broader net, the distortion framework is highly versatile and can shed a new light on quantitatively evaluating the effectiveness of more complex collective decision-making paradigms such as distributed elections [Filos-Ratsikas *et al.*, 2019], participatory budgeting [Benade *et al.*, 2021], and primaries [Borodin *et al.*, 2019]. An exciting direction for the future is to use this framework to analyze real-world decision-making paradigms such as sortition [Flanigan *et al.*, 2021] and liquid democracy [Brill, 2019].

References

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

[Anagnostides *et al.*, 2021] Ioannis Anagnostides, Dimitris Fotakis, and Panagiotis Patsilinakos. Metric-distortion bounds under limited information. In *Proc. of 14th SAGT*, pages 299–313, 2021.

[Anshelevich *et al.*, 2021] Elliot Anshelevich, Aris Filos-Ratsikas, Nisarg Shah, and Alexandros A. Voudouris. Distortion in social choice problems: The first 15 years and beyond. In *Proc. of 30th IJCAI*, pages 4294–4301, 2021. Survey Track.

[Arrow, 1951] Kenneth Arrow. *Social Choice and Individual Values*. Wiley, 1951.

[Aziz, 2020] Haris Aziz. Justifications of welfare guarantees under normalized utilities. *ACM SIGecom Exchanges*, 17(2):71–75, 2020.

[Benade *et al.*, 2021] Gerdus Benade, Swaprava Nath, Ariel D. Procaccia, and Nisarg Shah. Preference elicitation for participatory budgeting. *Management Science*, 67(5):2813–2827, 2021.

[Bentham, 1780] Jeremy Bentham. *An Introduction to the Principles of Morals and Legislation*. Dover Publications, 1780.

[Borodin *et al.*, 2019] Allan Borodin, Omer Lev, Nisarg Shah, and Tyrone Strangway. Primarily about primaries. In *Proc. of 33rd AAAI*, pages 1804–1811, 2019.

[Boutilier *et al.*, 2015] Craig Boutilier, Ioannis Caragiannis, Simir Haber, Tyler Lu, Ariel D. Procaccia, and Or Sheffet. Optimal social choice functions: A utilitarian view. *Artificial Intelligence*, 227:190–213, 2015.

[Brill, 2019] Markus Brill. Interactive democracy: New challenges for social choice theory. In *The Future of Economic Design*, pages 59–66. Springer, 2019.

[Budish *et al.*, 2013] Eric Budish, Yeon-Koo Che, Fuhito Kojima, and Paul Milgrom. Designing random allocation mechanisms: Theory and applications. *American economic review*, 103(2):585–623, 2013.

[Caragiannis *et al.*, 2017] Ioannis Caragiannis, Swaprava Nath, Ariel D. Procaccia, and Nisarg Shah. Subset selection via implicit utilitarian voting. *Journal of Artificial Intelligence Research*, 58:123–152, 2017.

[Cheng *et al.*, 2020] Yu Cheng, Zhihao Jiang, Kamesh Munagala, and Kangning Wang. Group fairness in committee selection. *ACM Transactions on Economics and Computation (TEAC)*, 8(4):1–18, 2020.

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

[Drummond and Boutilier, 2013] Joanna Drummond and Craig Boutilier. Elicitation and approximately stable matching with partial preferences. In *Proc. of 23rd IJCAI*, pages 97–105, 2013.

[Dwork *et al.*, 2001] Cynthia Dwork, Ravi Kumar, Moni Naor, and Dandapani Sivakumar. Rank aggregation methods for the web. In *Proc. of 10th WWW*, pages 613–622, 2001.

[Ebadian *et al.*, 2022] Soroush Ebadian, Anson Kahng, Dominik Peters, and Nisarg Shah. Optimized fairness and distortion in voting. In *Proc. of 23rd EC*, 2022. Available at https://arxiv.org/abs/2205.15760.

[Ephrati and Rosenschein, 1997] Eithan Ephrati and Jeffrey S. Rosenschein. A heuristic technique for multi-agent planning. *Annals of Mathematics and Artificial Intelligence*, 20(1):13–67, 1997.

[Fain *et al.*, 2017] Brandon Fain, Ashish Goel, Kamesh Munagala, and Sukolsak Sakshuwong. Sequential deliberation for social choice. In *Proc. of 13th WINE*, pages 177–190, 2017.

[Fain *et al.*, 2020] Brandon Fain, William Fan, and Kamesh Munagala. Concentration of distortion: The value of extra voters in randomized social choice. In *Proc. of 29th IJCAI*, pages 110–116, 7 2020.

[Filmus and Oren, 2014] Yuval Filmus and Joel Oren. Efficient voting via the top-k elicitation scheme: a probabilistic approach. In *Proc. of 15th EC*, pages 295–312, 2014.

[Filos-Ratsikas *et al.*, 2019] Aris Filos-Ratsikas, Evi Micha, and Alexandros A. Voudouris. The distortion of distributed voting. In *Proc. of 12th SAGT*, pages 312–325. Springer, 2019.

[Flanigan *et al.*, 2021] Bailey Flanigan, Paul Gölz, Anupam Gupta, Brett Hennig, and Ariel D. Procaccia. Fair algorithms for selecting citizens' assemblies. *Nature*, 596(7873):548–552, 2021.

[Gkatzelis *et al.*, 2020] Vasilis Gkatzelis, Daniel Halpern, and Nisarg Shah. Resolving the optimal metric distortion conjecture. In *Proc. of 61st FOCS*, pages 1427–1438. IEEE, 2020.

[Gross *et al.*, 2017] Stephen Gross, Elliot Anshelevich, and Lirong Xia. Vote until two of you agree: Mechanisms with small distortion and sample complexity. In *Proc. of 31st AAAI*, 2017.

[Halpern and Shah, 2021] Daniel Halpern and Nisarg Shah. Fair and efficient resource allocation with partial information. In *Proc. of 30th IJCAI*, pages 224–230, 2021.

[Haynes *et al.*, 1997] Thomas Haynes, Sandip Sen, Neeraj Arora, and Rajani Nadella. An automated meeting scheduling system that utilizes user preferences. In *Proc. of 1st AGENTS*, pages 308–315, 1997.

[Hosseini *et al.*, 2021] Hadi Hosseini, Vijay Menon, Nisarg Shah, and Sujoy Sikdar. Necessarily optimal one-sided matchings. In *Proc. of 35th AAAI*, pages 5481–5488, 2021.

[Kempe, 2020a] David Kempe. An analysis framework for metric voting based on lp duality. In *Proc. of 34th AAAI*, pages 2079–2086, 2020.

[Kempe, 2020b] David Kempe. Communication, distortion, and randomness in metric voting. In *Proc. of 34th AAAI*, pages 2087–2094, 2020.

[Lang and Xia, 2016] Jérôme Lang and Lirong Xia. Voting in combinatorial domains. In Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 9, pages 197–222. Cambridge University Press, 2016.

[Lee *et al.*, 2014] David T. Lee, Ashish Goel, Tanja Aitamurto, and Hélène Landemore. Crowdsourcing for participatory democracies: Efficient elicitation of social choice functions. In *Proc. of 2nd AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*, pages 133–142, 2014.

[Lesser *et al.*, 2003] Victor Lesser, Charles L. Ortiz Jr., and Milind Tambe. *Distributed sensor networks: A multiagent perspective*, volume 9. Springer Science & Business Media, 2003.

[Lu and Boutilier, 2011] Tyler Lu and Craig Boutilier. Vote elicitation with probabilistic preference models: Empirical estimation and cost tradeoffs. In *Proc. of 2nd ADT*, pages 135–149, 2011.

[Mandal *et al.*, 2019] Debmalya Mandal, Ariel D. Procaccia, Nisarg Shah, and David P. Woodruff. Efficient and thrifty voting by any means necessary. In *Proc. of 33rd NIPS*, pages 7178–7189, 2019.

[Mandal *et al.*, 2020] Debmalya Mandal, Nisarg Shah, and David P. Woodruff. Optimal communication-distortion tradeoff in voting. In *Proc. of 21st EC*, pages 795–813, 2020.

[Oren *et al.*, 2013] Joel Oren, Yuval Filmus, and Craig Boutilier. Efficient vote elicitation under candidate uncertainty. In *Proc. of 23rd AAAI*, 2013.

[Pitt *et al.*, 2006] Jeremy Pitt, Lloyd Kamara, Marek Sergot, and Alexander Artikis. Voting in multi-agent systems. *The Computer Journal*, 49(2):156–170, 2006.

[Procaccia and Rosenschein, 2006] Ariel D. Procaccia and Jeffrey S. Rosenschein. The distortion of cardinal preferences in voting. In *Proc. of 10th CIA*, pages 317–331, 2006.

[Procaccia *et al.*, 2012] Ariel D. Procaccia, Sashank J. Reddi, and Nisarg Shah. A maximum likelihood approach for selecting sets of alternatives. In *Proc. of 28th UAI*, pages 695–704, 2012.

[Shoham and Leyton-Brown, 2008] Yoav Shoham and Kevin Leyton-Brown. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2008.