Efficient Voting via The Top-k Elicitation Scheme: A Probabilistic Approach

Yuval Filmus and Joel Oren

Abstract

Top-k voting is a common form of preference elicitation due to its conceptual simplicity both on the voters' side and on the decision maker's side. In a typical setting, given a set of candidates, the voters are required to submit only the k-length prefixes of their intrinsic rankings of the candidates. The decision maker then tries to correctly predict the winning candidate with respect to the *complete* preference profile according to a prescribed voting rule. This raises a tradeoff between the communication cost (given the specified value of k), and the ability to correctly predict the winner.

We focus on *arbitrary* positional scoring rules in which the voters' scores for the candidates is given by a vector that assigns the ranks real values. We study the performance of top-k elicitation under three probabilistic models of preference distribution: a neutral distribution (impartial culture); biased distributions, such as the Mallows distribution; and worst-case (but fully known) distributions.

1 Introduction

The ongoing spread of large-scale, multi-user platforms has raised significant computational problems. One obvious example that frequently shows up in the context of recommendation and group decision making, is the need to efficiently aggregate user preferences. To elaborate on this point, consider a scenario with n individual agents (which we call *voters*), having preferences over a set C of candidates (or alternatives), in which a "consensus" (winning) candidate should be selected according to some predetermined rule. The need for efficiency sometimes dictates that we, the decision-makers, be judicious in the manner in which we elicit the preferences of the users. Many of these settings raise algorithmic questions that pertain to the extraction and aggregation of the votes, and to computing the correct winner.

We focus on the issue of efficient preference elicitation. In a system with a large collection of candidates to choose from, requiring the voters to submit complete rankings of the candidates is often ill-advised, and even infeasible, due to the resulting communication and cognitive overhead. Therefore, the task of devising protocols for obtaining the voters' preferences, while keeping the amount of communicated information down to a minimum, is imperative.

To contrast general communication complexity results, which state that in the worst-case, many voting rules require much information from the voters, empirical studies have shown that in practice, some elections are amenable to efficient voting protocols (e.g., [8]). One recent way of bridging this gap between the theoretical bounds and the empirical findings is to take a belief-based approach, by assuming that the preferences are distributed according to some specified prior. Given such probabilistic beliefs, the common goal is to design protocols for efficiently eliciting parts of the voters' preferences, and then deciding on the winner with a reasonable degree of confidence.

One straightforward method of elicitation is the top-k voting method: given a set C of m candidates, each of the n voters submits a ranking of their k most favored candidates (i.e., the k-length prefix of their intrinsic ranking of the candidates C). The decision maker then employs a prescribed voting rule for selecting a candidate based solely on the partially reported preference. The immediate question that this setting raises is: what is a sufficient bound on k that, would guarantee the selection of the correct candidate? (had he had the complete preference profile).

In this paper, we present a technique for studying the performance of this elicitation method based on a probabilistic analysis of the distribution of the scores. We primarily focus on a particular class of voting rules known as *positional scoring rules*. Given a non-increasing vector $\boldsymbol{\alpha} \in \mathbb{R}^m$ and a ranking of the candidates π_i, corresponding to voter i's preferences, candidate $c \in C$ receives a score of $\alpha(j)$ if c is ranked j'th in π_i. The winning candidate is the candidate with the maximal total score.

In the top-k voting scheme, each voter i reports only the k-length prefix of her intrinsic ranking $\pi_i^{-1}(1), \dots, \pi_i^{-1}(k)$. The decision maker, in turn, selects a candidate based solely on this partial view of the preference profiles.

Contributions We begin by studying the performance of top-k voting under the neutral, impartial culture distribution, in which the preferences are drawn uniformly at random (Section 3). Our study aims to find, for a given positional scoring rule, a closed-form criterion for the range of k for which it is possible to predict the winning candidate with high probability given only the k-length prefixes of the rankings. We state our results in terms of a measure we call the *partition variability ratio*, which is monotonically increasing in k. When this ratio is small, we show that no algorithm can predict the winning candidate with high probability. When the ratio crosses a certain threshold, we give a concrete algorithm (Algorithm FairCutoff) that predicts the winning candidate with high probability.

We demonstrate the use of our criterion on several scoring rules. This part of our work can be thought of as a direct extension of key aspects of the model studied by Oren et al. [15], in which we offer a more unified and general approach to top-k voting. In particular, we show that for the Borda scoring rule, no top-k can determine the correct winner w.h.p. unless $k = \Omega(m)$. This gives a $\log m$-factor improvement over the $\Omega(m/\log m)$ bound previously proved by Oren et al.

In Section 4, we further illustrate our general approach by providing a similar analysis for the Copeland voting rule (though the details differ significantly from the proof of Theorem 3.1). This results in a lower bound of $\Omega(m/\sqrt{\log m})$ (Theorem 4.1).

In Section 5, we proceed to analyze the limiting behavior of top-k voting under positional scoring rules and a class of biased distributions over preferences, in which there is a candidate that dominates all other candidates.

In Section 6, we take an adversarial learning approach, by considering arbitrary preference distributions (but fully known to the decision maker). We obtain a lower bound of $k = \Omega(m)$ for the harmonic positional scoring rule, where the score associated with rank i is $1/i$, by constructing an appropriate distribution over preferences. This contrasts our logarithmic bound for the impartial culture. We also show that under *any* preference distribution, an exponentially decaying score vector requires only $k = O(\log m)$ for correct winner determination, for sufficiently large m, and n.

In Section 7, we empirically demonstrate the efficacy of our top-k elicitation method, and illustrate the bounds obtained by our criteria.

Previous work There has been a growing body of literature in computational social choice that studies worst-case objectives pertaining to partial preference elicitation. These studies typically focus on heuristics for determining potential winners (the so-called possible winners) and the necessary winners; i.e., candidates who are guaranteed to win irrespective of any complete extension of the preferences (see e.g., [9, 16]). Baumeister et al. [2] have studied these complexity issues in the context of top-k voting.

A number of studies have shown that in the worst case, many of the common voting rules may require the voters to communicate a significant amount of information about their preferences for predicting the correct winner with absolute certainty; either in the communication complexity sense, or with respect to concrete elicitation protocols (e.g., [6, 16, 5]). This implies that top-k voting is ineffective for arbitrary preference profiles.

On the other hand, the practical efficacy of methods for these objectives, including top-k voting, has been empirically demonstrated by Kalech et al. [8]. This prompts the adoption of a probabilistic approach in which the votes are assumed to be drawn according to a probabilistic model. An important example of such a model is the Mallows ϕ-distribution [13, 14]. The so-called impartial model assumption, is a special case of the Mallows distribution, in which the preferences are assumed to be drawn uniformly at random from the complete set of rankings.

Lu and Boutilier [10, 11] adopted this approach, in settings where they took a regret-minimization towards optimizing the score of the selected candidate.

Our work is as a continuation of a model studied by Oren et al. [15], which provided an analysis of the top-k elicitation scheme under a distributional assumption on both the preferences, *and* the availability of the candidates, for predicting the correct Borda winner. They showed a lower bound of $\Omega(m/\log m)$ on k under the impartial culture assumption (improved in this paper). We generalize their probabilistic argument to handle *arbitrary* scoring rules.

In a recent study, Caragiannis et al. [4] studied the ability of scoring rules to reconstruct the underlying "true" ranking, based only on noisy rankings. Some of our results on biased distributions make similar generalizations of distributions such as the Mallows distribution, and employ similar techniques. They show that in the limiting case (where n goes to infinity), broad classes of scoring rules can correctly determine the underlying ranking.

2 Preliminaries

We consider a setting with a set $C = \{c_1, \ldots, c_m\}$ of m candidates. Let $\mathcal{L}$ the set of all possible ordinal preferences of C, where an ordering $\pi \in \mathcal{L}$ is a permutation $\pi\colon C \to [m]$, mapping candidates to *ranks*. That is, for a preference π and $1 \le j < j' \le m$, we say that $\pi_i^{-1}(j)$ is preferred over $\pi_i^{-1}(j')$ by a voter with preference π.

We also let $N = \{1, \ldots, n\}$ denote the set of voters, such that with each voter $i \in N$, there is an associated preferences $\pi_i \in \mathcal{L}$. We let $P = \{\pi_i\}_{i \in N} \in \mathcal{L}^n$ denote the *preference profile*; i.e., the set of all voter preferences. It is commonly assumed that $n \gg m$, and that the preferences are drawn according to some probabilistic model. We will describe some of these models below.

Another key component of our setting is a *voting rule* v, which is a function that selects a "winning" candidate based on the preference profile. Formally, we have that $v\colon \mathcal{L}^n \times 2^C \to C$ (in the literature, this is sometimes referred to as a *social welfare function*). We are particularly interested in a broad class of voting rules called *score based rules*. Given the preference profile P, a score based scoring rule relies on a function $sc\colon \mathcal{L}^n \to \mathbb{R}$ that assigns a score to each of the candidates. The election winner under such a rule is the candidate having the maximal score.

Positional Scoring rules A positional scoring rule is characterized by a *score vector* $\boldsymbol{\alpha} \in \mathbb{R}^m_{\ge 0}$ of non-increasing scores: $\alpha(j) \ge \alpha(j+1)$ for $1 \le j \le m-1$. The score given by a voter $i \in N$ for a candidate $c \in C$, ranked j'th in π_i, is $\alpha(j) = \alpha(\pi_i(c))$. We denote the average score of a candidate $c \in C$, by $sc_{\boldsymbol{\alpha}}(c) = \frac{1}{n} \sum_{i \in N} \alpha(\pi_i(c))$. When the score vector is known from context, we omit the subscript $\boldsymbol{\alpha}$, for notational convenience. The winner of the election is the candidate with the highest average score: $\arg\max_{c \in C} sc(c)$.

Examples of positional scoring rules include (1) the Borda scoring rule, for which the score vector is $\boldsymbol{\alpha}_B = (m-1, m-2, \ldots, 0)$, (2) the plurality (majority) scoring rule, in which corresponding score vector is $\boldsymbol{\alpha}_P = (1, 0, 0, \ldots, 0)$, (3) the k-approval rule, which is characterized by the score a vector with a prefix of k 1's followed by zeros; this allows each voter to specify which set of k candidates he "approves".

We also study the (non-positional) Copeland rule, which can be defined as follows. We say that c_i *beats* c_j in a pairwise election if the number of votes in P, in which c_i precedes c_j is larger than the number of votes in P which c_j precedes c_i. The score of a candidate c, $sc(c)$, is the number of candidates that she beats. As with all scoring rules, the candidate with maximal score wins the election. The Copeland scoring rule is tightly related to the notion of *Condorcet compatible voting rules:* the winning candidates receives the majority of the votes in a pairwise election with any other candidate. Indeed, a Condorcet winner is always a Copeland winner.

Top-k elicitation For a given integer k between 1 and m, the decision maker asks the voters to report only the k-length prefixes of their preference rankings, $(\pi_i^{-1}(1), \ldots, \pi_i^{-1}(k))$, for every $i \in N$, and has to make a decision based only on these prefixes. The goal of the decision maker is to recover the true winner given only the k-length prefixes.

Given the distribution of the preferences and a prescribed voting rule, we would like to determine the range of k for which the decision maker can predict the winner with high probability, that is with probability tending to 1 as m grows.

Distributional models of preferences We consider various models of distributions over preferences. Many of these models are characterized by an underlying "canonical" preference, the probabilities of the different preferences decaying monotonically with their dissimilarity to the canonical preference, as measured by some distance function.

A common such distance metric for permutations is the Kendall tau distance, defined by $d_{KT}(\pi_1, \pi_2) = |\{c, c' : \pi_1^{-1}(c) < \pi_1^{-1}(c') \text{ and } \pi_2^{-1}(c) > \pi_2^{-1}(c')\}|$. The popular Mallows distribution is specified by a fraction $\phi \in [0, 1]$, in addition to the reference ranking, and the probability of a preference decreases exponentially with its distance to the reference ranking: $Pr[\pi] = \phi^{d_{KT}(\pi, \hat{\pi})} / Z_m$, where Z_m is a normalizing term.

Whenever we state that the preferences are distributed according to a Mallows distribution $D(\hat{\pi}, \phi)$, we mean that the preferences are drawn i.i.d. from $D(\hat{\pi}, \phi)$.

A heavily used special instance of the Mallows distribution is the case $\phi = 1$, in which the preferences are sampled uniformly at random from $\mathcal{L}$ by each of the voters. This is also known as the *impartial culture* assumption (or succinctly, IC). We focus on this distribution in Section 3.

3 Top-k voting for positional scoring rules and a neutral prior

We begin with the model in which the preferences are assumed to be drawn from the uniform distribution over rankings $\mathcal{L}$. Our main goal is to provide a direct method for "mechanically" obtaining either upper or lower bounds on the minimum value of k necessary for determining the correct winning candidate, with higher probabilities.

Given the top-k part of the votes, our goal is to choose a candidate who will win with probability close to 1, if there is such a candidate. The "optimal" algorithm will compute (or estimate, if computational efficiency is required) the probability that each candidate wins, and choose the candidate with the maximal chance to win. However, such an algorithm doesn't seems to readily lend itself to systematic analysis. Instead, we consider the following

Algorithm FairCutoff: The top-k algorithm for positional scoring rules.

Input: Top-k votes: $(\pi_1^k, \ldots, \pi_n^k)$, where π_i^k denotes the top-k ranking of voter i over a set of k candidates. A score vector $\boldsymbol{\alpha}$.

1 **foreach** $c \in C$ **do**

2 Set $sc_i^T(c) = \begin{cases} \alpha(\pi_i(c)) & \text{if } \pi_i(c) \le k, \\ \frac{1}{m-k}\sum_{j=k+1}^m \alpha(j) & \text{otherwise.} \end{cases}$

3 **return** $\arg\max_{c \in C} \sum_{i=1}^n sc_i^T(c)$.

simple approach. For each candidate c, the top-k score that corresponds to voter i's vote, is the original score given in vector $\boldsymbol{\alpha}$, if the candidate is in i's top-k ranking. Otherwise, we assign it a score that corresponds to the expected score of c, had it been positioned uniformly at random in one of the bottom $m-k$ positions. The algorithm then selects the candidate with the maximal total score. The full details are given in Algorithm FairCutoff.[1] For the purpose of analysis, we define the complementary "bottom" score, given by: $sc^B(c) = \frac{1}{n}\sum_{i \in N} sc_i^B(c)$ where

$$sc_i^B(c) = \begin{cases} 0 & \text{if } \pi_i(c) \le k, \\ \alpha(\pi_i(c)) - \frac{1}{m-k}\sum_{j=k+1}^m \alpha(j) & \text{otherwise.} \end{cases}$$

We note that $sc(c) = sc^T(c) + sc^B(c)$.

We now present the main theorem of this section.

Theorem 3.1. *Define*

$$V_T = \frac{1}{m}\sum_{i=1}^k \alpha(i)^2 + \frac{1}{m(m-k)}\left(\sum_{i=k+1}^m \alpha(i)\right)^2 - \frac{1}{m^2}\left(\sum_{i=1}^m \alpha(i)\right)^2,$$

$$V_B = \frac{1}{m}\sum_{i=k+1}^m \alpha(i)^2 - \frac{1}{m(m-k)}\left(\sum_{i=k+1}^m \alpha(i)\right)^2.$$

*Furthermore, define the k-*partition variability ratio *to be $r_k = V_T/V_B$.*

Lower bound. *If $r_k = o(\log m)$ then no algorithm for predicting the overall winner given the top-k votes succeeds with probability $1 - o_m(1)$, when taking the limit $n \to \infty$. (That is, for each m we analyze the success probability for large enough n.)*

Upper bound. *If $r_k = \omega(\log^{4/3} m)$ then $c_{\max}$, i.e., the candidate with the maximum score based on Algorithm FairCutoff, is the overall winner with probability $1 - o_m(1)$, for large enough n.*

V_T measures the uncertainty coming from the top-k part of the votes (corresponding to sc^T), while V_B measures the uncertainty coming from the bottom part of the votes (corresponding to sc^B). When V_T/V_B is small, the bottom uncertainty dominates the top-k information, and so the winner cannot be determined given only the top-k part. When V_T/V_B is large, the top-k part dominates the "noise" coming from the bottom part of the votes. We defer the proof of the theorem to Subsection 3.2.

Theorem 3.1 gives a threshold phenomenon: as long as $V_T/V_B \ll \log m$, the winner cannot be predicted, while for $V_T/V_B \gg \log^{4/3} m$, Algorithm FairCutoff predicts the winner with

[1] We show that in practice, using our simulation results (Section 7), that the performance of the FairCutoff algorithm is quite comparable to the optimal algorithm.

high probability. Lemma 3.1 given below, shows that as k increases, the top uncertainty V_T grows while the bottom uncertainty V_B shrinks, and so the ratio V_T/V_B is increasing in k, implying that there has to exist such a threshold.

Lemma 3.1. *Let $V_T(k), V_B(k)$ be the quantities defined in Theorem 3.1. Then*

$$V_T(k) + V_B(k) = \frac{1}{m}\sum_{i=1}^{m}\alpha(i)^2 - \frac{1}{m^2}\left(\sum_{i=1}^{m}\alpha(i)\right)^2,$$

and $V_T(0) = 0$, $V_B(m-1) = 0$. If furthermore the scores $\alpha(1), \dots, \alpha(m)$ are non-increasing then $V_T(k)$ is non-decreasing and $V_B(k)$ is non-increasing.

Proof. A straightforward calculation gives the formula for $V_T(k) + V_B(k)$ and shows that $V_T(0) = V_B(m-1) = 0$. Suppose now that the scores are non-increasing. Since $V_T(k)+V_B(k)$ is independent of k, it is enough to show that $V_B(k)$ is non-increasing. We have

$$m(V_B(k-1) - V_B(k)) = \alpha(k)^2 - \frac{1}{m-k+1}\left(\sum_{i=k}^{m}\alpha(i)\right)^2 + \frac{1}{m-k}\left(\sum_{i=k+1}^{m}\alpha(i)\right)^2.$$

Let $S(k) = \sum_{i=k+1}^{m}\alpha(i)$. Then

$$\begin{aligned} m(V_B(k-1) - V_B(k)) &= \alpha(k)^2 - \frac{\alpha(k)^2 + S(k)^2 + 2\alpha(k)S(k)}{m-k+1} + \frac{S(k)^2}{m-k} \\ &= \frac{m-k}{m-k+1}\alpha(k)^2 + \frac{S(k)}{m-k+1}\left[\frac{S(k)}{m-k} - 2\alpha(k)\right]. \end{aligned}$$

Since $S(k) \le (m-k)\alpha(k)$,

$$m(V_B(k-1) - V_B(k)) \le \frac{m-k}{m-k+1}\alpha(k)^2 + \frac{(m-k)\alpha(k)}{m-k+1}\left[-\alpha(k)\right] = 0.$$

□

3.1 Application to common scoring rules

We now demonstrate its implications to the efficacy of the top-k voting method, when applied to different scoring rules. We begin with the Borda scoring rule. The following bound strengthens the bound given in [15]:

Theorem 3.2. *Suppose that the underlying election is held using the Borda voting rule. Then the top-k elicitation method requires $k = \Omega(m)$, in order to determine the correct Borda winner, with probability $1 - o_m(1)$, as as $n \to \infty$.*

Proof. We use the criterion given by Theorem 3.1. Calculating V_T and V_B, we obtain

$$\begin{aligned} V_T &= \frac{1}{m}\sum_{i=1}^{k}(m-i)^2 + \frac{1}{m(m-k)}\left(\sum_{i=k+1}^{m}(m-i)\right)^2 - \frac{1}{m^2}\left(\sum_{i=1}^{m}(m-i)\right)^2 \\ &= \frac{k(k^2 + 3m(m-k) - 1)}{12m}, \\ V_B &= \frac{1}{m}\sum_{i=k+1}^{m}(m-i)^2 - \frac{1}{m(m-k)}\left(\sum_{i=k+1}^{m}(m-i)\right)^2 = \frac{(m-k+1)(m-k)(m-k+1)}{12m}. \end{aligned}$$

Assuming $k \le m/2$, we have $V_T = \Theta(km)$ and $V_B = \Theta(m^2)$, so that $V_T/V_B = \Theta(k/m) = o(\log m)$. Therefore, no top-$k$ algorithm succeeds with probability at least $1 - o_m(1)$. □

Our next case study is the harmonic scoring rule, which was first proposed by Boutilier et al. [3].

Definition 1 (The harmonic scoring rule)**.** *The harmonic scoring rule is defined by the m-dimensional vector $\boldsymbol{\alpha_h}$, such that for $i \in [m]$, $\alpha_h(i) = 1/i$.*

As we now show, the harmonic tends to be quite amenable to efficient elicitation via our top-k elicitation method.

Theorem 3.3. *Consider the harmonic scoring rule. If $k = \omega(\log^{4/3} m)$ then FairCutoff selects the correct winner with probability $1 - o_m(1)$, for large enough n. On the other hand, if $k = o(\log m)$, no top-k algorithm can select the correct winner with probability $1 - o_m(1)$.*

Proof. We first calculate the two specified terms given in Theorem 3.1, assuming $k = o(m)$:

$$V_T = \frac{1}{m}\sum_{i=1}^{k}\frac{1}{i^2} + \frac{1}{m(m-k)}\Big(\sum_{i=k+1}^{m}\frac{1}{i}\Big)^2 - \frac{1}{m^2}\Big(\sum_{i=1}^{m}\frac{1}{m}\Big)^2$$
$$= \frac{\pi^2/6 - \Theta(\frac{1}{k})}{m} + \frac{\Theta(\log^2(\frac{m}{k}))}{m(m-k)} - \frac{\Theta(\log^2 m)}{m^2} = \Theta(\frac{1}{m}),$$

where the first equality follows from the elementary identities $\sum_{i=1}^{t}\frac{1}{i} = \log t \pm \Theta(1)$ and $\sum_{i=1}^{t}\frac{1}{i^2} = \frac{\pi^2}{6} - \Theta(\frac{1}{t})$. For the second equality we only used the fact that $k = o(m)$.

We similarly derive the second term:

$$V_B = \frac{1}{m}\sum_{i=k+1}^{m}\frac{1}{i^2} - \frac{1}{m(m-k)}\left(\sum_{i=k+1}^{m}\frac{1}{i}\right)^2$$
$$= \frac{\Theta(\frac{1}{k}) - \Theta(\frac{1}{m})}{m} - \frac{\Theta(\log^2(\frac{m}{k}))}{m(m-k)} = \Theta(\frac{1}{mk}).$$

Therefore when $k = o(m)$, we obtain $r_k = V_T/V_B = \Theta(k)$. The bounds follow by an application of Theorem 3.1. □

Our final case study is geometric scoring rules.

Definition 2 (Geometric scoring rules)**.** *The geometric scoring rule with parameter ρ is given by the m-dimensional vector $\alpha_\rho(i) = \rho^i$.*

Theorem 3.4. *Consider the geometric scoring rule with parameter ρ (not depending on m). If $k = \omega(\log\log m)$ then FairCutoff selects the correct winner with probability $1 - o_m(1)$, for large enough n. On the other hand, if $k = o(\log\log m)$, no top-k algorithm can select the correct winner with probability $1 - o_m(1)$.*

Proof. We calculate the specified terms given in Theorem 3.1, assuming $k \le m - 2$:

$$V_T = \frac{1}{m}\sum_{i=1}^{k}\rho^{2i} + \frac{1}{m(m-k)}\left(\sum_{i=k+1}^{m}\rho^i\right)^2 - \frac{1}{m^2}\left(\sum_{i=1}^{m}\rho^i\right)^2$$
$$= \frac{\Theta(1)}{m} + \frac{\Theta(\rho^{2k})}{m(m-k)} - \frac{\Theta(1)}{m^2} = \frac{\Theta(1)}{m},$$
$$V_B = \frac{1}{m}\sum_{i=k}^{m}\rho^{2i} - \frac{1}{m(m-k)}\left(\sum_{i=k+1}^{m}\rho^i\right)^2$$
$$= \frac{\Theta(\rho^{2k})}{m} - \frac{\Theta(\rho^{2k})}{m(m-k)} = \frac{\Theta(\rho^{2k})}{m}.$$

Therefore $V_T/V_B = \Theta(\rho^{-2k})$. The bounds follow from an application of Theorem 3.1. □

3.2 Proving Theorem 3.1

Before proceeding with the proof of the theorem, we define a few pieces of notation. Given a set of pre-defined random variables, $x_1, \ldots, x_m$, we let $x_{\max}$, and $x_{2\max}$ denote highest and second highest x_i values, respectively (note that they may be equal). Similarly, we abuse our notation a bit, by letting $c_{\max}$ and $c_{2\max}$ denote the candidates with the highest and second highest $sc^T(\cdot)$ values, among the candidates in C. Similar notations will be used for other sets of variables.

At a high-level, our approach is the following: For two distinct candidates $c, c' \in C$, let $D^T(c, c') = sc^T(c) - sc^T(c')$; i.e., the difference in their top-k scores (note that $D^T(c_{\max}, c_{2\max})$ is always non-negative). We will first aim to characterize the limiting behaviour of $D^B(c_{\max}, c_{2\max})$, for sufficiently large voter populations. Then, we will provide a similar characterization on the analogously defined $D^B(c_{\max}, c_{2\max}) = sc^B(c_{\max}) - sc^B(c_{2\max})$. Our bounds will then follow as a result of bounding the probability of the event in which $D^T(c_{\max}, c_{2\max}) + D^B(c_{\max}, c_{2\max}) < 0$. The first step in the proof is estimating $D^T(c_{\max}, c_{2\max})$. The strategy (due to Yury Makarychev [12]) is to reduce this to a question regarding the difference between the two largest elements in a vector of i.i.d. normal random variables.

We start by computing the mean, variance and covariance of the scores due to a *single* voter, and the corresponding data for the aggregated scores.

Lemma 3.2. *Define*

$$E_T = \frac{1}{m}\sum_{i=1}^{m} \alpha(i),$$
$$V_T = \frac{1}{m}\sum_{i=1}^{k} \alpha(i)^2 + \frac{1}{m(m-k)}\left(\sum_{i=k+1}^{m} \alpha(i)\right)^2 - \frac{1}{m^2}\left(\sum_{i=1}^{m} \alpha(i)\right)^2.$$

The mean, variance and covariance of the scores of a single voter i are $\mathbb{E}[sc_i^T(c)] = E_T$, $\mathrm{Var}[sc_i^T(c)] = V_T$ and $\mathrm{Cov}(sc_i^T(c), sc_i^T(c')) = -V_T/(m-1)$.

The mean, variance and covariance of the aggregated scores are $\mathbb{E}[sc^T(c)] = E_T$, $\mathrm{Var}[sc^T(c)] = V_T/n$ and $\mathrm{Cov}(sc^T(c), sc^T(c')) = -V_T/(n(m-1))$.

Proof. The average score is

$$\mathbb{E}[sc_i^T(c)] = \frac{1}{m}\left(\sum_{i=1}^{k} \alpha(i) + (m-k)\frac{1}{m-k}\sum_{i=k+1}^{m} \alpha(i)\right)$$
$$= \frac{1}{m}\sum_{i=1}^{m} \alpha(i).$$

In order to compute the variance, we first compute the second moment:

$$\mathbb{E}[sc_i^T(c)^2] =$$
$$\frac{1}{m}\left(\sum_{i=1}^{k} \alpha(i)^2 + (m-k)\left(\frac{1}{m-k}\sum_{i=k+1}^{m} \alpha(i)\right)^2\right)$$
$$= \frac{1}{m}\sum_{i=1}^{k} \alpha(i)^2 + \frac{1}{m(m-k)}\left(\sum_{i=k+1}^{m} \alpha(i)\right)^2.$$

The formula for the variance now immediately follows.

As for the covariance, let $\sigma = \mathrm{Cov}(sc_i^T(c), sc_i^T(c'))$ for any $c \neq c'$, and note that $\mathrm{Cov}(sc_i^T(c), sc_i^T(c)) = V_T$. Since $\sum_{c\in C} sc_i^T(c)$ is constant,

$$\begin{aligned}0 &= \mathrm{Cov}\left(\sum_{c\in C} sc_i^T(c), \sum_{c'\in C} sc_i^T(c')\right)\\ &= \sum_{c,c'\in C} \mathrm{Cov}(sc_i^T(c), sc_i^T(c'))\\ &= m(m-1)\sigma + mV_T.\end{aligned}$$

Therefore $\sigma = -V_T/(m-1)$.

Finally, we have $\mathbb{E}[sc^T(c)] = n\,\mathbb{E}[sc_i^T(c)/n] = E_T$ and $\mathrm{Var}[sc^T(c)] = n\,\mathrm{Var}[sc_i^T(c)/n] = n(V_T/n^2) = V_T/n$, and similarly $\mathrm{Cov}(sc^T(c), sc^T(c')) = -V_T/(n(m-1))$. ☐

We can now use the central limit theorem to reduce the estimation of $sc^T(c_{\max}) - sc^T(c_{2\max})$ to a question about Gaussians.

Lemma 3.3. *Let $r \sim \mathcal{N}(0, V_T/(n(m-1)))$, and let $y_j = sc^T(c_j) + r$. Then $\frac{1}{\sqrt{mV_t/(n(m-1))}}(y_1 - E_T, \dots, y_m - E_T)$ converges in distribution (as $n \to \infty$) to a standard multivariate normal distribution of dimension m (with zero mean and covariance matrix I_m).*

Proof. Let $r_i \sim \mathcal{N}(0, V_T/(m-1))$, and note that r has the same distribution as $(r_1 + \cdots + r_n)/n$. Therefore we can define $r = (r_1 + \cdots + r_n)/n$. We have $\mathbb{E}[y_j] = \mathbb{E}[sc^T(c_j)] = E_T$, $\mathrm{Var}[y_j] = \mathrm{Var}[sc^T(c_j)] + \mathrm{Var}[r] = V_T/n + V_T/(n(m-1)) = mV_T/(n(m-1))$ and $\mathrm{Cov}(y_j, y_k) = \mathrm{Cov}(sc^T(c_j), sc^T(c_k)) + \mathrm{Var}(r) = 0$. Since $(y_1, \dots, y_m)$ is an average of n i.i.d. well-behaved random variables $(sc_i^T(1) + r_i, \dots, sc_i^T(m) + r_i)$, the central limit theorem applies and shows that $(y_1, \dots, y_m)$ converges in distribution to m i.i.d. copies of $\mathcal{N}(E_T, mV_T/(n(m-1)))$. This implies the lemma. ☐

The trick here is that $y_j - y_k = sc^T(c_j) - sc^T(c_k)$. The question we need to solve now is the following: Suppose that $x_1, \dots, x_m$ are i.i.d. standard random variables; what is the typical value of $x_{\max} - x_{2\max}$? In order to obtain a concentration bound on this difference, we will seek to bound on both $x_{\max}^2 - x_{2\max}^2$ and $x_{\max} + x_{2\max}$, knowing that the ratio of these two terms will give us our desired bound.

We let $u_c = \overline{\Phi}(x_c)$, where $\overline{\Phi}$ is the *complementary* cumulative distribution function of a standard normal variable. The idea is to use the fact that $u_c = \overline{\Phi}(x_c) \sim U(0,1)$, and to analyze the typical values of $u_{\min} = \overline{\Phi}(x_{\max})$ and $u_{2\min} = \overline{\Phi}(x_{2\max})$ as well as the *ratio* $u_{2\min}/u_{\min}$. We are interested in the ratio since it is well known that

$$\log \overline{\Phi}(x) \approx -\frac{x^2}{2}.$$

and in particular,

$$\log \frac{u_{2\min}}{u_{\min}} \approx \frac{x_{\max}^2 - x_{2\max}^2}{2}.$$

We start our analysis with $u_{2\min}/u_{\min}$.

Lemma 3.4. *Let $1 \leq \ell_1 \leq \ell_2 \leq \infty$.*

$$\Pr[\ell_1 \leq \tfrac{u_{2\min}}{u_{\min}} \leq \ell_2] = \frac{1}{\ell_1} - \frac{1}{\ell_2}.$$

Proof. The cumulative distribution function of $u_{\min}$ is easily calculated to be $1-(1-u)^m$, and therefore its density is $m(1-u)^{m-1}$. Given $u_{\min}$, the other u_c's have distribution $U(u_{\min},1)$. Therefore the cumulative distribution function of $u_{2\min}$ is $1-(\frac{1-u}{1-u_{\min}})^{m-1}$. Therefore for $1 \leq \ell \leq \infty$,

$$\begin{aligned}
&\Pr[\tfrac{u_{2\min}}{u_{\min}} \geq \ell] \\
&= \Pr[u_{2\min} \geq \ell u_{\min}] \\
&= \int_0^{1/\ell} \Pr[u_{2\min} \geq \ell u | u_{\min} = u] m(1-u)^{m-1}\, du \\
&= \int_0^{1/\ell} \left(\frac{1-\ell u}{1-u}\right)^{m-1} m(1-u)^{m-1}\, du \\
&= \int_0^{1/\ell} m(1-\ell u)^{m-1}\, du \\
&= -\frac{(1-\ell u)^m}{\ell}\bigg|_0^{1/\ell} = \frac{1}{\ell}.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\Pr[\ell_1 \leq \tfrac{u_{2\min}}{u_{\min}} \leq \ell_2] \\
=&\Pr[\tfrac{u_{2\min}}{u_{\min}} \geq \ell_1] - \Pr[\tfrac{u_{2\min}}{u_{\min}} \geq \ell_2] \\
=&\frac{1}{\ell_1} - \frac{1}{\ell_2}.
\end{aligned}$$

□

Using this lemma, we can show that with high probability (with respect to m), both $x_{\max}$ and $x_{2\max}$ are $\Theta(\sqrt{\log m})$. We will need to use some estimates on tails of the normal distribution, starting with the following well-known estimate (e.g., [7]):

$$\frac{e^{-x^2/2}}{x\sqrt{2\pi}}\left(1-\frac{1}{x^2}\right) \leq \overline{\Phi}(x) \leq \frac{e^{-x^2/2}}{x\sqrt{2\pi}}. \tag{3.1}$$

Lemma 3.5.

$$-\frac{d}{dx}\log\overline{\Phi}(x) = x + O\left(\frac{1}{x}\right).$$

Proof. Since $\overline{\Phi}'(x) = -e^{-x^2/2}/\sqrt{2\pi}$ and $(\log\overline{\Phi}(x))' = \overline{\Phi}'(x)/\overline{\Phi}(x)$, we deduce from (3.1) that

$$x \leq -\frac{d}{dx}\log\overline{\Phi}(x) \leq \frac{x}{1-1/x^2} = x(1+O(1/x^2)).$$

□

Lemma 3.6. *With probability $1-o_m(1)$, both $x_{\max} = \Theta(\sqrt{\log m})$ and $x_{2\max} = \Theta(\sqrt{\log m})$.*

Proof. We start with a concentration estimate for $u_{\min}$:

$$\begin{aligned}
\Pr\left[\frac{1}{m^2} \leq u_{\min} \leq \frac{1}{\sqrt{m}}\right] &= \left(1-\frac{1}{\sqrt{m}}\right)^m - \left(1-\frac{1}{m^2}\right)^m \\
&=(1-o_m(1)) - o_m(1) = 1-o_m(1).
\end{aligned}$$

The estimate for $x_{\max}$ is immediate from (3.1). In order to handle $x_{2\max}$, we use Lemma 3.4. Choosing $\ell_1 = 1$ and $\ell_2 = m^{1/3}$, we see that $u_{2\min}/u_{\min} \leq m^{1/3}$ with probability $1 - 1/m^{1/3} = 1-o_m(1)$. Therefore with probability $1-o_m(1)$, $1/m^2 \leq u_{\min} \leq u_{2\min} \leq m^{1/3}u_{\min} \leq 1/m^{1/6}$. The estimate for $x_{2\max}$ is now immediate from (3.1). □

Putting everything together, we can prove our estimate on $x_{\max} - x_{2\max}$.

Lemma 3.7. *Let $1 \le \ell_1 \le \ell_2 \le \infty$. With probability $1/\ell_1 - 1/\ell_2 - o_m(1)$,*

$$\Omega\left(\frac{\log \ell_1}{\sqrt{\log m}}\right) \le x_{\max} - x_{2\max} \le O\left(\frac{\log \ell_2}{\sqrt{\log m}}\right).$$

Proof. Lemma 3.4 and Lemma 3.6 show that with probability $1/\ell_1 - 1/\ell_2 - o_m(1)$, the following estimates hold: $x_{\max} = \Theta(\sqrt{\log m})$, $x_{2\max} = \Theta(\sqrt{\log m})$, and $\ell_1 \le u_{2\min}/u_{\min} \le \ell_2$. We can restate the latter fact as

$$\log \ell_1 \le \log \overline{\Phi}(x_{2\max}) - \log \overline{\Phi}(x_{\max}) \le \log \ell_2.$$

The mean value theorem shows that

$$\frac{\log \overline{\Phi}(x_{2\max}) - \log \overline{\Phi}(x_{\max})}{x_{\max} - x_{2\max}} = -\frac{d}{dx} \log \overline{\Phi}(x^*)$$

for some $x_{2\max} \le x^* \le x_{\max}$. Clearly $x^* = \Theta(\sqrt{\log m})$, and so Lemma 3.5 shows that $-(d/dx) \log \overline{\Phi}(x^*) = \Theta(\sqrt{\log m})$. Therefore

$$x_{\max} - x_{2\max} = \frac{\log \overline{\Phi}(x_{2\max}) - \log \overline{\Phi}(x_{\max})}{\Theta(\sqrt{\log m})}.$$

The lemma easily follows. □

Combining this with Lemma 3.3, we obtain a similar result on $sc^T_{\max} - sc^T_{2\max}$.

Lemma 3.8. *Let $1 \le \ell_1 \le \ell_2 \le \infty$. With probability $1/\ell_1 - 1/\ell_2 - o_m(1) - o_n(1)$,*

$$\Omega\left(\log \ell_1 \sqrt{\frac{V_T}{n \log m}}\right) \le sc^T(c_{\max}) - sc^T(c_{2\max}) \le O\left(\log \ell_2 \sqrt{\frac{V_T}{n \log m}}\right).$$

Proof. First, note that $sc^T(c_{\max}) - sc^T(c_{2\max}) = y_{\max} - y_{2\max} = (y_{\max} - E_T) - (y_{2\max} - E_T)$. Since the mapping $(x_1, \ldots, x_m) \mapsto x_{\max} - x_{2\max}$ is continuous, Lemma 3.3 shows that $\frac{sc^T(c_{\max}) - sc^T(c_{2\max})}{\sqrt{mV_t/(n(m-1))}}$ converges in distribution to the distribution of $x_{\max} - x_{2\max}$. That means that up to an error factor of $o_n(1)$, we can translate the results of Lemma 3.7 to results about scores by multiplying throughout by $\Theta(\sqrt{V_t/n})$, which gives the lemma. □

As a corollary, we can show that $sc^T(c_{\max}) - sc^T(c_{2\max})$ is "roughly" $\sqrt{V_T/(n \log m)}$.

Lemma 3.9. *Let $\tau_1(m) = o_m(1)$ and $\tau_2(m) = \omega_m(1)$. For large enough n, m,*

$$\Omega(\tau_1(m)\sqrt{V_T/(n \log m)}) \le sc^T(c_{\max}) - sc^T(c_{2\max}) \le O(\tau_2(m)\sqrt{V_T/(n \log m)})$$

with probability $1 - o_m(1) - o_n(1)$.

Proof. Choose $\ell_1 = \exp \tau_1(m)$ and $\ell_2 = \exp \tau_2(m)$ in Lemma 3.8 to obtain the stated bound, which holds with probability $1/\ell_1 - 1/\ell_2 - o_m(1) - o_n(1)$. The lemma follows since $1/\ell_1 \to 1$ and $1/\ell_2 \to 0$. □

This lemma is good enough to prove a lower bound on k. In order to prove a good upper bound, we need to estimate the difference $sc^T(c_{\max}) - sc^T(c_{p-\max})$, for other values of p; here $c_{1-\max} = c_{\max}$, $c_{2-\max} = c_{2\max}$, and so on.

Lemma 3.10. *Suppose $p = o(\sqrt{m}/\log m)$ satisfies also $p = \omega_m(1)$. Then $sc^T(c_{\max}) - sc^T(c_{p-\max}) = \Theta(\log p\sqrt{V_T/(n\log m)})$ with probability $1 - o_m(1) - o_n(1)$.*

The analysis is similar (albeit more involved), and is deferred to the end of this section (Subsection 3.3).

Now, we take a similar approach by estimating $sc^B(c_{\max}) - sc^B(c_{2\max})$ (the direction of the bound will depend on the type of bound on k). If $c_{\max}, c_{2\max}$ were two arbitrary candidates then we could use the central limit theorem to directly estimate $sc^B(c_{\max}) - sc^B(c_{2\max})$. The expectation would be 0 because of symmetry, and the variance is given by the following lemma.

Lemma 3.11. *Let $c \in C$ be an arbitrary candidate. The mean of $sc_i^B(c)$ is 0, and its variance is*

$$V_B = \frac{1}{m}\sum_{i=k+1}^{m}\alpha(i)^2 - \frac{1}{m(m-k)}\left(\sum_{i=k+1}^{m}\alpha(i)\right)^2.$$

The variance of $sc^B(c) - sc^B(c')$ is $2\frac{m}{m-1}V_B/n$.

Proof. The expectation is given by

$$\begin{aligned}\mathbb{E}[sc_i^B(c)] &= \frac{1}{m}\sum_{i=k+1}^{m}\left(\alpha(i) - \frac{1}{m-k}\sum_{j=k+1}^{m}\alpha(j)\right)\\ &= \frac{1}{m}\sum_{i=k+1}^{m}\alpha(i) - \frac{m-k}{m(m-k)}\sum_{j=k+1}^{m}\alpha(j) = 0.\end{aligned}$$

Since $\mathbb{E}[sc_i^B(c)] = 0$,

$$\begin{aligned}V_B &= \mathbb{E}[sc_i^B(c)^2]\\ &= \frac{1}{m}\sum_{i=k+1}^{m}\left(\alpha(i) - \frac{1}{m-k}\sum_{j=k+1}^{m}\alpha(j)\right)^2\\ &= \frac{1}{m}\sum_{i=k+1}^{m}\alpha(i)^2\\ &+ \frac{1}{m}\left[-\frac{2}{m-k} + \frac{m-k}{(m-k)^2}\right]\left(\sum_{i=k+1}^{m}\alpha(i)\right)^2\\ &= \frac{1}{m}\sum_{i=k+1}^{m}\alpha(i)^2 - \frac{1}{m(m-k)}\left(\sum_{i=k+1}^{m}\alpha(i)\right)^2.\end{aligned}$$

Next, as in the proof of Lemma 3.2, $\mathrm{Cov}(sc_i^B(c), sc_i^B(c')) = -\frac{V_B}{m-1}$. Since $\mathbb{E}[sc_i^B(c)] = 0$,

$$\begin{aligned}\mathbb{E}[(sc_i^B(c) - sc_i^B(c'))^2] &= 2V_B - 2\,\mathrm{Cov}(sc_i^B(c), sc_i^B(c'))\\ &= 2\frac{m}{m-1}V_B.\end{aligned}$$

Therefore $\mathrm{Var}[sc_i^B(c) - sc_i^B(c')] = 2\frac{m}{m-1}V_B$. Finally, $\mathrm{Var}[sc^B(c) - sc^B(c')] = \mathrm{Var}[sc^B(c) - sc^B(c')]/n = 2\frac{m}{m-1}V_B/n$. □

Notice however that, $c_{\max}$ and $c_{2\max}$ are *not* arbitrary candidates. We will show that the effect of this issue on the difference in the scores is quite negligible. The idea is to eliminate the dependence on the choosing rule by dividing the voters into four groups, according to whether $c_{\max}$ came up in the bottom or top, and whether $c_{2\max}$ came up in the bottom or top:

$$P_1 = \{\sigma \in P : \sigma(c_{\max}), \sigma(c_{2\max}) \le k\}, \quad P_2 = \{\sigma \in P : \sigma(c_{\max}) > k, \sigma(c_{2\max}) \le k\},$$
$$P_3 = \{\sigma \in P : \sigma(c_{\max}) \le k, \sigma(c_{2\max}) > k\}, \quad P_4 = \{\sigma \in P : \sigma(c_{\max}), \sigma(c_{2\max}) > k\}.$$

The voters in each of these groups behave as if $c_{\max}, c_{2\max}$ were arbitrary candidates, under the condition that some of them are at the top k and some not. The number of voters in these groups $n_i = |P_i|$ are strongly concentrated around their means $\nu_1, \nu_2, \nu_3, \nu_4$ due to a Chernoff bound. Given n_1, n_2, n_3, n_4, we can use the central limit theorem to approximate the distribution of $sc^B(c_{\max}) - sc^B(c_{2\max})$.

We start by analyzing the distribution of $sc^B(c_{\max}) - sc^B(c_{2\max})$ given the deviation parameters $\epsilon_i = n_i - \nu_i$. We present here the (easily verifiable) values of $\nu_1, \nu_2, \nu_3, \nu_4$:

$$\nu_1 = \frac{k(k-1)}{m(m-1)} n, \quad \nu_2 = \nu_3 = \frac{k(m-k)}{m(m-1)} n \quad \nu_4 = \frac{(m-k)(m-k-1)}{m(m-1)} n.$$

First, we establish the mean and variance of $sc^B(c_{\max}) - sc^B(c_{2\max})$.

Lemma 3.12. *Suppose $\epsilon_i = n_i - \nu_i$ are given. Then*

$$\mathbb{E}[sc^B(c_{\max}) - sc^B(c_{2\max})] = 0,$$
$$\mathrm{Var}[sc^B(c_{\max}) - sc^B(c_{2\max})] = \frac{2\frac{m}{m-1} V_B}{n} \pm O_\alpha\left(\frac{\max(\epsilon_2, \epsilon_3, \epsilon_4)}{n^2}\right).$$

Here $O_\alpha(\cdot)$ means that the constant depends on the weights $\alpha(1), \ldots, \alpha(m)$.

Proof. We start with the mean. If $i \in P_1 \cup P_3$ then $sc_i^B(c_{\max}) = 0$. If $i \in P_2 \cup P_4$ then as in the proof of Lemma 3.11, $\mathbb{E}[sc_i^B(c_{\max})] = 0$. We conclude that $\mathbb{E}[sc^B(c_{\max})] = 0$, and similarly $\mathbb{E}[sc^B(c_{2\max})] = 0$.

As for the variance, let v_1, v_2, v_3, v_4 be the variance arising from a single voter in P_1, P_2, P_3, P_4, respectively. Note that $v_1 = 0$ and $v_2 = v_3$. Thus

$$\mathrm{Var}[sc^B(c_{\max}) - sc^B(c_{2\max})] = \frac{n_2 + n_3}{n^2} v_3 + \frac{n_4}{n^2} v_4.$$

We know that when $\epsilon_2 = \epsilon_3 = \epsilon_4 = 0$, the above must equal $2\frac{m}{m-1} V_B / n$. Therefore

$$\mathrm{Var}[sc^B(c_{\max}) - sc^B(c_{2\max})] = \frac{2\frac{m}{m-1} V_B}{n} \pm O_\alpha\left(\frac{\max(\epsilon_2, \epsilon_3, \epsilon_4)}{n^2}\right).$$

□

This allows us to conclude that $sc^B(c_{\max}) - sc^B(c_{2\max})$ is close in distribution to a normal random variable.

Lemma 3.13. *Suppose $k \neq 1, m$. The random variable $sc^B(c_{\max}) - sc^B(c_{2\max})$ converges in distribution to a Gaussian $\mathcal{N}(0, 2\frac{m}{m-1} V_B / n)$.*

Proof. Given $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4$, the random variable $sc^B(c_{\max}) - sc^B(c_{2\max})$ is the average of n_1 constant random variables and $n_2 + n_3 + n_4$ non-constant random variables with one of three given bounded distributions. Since $k \neq m$, $n_2 + n_3 + n_4 = \Omega(n)$ with probability $1 - o_n(1)$, and in that case the Berry–Esseen theorem shows that $sc^B(c_{\max}) - sc^B(c_{2\max})$ is $o_n(1)$-close in distribution to a Gaussian $\mathcal{N}(0, V)$, where $V = \mathrm{Var}[sc^B(c_{\max}) - sc^B(c_{2\max})]$; the expectation vanished due to Lemma 3.12. Now $\epsilon_2, \epsilon_3, \epsilon_4$ are all $o(n)$ with probability $1 - o_n(1)$, and so the lemma shows that $V = 2\frac{m}{m-1}V_B/n(1 + o_n(1))$ in that case. Therefore with probability $1 - o_n(1)$, $sc^B(c_{\max}) - sc^B(c_{2\max})$ is $o_n(1)$-close in distribution to $\mathcal{N}(0, 2\frac{m}{m-1}V_B/n)$. The lemma follows. □

Nothing in the proof of Lemma 3.13 used any special properties of $c_{\max}, c_{2\max}$; rather, they were arbitrary candidates. Therefore the lemma holds for any two candidates.

Combining Lemma 3.9 with Lemma 3.13, we can prove our main theorem.

We are now ready to prove Theorem 3.1.

Proof. **Lower bound.** Let $\tau_2(m) = \sqrt{V_B \log m / V_T} \to \infty$. Lemma 3.9 shows that $sc^T(c_{\max}) - sc^T(c_{2\max}) = O(\tau_2(m)\sqrt{V_T/(n \log m)}) = O(\sqrt{V_B/n})$ with probability $1 - o_m(1) - o_n(1)$, and Lemma 3.13 shows that $sc^B(c_{\max}) - sc^B(c_{2\max})$ converges in distribution to $\mathcal{N}(0, 2\frac{m}{m-1}V_B/n)$. Therefore:

- With constant probability, $sc^B(c_{\max}) - sc^B(c_{2\max}) < -(sc^T(c_{\max}) - sc^T(c_{2\max}))$, and so $sc(c_{\max}) < sc(c_{2\max})$ (according to the properties of the Gaussian distribution, the difference can be a constant multiple of standard deviations away from its mean). In particular, with constant probability $c_{\max}$ is not the overall winner.
- With constant probability, $sc^B(c_{\max}) - sc^B(c_{2\max}) \geq 0$, and so $sc(c_{\max}) \geq sc(c_{2\max})$. In particular, with constant probability $c_{2\max}$ is not the overall winner.
- Let c be any other candidate. The proof of Lemma 3.13 used no special properties of $c_{\max}$ or $c_{2\max}$, and so it applies to $sc^B(c_{\max}) - sc^B(c)$ as well. Therefore with constant probability, $sc^B(c_{\max}) - sc^B(c) \geq 0$, and so $sc(c_{\max}) \geq sc(c)$. In particular, with constant probability c is not the overall winner.

We conclude that each candidate fails to be the overall winner with some constant probability.

Upper bound. Let $p = e^{\log^{1/3} m}$. Define $\tau(m) = \sqrt{V_B \log^{4/3} m / V_T} = o_m(1)$ and $\tau_2(m) = \sqrt{\tau(m)} = o_m(1)$. We have the following:

- Lemma 3.9 shows that with probability $1 - o_m(1) - o_n(1)$ and all $q \geq 2$, $sc^T(c_{\max}) - sc^T(c_{q-\max}) \geq sc^T(c_{\max}) - sc^T(c_{2\max}) = \Omega(\tau_2(m)\sqrt{V_T/(n \log m)}) = \Omega(\tau(m)^{-1/2}\tau(m)\sqrt{V_T/(n \log m)}) = \Omega(\tau(m)^{-1/2}\sqrt{(V_B/n) \log p}) = \omega(\sqrt{(V_B/n) \log p})$.
- Hence Lemma 3.13, together with the tail bound (3.1), shows that $sc(c_{\max}) > sc(c_{q-\max})$ for all $2 \leq q \leq p$ with probability $1 - o_m(1)$ for large enough n.
- Lemma 3.10 shows that $sc^T(c_{\max}) - sc^T(c_{p-\max}) = \Theta(\log p\sqrt{V_T/(n \log m)}) = \omega(\sqrt{(V_B/n) \log m})$ with probability $1 - o_m(1) - o_n(1)$.
- Hence Lemma 3.13, together with the tail bound (3.1), shows that $sc(c_{\max}) > sc(c_{q-\max})$ for all $q \geq p$ with probability $1 - o_m(1)$ for large enough n.

We conclude that with probability $1 - o_m(1)$ and large enough n, $c_{\max}$ has the largest overall score. □

3.3 Proving Lemma 3.10

We start by showing that when $p = o(\sqrt{m}/\log m)$, the corresponding uniform random statistics are smaller than $1/\sqrt{m}$, and so $x_{p-\max} = \Theta(\sqrt{\log m})$.

Lemma 3.14. *Suppose* $p = o(\sqrt{m}/\log m)$. *With probability* $1 - o_m(1)$, $u_{p-\min} \leq 1/\sqrt{m}$ *and* $x_{p-\max} = \Theta(\sqrt{\log m})$.

Proof. It is well-known (e.g. [1, (2.2.2)]) that the density of $u_{p-\min}$ is $\frac{m!}{(p-1)!(m-p)!} u^{p-1}(1-u)^{m-p}$. Therefore

$$\begin{aligned}
&\Pr[u_{p-\min} \geq 1/\sqrt{m}] \\
=&\int_{1/\sqrt{m}}^{1} \frac{m!}{(p-1)!(m-p)!} u^{p-1}(1-p)^{m-p}\, du \\
\leq&\int_{1/\sqrt{m}}^{1} \frac{m!}{(p-1)!(m-p)!}(1-p)^{m-p}\, du \\
=&\binom{m}{p-1}(1-p)^{m-p+1}\Big|_{1/\sqrt{m}}^{1} \\
=&\binom{m}{p-1}(1-1/\sqrt{m})^{m-p+1} \\
\leq& m^{p-1}(1-1/\sqrt{m})^{m-p+1}.
\end{aligned}$$

Taking the logarithm,

$$\begin{aligned}
&\log \Pr[u_{p-\min} \geq 1/\sqrt{m}] \\
\leq&(p-1)\log m - \frac{m-p+1}{\sqrt{m}} \\
=&(p-1)(\log m + \tfrac{1}{\sqrt{m}}) - \sqrt{m}.
\end{aligned}$$

Where the inequality follows from the previous bound and the bound $\ln(1-x) \leq -x$, for $1 < x < 1$. Since $p = o(\sqrt{m}/\log m)$, the logarithm tends to $-\infty$, and so the probability is $o_m(1)$. The corresponding result for $x_{p-\max}$ follows from estimate (3.1) for the lower bound, and Lemma 3.6 for the upper bound. □

Next, we extend Lemma 3.4.

Lemma 3.15. *Let* $1 \leq \ell_1 \leq \ell_2 \leq \infty$.

$$\Pr[\ell_1 \leq \tfrac{u_{p-\min}}{u_{\min}} \leq \ell_2] = \left(1 - \frac{1}{\ell_2}\right)^{p-1} - \left(1 - \frac{1}{\ell_1}\right)^{p-1}.$$

Proof. It is well-known (e.g. [1, (2.3.9)]) that the joint density function of $u = u_{\min}$ and

$v = u_{p-\min}$ is $\frac{n!}{(p-2)!(n-j)!}(v-u)^{p-2}(1-v)^{n-p}$. Therefore

$$\begin{aligned}
&\Pr[u_{p-\min} \geq \ell u_{\min}] \\
=&\int_0^1 \int_0^{v/\ell} \frac{n!}{(p-2)!(n-p)!}(v-u)^{j-2}(1-v)^{n-p}\, dudv \\
=&-\int_0^1 \frac{n!}{(p-1)!(n-p)!}(v-u)^{p-1}(1-v)^{n-p}\Big|_0^{v/\ell} dv \\
=&\int_0^1 \frac{n!}{(j-1)!(n-p)!}v^{p-1}(1-v)^{n-p}\left[1-\left(1-\frac{1}{\ell}\right)^{p-1}\right] dv \\
=&1-\left(1-\frac{1}{\ell}\right)^{p-1}.
\end{aligned}$$

Where the last equality follows from two of the definitions of the beta function; i.e., $\int_0^1 x^a(1-x)^b dx = \frac{1}{(a+b+1)\binom{a+b}{a}}$. The lemma easily follows. □

Next in turn is a generalization of Lemma 3.7 and Lemma 3.8.

Lemma 3.16. *Let $1 \leq \ell_1 \leq \ell_2 \leq \infty$, and suppose that $p = o(\sqrt{m}/\log m)$. With probability $(1-1/\ell_2)^{p-1} - (1-1/\ell_1)^{p-1} - o_m(1)$,*

$$\Omega\left(\frac{\log \ell_1}{\sqrt{\log m}}\right) \leq x_{\max} - x_{p-\max} \leq O\left(\frac{\log \ell_2}{\sqrt{\log m}}\right),$$

and with probability $(1-1/\ell_2)^{p-1} - (1-1/\ell_1)^{p-1} - o_m(1) - o_n(1)$,

$$\Omega\left(\log \ell_1 \sqrt{\frac{V_T}{n \log m}}\right) \leq sc^T(c_{\max}) - sc^T(c_{p-\max}) \leq O\left(\log \ell_2 \sqrt{\frac{V_T}{n \log m}}\right).$$

Proof. The proof is very similar to the proofs of Lemma 3.7 and Lemma 3.8. □

Lemma 3.10 can therefore be thought of as a corollary of the above lemma, and furthermore it is an analogue of Lemma 3.9:

Lemma 3.10. *Suppose $p = o(\sqrt{m}/\log m)$ satisfies also $p = \omega_m(1)$. Then $sc^T(c_{\max}) - sc^T(c_{p-\max}) = \Theta(\log p\sqrt{V_T/(n\log m)})$ with probability $1 - o_m(1) - o_n(1)$.*

Proof. Choose $\ell_1 = \sqrt{p-1}$ and $\ell_2 = (p-1)^2$ to obtain the stated bound, which holds with probability $(1-1/\ell_2)^{p-1} - (1-1/\ell_1)^{p-1} - o_m(1) - o_n(1) = 1 - o_m(1) - o_n(1)$. □

4 Copeland's Voting Rule

Having considered positional scoring rules, we now further demonstrate the applicability of our of approach by considering the (non-positional) Copeland scoring rule. We give a lower bound on k that corresponds to any top-k algorithm:

Theorem 4.1. *For $k \leq m/\sqrt{\log m}$, no algorithm can predict the winner under Copeland with probability better than $1 - \Omega(1)$.*

We first give an outline of the proof, and afterwards provide the detailed proof. Relating to the definition of the scoring rule, for an individual vote $i \in N$ and two distinct candidates $c, c' \in C$, we set $sc_i(c, c') = 1$ if $\pi_i(c) < \pi_i(c')$, and $sc_i(c, c') = -1$ if $\pi_i(c) > \pi_i(c')$. Note that a candidate c beats a candidate c' exactly when $sc(c, c') > 0$. As done for positional scoring rules, we can rewrite $sc_i(c, c')$ as the sum of two pairwise scores $sc_i^T(c, c')$ and $sc_i^B(c, c')$. The score $sc_i^T(c, c')$ behaves like $sc_i(c, c')$ if at least one of the candidates is positioned in the top-k ranking of voter i (thus allowing us to deduce the relation of c and c'), and is zero otherwise. The score $sc_i^B(c, c')$ is defined so that $sc_i(c, c') = sc_i^T(c, c') + sc_i^B(c, c')$.

The idea of the analysis is to show that for small enough k, each candidate c has a constant probability of losing. The top and bottom scores are both roughly normally distributed (with correlations). In contrast to the case of positional scoring rules, dealing with correlations is simpler in our case: for every three distinct candidates $c, c', c'' \in C$, it can be shown that $sc_i^T(c, c')$ and $sc_i^T(c, c'')$ are positively correlated. Treating the score of a candidate as the sum of the entries of the vector $(sc_i^T(c, c'))_{c' \neq c}$, we use this fact to decompose this (random) vector into two, more tractable, vectors. This allows us to bound the advantage of c over most other candidates c' in the top score.

Similarly, the bottom scores are positively correlated, due to a "bias" corresponding to the average position of c in voters in which it appears outside the top-k. With constant probability, this bias is negative, and so the total score is roughly binomially distributed with a negative bias. This shows that c could lose with constant probability.

We suspect that the true lower bound for k is, in fact, $\Omega(m)$.

We now proceed with the detailed proof of Theorem 4.1. For completeness, we give the complete the definitions of the voter-specific, Copeland score, as well as the top-k scores and bottom-$(m-k)$ scores, $sc^T(\cdot, \cdot), sc^B(\cdot, \cdot)$, that correspond to the definition of the Copeland scoring rule:

$$sc_i(c, c') = \begin{cases} +1 & \text{if } \pi_i(c) < \pi_i(c'), \\ -1 & \text{if } \pi_i(c') < \pi_i(c). \end{cases}$$

$$sc_i^T(c, c') = \begin{cases} +1 & \text{if } \pi_i(c) < \pi_i(c') \text{ and } \pi_i(c) \leq k, \\ -1 & \text{if } \pi_i(c') < \pi_i(c) \text{ and } \pi_i(c') \leq k, \\ 0 & \text{if } \pi_i(c), \pi_i(c') > k. \end{cases}$$

$$sc_i^B(c, c') = \begin{cases} +1 & \text{if } k < \pi_i(c) < \pi_i(c'), \\ -1 & \text{if } k < \pi_i(c') < \pi_i(c), \\ 0 & \text{if } \pi_i(c) \leq k \text{ or } \pi_i(c') \leq k. \end{cases}$$

By definition, we have $sc_i(c, c') = sc_i^T(c, c') + sc_i^B(c, c')$.

As done for positional scoring rules, we will consider the *normalized* sum $sc(c, c') = \frac{1}{\sqrt{n}} \sum_{i=1}^n sc_i(c, c')$. Recall that for positional scoring rules, we were concerned with the *average* score; this slightly different normalization is used to make the proof less cumbersome.

Fix a candidate c, and let $p = k/m$, $q = 1 - p$. In order to arrive at simpler terms, from now on, whenever we write $A \backsim B$, we mean that A differs from B by a multiplicative (and negligible) error of $1 \pm o_m(1)$, assuming that $p = o_m(1)$. We start by approximating the distributions of the vectors $sc^T(c, c')_{c' \neq c}$ and $sc^B(c, c')_{c' \neq c}$, for two distinct candidates $c, c' \in C$.

Lemma 4.1. *We have* $\mathbb{E}[sc^T(c, c')] = \mathbb{E}[sc^B(c, c')] = 0$, *and for* $c' \neq c''$,

$$V_T = \mathrm{Var}[sc^T(c, c')] \approx 1 - q^2 \backsim 2p, \qquad \eta_T = \mathrm{Cov}(sc^T(c, c'), sc^T(c, c'')) \approx \frac{1 - q^3}{3} \backsim p,$$

$$V_B = \mathrm{Var}[sc^B(c, c')] \approx q^2 \backsim 1, \qquad \eta_B = \mathrm{Cov}(sc^B(c, c'), sc^B(c, c'')) \approx \frac{q^3}{3} \backsim \frac{1}{3}.$$

Proof. Clearly $\mathbb{E}[sc^T(c,c')] = \mathbb{E}[sc^B(c,c')] = 0$. Due to our choice of normalization, sc^T, sc^B have the same variance and covariance as sc_i^T, sc_i^B. The top variances are

$$\begin{aligned}\mathrm{Var}[sc_i^T(c,c')] &= \Pr[\pi_i^{-1}(c) \le k \text{ or } \pi_i^{-1}(c') \le k] \\ &= 1 - \frac{(m-k)(m-k-1)}{m(m-1)}.\end{aligned}$$

The top covariances are

$$\begin{aligned}\mathrm{Cov}(sc_i^T(c,c'), sc_i^T(c,c'')) &= \Pr[\pi^{-1}(c) < \pi^{-1}(c'), \pi^{-1}(c'') \text{ and } \pi^{-1}(c) \le k] \\ &+ \Pr[\pi^{-1}(c'), \pi^{-1}(c'') < \pi^{-1}(c) \text{ and } \pi^{-1}(c'), \pi^{-1}(c'') \le k] \\ &- \Pr[\pi^{-1}(c') < \pi^{-1}(c) < \pi^{-1}(c'') \text{ and } \pi^{-1}(c) \le k] \\ &- \Pr[\pi^{-1}(c'') < \pi^{-1}(c) < \pi^{-1}(c') \text{ and } \pi^{-1}(c) \le k] \\ &= \frac{1}{3}\left(1 - \frac{(m-k)(m-k-1)(m-k-2)}{m(m-1)(m-2)}\right) \\ &+ \frac{1}{3}\frac{k(k-1)(k-2)}{m(m-1)(m-2)} + \frac{k(k-1)(m-k)}{m(m-1)(m-2)} \\ &- \frac{2}{6}\frac{k(k-1)(k-2)}{m(m-1)(m-2)} - \frac{2}{2}\frac{k(k-1)(m-k)}{m(m-1)(m-2)} \\ &= \frac{1}{3}\left(1 - \frac{(m-k)(m-k-1)(m-k-2)}{m(m-1)(m-2)}\right).\end{aligned}$$

The bottom variances are

$$\begin{aligned}\mathrm{Var}[sc_i^B(c,c')] &= \Pr[\pi_i^{-1}(c), \pi_i^{-1}(c') > k] \\ &= \frac{(m-k)(m-k-1)}{m(m-1)}.\end{aligned}$$

Finally, the bottom covariances are

$$\begin{aligned}\mathrm{Cov}(sc_i^B(c,c'), sc_i^B(c,c'')) &= \Pr[k < \pi^{-1}(c) < \pi^{-1}(c'), \pi^{-1}(c'')] + \Pr[k < \pi^{-1}(c'), \pi^{-1}(c'') < \pi^{-1}(c)] \\ &- \Pr[k < \pi^{-1}(c') < \pi^{-1}(c) < \pi^{-1}(c'')] - \Pr[k < \pi^{-1}(c'') < \pi^{-1}(c) < \pi^{-1}(c')] \\ &= \frac{(m-k)(m-k-1)(m-k-2)}{m(m-1)(m-2)}\left(\frac{1}{3} + \frac{1}{3} - \frac{2}{6}\right) \\ &= \frac{1}{3}\frac{(m-k)(m-k-1)(m-k-2)}{m(m-1)(m-2)}.\end{aligned}$$

□

The distributions of the vectors $sc^T(c,c')_{c'\neq c}$ and $sc^B(c,c')_{c'\neq c}$ approach normal distributions with the given variance and covariance. Note that in contrast to the case of positional scoring rules, the pairwise scores exhibit *positive* correlations. In order to handle these correlations, we will decompose these vectors, and treat them as sums of independent random variables. This will make the analysis simpler, as it allows us to deal with these vectors as sums of two, easier to work with, random vectors.

Consider the following decomposition, which relates to the distribution of the vectors $sc^T(c,c')_{c'\neq c}, sc^B(c,c')_{c'\neq c}$. For $\ell \in \{T, B\}$, sample a *single* normal random variable z_ℓ from $\mathcal{N}(0, \eta_\ell)$. Then, sample $m-1$ i.i.d. normal random variables from the distribution $\mathcal{N}(0, V_\ell - \eta_\ell)$, and denote the resulting $(m-1)$-dimensional vector by $\mathbf{y}_\ell$. Finally, let $\mathbf{x}_\ell$

denote the vector that results from adding z_ℓ to every entry in $\mathbf{y}_\ell$. For convenience, we denote the entries corresponding to candidate c' ($\neq c$) in $\mathbf{y}_\ell$ and $\mathbf{x}_\ell$ by $y_\ell(c')$ and $x_\ell(c')$. First, sample a *single* zero mean normal vector with individual variance V and pairwise covariance η can be generated by first sampling a normal random variable from $\mathcal{N}(0,\eta)$ normal variable to a vector of $\mathcal{N}(0,V-\eta)$ normal variables. The following lemma shows that the above sum of two vectors essentially describes a decomposition of the vectors $sc^T(c,c')_{c'\neq c}, sc^B(c,c')_{c'\neq c}$.

Lemma 4.2. *The random vectors $sc^T(c,c')_{c'\neq c}$ and $sc^B(c,c')_{c'\neq c}$ converge in distribution to the distribution of the random vectors $\mathbf{x_T}$ and $\mathbf{x_B}$ (respectively).*

Proof. The lemma follows from the central limit theorem, once we notice that the expectation and covariance matrices match in both cases. □

As the vector $\mathbf{x_T}$ is obtained by adding a uniform bias $\mathcal{N}(0,(V_T-\eta_T)\backsim p)$ to i.i.d. samples of the distribution $\mathcal{N}(0,\eta_T\backsim p)$, we would expect that most of the vector is $O(\sqrt{p})$, with high probability. This is stated formally in the following lemma.

Lemma 4.3. *Let $\epsilon>0$ be given. For $M=2\sqrt{\log\frac{2}{\epsilon}}$ the following holds, assuming $M\geq 1$. For large enough n, with probability $1-\epsilon$ at least a $1-\epsilon$ fraction of c' satisfy $sc^T(c,c')\leq M\sqrt{V_T}$.*

Proof. We show that this holds for the vector $\mathbf{x_T}$ with probability $1-\epsilon/2$, whence the lemma follows from convergence in distribution (Lemma 4.2). The bound on the cdf of the normal distribution (Eq. (3.1)) shows that $\Pr[z_T\leq M\sqrt{\eta_T/2}]\geq 1-\epsilon/2$ and $\Pr[y_T(c')\leq M\sqrt{(V_T-\eta_T)/2}]\geq 1-\epsilon/2$. Hence with probability $1-\epsilon/2$, a $1-\epsilon/2$ fraction of c' satisfy $x_T(c')=z_T+y_T(c')\leq M(\sqrt{\eta_T/2}+\sqrt{(V_T-\eta_T)/2})\leq M\sqrt{V_T}$. □

On the other hand, conditioned on z_B, the number of candidates c' such that $x_B(c')+M\sqrt{V_T}>0$ is binomially distributed. In particular, if $z_B<-M\sqrt{V_T}$ then it is extremely likely that c' loses. This argument (which has to be adjusted to handle the ϵ fraction of "bad" candidates c') is given in the following lemma.

Lemma 4.4. *There is a global constant $\epsilon_0>0$ such that the following holds. Suppose that a $1-\epsilon_0$ fraction of c' satisfy $sc^T(c,c')\leq M\sqrt{V_T}$, for some $M>0$. Let $\sigma=M\sqrt{V_T}/\sqrt{\eta_B}+\sqrt{V_B-\eta_B}/\sqrt{\eta_B}\backsim O(M\sqrt{p}+1)$. For large enough n, candidate c loses with probability approaching $\overline{\Phi}(\sigma)\backsim O(1)$.*

Proof. With probability $\overline{\Phi}(\sigma)=\Phi(-\sigma)$, we have $z_B<-\sigma\cdot\sqrt{\eta_B}=-M\sqrt{V_T}-\sqrt{V_B-\eta_B}$. Therefore for a $1-\epsilon_0$ fraction of the candidates, $sc^T(c,c')+x_B(c')$ is a normal random variable with expectation at most $-\sqrt{V_B-\eta_B}$. The number N of these candidates satisfying $sc^T(c,c')+x_B(c')<0$ is thus stochastically bounded from below by $\mathrm{Bin}((1-\epsilon_0)m,\Phi(1))$. In particular, we have $N\geq(1-\epsilon_0)\Phi(0.9)m$ with probability $1-o_n(1)$. This guarantees that c loses as long as $(1-\epsilon_0)\Phi(0.9)>1/2$, which holds for small enough $\epsilon_0>0$. The proof is complete by taking the normal approximation via Lemma 4.2. □

We can now prove the main theorem.

Proof of Theorem 4.1. Choose $\epsilon=1/m^2$ in Lemma 4.3. Applying the union bound, we obtain that with probability $1-1/m$, for all candidates c it holds that a $1-1/m^2$ fraction of other candidates $c'\neq c$ satisfies $sc^T(c,c')\leq M\sqrt{V_T}$, where $M=O(\sqrt{\log m})$. Applying Lemma 4.4, we see that each candidate loses with probability approaching $\overline{\Phi}(\sigma)$, where $\sigma=O(M\sqrt{k/m}+1)=O(1)$. The lemma follows since $\sigma=O(1)$ implies $\overline{\Phi}(\sigma)=1-\Omega(1)$. □

5 Mallows distribution

Theorem 3.1 shows that top-k allows for efficient elicitation under the harmonic and geometric positional scoring rules, even under the most neutral preferences distribution. For the Borda and Copeland scoring rules, we've shown that it is not the case (this is confirmed empirically in our simulation results, presented in Section 7). This motivates the following question: are there any classes of preference distributions for which top-k performs well under these supposedly inefficient scoring rules? The purpose of our following discussion is to provide such general distributions, and to argue that in the limiting case where $n \to \infty$, only a constant k is sufficient.

The following piece of notation would be useful: given a distribution D over $\mathcal{L}$ and a candidate $c \in C$, we let $q_t(c) = Pr_{\pi \sim D}[\pi(c) \leq t]$; i.e., the probability that c_i is positioned in the first t positions.

Consider the following class of distributions:

Definition 3. *Let D be a distribution over the set of preferences $\mathcal{L}$. Then D is said to be* positionally-biased *(PoB) if there exists a distinguished candidate $c \in C$ such that $q_t(c) > q_t(c')$ for all candidates $c' \neq c$ and $1 \leq t < m$. Furthermore, we call the said candidate c the* favored *candidate.*

Theorem 5.1. *Let D be a positionally-biased distribution over $\mathcal{L}$, and let c be its favored candidate. Suppose that the election is defined by a non-constant positional scoring rule. Then candidate c wins with probability $1-o_n(1)$, and so the overall winner under distribution D can be predicted without looking at the votes at all.*

Sketch of Proof First, by a majorization argument, it follows that the expected score of c is strictly higher than that of all other candidates. The statement of the theorem follows by a straightforward application of the Chernoff bound. □

We now argue that the Mallows distribution is PoB, and that furthermore, natural generalizations of it are also PoB. To do so, we will need the following simple properties:

Definition 4 (Swap increasing distance)**.** *A distance function $d\colon \mathcal{L} \times \mathcal{L} \to \mathbb{Z}_{\geq 0}$ is* swap-increasing *if for any two $\pi_1, \pi_2 \in \mathcal{L}$ and any two $c, c' \in C$ such that $\pi_1(c) < \pi_1(c')$ and $\pi_2(c) < \pi_2(c')$ we have $d(\pi_1, \pi_2') > d(\pi_1, \pi_2)$, where π_2' is obtained from π_2 by switching c and c'.* [2]

Definition 5 (Monotone distributions)**.** *Let D be a distribution over $\mathcal{L}$, parametrized by some fixed reference ranking $\hat{\pi}$ and a swap-increasing distance function $d(\cdot,\cdot)$. Then D is said to be* monotone *if $Pr_{\pi \sim D}[\pi]$ is decreasing with $d(\pi, \hat{\pi})$.*

Lemma 5.1. *A monotone distribution is positionally-biased, with $c = \hat{\pi}(1)$ being the favored candidate.*

Sketch of Proof Let $c' \neq c$. If $\sigma^{-1}(c) < \sigma^{-1}(c')$ and σ' is obtained from σ by switching c and c', then $\Pr[\sigma] > \Pr[\sigma']$. This implies that $q_t(c) > q_t(c')$ for all $1 \leq t < m$. □

The following is a well-known folk theorem (e.g., [4]):

Lemma 5.2. *The Kendall tau distance function is swap-increasing.*

[2]Our definition is a weakening of a similar definition in [4] (every distance function satisfying their definition also satisfies ours).

As a corollary, we deduce that Mallows distributions with dispersion parameter $\phi < 1$ are positionally-biased, and so Theorem 5.1 applies to them.

Theorem 5.1 shows that if the preference distribution is positionally-biased then there is no need to elicit votes at all, for large enough n. However, that may be an unrealistic assumption. It could be, for example, that the preferences are known to be distributed according to a Mallows distribution, but the reference profile $\hat{\pi}$ is not known in advance. It is not hard to show that even in this case, for large enough n, $k = 1$ is sufficient to recover $\hat{\pi}(1)$ and so predict the winner with high probability.

We note that Caragiannis et al. made a very similar set of arguments in the context of predicting the underlying ranking using scoring rules in [4].

We now generalize all the foregoing for the case of Copeland, and more generally Condorcet-compatible rules. Recall that a voting rule is said to be Condorcet-compatible if the candidate who beats all other candidates in pairwise elections, always wins the elections. First, we define a corresponding class of distributions:

Definition 6. *A distribution D over the set of preferences $\mathcal{L}$ is* pairwise-biased *(PwB) if there exists a distinguished candidate $c \in C$ (the favored candidate) such that for every other candidate $c' \in C \setminus \{c\}$, $Pr_{\pi \sim D}[\pi(c) < \pi(c')] > 1/2$; i.e., c is more likely to precede c' than the other way around.*

Theorem 5.2. *Suppose that the voter preferences are drawn from a pairwise-biased distribution D, with a favored candidate $c \in C$. Then candidate c is the Condorcet winner with probability $1 - o_n(1)$, and so the overall winner under distribution D can be predicted without looking at the votes at all for any Condorcet-compatible rule.*

Proof. For $c_i \neq c$, let $c_i > 1/2$ be the probability that c precedes c_i. Chernoff's bound shows that c beats c_i in a pairwise election with probability $1 - o_n(1)$. As this is true for all $c_i \neq c$, we deduce that c is a Condorcet winner with probability $1 - o_n(1)$. □

Note that the Copeland voting rule is indeed Condorcet compatible, and so this result contrasts strongly with the setting of an impartial culture.

6 Worst-case distributions

Having shown a contrast between the neutral distribution (IC), and the far less demanding (in terms of the bounds on k) Mallows distribution and its generalization, it would be interesting to consider the following worst-approach: suppose that an adversary chooses a distribution D, and makes its full details public. In an analogy to the previous input models, we are interested in studying the limitations (or sometimes, capabilities), for this worst-case, *fully-known* distribution D.

We focus on two of the scoring rules that were shown to be the *least* demanding, under the impartial culture assumption. For the harmonic scoring rule, we construct a worst case distribution, giving a worst-case lower bound of $\Omega(m)$. Note that this distribution admits this lower bound despite of its exhibiting a significant amount of noise. Then, we prove the robustness of the geometric positional scoring rule, proving an upper bound of $k = O(\log m)$ for any distribution, for sufficiently high n.

We start by arguing that the harmonic rule is difficult under this model.

Theorem 6.1. *There is a distribution D_H (more properly, a family of distributions depending on m) such that predicting the winner (with respect to the harmonic weights) with probability $1 - o_m(1)$ requires $k = \Omega(m)$.*

Proof. The distribution D_H is a $1/2 - 1/2$ mixture of two distributions D_1, D_2. In distribution D_1, candidate c_1, c_2 are given the positions $m/(10 \log m)$ and $m/(10 \log m) + 1$ (at random), and the rest of the candidates are distributed randomly. In distribution D_2, candidates c_1, c_2 are given the positions $m/2$ and m (at random), and the rest of the candidates are distributed randomly. It is easy to check that the expected score of candidates c_1, c_2 is roughly $5 \log m/m$, while the expected score of all other candidates is only roughly $\log m/m$. Therefore one of c_1, c_2 must win.

If $k < m/2$ then the top-k votes only reveal information for D_1-voters. Let $sc_i^T(c_1), sc_i^T(c_2)$ be the scores revealed in the top-k choices of voter i, and let $sc_i^B(c_1), sc_i^B(c_2)$ be the scores revealed in the rest of the profile. We have

$$\mathbb{E}[(sc_i^T(c_1) - sc_i^T(c_2))^2] = \frac{1}{2}\left(\frac{10 \log m}{m} - \frac{10 \log m}{m + 10 \log m}\right)^2 = \Theta\left(\frac{\log^4 m}{m^4}\right),$$
$$\mathbb{E}[(sc_i^B(c_1) - sc_i^B(c_2))^2] = \frac{1}{2}\left(\frac{2}{m} - \frac{1}{m}\right)^2 = \Theta\left(\frac{1}{m^2}\right).$$

Let $\Delta^T = \sum_i (sc_i^T(c_1) - sc_i^T(c_2))$ and $\Delta^B = \sum_i (sc_i^B(c_1) - sc_i^B(c_2))$. Individually, the quantities Δ^T, Δ^B have an approximately normal distribution. Furthermore, if we condition on the number of D_1-voters, then the quantities become independent. Since the number of D_1-voters is strongly concentrated around its mean, Δ^T, Δ^B are asymptotically independent. Since $\mathrm{Var}[\Delta^B] \gg \mathrm{Var}[\Delta^T]$, this shows that the information in the top-k part isn't enough to predict the winner: with high probability $|\Delta^T| \le \log m \sqrt{\mathrm{Var}[\Delta^T]}$, while there is constant probability (close to 1/2) that $\Delta^B > \log m \sqrt{\mathrm{Var}[\Delta^T]}$, and constant probability (close to 1/2) that $\Delta^B < -\log m \sqrt{\mathrm{Var}[\Delta^T]}$. □

Next, we show that the geometric rule is *not* difficult under this model.

Theorem 6.2. *Fix ρ, and consider the geometric scoring rule with a constant decay factor of ρ. There is a distribution D_ρ such that predicting the winner with probability $1 - o_m(1)$ requires $k = \Omega(\log m)$. Conversely, there is a constant $\beta > 0$ such that if $k \ge \beta \log m$ then top-k suffices to predict the winner with probability $1 - o_m(1)$ for* every *distribution; we stress that the distribution is known to the algorithm.*

We start with an outline of the proof. The idea is to use a generalization of Algorithm FairCutoff. Fix a scoring rule α (in this case, a geometric rule), a distribution D and an integer k. The algorithm will compute for each voter i and candidate c_j a "top" score $sc_i^T(c_j)$ based only on the top-k part of voter i's vote:

$$sc_i^T(c_j) = \begin{cases} \alpha(t) & \text{if } \pi_i^{-1}(t) = c_j \text{ for some } t \le k, \\ \mathbb{E}[\alpha(\pi_i(c_j)) | \pi_i^{-1}(1), \dots, \pi_i^{-1}(k)] & \text{otherwise.} \end{cases}$$

Here the expectation is taken according to D. The "bottom" score $sc_i^B(c_j)$ complements the top score so that $sc_i^T(c_j) + sc_i^B(c_j) = sc_i(c_j)$:

$$sc_i^B(c_j) = \begin{cases} \alpha(t) - \mathbb{E}[\alpha(\pi_i(c_j)) | \pi_i^{-1}(1), \dots, \pi_i^{-1}(k)] & \text{if } \pi_i^{-1}(t) = c_j \text{ for some } t > k, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\mathbb{E}[sc_i^B(c_j)] = 0$, and so $\mathbb{E}[sc_i^T(c_j)] = \mathbb{E}[sc_i(c_j)]$. As in Section 3, we define $sc^T(c_j), sc^B(c_j), sc(c_j)$ to be averages of $sc_i^T(c_j), sc_i^B(c_j), sc_i(c_j)$ over all voters i.

The difficult part of the proof of Theorem 6.2 is showing that $k \ge C \log m$ suffices to predict the winner with high probability. The idea is to use the algorithm just described.

The only real competition is among the set of candidates S obtaining the maximal expected score. For there to be a competition, S needs to have more than one candidate. Since the average score over all players is $\Theta(1/m)$, any candidate in S has expected score at least $\Omega(1/m)$. We show that this implies a lower bound of $\Omega(1/m^3\sqrt{n})$ on the variance of $sc^T(c_1) - sc^T(c_2)$ for any two $c_1, c_2 \in S$. Since $sc^T(c_1) - sc^T(c_2)$ is roughly normal, this implies anticoncentration of the random variable $sc^T(c_1) - sc^T(c_2)$. In other words, there is some gap between the top scores of any two candidates in S. We show that with probability $1 - o_m(1)$, this gap is at least $\Omega(1/m^{4.5}\sqrt{n})$. On the other hand, $\mathbb{E}[sc^B(c_1) - sc^B(c_2)] = 0$, and the corresponding random variable is roughly normal with variance $O(1/m^5)$; the bound on the variance follows from the fact that $|sc^B(c_1)| = O(1/m^{2.5})$ since k is large enough. Therefore the contribution of the bottom scores is not enough to overturn the winner as judged from the top scores.

We proceed with the full proof of Theorem 6.2.

Proof of Theorem 6.2. The first part is simple. The distribution D_ρ puts candidates c_1, c_2 in places $\log_\rho(1/\sqrt{m}), \log_\rho(1/\sqrt{m}) + 1$ (at random), and distributes the rest of the candidates randomly. The expected score of candidates c_1, c_2 is $\Theta(1/\sqrt{m}) = \omega(1/m)$, whereas the expected score of the other candidates is $\Theta(1/m)$. Therefore with probability $1 - o_m(1)$, one of c_1, c_2 wins the elections. If $k < \log_\rho \sqrt{m}$ then the winner isn't determined by the top-k part of the votes, and so $k \geq \log_\rho \sqrt{m} = \Omega(\log m)$ is required.

The second part is more involved. Suppose that we are given a distribution D, and let S be the set of candidates which have the maximal expected score. A Chernoff bound shows that with probability $1 - o_n(1)$, one of the candidates in S wins the elections. If $|S| = 1$ then the winner can be determined without eliciting any votes, so we can assume that $|S| \geq 2$. Consider any two candidates $c_1, c_2 \in S$, and let their positions under D be the (correlated) random variables t_1, t_2. Since the expected score of a random candidate is $\Theta(1/m)$, we know that the expected score of c_1, c_2 is $\Omega(1/m)$. We would like to lower bound $\mathbb{E}[(sc_i^T(c_1) - sc_i^T(c_2))^2]$. Up to constant factors, this quantity is equal to

$$\sum_{i \leq k} \Pr[\min(t_1, t_2) = i]\rho^{2i}.$$

Since $\mathbb{E}[sc_i(c_1)] = \mathbb{E}[\rho^{t_1}] = \Omega(1/m)$, we know that $\Pr[t_1 \leq \log_\rho(1/m^2)] = \Omega(1/m)$. Since $k \geq \beta \log m \geq \log_\rho(1/m^2)$ (for large enough β),

$$\begin{aligned}\sum_{i \leq k} \Pr[\min(t_1, t_2) = i]\rho^{2i} &\geq \frac{1}{m^2} \sum_{i \leq \log_\rho(1/m^2)} \Pr[\min(t_1, t_2) = i]\rho^i \\ &\geq \frac{1}{m^2} \sum_{i \leq \log_\rho(1/m^2)} \Pr[t_1 = i]\rho^i \geq \Omega\left(\frac{1}{m^3}\right).\end{aligned}$$

The first inequality follows from $\rho^i \geq 1/m^2$ for all $i \leq \log_\rho(1/m^2)$. The second inequality follows from the fact that the distribution of $\min(t_1, t_2)$ majorizes the distribution of t_1. The third inequality follows from $\mathbb{E}[\rho^{t_1}] = \Omega(1/m)$ and the fact that the contribution of terms $i > \log_\rho(1/m^2)$ to the expectation is at most $1/m^2$.

We conclude that $\mathbb{E}[(sc_i^T(c_1) - sc_i^T(c_2))^2] = \Omega(1/m^3)$.

The distribution of $sc_i^T(c_1) - sc_i^T(c_2)$ is asymptotically normal, and since $c_1, c_2 \in S$, its expectation is 0. Therefore it is asymptotically $\mathcal{N}(0, \sigma^2/n)$ for some $\sigma^2 = \Omega(1/m^3)$. Since the density function of $\mathcal{N}(0, \sigma^2/n)$ is at most $1/\sqrt{2\pi\sigma^2/n}$, $\Pr[|\mathcal{N}(0, \sigma^2/n)| \leq \delta] \leq 2\delta/\sqrt{2\pi\sigma^2/n} = O(\delta/\sqrt{\sigma^2/n})$. Taking $\delta = \sqrt{\sigma^2/n}/m^3$, we deduce that with probability $1 - O(1/m^3)$, $|sc^T(c_1) - sc^T(c_2)| \geq \sqrt{\sigma^2/n}/m^3 = \Omega(1/m^{4.5}\sqrt{n})$. Since there are at most

m^2 pairs of elements in S, by taking the union bound, we can conclude that $sc^T(c_{\max}) - sc^T(c_j) = \Omega(1/m^{4.5}\sqrt{n})$ with probability $1 - o_m(1)$ for all $c_j \in S$ other than $c_{\max}$, where $c_{\max} \in S$ is the candidate obtaining the highest top score sc^T among the candidates in S.

On the other hand, for all pairs of distinct candidates $c_1, c_2 \in S$, we have $|sc_i^B(c_1) - sc_i^B(c_2)| \le \rho^{\beta \log m} = O(\rho^k) = O(1/m^5)$ (for large enough β), implying an upper bound of $O(1/m^{10})$ on the variance of this difference. Using the central limit theorem again, we get that $|sc_i^B(c_1) - sc_i^B(c_2)|$ is asymptotically distributed according to $\mathcal{N}(0, \tau^2/n)$ for some $\tau^2 = O(1/m^{10})$. Applying (3.1), we see that with probability $1 - 1/m^3$, $|sc^B(c_1) - sc^B(c_2)| = O(\log m\sqrt{\tau^2/n}) = O(\log m/m^5\sqrt{n})$. After taking the union bound, we get that this is true for all distinct $c_1, c_2 \in S$ with probability $1 - o_m(1)$. Since $O(\log m/m^5\sqrt{n}) < \Omega(1/m^{4.5}\sqrt{n})$ for large enough m, this shows that with probability $1 - o_m(1)$, candidate $c_{\max}$ wins the elections. □

7 Empirical Results

We ran several simulations to verify the results proved in the previous sections. Our first set of simulations is designed to verify Theorem 3.1. For various values of n (number of voters), m (number of candidates) and k (the top-k parameter), and several scoring rules, we compared three algorithms: (1) the algorithm from [15], that assigns 0 points to the bottom $m - k$ candidates in a given vote (labeled as *Naive*), (2) Algorithm FairCutoff, and (3) the optimal algorithm, which calculates the probability that each candidate wins (given the top-k portion of the votes), and chooses the candidate with the maximal winning probability (computing the probabilities was done by sampling).

In order to test the efficacy of top-k voting for the Copeland rule, we ran two different algorithms, Algorithm FairPWCutoff and the naive algorithm, defined as follows. For every pair of candidates $c, c' \in C$ and a top-k vote, if both appear in the top-k ranking, then the higher ranked receives $+1$ points, whereas the lower ranked one receives -1 points. Algorithm FairPWCutoff does the same if only one of them appears in the top-k ranking (implying that the other candidate is ranked lower), whereas the naive algorithm does not award any points in this case. When both candidates do not appear in the top-k ranking, no points are awarded in both algorithms.

Figure 1 gives the success probabilities of these algorithms in the case of 20 candidates and 2,000 voters for four different scoring rules: Borda, the harmonic rule, the geometric scoring rule with parameter $\rho = 1/2$, and the Copeland method. Figure 2 gives the success probabilities of Algorithm FairCutoff and Algorithm FairPWCutoff for 50 candidates and 10^4 voters.

Figure 1 shows that Algorithm FairCutoff outperforms the naive algorithm, and in most cases matches the performance of the optimal algorithm. The optimal algorithm performs significantly better only for Copeland. Figure 2 shows very clearly that Borda and Copeland are the hardest rules whereas the geometric scoring rule is the easiest. The success probability of Borda is closely related to the partition variability ratio r_k, as calculated in Theorem 3.2.

Our second set of simulations is designed to verify Theorem 5.1 and its extension to the case where the reference ranking is unknown (using $k = 1$). For various values of n and m, several scoring rules, and several values of the Mallows parameter ϕ, we computed the probability that the winner matches the first ranked candidate in the reference ranking, and the probability that the same candidate also appeared the most times as the first choice of the voters (marked *First* in the figure). The results for 20 candidates and 2,000 voters appear in Figure 3. The results displayed in the figure show that unless ϕ is very high (larger than roughly 0.8), the first ranked candidate almost always wins, and is almost

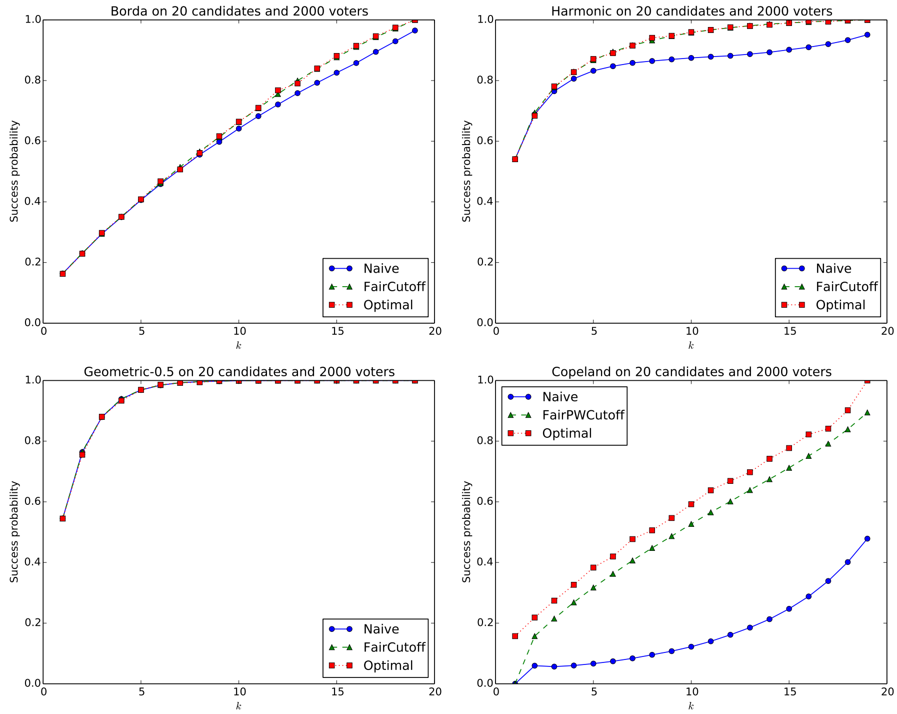

Figure 1: Success probabilities of various algorithms and various values of k in the case of 20 candidates and 2,000 voters

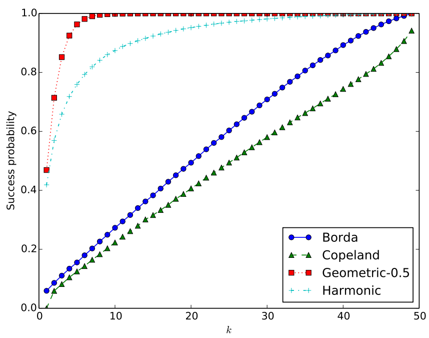

Figure 2: Success probabilities of Algorithm FairCutoff and Algorithm FairPWCutoff for various values of k in the case of 50 candidates and 10^4 voters

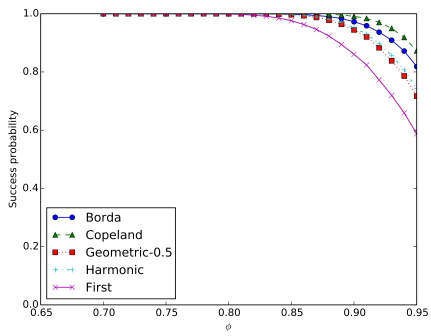

Figure 3: Recovery of first-ranked candidate in the case of 20 candidates and 10^4 voters

always identifiable by looking at the top votes. It also shows that our scoring rules are more reliable at recovering the first ranked candidate, compared to plurality (which corresponds to looking at the top votes).

8 Conclusions

We have studied a well-known method of preference elicitation. As we have shown, the approaches needed for the different input models that were considered differ substantially. For the neutral prior (impartial culture), we have presented a general technique for analyzing the bounds on the amount of information needed for correct winner selection, and demonstrated it on both positional scoring rules and the Copeland scoring rule. We also analyzed biased distributions, showing that the it is possible to predict the winner given only the biased distribution, and studied the limitations of the top-k scheme in the context of arbitrary distributions.

Our study raises a number of natural questions. To begin with the neutral prior, can we apply our technique to other scoring rules? Also, as mentioned in the paper, we believe that our bound for Copeland's voting rule can be improved.

As a different direction, it would be interesting to consider other elicitation schemes, and see whether analogous approaches can be applied to them. In particular, various iterative methods, as well as methods that rely on pairwise comparisons, have been studied extensively both empirically and from the perspective of rank aggregation. It would be interesting to obtain theoretically proven bounds for such schemes.

References

[1] Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (2008). *A first course in order statistics*, volume 54 of *Classics in Applied Mathematics*. SIAM.

[2] Baumeister, D., Faliszewski, P., Lang, J., and Rothe, J. (2012). Campaigns for lazy voters: Truncated ballots. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems - Volume 2*, AAMAS '12, pages 577–584, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.

[3] Boutilier, C., Caragiannis, I., Haber, S., Lu, T., Procaccia, A. D., and Sheffet, O. (2012). Optimal social choice functions: A utilitarian view. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, EC '12, pages 197–214, New York, NY, USA. ACM.

[4] Caragiannis, I., Procaccia, A. D., and Shah, N. (2013). When do noisy votes reveal the truth? In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*, EC '13, pages 143–160, New York, NY, USA. ACM.

[5] Conitzer, V. and Sandholm, T. (2002). Vote elicitation: Complexity and strategy-proofness. In *Eighteenth National Conference on Artificial Intelligence*, pages 392–397, Menlo Park, CA, USA. American Association for Artificial Intelligence.

[6] Conitzer, V. and Sandholm, T. (2005). Communication complexity of common voting rules. In *Proceedings of the 6th ACM conference on Electronic commerce*, pages 78–87. ACM.

[7] Dümbgen, L. (2010). Bounding standard gaussian tail probabilities. *arXiv preprint arXiv:1012.2063*.

[8] Kalech, M., Kraus, S., Kaminka, G. A., and Goldman, C. V. (2011). Practical voting rules with partial information. *Journal of Autonomous Agents and Multi-Agent Systems*, **22**(1), 151–182.

[9] Konczak, K. (2005). Voting procedures with incomplete preferences. In *in Proc. IJCAI-05 Multidisciplinary Workshop on Advances in Preference Handling*.

[10] Lu, T. and Boutilier, C. (2011a). Robust approximation and incremental elicitation in voting protocols. In *IJCAI*, pages 287–293.

[11] Lu, T. and Boutilier, C. (2011b). Vote elicitation with probabilistic preference models: Empirical estimation and cost tradeoffs. In R. Brafman, F. Roberts, and A. Tsoukis, editors, *Algorithmic Decision Theory*, volume 6992 of *Lecture Notes in Computer Science*, pages 135–149. Springer Berlin Heidelberg.

[12] Makarychev, Y. (2012). Balls and Bins analysis in the $m \gg n$ regime: gaps (answer). http://cstheory.stackexchange.com/a/14539/40.

[13] Mallows, C. L. (1957). Non-null ranking models. In *Biometrika*, volume 44, pages 114–130.

[14] Marden, J. I. (1995). *Analyzing and modeling rank data*. Chapman and Hall, London.

[15] Oren, J., Filmus, Y., and Boutilier, C. (2013). Efficient vote elicitation under candidate uncertainty. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, IJCAI'13, pages 309–316. AAAI Press.

[16] Xia, L. and Conitzer, V. (2008). Determining possible and necessary winners under common voting rules given partial orders. In *Proceedings of the 23rd National Conference on Artificial Intelligence — Volume 1*, AAAI'08, pages 196–201. AAAI Press.

Yuval Filmus
Institute of Advanced Study
Princeton, USA
Email: `yfilmus@ias.edu`

Joel Oren
University of Toronto
Toronto, Canada
Email: `oren@cs.toronto.edu`