Optimal Aggregation of Uncertain Preferences*

Ariel D. Procaccia
Computer Science Department
Carnegie Mellon University
arielpro@cs.cmu.edu

Nisarg Shah
Computer Science Department
Carnegie Mellon University
nkshah@cs.cmu.edu

Abstract

A paradigmatic problem in social choice theory deals with the aggregation of subjective preferences of individuals — represented as rankings of alternatives — into a social ranking. We are interested in settings where individuals are uncertain about their own preferences, and represent their uncertainty as distributions over rankings. Under the classic objective of minimizing the (expected) sum of Kendall tau distances between the input rankings and the output ranking, we establish that preference elicitation is surprisingly straightforward and near-optimal solutions can be obtained in polynomial time. We show, both in theory and using real data, that ignoring uncertainty altogether can lead to suboptimal outcomes.

1 Introduction

Recent years have seen a growing interest in the problem of predicting the objective quality of alternatives based on noisy votes over them (Conitzer and Sandholm 2005; Lu and Boutilier 2011; Conitzer, Rognlie, and Xia 2009; Elkind, Faliszewski, and Slinko 2010; Azari Soufiani, Parkes, and Xia 2012; Azari Soufiani et al. 2013; Azari Soufiani, Parkes, and Xia 2014; Procaccia, Reddi, and Shah 2012; Jiang et al. 2014). Research on this problem is partly driven by applications to crowdsourcing (Procaccia, Reddi, and Shah 2012; Caragiannis, Procaccia, and Shah 2013) and multiagent systems (Jiang et al. 2014) — domains where uncertainty stems from the limited ability of voters to identify an objective ground truth. In contrast, we study situations where voters are uncertain about their own subjective preferences.

Taking a step back, we note that the rigorous study of social choice dates back to the late 18th Century, but, more than two centuries later, its most prominent application — political elections — is just as relevant. In an election, voters express their subjective preferences over the candidates, and the goal is to reach social consensus among these possibly conflicting preferences. However, it is often difficult for voters to accurately determine their preferences due to missing information. This issue is exacerbated as candidates often try to hide their position on the sensitive issues that voters care about. To illustrate this point, (Shepsle 1972) gives the fascinating — albeit disturbing — example of the campaign strategy devised by Nicholas Biddle, the manager of William Henry Harrison's campaign for president of the United States: *"Let him say not one single word about his principles, or his creed — let him say nothing — promise nothing. Let no Committee, no convention — no town meeting ever extract from him a single word, about what he thinks now, or what he will do hereafter. Let the use of pen and ink be wholly forbidden as if he were a mad poet in Bedlam."*

Nevertheless, when voters actually vote in an election, they are required to distill a deterministic vote. Aggregating such votes ignores the underlying uncertainty. Sadly, even if one had a way of taking uncertainty into account, political election procedures are notoriously hard to change.

In contrast, the Internet has given rise to flexible voting platforms that draw on academic research; examples include All Our Ideas (`www.allourideas.org`) — which has already collected more than seven million votes, Pnyx (`pnyx.dss.in.tum.de`), and Whale (`http://whale3.noiraudes.net/whale3/index.do`). Much like political elections, these platforms currently ignore voters' uncertainty regarding their own subjective preference. But uncertainty does exist. Even in small-scale settings such as a group of friends choosing a restaurant or a movie, uncertainty stems from lack of information, knowledge, or deliberation time. This is also true in large-scale settings. For example, All Our Ideas hosted a popular survey conducted by the New York City Mayor's office to choose from various ideas to make NYC "greener and greater". But it is hard to accurately compare such ideas due to their unpredictable overall impact on the city.

Motivated by the foregoing observations, the goal of this paper is to *study social consensus in the presence of voters' uncertainty about their own subjective preferences, design effective rules for eliciting and aggregating uncertain votes, and quantify the benefit of doing so.*

Our approach. We use an expressive model of uncertainty: We represent an uncertain vote as a distribution over rankings. On one end of the spectrum, a supremely confident voter will report a distribution with singleton support. On the other end of the spectrum, a clueless voter will report a uniform distribution. While in general this type of information seems difficult to elicit, we will show that, for our purposes, it is sufficient to ask queries of the form "how likely is it that you prefer x to y?" Importantly, this is close in spirit to

*A preliminary version of this paper appeared in Proceedings of the 30th AAAI Conference on Artificial Intelligence, 2016.

pairwise comparison queries of the form "do you prefer x to y?", which platforms such as All Our Ideas already use.

It remains to define a *good* output ranking with respect to uncertain input votes. To this end, let us define the *Kendall tau (KT) distance* between two rankings as the number of pairs of alternatives on which the two rankings disagree; it is equal to the number of swaps bubble sort would require to convert one ranking into the other. Our goal is to find a ranking that minimizes the expected sum of KT distances to the voters' actual rankings, where the expectation is taken over the voters' uncertainty about their own preferences.

Being able to quantify the quality of an output ranking is important because we would like to measure the loss in quality when uncertainty is not taken into account. But why this specific measure — the sum of KT distances? It is an extremely well studied measure in the classical setting (with no uncertainty); its minimizer — the Kemeny rule — has many virtues. It is characterized by a number of desirable axiomatic properties (Young and Levenglick 1978), and also has alternative justifications in the *distance rationalizability* framework (Meskanen and Nurmi 2008) and in the *maximum likelihood estimation* framework (Young 1988).

Our Results. Our first result (Theorem 1) shows that to compute the optimal ranking, from the elicitation viewpoint we only need to ask voters to report their likelihood of preferring one alternative to another, and from the computational viewpoint the problem can be formulated as the popular $\mathcal{NP}$-hard problem of finding the minimum feedback arc set of a tournament (FAST). Our next result (Theorem 2) offers two methods to reduce the computation and elicitation burden — at the cost of only computing an approximate solution with high confidence — and provides a tradeoff between the two measures: one method draws on the existing polynomial-time approximation scheme for FAST (Kenyon-Mathieu and Schudy 2007), and the other leverages a novel result (Lemma 1) about feedbacks of approximate tournaments, which may be of independent interest. We also investigate the structure of the optimal rule in a special case, and show that it can be highly counterintuitive (Theorem 3).

Finally, we show (Theorem 4) that ignoring uncertainty altogether — as is done today — can lead to moderately or severely suboptimal outcomes in the worst case, depending on the way the objective function is defined. Our experimental results in Section 5 indicate that this is true even with preferences from real-world datasets.

Additional related work. Perhaps the most closely related work to ours is that of Enelow and Hinich (1981), who propose ways of estimating the level of uncertainty in subjective preferences in political elections from survey data. However, they do not address aggregation, which is the focus of our work. Our proposal of asking voters the *likelihood* of preferring one alternative to another is reminiscent of similar proposals (Burden 1997; Alvarez and Franklin 1994), but our work is driven by optimal aggregation. Ok et al. (2012) decouple two sources of uncertainty in subjective preferences: indecisiveness in beliefs versus tastes. We emphasize that the uncertainty studied in this work is different from many other forms of uncertainty studied in social choice (Nurmi 2002; Caragiannis, Procaccia, and Shah 2013).

Uncertainty in rank aggregation has recently been explored in machine learning (Niu et al. 2013; Soliman and Ilyas 2009). In a closely related paper, Niu et al. (2013) propose aggregating web page rankings into a single ranking by first *artificially* converting the deterministic inputs into distributions, and then aggregating these distributions. In contrast, in our setting the distributions are inputs from the voters themselves. Further, unlike us, Niu et al. (2013) do not focus on the elicitation or computational aspects. A separate line of work in crowdsourcing deals with estimating the quality of workers, and using it as an indicator of their accuracy while aggregating their opinions to pinpoint an objective ground truth (Welinder et al. 2010; Dekel and Shamir 2009). In contrast, in our work there is no objective ground truth, and the uncertainty of the voters is part of the input.

2 Model

Let $[k] \triangleq \{1, \ldots, k\}$. Let A denote a set of m *alternatives*, and $\mathcal{L}(A)$ be the set of rankings of the alternatives. For $\sigma \in \mathcal{L}(A)$, $a \succ_\sigma b$ denotes that alternative a is preferred to alternative b in σ. Let the set of *voters* be $N = \{1, \ldots, n\}$.

In the classical setting, each voter $i \in [n]$ submits a vote $\sigma_i \in \mathcal{L}(A)$, which is a ranking representing the voter's subjective preferences over the alternatives. A collection of submitted votes is called a (preference) *profile*, and is typically denoted by π. A *voting rule* (technically, a *social welfare function*) $f : \mathcal{L}(A)^n \to \mathcal{L}(A)$ is a function that aggregates the input rankings into a social ranking. A well-known voting rule is the Kemeny rule, which finds the social ranking (called the *Kemeny ranking*, and denoted σ_{KEM}) that minimizes the sum of *Kendall tau (KT)* distances from the input rankings. The Kendall tau distance, denoted d_{KT}, counts the number of pairs of alternatives on which two rankings disagree. Hence, $\sigma_{\text{KEM}} = \arg\min_{\sigma \in \mathcal{L}(A)} \sum_{i=1}^{n} d_{KT}(\sigma, \sigma_i)$.

Let us describe an alternative way of thinking about the Kemeny rule. A *weighted tournament* (hereinafter, simply a *tournament*) is a complete directed graph with two weighted edges between each pair of vertices (one in each direction). *Minimum feedback arc set for tournaments (FAST)* is the problem of finding a ranking of the vertices (hereinafter, the *minimum feedback ranking*, or the *optimal ranking*) that minimizes the sum of weights of edges that disagree with the ranking (i.e., that go from a lower-ranked vertex to a higher-ranked vertex). Given a profile π, its *weighted pairwise majority tournament* is the tournament whose vertices are the alternatives, and the weight of the edge from alternative a to alternative b, denoted w_{ab}, is the number of voters who prefer a to b. It is easy to check that the minimum feedback ranking of this tournament is the Kemeny ranking of π.

In this paper, we consider voters who are uncertain about their own subjective preferences. The uncertain preferences of voter i are a distribution over rankings D_i from which the voter's actual preferences σ_i are drawn. Let $\Delta(\mathcal{L}(A))$ denote the set of distributions over rankings of alternatives in A. In our setting, a voting rule $f : \Delta(\mathcal{L}(A))^n \to \mathcal{L}(A)$ aggregates the uncertain votes of the voters into a social ranking. Extending the reasoning behind the Kemeny rule, let the objective function $h(\sigma) = \mathbb{E}_{\sigma_i \sim D_i, \forall i \in [n]} \sum_{i=1}^{n} d_{KT}(\sigma, \sigma_i)$ be the *expected* sum of Kendall tau distances from the uncertain input votes, and let the optimal ranking $\sigma_{\text{OPT}} = \arg\min_{\sigma \in \mathcal{L}(A)} h(\sigma)$ be its minimizer.

3 Computation and Elicitation

To naïvely compute the optimal ranking, one would need to elicit the complete distribution D_i from each voter i, compute the objective function value (the expected sum of KT distances from the input distributions) for every possible ranking, and then select the ranking with the smallest objective function value. However, this is nastily expensive: it requires $\Omega(m!)$ communication and $\Omega((m!)^2)$ computations! Fortunately, both requirements can be reduced significantly.

Theorem 1. *Given the voters' uncertain subjective preferences as distributions over rankings, the social ranking that minimizes the expected sum of Kendall tau distances from the voters' preferences is the minimum feedback ranking of the tournament over the alternatives where the weight of the edge from alternative a to alternative b is the sum of probabilities of the voters preferring a to b.*

Proof. First, using linearity of expectation and the definition of the KT distance, we get

$$h(\sigma) = \sum_{i=1}^{n} \mathbb{E}_{\sigma_i \sim D_i} \sum_{a,b \in A : a \succ_\sigma b} \mathbb{I}[b \succ_{\sigma_i} a]$$
$$= \sum_{a,b \in A : a \succ_\sigma b} \sum_{i=1}^{n} \mathbb{E}_{\sigma_i \sim D_i} \left[\mathbb{I}[b \succ_{\sigma_i} a] \right],$$

where $\mathbb{I}$ is the indicator function, and the final transition follows from linearity of expectation and by interchanging summations. In the final expression, $\mathbb{E}_{\sigma_i \sim D_i}[\mathbb{I}[b \succ_{\sigma_i} a]] = \Pr_{D_i}[b \succ a]$ is the probability that voter i prefers b to a. Note that $h(\sigma)$ is simply the feedback of σ for the required tournament, and $\sigma_{\text{OPT}} = \arg\min_{\sigma \in \mathcal{L}(A)} h(\sigma)$ is the minimum feedback ranking. ■

Due to Theorem 1, the communication requirement for computing σ_{OPT} can be substantially reduced to eliciting only $O(n \cdot m^2)$ pairwise comparison probabilities from the voters, and the computational requirement can be substantially reduced to $O((n + m!) \cdot m^2)$ (computing the required tournament and finding its minimum feedback ranking). However, the running time is still exponential in the number of alternatives m, which is unavoidable because FAST is $\mathcal{NP}$-hard (Bartholdi, Tovey, and Trick 1989).

3.1 Computationally Efficient Approximations

Interestingly, there is a polynomial-time approximation scheme (PTAS) for FAST (Kenyon-Mathieu and Schudy 2007), which means that for any constant $\epsilon > 0$, one can find a ranking whose feedback is at most $(1 + \epsilon)$ times the minimum feedback, in time that is polynomial in m. This is serendipitous, as the minimum feedback in general (non-tournament) graphs is hard to approximate to a factor of 1.361 (Kenyon-Mathieu and Schudy 2007).

For a constant $\epsilon > 0$, the PTAS, when applied to the tournament from Theorem 1, returns a ranking $\widehat{\sigma}$ with $h(\widehat{\sigma}) \leq (1 + \epsilon) \cdot h(\sigma_{\text{OPT}})$ in time that is polynomial in m (and exponential in $1/\epsilon$). Interestingly, the PTAS only requires the edge weights up to an additive accuracy of $n\epsilon/m^2$.[1] We leverage this flexibility to further reduce the burden of elicitation. Using Hoeffding's inequality, one can easily show that eliciting each pairwise comparison probability from only $O(m^4 \log(m/\delta)/\epsilon^2)$ voters is sufficient to estimate the edge weights such that with probability at least $1 - \delta$, every estimate has an additive error of at most $n\epsilon/m^2$. Crucially, the number of voters required to compare a given pair of alternatives is independent of the total number of voters n.

While the PTAS helps achieve polynomial running time, in practice intractability is often not a serious concern in the first place due to a growing array of fast, exact algorithms developed for FAST (Conitzer, Davenport, and Kalagnanam 2006; Betzler et al. 2009). We show that with an exact solver for FAST, we can further reduce the elicitation burden.

Lemma 1. *Let $0 \leq \epsilon \leq 1/3$. If there exists a constant $c > 0$ such that the edge weights of a tournament satisfy $w_{ab} + w_{ba} \geq c$ for all pairs of vertices (a, b), then the minimum feedback of the tournament in which the weights are approximated up to an additive accuracy of $c \cdot \epsilon$ provides a $(1 + 12\epsilon)$-multiplicative approximation for the minimum feedback in the original tournament.*

Before we dive into the proof, let us compare Lemma 1 to the PTAS. While Lemma 1 allows more error in the weights (by a factor of m^2), it requires an exact solution to the tournament with approximate weights (thus keeping the problem computationally intractable). In contrast, the PTAS only approximately solves the tournament with approximate weights (thus allowing polynomial running time). Because minimum feedback arc set is an important optimization problem with numerous applications, we believe that Lemma 1 may be of independent interest.

Proof of Lemma 1. Let V denote the set of vertices of the tournament. Let w_{ab} and $\widehat{w}_{ab}$ denote the true and the approximate weights, respectively, on the edge from a to b. Let $F(\cdot)$ and $\widehat{F}(\cdot)$ denote the feedbacks in the tournaments with the true and the approximate weights, respectively. Finally, let σ and $\widehat{\sigma}$ denote the minimizers of $F(\cdot)$ and $\widehat{F}(\cdot)$, respectively. Then, we wish to prove that given $|w_{ab} - \widehat{w}_{ab}| \leq c \cdot \epsilon$ for all distinct $a, b \in V$, we have $F(\widehat{\sigma}) \leq (1 + 12\epsilon) \cdot F(\sigma)$.

Let M be the set of pairs of vertices on which σ and $\widehat{\sigma}$ agree, and let N be the set of remaining pairs of vertices. For a ranking τ, let $F_M(\tau)$ and $F_N(\tau)$ (resp., $\widehat{F}_M(\tau)$ and $\widehat{F}_N(\tau)$) denote the parts of its feedback $F(\tau)$ (resp., $\widehat{F}(\tau)$) over the pairs of alternatives in M and N, respectively. Then

$$F_M(\sigma) = F_M(\widehat{\sigma}), \quad \widehat{F}_M(\sigma) = \widehat{F}_M(\widehat{\sigma})$$
$$0 \leq F(\widehat{\sigma}) - F(\sigma) = F_N(\widehat{\sigma}) - F_N(\sigma)$$
$$0 \leq \widehat{F}(\sigma) - \widehat{F}(\widehat{\sigma}) = \widehat{F}_N(\sigma) - \widehat{F}_N(\widehat{\sigma}).$$

[1] It is obvious that the PTAS can only access polynomially many bits, but this result is *much* more flexible.

Because each approximate edge weight has an additive error of at most $c \cdot \epsilon$, we have $|F_N(\tau) - \widehat{F}_N(\tau)| \leq |N| \cdot c \cdot \epsilon$ for all rankings τ. We now obtain an upper bound on $F(\widehat{\sigma})$.

$$
\begin{aligned}
F(\widehat{\sigma}) &= F_M(\widehat{\sigma}) + F_N(\widehat{\sigma}) \leq F_M(\widehat{\sigma}) + (\widehat{F}_N(\widehat{\sigma}) + |N| \cdot c \cdot \epsilon) \\
&\leq F_M(\sigma) + \widehat{F}_N(\sigma) + |N| \cdot c \cdot \epsilon \\
&\leq F_M(\sigma) + F_N(\sigma) + 2 \cdot |N| \cdot c \cdot \epsilon \\
&= F(\sigma) + 2 \cdot |N| \cdot c \cdot \epsilon. \qquad (1)
\end{aligned}
$$

To convert the additive bound into a multiplicative bound, we analyze the sum of the two quantities.

$$F(\widehat{\sigma}) + F(\sigma) \geq F_N(\widehat{\sigma}) + F_N(\sigma) = \sum_{(a,b)\in N} w_{ab} + w_{ba} \geq |N| \cdot c,$$

where the second transition holds because σ and $\widehat{\sigma}$ rank the pair $(a, b) \in N$ differently, and the final transition follows from the assumption $w_{ab} + w_{ba} \geq c$. Substituting this upper bound on $|N| \cdot c$ into the right hand side of Equation (1), and simplifying, we get that

$$F(\widehat{\sigma}) \leq F(\sigma) \cdot \frac{1 + 2\epsilon}{1 - 2\epsilon} \leq (1 + 12\epsilon)F(\sigma).$$

In the last inequality, $(1+2\epsilon)/(1-2\epsilon) = 1+4\epsilon/(1-2\epsilon) \leq 1 + 12\epsilon$ holds because $\epsilon \leq 1/3$. ■

Due to Lemma 1, a $(1 + \epsilon)$-approximation to the optimal ranking only requires edge weights up to an additive accuracy of $\Theta(n \cdot \epsilon)$. Combining with Hoeffding's inequality, a $(1 + \epsilon)$-approximate solution can be computed with probability at least $1 - \delta$ by asking only $O(\log(m/\delta)/\epsilon^2)$ voters to compare each pair of alternatives, thus improving the previous bound by a significant factor of m^4.

To conclude, the PTAS and our novel approach for handling additive error bounds (Lemma 1) provide approximate solutions to our problem by offering different tradeoffs between the computational and communication requirements.

Theorem 2. *It is possible to compute a ranking that provides a $(1 + \epsilon)$-approximation to the optimal ranking with probability at least $1 - \delta$ in two ways:*

1. *Asking $O(m^4 \log(m/\delta)/\epsilon^2)$ random voters for each pairwise comparison, and running a PTAS for FAST on the tournament with estimated weights, which guarantees polynomial running time.*
2. *Asking $O(\log(m/\delta)/\epsilon^2)$ random voters for each pairwise comparison, and running an exact solver for FAST on the tournament with estimated weights (assuming $\epsilon < 1/3$), which does not guarantee polynomial running time.*

3.2 Special Case: The Mallows Model

The characterization of the optimal ranking in Theorem 1 is concise, but does not provide any deep intuition behind how a single voter's level of uncertainty affects the outcome. In this section we examine a setting where a voter's uncertain preferences can be represented by a single ranking together with a single real-valued confidence parameter.

In more detail, we represent a voter's uncertain preferences using the popular Mallows model (1957), which is parametrized by a central ranking $\sigma^* \in \mathcal{L}(A)$, and a noise parameter $\varphi \in [0, 1]$. Given these parameters, the probability of drawing a ranking σ is $\Pr[\sigma|\sigma^*, \varphi] = \varphi^{d_{KT}(\sigma,\sigma^*)}/Z_\varphi^m$, where Z_φ^m is a normalization constant that can be shown to be independent of the central ranking σ^*. The noise parameter φ — which can be thought of as the level of uncertainty — can be varied smoothly from $\varphi = 0$, which represents perfect confidence, to $\varphi = 1$, which represents the uniform distribution (which has the greatest amount of uncertainty). The Mallows model has been used extensively in social choice and machine learning applications (see, e.g., Lebanon and Lafferty 2002; Procaccia, Reddi, and Shah 2012; Lu and Boutilier 2011). We remark that under the Mallows model pairwise comparison probabilities required to compute the edge weights in Theorem 1 have a closed form that can be evaluated easily (Mao et al. 2014).

To distill the essence of the optimal solution under the Mallows model, let us simplify the problem by focusing on the case where all voters have identical noise parameter φ. Intuitively, the optimal ranking should coincide with the Kemeny ranking of the profile consisting of the central rankings of the Mallows models representing the uncertain votes. Indeed, as $\varphi \to 0$, this is obvious because the uncertain votes converge to point distributions around the central rankings. Hence, optimizing the sum of expected KT distances from these votes converges to optimizing the sum of KT distances from their central rankings. However, we show that this intuition breaks down when we consider the case of high uncertainty ($\varphi \to 1$). The case of $\varphi \to 1$ is especially important because, in a sense, it maximizes the effect of uncertainty on the optimization objective, and, indeed, this case has also received special attention in the past (Procaccia, Reddi, and Shah 2012).

Theorem 3. *When the uncertain preferences of the voters are represented using the Mallows model with central rankings $\{\sigma_i^*\}_{i\in[n]}$ and a common noise parameter φ, then:*

1. *There exists $\varphi^* < 1$ such that for all $\varphi \geq \varphi^*$, the optimal ranking minimizes the objective*

$$
\begin{aligned}
h'(\sigma) = &\sum_{a\in A} 2 \cdot \mathit{rank}(a, \sigma) \cdot \sum_{i=1}^{n} (m - \mathit{rank}(a, \sigma_i^*)) \\
&+ \sum_{a,b\in A: a\succ_\sigma b} \sum_{i=1}^{n} \mathbb{I}[b \succ_{\sigma_i^*} a],
\end{aligned}
$$

where $rank(a, \sigma)$ denotes the rank of alternative a in σ, and $\mathbb{I}$ is the indicator function.

2. *There exist values for $\{\sigma_i^*\}_{i\in[n]}$ and φ for which the optimal ranking does not minimize the Kemeny objective (the sum of distances from the central rankings).*

Proof. Let $\varphi = 1 - \epsilon$. Thus, as $\varphi \to 1$, we have $\epsilon \to 0$. Let σ_i^* denote the central ranking of the Mallows model for voter $i \in [n]$. Let us analyze our objective function for a ranking σ in this case.

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbb{E}_{\sigma_i \sim D_i} d_{KT}(\sigma, \sigma_i) \\
&= \sum_{i=1}^{n} \sum_{\sigma_i \in \mathcal{L}(A)} d_{KT}(\sigma, \sigma_i) \cdot \Pr[\sigma_i | \sigma_i^*, \varphi] \\
&\propto \sum_{i=1}^{n} \sum_{\sigma_i \in \mathcal{L}(A)} d_{KT}(\sigma, \sigma_i) \cdot \varphi^{d_{KT}(\sigma_i, \sigma_i^*)} \\
&\approx \sum_{i=1}^{n} \sum_{\sigma_i \in \mathcal{L}(A)} d_{KT}(\sigma, \sigma_i) \cdot (1 - \epsilon \cdot d_{KT}(\sigma_i, \sigma_i^*)), \quad (2)
\end{aligned}$$

where the last transition holds because $1 - \epsilon \cdot k$ is the first order approximation of $\varphi^k = (1-\epsilon)^k$ when $\epsilon \to 0$. It is thus clear that when $\epsilon \to 0$, the optimal ranking must minimize the final expression in Equation (2) (although, not every ranking minimizing the expression is optimal). Next, notice that in Equation (2), the sum $\sum_{\sigma_i \in \mathcal{L}(A)} d_{KT}(\sigma, \sigma_i)$ is constant for every $i \in [n]$ because the Kendall tau distance is neutral. Hence, minimizing the expression in Equation (2) amounts to *maximizing* the following objective function.

$$\begin{aligned}
&\sum_{i=1}^{n} \sum_{\sigma_i \in \mathcal{L}(A)} d_{KT}(\sigma, \sigma_i) \cdot d_{KT}(\sigma_i, \sigma_i^*) \\
&= \sum_{\tau \in \mathcal{L}(A)} \sum_{i=1}^{n} d_{KT}(\sigma, \tau) \cdot d_{KT}(\tau, \sigma_i^*) \\
&= \sum_{\tau \in \mathcal{L}(A)} d_{KT}(\sigma, \tau) \cdot d_{KT}(\tau, \pi), \quad (3)
\end{aligned}$$

where π is the preference profile consisting of the central rankings $\{\sigma_i^*\}_{i \in [n]}$, and $d_{KT}(\tau, \pi) = \sum_{i=1}^{n} d_{KT}(\tau, \sigma_i^*)$. Let n_{ab} denote the number of rankings in π in which a is preferred to b. Then, we expand the objective function in Equation (3) using the definition of the Kendall tau distance.

$$\begin{aligned}
&\sum_{\tau \in \mathcal{L}(A)} d_{KT}(\sigma, \tau) \cdot d_{KT}(\tau, \pi) \\
&= \sum_{\tau \in \mathcal{L}(A)} \left[\sum_{\substack{a,b \in A: \\ a \succ_\sigma b}} \mathbb{I}[b \succ_\tau a] \right] \times \\
&\qquad \left[\sum_{c,d \in A} \mathbb{I}[c \succ_\tau d] \cdot n_{dc} + \mathbb{I}[d \succ_\tau c] \cdot n_{cd} \right] \\
&= \sum_{\substack{a,b \in A: \\ a \succ_\sigma b}} \sum_{c,d \in A} \sum_{\tau \in \mathcal{L}(A)} \mathbb{I}[b \succ_\tau a] \times \\
&\qquad \Big[\mathbb{I}[c \succ_\tau d] \cdot n_{dc} + \mathbb{I}[d \succ_\tau c] \cdot n_{cd} \Big] \\
&= \sum_{\substack{a,b \in A: \\ a \succ_\sigma b}} \sum_{\substack{\tau \in \mathcal{L}(A): \\ b \succ_\tau a}} \sum_{c,d \in A} \mathbb{I}[c \succ_\tau d] \cdot n_{dc} + \mathbb{I}[d \succ_\tau c] \cdot n_{cd}. \quad (4)
\end{aligned}$$

Observe that the summation over τ contains all $m!/2$ rankings under which b is preferred to a. Also, note that in Equation (4), in the innermost summation over $c, d \in A$, each pair of alternatives is taken only once (i.e., the order of c and d does not matter). Let us focus on this summation. Define $sc(q) = \sum_{r \in A, r \neq q} n_{qr}$. There are three cases:

1. $c = a$ and $d = b$: In this case, $\mathbb{I}[b \succ_\tau a] = 1$ due to the restriction on τ in its summation. Hence, the overall contribution of this case to the sum is $m!/2 \cdot n_{ab}$.
2. $c = a$ and $d \in A \setminus \{a, b\}$: In this case, only one-thirds of the rankings which satisfy $b \succ a$ also satisfy $a \succ d$. Hence, we get $m!/2 \cdot (2/3 \cdot n_{ad} + 1/3 \cdot n_{da})$. Summing over all $d \in A \setminus \{a, b\}$ and observing that $n_{ad} + n_{da} = n$, we get that (up to an additive constant) the contribution of this case to the sum is $m!/2 \cdot 1/3 \cdot (sc(a) - n_{ab})$.
3. $c = b$ and $d \in A \setminus \{a, b\}$: This is similar to the previous case, but only one-thirds of the rankings which satisfy $b \succ a$ also satisfy $d \succ b$. Hence, we get $m!/2 \cdot (1/3 \cdot n_{bd} + 2/3 \cdot n_{db})$. Thus, summing over all $d \in A \setminus \{a, b\}$, we get that (up to an additive constant) the contribution of this case to the sum is $-m!/2 \cdot 1/3 \cdot (sc(b) - n_{ba})$.
4. $c, d \notin \{a, b\}$. In this case, $b \succ a$ and $c \succ d$ are independent events. Hence, exactly half of the $m!/2$ rankings that satisfy $b \succ a$ also satisfy $c \succ d$, and the rest half satisfy $d \succ c$. Combining with the fact that $n_{cd} + n_{dc} = n$, rthe overall contribution of this case to the sum is a constant.

Since additive constants can be ignored in maximizing the objective function, the objective function reduces to

$$\begin{aligned}
&\sum_{a,b \in A: a \succ_\sigma b} \frac{m!}{2} \cdot n_{ab} + \frac{m!}{2} \cdot \frac{1}{3} \cdot (sc(a) - n_{ab}) \\
&\qquad - \frac{m!}{2} \cdot \frac{1}{3} \cdot (sc(b) - n_{ba}) \\
&\propto \sum_{a,b \in A: a \succ_\sigma b} n_{ab} + \frac{1}{3} \cdot (sc(a) - n_{ab}) - \frac{1}{3} \cdot (sc(b) - n_{ba}) \\
&\propto \sum_{a,b \in A: a \succ_\sigma b} 2 \cdot n + (sc(a) - sc(b)) - n_{ba} \\
&= m \cdot (m-1) \cdot n + \sum_{a \in A} (m + 1 - 2 \cdot \text{rank}(a, \sigma)) \cdot sc(a) \\
&\qquad - \sum_{a,b \in A: a \succ b} n_{ba} \\
&= \frac{3}{2} \cdot m \cdot (m-1) \cdot n - \sum_{a \in A} 2 \cdot \text{rank}(a, \sigma) \cdot sc(a) \\
&\qquad - \sum_{a,b \in A: a \succ_\sigma b} n_{ba},
\end{aligned}$$

where the second transition uses the fact that $n_{ab} = n - n_{ba}$, and the last transition follows by observing that $sc(a)$ appears with a positive sign for each of $m - \text{rank}(a, \sigma)$ alternatives that it defeats in σ, and with a negative sign for each of $\text{rank}(a, \sigma) - 1$ alternatives that defeat it. Finally, we can again ignore the additive constant in maximization. Now, we can show that maximizing this objective function is equivalent to minimizing the objective function $h'(\sigma)$ given in the theorem statement by observing that

$$sc(a) = \sum_{t \in A\setminus\{a\}} n_{at} = \sum_{t \in A\setminus\{a\}} \sum_{i=1}^{n} \mathbb{I}[a \succ_{\sigma_i^*} t]$$
$$= \sum_{i=1}^{n} \sum_{t \in A\setminus\{a\}} \mathbb{I}[a \succ_{\sigma_i^*} t] = \sum_{i=1}^{n} m - \text{rank}(a, \sigma_i^*),$$

and

$$\sum_{a,b \in A: a \succ_\sigma b} n_{ba} = \sum_{a,b \in A: a \succ_\sigma b} \sum_{i=1}^{n} \mathbb{I}[b \succ_{\sigma_i^*} a].$$

As noted before, we have used first-order approximation in $(1-\epsilon)^k \approx 1 - \epsilon \cdot k$. Hence, every optimal ranking minimizes the objective function $h'(\cdot)$, but not every minimizer of $h'(\cdot)$ is an optimal ranking. Since $h'(\cdot)$ is a linear combination of the Borda and the Kemeny objectives, we can conclude that when the intersection of the set of Borda rankings and the set of Kemeny rankings is non-empty, the set of optimal ranking must be a subset of this intersection.

However, we give an example profile for which this intersection is empty, and *none of the optimal rankings* in our case is a Kemeny ranking. Consider the profile with 5 votes where 3 of the votes are $\sigma = 1 \succ 2 \succ 3 \succ 4 \succ \ldots \succ m$, and the remaining 2 votes are $\sigma' = 3 \succ 1 \succ 2 \succ 4 \succ \ldots \succ m$. In this case, the unique Kemeny ranking is σ, whereas it can be checked that the unique ranking minimizing the objective function $h'(\cdot)$ is σ' (which is the unique Borda ranking in this case). Hence, from our previous analysis it follows that for φ sufficiently close to 1, σ' should be the unique optimal ranking. ■

Curiously, in the objective function in part 1 of Theorem 3, $\sum_{i=1}^{n}(m - \text{rank}(a, \sigma_i^*))$ is known as the Borda score of alternative a in profile $\{\sigma_i^*\}_{i\in[n]}$. Hence, the rearrangement inequality implies that the ranking returned by the popular voting rule (in the classic sense) known as Borda count on the profile $\{\sigma_i^*\}_{i\in[n]}$, in which the alternatives are ranked in a non-increasing order of their Borda scores, minimizes the first term of the objective function. Next, the second term is the Kemeny objective (i.e., the sum of Kendall tau distances) of the profile $\{\sigma_i^*\}_{i\in[n]}$, which is minimized by the Kemeny ranking of the profile. Thus, optimizing the *expected Kemeny objective* on the profile consisting of uncertain votes reduces to optimizing a linear combination of the Borda and the Kemeny objectives on the profile consisting of the (deterministic) central rankings of these uncertain votes.

4 Ignoring Uncertainty

While voters are typically uncertain about their subjective preferences, and despite the fact that such uncertainty can be taken into account with minimal additional effort (Theorems 1 and 2), the fact remains that uncertainty is ignored in almost all real-world voting scenarios today. This raises a natural question: *How much do we lose by not taking uncertainty into account?* Our analysis in this section answers this question. In fact, our analysis applies to the objective of minimizing the sum of expected distances from uncertain votes, where the distance is measured using *any metric* over the space of rankings (and not just the Kendall tau distance that we used hereinbefore).

Let d be a distance metric over the space rankings. Given uncertain votes $\{D_i\}_{i\in[n]}$, let $h(\sigma) = \sum_{i=1}^{n} \mathbb{E}_{\sigma_i \sim D_i} d(\sigma, \sigma_i)$ be the objective function to be minimized, and let the optimal ranking σ_{OPT} be its minimizer. To measure the loss due to ignoring uncertainty, we need to know how the voters would vote if they were asked to report a single ranking (in the classic voting setting) when their subjective preferences are uncertain. There are many promising approaches (e.g., the voter may report the ranking with the highest probability); it is hard to make an objective choice. Hence, we take a more structured approach. Following Caragiannis, Procaccia, and Shah (2013), we say that a distribution over rankings is d-monotonic around a central ranking σ^* if $d(\sigma, \sigma^*) < d(\sigma', \sigma^*)$ implies $\Pr[\sigma|\sigma^*] \geq \Pr[\sigma'|\sigma^*]$.[2] We assume that the uncertain vote D_i of voter i is a d-monotonic distribution around a central ranking σ_i^*. When required to submit a single ranking, the voter simply reports σ_i^*.

For a voting rule f in the classical setting, we are interested in the worst-case (over the uncertain votes $\{D_i\}_{i\in[n]}$) multiplicative approximation ratio $h(f(\{\sigma_i^*\}_{i\in[n]}))/h(\sigma_{\text{OPT}})$. Before turning to our main result of this section, we need one more definition: a distance metric over rankings is called *neutral* if relabeling alternatives in two rankings in the same fashion does not alter the distance between them. That is, the distance metric is invariant to the labels of the alternatives. Neutrality is an extremely mild assumption satisfied by all reasonable distance metrics (including the KT distance).

Theorem 4. *When the voters' uncertain votes are d-monotonic distributions for a neutral metric d:*

1. *Minimizing the sum of distances from the central rankings of the uncertain votes provides a 3-approximation to minimizing the sum of expected distances from the actual preferences.*
2. *In the worst case, the approximation ratio can be at least $1 + \text{diam}(d)/\text{avg}(d)$, where $\text{diam}(d) = \max_{\sigma,\sigma' \in \mathcal{L}(A)} d(\sigma, \sigma')$ is the diameter, and $\text{avg}(d) = \sum_{\sigma' \in \mathcal{L}(A)} d(\sigma, \sigma')/m!$ (which is independent of σ for a neutral d) is the average distance. In particular, for the Kendall tau distance the factor of 3 is tight.*

Proof. First, it is easy to verify that the central ranking σ_i^* of the distribution D_i minimizes the expected distance (measured by d) from D_i, i.e., $\sigma_i^* = \arg\min_{\sigma \in \mathcal{L}(A)} \mathbb{E}_{\sigma_i \sim D_i} d(\sigma, \sigma_i)$. This follows from the rearrangement inequality and the definition of d-monotonicity.

Let $\sigma_{\text{KEM}} = \arg\min_{\sigma \in \mathcal{L}(A)} \sum_{i=1}^{n} d(\sigma, \sigma_i^*)$ denote the Kemeny ranking of the profile of central rankings $\{\sigma_i^*\}_{i\in[n]}$.

[2] Our definition is more general. For instance, it includes uniform and point distributions while the definition of Caragiannis, Procaccia, and Shah (2013) excludes them as it requires $\Pr[\sigma|\sigma^*]$ to be monotonically increasing in $d(\sigma, \sigma^*)$.

Let $\sigma_{\text{OPT}} = \arg\min_{\sigma \in \mathcal{L}(A)} h(\sigma)$ denote the optimal ranking, where $h(\sigma) = \sum_{i=1}^n E_{\sigma_i \sim D_i} d(\sigma, \sigma_i)$. We are interested in $h(\sigma_{\text{KEM}})/h(\sigma_{\text{OPT}})$. Define $K = \sum_{i=1}^n d(\sigma_{\text{KEM}}, \sigma_i^*)$ and $X = \sum_{i=1}^n \mathbb{E}_{\sigma_i \sim D_i} d(\sigma_i, \sigma_i^*)$. Then, we have

$$h(\sigma_{\text{KEM}}) \leq \sum_{i=1}^n \mathbb{E}_{\sigma_i \sim D_i}[d(\sigma_{\text{KEM}}, \sigma_i^*) + d(\sigma_i^*, \sigma_i)] = K + X,$$

$$h(\sigma_{\text{OPT}}) \geq \sum_{i=1}^n \mathbb{E}_{\sigma_i \sim D_i} d(\sigma_i^*, \sigma_i) = X,$$

where the first equation follows due to the triangle inequality, and the second equation holds because $\sigma_i^* = \arg\min_{\sigma \in \mathcal{L}(A)} E_{\sigma_i \sim D_i} d(\sigma_i, \sigma)$. If $K \leq X$, we already have $h(\sigma_{\text{KEM}})/h(\sigma_{\text{OPT}}) \leq (K+X)/X \leq 2$.[3] Suppose $K > X$. Then, by the triangle inequality, we have

$$h(\sigma_{\text{OPT}}) \geq \sum_{i=1}^n \mathbb{E}_{\sigma_i \sim D_i}[d(\sigma_{\text{OPT}}, \sigma_i^*) - d(\sigma_i^*, \sigma_i)] \geq K - X,$$

where the last inequality follows from the definitions of K and σ_{KEM}. Hence, $h(\sigma_{\text{OPT}}) \geq \max(X, K-X)$. Substituting this, we get that the approximation ratio is at most

$$\begin{aligned} \frac{K+X}{\max(X, K-X)} &\leq \frac{K}{\max(X, K-X)} + 1 \\ &= \min\left(\frac{X}{K-X}, \frac{K-X}{X}\right) + 2 \leq 3, \end{aligned}$$

where the final step holds because either a number or its inverse must be at most 1.[3]

For the lower bound of part 2, let us construct a specific instance. Consider the two rankings σ and σ' that are at the greatest distance under d, i.e., $d(\sigma, \sigma') = \text{diam}(d)$. Suppose that the uncertain votes of $n/2$ voters coincide with a point distribution concentrated at ranking σ. Suppose the uncertain votes of the remaining $n/2$ voters coincide with the uniform distribution. Without loss of generality, let the central ranking for the uniform distribution be σ' such that $d(\sigma, \sigma') = \text{diam}(d)$. This assumption is without loss of generality because the effect can be achieved by considering d-monotonic distributions with central ranking σ' that are arbitrarily close to the uniform distribution. Next, note that there are several minimizers of the sum of distances from the central rankings, including σ and σ' themselves. Without loss of generality, assume that the minimizer returned by the voting rule is σ'. Again, this assumption is without loss of generality because one can consider profiles where the fractions of voters with central rankings σ and σ' are $1/2-\epsilon$ and $1/2+\epsilon$, respectively, for an arbitrarily small $\epsilon > 0$.

It is now easy to check that the objective function value achieved by σ' is $n/2 \cdot \text{diam}(d) + n/2 \cdot \text{avg}(d)$, whereas the minimum objective function value achieved by $\sigma_{\text{OPT}} = \sigma$ is $n/2 \cdot 0 + n/2 \cdot \text{avg}(d)$. Hence, the approximation ratio is $1 + \text{diam}(d)/\text{avg}(d)$, as required.

For the Kendall tau distance, $\text{diam}(d_{KT}) = \binom{m}{2}$, whereas $\text{avg}(d_{KT}) = \binom{m}{2}/2$ because any given ranking disagrees with a ranking chosen uniformly at random on half of the pairs of alternatives, in expectation. Hence, $1 + \text{diam}(d_{KT})/\text{avg}(d_{KT}) = 3$, meaning that the upper bound is tight. ■

[3]Let $0/0 = 1$, because if $h(\sigma_{\text{KEM}}) = h(\sigma_{\text{OPT}}) = 0$, then σ_{KEM} is optimal, i.e., a 1-approximation of σ_{OPT}.

While a factor of 3 is not terribly high from a theoretical-computer-science viewpoint, it can be significant in practical applications. Further, even in settings where this is acceptable, the result should be taken with a grain of salt. The feedback of a tournament is at least $\sum_{(a,b) \in A} \min(w_{ab}, w_{ba})$. Hence, the tournament in Theorem 1 (whose feedback is the objective function in Theorem 4) intrinsically has high feedback, leading to the relatively low approximation ratio.

Alternatively, we can define the objective function $\widehat{h}(\sigma) = h(\sigma) - \sum_{(a,b) \in A} \min(w_{ab}, w_{ba})$. By the foregoing discussion, clearly $\widehat{h}(\sigma) \geq 0$. This objective function is also intuitive, in the following sense: While h gives moderate weight to a voter who is completely uncertain about his preferences (and hence reports a uniform distribution), $\widehat{h}$ essentially ignores such a voter because he is going to be equally happy with any social ranking.

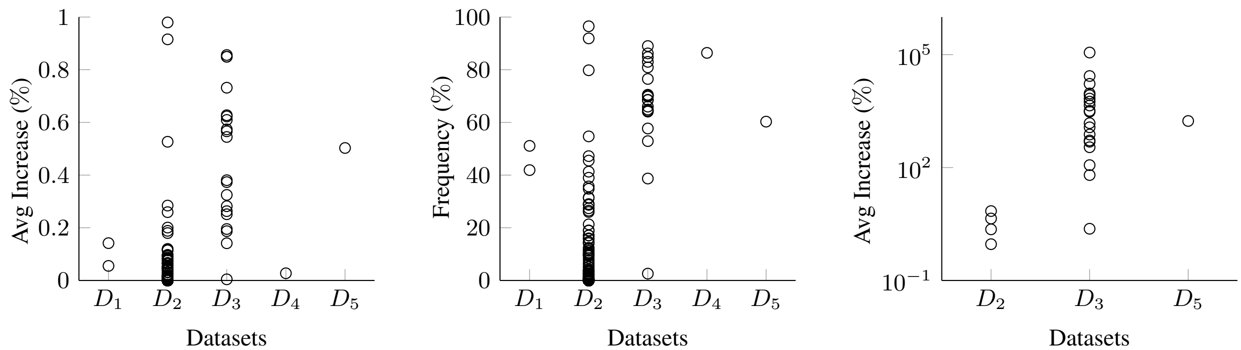
Interestingly, on the example we use to establish the lower bound in the proof of Theorem 4, minimizing the sum of distances from the central rankings has an infinite approximation ratio according to $\widehat{h}$. Thus, as usual, the multiplicative approximation factor is sensitive to the way the objective function is defined. Finally, observe that our optimality and approximation results (Theorems 1 and 2) also apply to optimizing the alternative objective function $\widehat{h}$.

5 Experimental Results

In Section 4 we demonstrated that in the *worst case*, ignoring uncertainty in the preferences can blow up the objective function value (which we seek to minimize) by a factor of 3 when using the standard objective function h or unboundedly when using the alternative objective function $\widehat{h}$. We now explore the impact of ignoring uncertainty using realistic, rather than worst case, preferences.

Due to the lack of real-world datasets with uncertain subjective preferences, we rely on datasets from Preflib (Mattei and Walsh 2013) that have subjective rankings, and introduce simulated uncertainty. Specifically, we use five datasets from Preflib: AGH Course Selection (D_1), Netflix (D_2), Skate (D_3), Sushi (D_4), and T-Shirt (D_5). Each dataset contains multiple preference profiles, with as many as $14,000$ voters (more than 750 on average) and as many as 30 alternatives (more than 5 on average). For each preference profile, we compute the approximation ratio averaged over 1000 simulations. In each simulation each vote in the profile is converted into an uncertain vote, represented as the Mallows model whose central ranking is the vote itself and whose noise parameter φ is chosen uniformly at random from $[0, 1]$. We use CPLEX to find the minimum feedback arc set through integer linear programming.

Figure 1(a) shows that there is up to 1% (that is, a mild) increase in the standard objective function h when ignoring uncertainty. With the alternative objective function $\widehat{h}$, however, the increase can be infinite. Figure 1(b) shows the



(a) Average increase in h (b) Frequency of infinite increase in $\widehat{h}$ (c) Average finite increase in $\widehat{h}$

Fig. 1: Effect of ignoring uncertainty in real preferences. Each dot represents a preference profile.

chances of observing infinite increase, which is significant for most preference profiles. While in datasets AGH and Sushi the multiplicative approximation factor always seems to be either infinity or 1, profiles from the other datasets exhibit a large increase in the alternative objective function even when averaged over simulations where it is not infinite; this is shown in Figure 1(c). Simulated profiles (simulating the central rankings, too, from either the Mallows model or the uniform distribution) give similar results.

6 Discussion

While the model of uncertainty we use — a general distribution over rankings — is very expressive, one may wish to generalize it further. Inspired by the random utility model (RUM), which has recently gained popularity in the machine learning literature (Azari Soufiani, Parkes, and Xia 2012, 2013; Azari Soufiani et al. 2013; Soufiani et al. 2013; Oh and Shah 2014), one may model the real subjective preferences of a voter as a utility for each alternative, and the uncertainty as a distribution over these utilities. It is unclear if restricted elicitation can lead to (approximately) optimal outcomes in this real-valued domain.

In addition, we would like to emphasize that uncertainty is ubiquitous, and our work opens doors to a variety of related domains. For example, in the closely related social choice setting with an underlying ground truth and objective (rather than subjective) votes, which is popular in the analysis of crowdsourcing systems, Shah and Zhou (2015) study mechanisms for incentivizing workers to correctly report their confidence levels. Optimal aggregation of the reported confidences is an open question; its analysis may lead to more accurate estimates of the ground truth, and thus to more effective crowdsourcing systems.

Acknowledgments

This work was partially supported by the NSF under grants CCF-1525932, CCF-1215883, and IIS-1350598, and by a Sloan Research Fellowship.

References

Alvarez, R. M., and Franklin, C. H. 1994. Uncertainty and political perceptions. *The Journal of Politics* 56(03):671–688.

Azari Soufiani, H.; Chen, W. Z.; Parkes, D. C.; and Xia, L. 2013. Generalized method-of-moments for rank aggregation. In *Proc. of 27th NIPS*, 2706–2714.

Azari Soufiani, H.; Parkes, D. C.; and Xia, L. 2012. Random utility theory for social choice. In *Proc. of 26th NIPS*, 126–134.

Azari Soufiani, H.; Parkes, D. C.; and Xia, L. 2013. Preference elicitation for general random utility models. In *Proc. of 29th UAI*, 596–605.

Azari Soufiani, H.; Parkes, D. C.; and Xia, L. 2014. Computing parametric ranking models via rank-breaking. In *Proc. of 31st ICML*, 360–368.

Bartholdi, J.; Tovey, C. A.; and Trick, M. A. 1989. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare* 6:157–165.

Betzler, N.; Fellows, M. R.; Guo, J.; Niedermeier, R.; and Rosamond, F. A. 2009. Fixed-parameter algorithms for Kemeny rankings. *Theoretical Computer Science* 410(45):4554–4570.

Burden, B. C. 1997. Deterministic and probabilistic voting models. *American Journal of Political Science* 1150–1169.

Caragiannis, I.; Procaccia, A. D.; and Shah, N. 2013. When do noisy votes reveal the truth? In *Proc. of 14th EC*, 143–160.

Conitzer, V., and Sandholm, T. 2005. Common voting rules as maximum likelihood estimators. In *Proc. of 21st UAI*, 145–152.

Conitzer, V.; Davenport, A.; and Kalagnanam, H. 2006. Improved bounds for computing Kemeny rankings. In *Proc. of 21st AAAI*, 620–626.

Conitzer, V.; Rognlie, M.; and Xia, L. 2009. Preference functions that score rankings and maximum likelihood estimation. In *Proc. of 21st IJCAI*, 109–115.

Dekel, O., and Shamir, O. 2009. Vox populi: Collecting high-quality labels from a crowd. In *Proc. of 22nd COLT*, 377–386.

Elkind, E.; Faliszewski, P.; and Slinko, A. 2010. Good rationalizations of voting rules. In *Proc. of 24th AAAI*, 774–779.

Enelow, J., and Hinich, M. J. 1981. A new approach to voter uncertainty in the Downsian spatial model. *American Journal of Political Science* 25(3):483–493.

Jiang, A. X.; Marcolino, L. S.; Procaccia, A. D.; Sadholm, T.; Shah, N.; and Tambe, M. 2014. Diverse randomized agents vote to win. In *Proc. of 28th NIPS*, 2573–2581.

Kenyon-Mathieu, C., and Schudy, W. 2007. How to rank with few errors. In *Proc. of 39th STOC*, 95–103.

Lebanon, G., and Lafferty, J. 2002. Cranking: Combining rankings using conditional probability models on permutations. In *Proc. of 19th ICML*, 363 – 370.

Lu, T., and Boutilier, C. 2011. Learning Mallows models with pairwise preferences. In *Proc. of 28th ICML*, 145–152.

Mallows, C. L. 1957. Non-null ranking models. *Biometrika* 44:114–130.

Mao, A.; Soufiani, H. A.; Chen, Y.; and Parkes, D. C. 2014. Capturing variation and uncertainty in human judgment. arXiv:1311.0251.

Mattei, N., and Walsh, T. 2013. Preflib: A library of preference data. In *Proc. of 3rd ADT*, 259–270.

Meskanen, T., and Nurmi, H. 2008. Closeness counts in social choice. In Braham, M., and Steffen, F., eds., *Power, Freedom, and Voting*. Springer-Verlag.

Niu, S.; Lan, Y.; Guo, J.; and Cheng, X. 2013. Stochastic rank aggregation. In *Proc. of 29th UAI*, 478–487.

Nurmi, H. 2002. *Voting Procedures Under Uncertainty*. Springer.

Oh, S., and Shah, D. 2014. Learning mixed multinomial logit model from ordinal data. In *Proc. of 28th NIPS*, 595–603.

Ok, E. A.; Ortoleva, P.; and Riella, G. 2012. Incomplete preferences under uncertainty: Indecisiveness in beliefs versus tastes. *Econometrica* 80(4):1791–1808.

Procaccia, A. D.; Reddi, S. J.; and Shah, N. 2012. A maximum likelihood approach for selecting sets of alternatives. In *Proc. of 28th UAI*, 695–704.

Shah, N. B., and Zhou, D. 2015. Double or nothing: Multiplicative incentive mechanisms for crowdsourcing. In *Proc. of 29th NIPS*. Forthcoming.

Shepsle, K. A. 1972. The strategy of ambiguity: Uncertainty and electoral competition. *American Political Science Review* 66(02):555–568.

Soliman, M. A., and Ilyas, I. F. 2009. Ranking with uncertain scores. In *Proc. of 25th ICDE*, 317–328.

Soufiani, H. A.; Diao, H.; Lai, Z.; and Parkes, D. C. 2013. Generalized random utility models with multiple types. In *Proc. of 27th NIPS*, 73–81.

Welinder, P.; Branson, S.; Belongie, S.; and Perona, P. 2010. The multidimensional wisdom of crowds. In *Proc. of 24th NIPS*, 2424–2432.

Young, H. P., and Levenglick, A. 1978. A consistent extension of Condorcet's election principle. *SIAM Journal on Applied Mathematics* 35(2):285–300.

Young, H. P. 1988. Condorcet's theory of voting. *The American Political Science Review* 82(4):1231–1244.