Journal of Artificial Intelligence Research 00 (0000) 0-0 Submitted 11/14; published 0/0

Modal Ranking: A Uniquely Robust Voting Rule*

Ioannis Caragiannis CARAGIAN@CEID.UPATRAS.GR
University of Patras

Ariel D. Procaccia ARIELPRO@CS.CMU.EDU
Carnegie Mellon University

Nisarg Shah NKSHAH@CS.CMU.EDU
Carnegie Mellon University

Abstract

Motivated by applications to crowdsourcing, we study voting rules that output a correct ranking of alternatives by quality from a large collection of noisy input rankings. We seek voting rules that are supremely robust to noise, in the sense of being correct in the face of any "reasonable" type of noise. We show that there is such a voting rule, which we call the *modal ranking* rule. Moreover, we establish that the modal ranking rule is the unique rule with the preceding robustness property within a large family of voting rules, which includes a slew of well-studied rules.

1. Introduction

The emergence of crowdsourcing platforms and human computation systems (Law & von Ahn, 2011) motivates a reexamination of an approach to voting that dates back to the Marquis de Condorcet (1785). He suggested that voters should be viewed as noisy estimators of a ground truth — a ranking of the candidates by their true quality. A *noise* model governs how voters make mistakes. For example, under the noise model suggested by Condorcet — also known today as the Mallows (1957) noise model — each voter ranks each pair of alternatives in the correct order with probability $p > 1/2$, and in the wrong order with probability $1 - p$ (roughly speaking). This specific noise model is quite unrealistic, and, more generally, the very idea of objective noise is arguable in the context of political elections, where opinions are subjective and there is no ground truth. However, the noisy voting setting is a perfect fit for crowdsourcing, where objective estimates provided by workers — often as votes (Little, Chilton, Goldman, & Miller, 2010; Mao, Procaccia, & Chen, 2013) — must be aggregated.

From this viewpoint, Condorcet and, more eloquently, Young (1988), argued that the goal of a voting rule — which aggregates input rankings into a single output ranking — should be to output the ranking that is most likely to be the ground truth ranking, under the given noise model. This approach has inspired a significant number of recent papers by AI researchers (Conitzer & Sandholm, 2005; Conitzer, Rognlie, & Xia, 2009; Elkind, Faliszewski, & Slinko, 2010; Xia, Conitzer, & Lang, 2010; Xia & Conitzer, 2011; Lu & Boutilier, 2011; Procaccia, Reddi, & Shah, 2012; Mao et al., 2013), some of which aim to design voting rules that are *maximum likelihood estimators (MLEs)* specifically for crowdsourcing settings.

But the maximum likelihood estimation requirement may be too stringent. Indeed, our recent work (Caragiannis, Procaccia, & Shah, 2013) points out that a voting rule may be an MLE for a

*. A preliminary version of this paper appeared in the Proceedings of the 28th AAAI Conference on Artificial Intelligence, pages 616-622, 2014.

specific noise model, but in realistic settings the noise can take unpredictable forms; see also (Mao et al., 2013). Instead, they propose the following robustness property, called *accuracy in the limit*: as the number of votes grows, the voting rule should output the ground truth ranking with high probability, i.e., with probability approaching one.[1] This allows a single voting rule to be robust against multiple noise models. Moreover, the focus on a large number of votes is natural in the context of crowdsourcing systems — the whole point is to aggregate information provided by a massive crowd! For example, social networks are enabling organizations to solicit noisy information from millions of users. Indeed, think of a technology company that asks fans to rank product prototypes by their perceived chance of success. While a large company can expect millions of votes, these votes are noisy and the type of noise is unpredictable.

In this paper, we seek voting rules that are robust against such unpredictable noise. Our research challenge is to

> *... find voting rules that are robust (in the accuracy in the limit sense) against* any *"reasonable" noise model.*

1.1 Overview of Results

We give a rather clear-cut solution to the preceding research challenge: There is a voting rule that is robust against any "reasonable" noise model, and it is *unique* within a huge family of voting rules. We call this supremely robust voting rule the *modal ranking* rule. Given a collection of input rankings, the modal ranking rule simply selects the most frequent ranking as the output. To the best of our knowledge, this strikingly basic voting rule has not received any attention in the literature, and for good reason: when the number of voters is not huge compared to the number of alternatives, it is likely that every ranking would appear at most once, so the modal ranking rule does not provide any useful guidance. However, when the number of voters is very large, the modal ranking rule is quite sensible; we will prove this intuitive claim formally.

To better understand this result (still on an informal level), we need to clarify two points: What do we mean by "reasonable" noise model? And what is the huge family of voting rules? Starting from the noise model, we employ additional notions introduced in our recent work (Caragiannis et al., 2013). We are interested in noise models that are *d-monotone* with respect to a distance function d on rankings, in the sense that the probability of a ranking increases as its distance according to d from the ground truth ranking decreases. A voting rule that is accurate in the limit with respect to any d-monotone noise model is said to be *monotone-robust* with respect to d. So, slightly more formally, the requirement is that the rule be monotone-robust with respect to any distance metric d.

Regarding the family of voting rules in which we prove that modal ranking is the unique robust rule, it is formed by the union of three families of rules: generalized scoring rules (GSRs) (Xia & Conitzer, 2008, 2009; Xia, 2013) with a "no holes" property, pairwise majority consistent (PM-c) rules, and position dominance consistent (PD-c) rules (Caragiannis et al., 2013). GSRs are a large family of voting rules that is known to capture almost all commonly-studied voting rules. Theorem 1 in Section 3 asserts that a GSR with no holes is monotone-robust with respect to all distance metrics if and only if it is the modal ranking rule. The no holes property is a technical restriction, but we show (Theorem 2 in Section 4) that it is quite mild by establishing that all prominent rules that are

1. In statistics, this property is known as *consistency*, but we avoid this terminology as it has completely different interpretations in social choice theory.

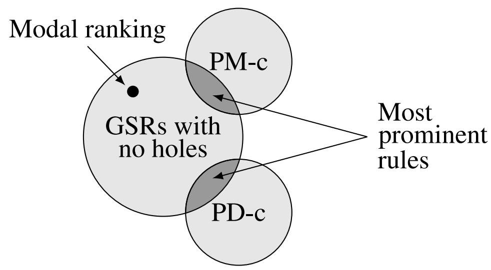

Figure 1: The modal ranking rule is uniquely robust within the union of three families of rules.

known to be GSRs have the no holes property. We believe that the concepts and proof techniques introduced in Section 4 may be of independent interest to the social choice community.

PM-c and PD-c rules together also contain most prominent voting rules (Caragiannis et al., 2013). These two families are disjoint, and neither is contained in the family of GSRs. Theorem 3 in Section 5 asserts that no PM-c or PD-c rule has the desired robustness property, thereby further extending the scope of the modal ranking rule's uniqueness to all rules that are either PM-c or PD-c or GSRs with no holes. See Figure 1 for a Venn diagram that illustrates the relation between these families of rules.

1.2 Related Work

This paper is most closely related to our recent work (Caragiannis et al., 2013), where we introduced the classes of PM-c and PD-c rules as well as the notions of d-monotone noise models, accuracy in the limit, and monotone-robustness. The main result in that paper is a characterization of the distance metrics d for which all PM-c and PD-c rules are monotone-robust. In other words, we *fixed the family of voting rules* to be PM-c or PD-c rules, and asked which distance metrics induce noise models for which *all* the rules in these families are robust. While the answer is a family of distance metrics that contains three popular distance metrics, it does not contain several other prominent distance metrics — moreover, it is by no means clear that natural distance metrics are the ones that induce the noise one encounters in practice. In contrast, in the current paper, instead of fixing the family of rules, we *fix the family of distances* to be all possible distance metrics d, and characterize the "family" of voting rules that are monotone-robust with respect to any d (this family turns out to be a singleton set).

On a conceptual level, unlike common voting rules, our modal ranking rule always returns a ranking from the given profile. Endriss and Grandi (2014) investigated a similar idea in the context of aggregation of binary opinions where each voter provides a yes/no opinion for a number of issues. In contrast to the modal ranking rule which (in our context) chooses the most frequent vote in the input profile, they investigated rules that return the input vote that is closest to the average or the majority vote.

On a technical level, we view our input profile (vector of rankings) as a point in $\mathbb{Q}^{m!}$ ($m!$ is the number of possible rankings), where each coordinate represents the fraction of times a ranking appears in the input profile. This geometric approach to the analysis of voting rules was initiated by Young (1975), and was later used by various other authors (Saari, 1995, 2008; Xia & Conitzer, 2009; Conitzer et al., 2009; Obraztsova, Elkind, Faliszewski, & Slinko, 2013; Mossel, Procaccia, & Rácz, 2013).

2. Preliminaries

Let A be the set of alternatives, where $|A| = m$. Let $\mathcal{L}(A)$ be the set of rankings (linear orders) over A, and $\mathcal{D}(\mathcal{L}(A))$ be the set of probability distributions over $\mathcal{L}(A)$. A vote σ is a ranking in $\mathcal{L}(A)$, and a profile π is a collection of votes. A voting rule (sometimes also known as a "rank aggregation rule") is formally a deterministic (resp., randomized) *social welfare function (SWF)* that maps every profile to a ranking (resp., a distribution over rankings). In this paper, we focus on randomized SWFs. Deterministic SWFs are a special case where the output distributions are centered at a single ranking. We do not study social choice functions (SCFs), which map each profile to a (single) selected alternative.

Families of SWFs. In order to capture many SWFs simultaneously, our results employ the definitions of three broad families of SWFs.

- *PM-c rules* (Caragiannis et al., 2013): For a profile π, the *pairwise-majority (PM) graph* is a directed graph whose vertices are the alternatives, and there exists an edge from $a \in A$ to $b \in A$ if a strict majority of the voters prefer a to b. A randomized SWF f is called *pairwise-majority consistent (PM-c)* if for every profile π with a complete acyclic PM graph whose vertices are ordered according to $\sigma \in \mathcal{L}(A)$, we have $\Pr[f(\pi) = \sigma] = 1$.

- *PD-c rules* (Caragiannis et al., 2013): In a profile π, alternative a is said to *position-dominate* alternative b if for every $k \in \{1, \ldots, m-1\}$, (strictly) more voters rank a in the first k positions than b. The *position-dominance (PD) graph* is a directed graph whose vertices are the alternatives, and there exists an edge from a to b if a position-dominates b. A randomized SWF f is called *position-dominance consistent (PD-c)* if for every profile π with a complete acyclic PD graph whose vertices are ordered according to $\sigma \in \mathcal{L}(A)$, we have $\Pr[f(\pi) = \sigma] = 1$.

- *GSRs* (Xia & Conitzer, 2008): We say that two vectors $y, z \in \mathbb{R}^k$ are *equivalent* (denoted $y \sim z$) if for every $i, j \in [k]$ we have $y_i \geq y_j \Leftrightarrow z_i \geq z_j$. We say that a function $g : \mathbb{R}^k \to \mathcal{D}(\mathcal{L}(A))$ is *compatible* if $y \sim z$ implies $g(y) = g(z)$. A *generalized scoring rule (GSR)* is given by a pair of functions (f, g), where $f : \mathcal{L}(A) \to \mathbb{R}^k$ maps every ranking to a k-dimensional vector and $g : \mathbb{R}^k \to \mathcal{D}(\mathcal{L}(A))$ is a compatible function that maps every k-dimensional vector to a distribution over rankings, and the output of the rule on a profile $\pi = (\sigma_1, \ldots, \sigma_n)$ is given by $g\left(\sum_{i=1}^{n} f(\sigma_i)\right)$. GSRs are characterized by two social choice axioms (Xia & Conitzer, 2009), and have interesting connections to machine learning (Xia, 2013). While GSRs were originally introduced as deterministic SCFs, the definition naturally extends to (possibly randomized) SWFs.

Popular SWFs. Let us define some popular SWFs that are captured by the families of SWFs defined above. First, define the weighted pairwise majority (PM) graph of a profile as the graph where the alternatives are the vertices and there is an edge from every alternative a to every other alternative b with weight equal to the fraction of voters that prefer a to b. Its main difference from the unweighted PM graph defined above is that the latter has at most one unweighted directed edge between two alternatives a and b, indicating which alternative is preferred by the (strict) majority of the voters.

- *The Kemeny rule*: Given a profile π, the Kemeny score of a ranking is the total weight of the edges of the weighted PM graph of π in its direction. The Kemeny rule selects the ranking with the highest Kemeny score. Tie-breaking is used to choose among all the rankings with identical highest Kemeny score.

- *Positional scoring rules*: A scoring rule is given by a scoring vector $\alpha = (\alpha_1, \ldots, \alpha_m)$ where $\alpha_i \geq \alpha_{i+1}$ for all $i \in [m]$ and $\alpha_1 > \alpha_m$. Under this rule, for each vote σ in π and $i \in [m]$, α_i points are awarded to the alternative in the i^{th} position. The alternatives are then sorted in descending order of their total points. Tie-breaking is used to sort alternatives with identical total points. Some example of positional scoring rules include plurality, Borda count, veto, and k-approval.

- *Single transferable vote (STV)*: STV proceeds in rounds, where in each round the alternative with lowest plurality score is eliminated, until only one alternative remains. The rule ranks the alternatives in the reverse order of their elimination. At each stage, tie-breaking is used to choose among the alternatives with identical plurality score in that stage.

- *Copeland's method*: The Copeland score of an alternative a in a profile π, denoted $PW^\pi(a)$, is the number of outgoing edges from a in the unweighted PM graph of π, i.e., the number of other alternatives that a defeats in a pairwise election. Copeland's method ranks the alternative in non-increasing order of their Copeland scores. Tie-breaking is used for sorting alternatives with identical Copeland scores.

- *The maximin rule*: Given a profile π, the maximin score of an alternative a is the minimum of the weights of the alternative's outgoing edges in the weighted PM graph of π. The maximin rule returns the alternatives in descending order of their maximin score. Tie-breaking is used to sort alternatives with identical maximin scores.

- *The Slater rule*: Given a profile π, the Slater rule selects the ranking which minimizes the number of pairs of alternatives on which it disagrees with the unweighted PM graph of π. Note that this is, in some sense, the unweighted version of the Kemeny rule, which, as defined above, minimizes disagreement with the weighted PM graph of π. Tie-breaking is used to choose among rankings having equal disagreement with the unweighted PM graph of π.

- *The Bucklin rule*: The Bucklin score of an alternative a is the minimum k such that a is ranked among the first k positions by a majority of the voters. The Bucklin rule sorts the alternatives in a non-decreasing order according to their Bucklin scores. Tie-breaking is used to sort alternatives with identical Bucklin scores.[2]

- *The ranked pairs method*: Under the ranked pairs method, given a profile π, all ordered pairs of alternatives (a, a') are sorted in a non-increasing order of the weight of the edge from a to a' in the weighted PM graph of π. Then, starting with the first pair in the list, the method "locks in" the outcome using the result of the pairwise comparison. It proceeds with subsequent pairs and locks in every pairwise result that does not contradict (by forming a cycle) the partial ordering established so far. Finally, the method outputs the total order obtained. Tie-breaking is used initially to sort ordered pairs of alternatives with identical weight in the weighted PM graph of π.

2. Sometimes, a deterministic scheme is used to break ties by the number of votes that rank an alternative among the first k positions, where k is the Bucklin score of the alternative. However, we cling to our assumption of an inclusive tie-breaking scheme for uniformity.

Noise models. A *noise model* G is a collection of probability distributions over rankings. For every $\sigma^* \in \mathcal{L}(A)$, $G(\sigma^*)$ denotes the distribution from which noisy estimates are generated when the ground truth is σ^*. The probability of sampling $\sigma \in \mathcal{L}(A)$ from this distribution is denoted by $\Pr_G[\sigma; \sigma^*]$.

In order to rule out noise models that are completely outlandish, we focus on d-monotonic noise models with respect to a distance metric d, using definitions from our recent work (Caragiannis et al., 2013). In more detail, a *distance metric* over $\mathcal{L}(A)$ is a function $d(\cdot, \cdot)$ that satisfies the following properties for all $\sigma, \sigma', \sigma'' \in \mathcal{L}(A)$:

- $d(\sigma, \sigma') \geq 0$, and $d(\sigma, \sigma') = 0$ if and only if $\sigma = \sigma'$.
- $d(\sigma, \sigma') = d(\sigma', \sigma)$.
- $d(\sigma, \sigma'') + d(\sigma'', \sigma') \geq d(\sigma, \sigma')$.

A noise model G is called *d-monotone* for a distance metric d if for all $\sigma, \sigma', \sigma^* \in \mathcal{L}(A)$, $\Pr_G[\sigma; \sigma^*] \geq \Pr_G[\sigma'; \sigma^*]$ if and only if $d(\sigma, \sigma^*) \leq d(\sigma', \sigma^*)$. That is, the closer a ranking is to the ground truth, the higher its probability.

Robust SWFs. We are interested in SWFs that can recover the ground truth from a large number of i.i.d. noisy estimates. Formally, an SWF f is called *accurate in the limit with respect to a noise model G* if, given an arbitrarily large number of samples from G with any ground truth σ^*, the rule outputs σ^* with arbitrarily high accuracy. That is, for every $\sigma^* \in \mathcal{L}(A)$, $\lim_{n\to\infty} \Pr[f(\pi^n) = \sigma^*] = 1$, where π^n denotes a profile consisting of n i.i.d. samples from $G(\sigma^*)$. A voting rule f is called *monotone-robust* with respect to a distance metric d if it is accurate in the limit for all d-monotonic noise models.

3. Modal Ranking is Unique Within GSRs

In this section, we characterize the modal ranking rule — which selects the most common ranking in a given profile — as the unique rule that is monotone-robust with respect to all distance metrics, among a wide sub-family of GSRs. For this, we use a geometric equivalent of GSRs introduced by Mossel et al. (2013) called "hyperplane rules". Like GSRs, hyperplane rules were also originally defined as deterministic SCFs. Below, we give the natural extension of the definition to (possibly randomized) SWFs.

Given a profile π, let x^π_σ denote the fraction of times the ranking $\sigma \in \mathcal{L}(A)$ appears in π. Hence, the point $x^\pi = (x^\pi_\sigma)_{\sigma\in\mathcal{L}(A)}$ lies in a probability simplex $\Delta^{m!}$. This allows us to use rankings from $\mathcal{L}(A)$ to index the $m!$ dimensions of every point in $\Delta^{m!}$. Formally,

$$\Delta^{m!} = \left\{ x \subseteq \mathbb{Q}^{m!} \;\middle|\; \sum_{\sigma\in\mathcal{L}(A)} x_\sigma = 1 \right\}.$$

Importantly, note that $\Delta^{m!}$ contains only points with rational coordinates. Weights $w_\sigma \in \mathbb{R}$ for all $\sigma \in \mathcal{L}(A)$ define a hyperplane H where $H(x) = \sum_{\sigma\in\mathcal{L}(A)} w_\sigma \cdot x_\sigma$ for all $x \in \Delta^{m!}$. This hyperplane divides the simplex into three regions; the set of points on each side of the hyperplane, and the set of points on the hyperplane.

Definition 1 (Hyperplane Rules)**.** A hyperplane rule is given by $r = (\mathcal{H}, g)$, where $\mathcal{H} = \{H_i\}_{i=1}^{l}$ is a finite set of hyperplanes, and $g : \{+, 0, -\}^l \to \mathcal{D}(\mathcal{L}(A))$ is a function that takes as input the signs of all the hyperplanes at a point and returns a distribution over rankings. Thus, $r(\pi) = g(\mathtt{sgn}(\mathcal{H}(x^\pi)))$, where

$$\mathtt{sgn}(\mathcal{H}(x^\pi)) = (\mathtt{sgn}(H_1(x^\pi)), \ldots, \mathtt{sgn}(H_l(x^\pi))),$$

and $\mathtt{sgn} : \mathbb{R} \to \{+, 0, -\}$ is the signum function given by

$$\mathtt{sgn}(x) = \begin{cases} + & x > 0 \\ 0 & x = 0 \\ - & x < 0 \end{cases}$$

Next, we state the equivalence between hyperplane rules and GSRs in the case of randomized SWFs. This equivalence was established by Mossel et al. (2013) for deterministic SCFs; it uses the output of a given GSR for each set of compatible vectors to construct the output of its corresponding hyperplane rule in each region, and vice-versa. Simply changing the output of the g functions of both the GSR and the hyperplane rule from a winning alternative (for deterministic SCFs) to a distribution over rankings (for randomized SWFs) and keeping the rest of the proof intact shows the equivalence for randomized SWFs.

Lemma 1 (Mossel et al., 2013)**.** *For randomized social welfare functions, the class of generalized scoring rules coincides with the class of hyperplane rules.*

We impose a technical restriction on GSRs that has a clear interpretation under the geometric hyperplane equivalence. Intuitively, it states that if the rule outputs the same ranking (without ties) almost everywhere around a point x^π in the simplex, then the rule must output the same ranking (without ties) on π as well. More formally, consider the regions in which the simplex is divided by a set of hyperplanes $\mathcal{H}$. We say that a region is *interior* if none of its points lie on any of the hyperplanes in $\mathcal{H}$, that is, if for every point x in the region, $\mathtt{sgn}(\mathcal{H}(x))$ does not contain any zeros. For $x \in \Delta^{m!}$, let

$$\mathcal{S}(x) = \left\{ y \in \Delta^{m!} \;\middle|\; \forall \sigma \in \mathcal{L}(A),\ x_\sigma = 0 \Rightarrow y_\sigma = 0 \right\}$$

denote the subspace of points that are zero in every coordinate where x is zero. We say that an interior region is *adjacent* to x if its intersection with $\mathcal{S}(x)$ contains points arbitrarily close to x.

Definition 2 (No Holes Property)**.** We say that a hyperplane rule (generalized scoring rule) has no holes if it outputs a ranking σ with probability 1 on a profile π whenever it outputs σ with probability 1 in all interior regions adjacent to x^π.

When this property is violated, we have a point x^π such that the output of the rule on x^π is different from the output of the rule almost everywhere around x^π, creating a hole at x^π. We later show (in Section 4; see Theorem 2) that the no holes property is a very mild restriction on GSRs. We are now ready to formally state our main result.

Theorem 1. *Let r be a (possibly randomized) generalized scoring rule with no holes. Then, r is monotone-robust with respect to all distance metrics if and only if r coincides with the modal ranking rule on every profile with no ties (i.e., r outputs the most frequent ranking with probability 1 on every profile where it is unique).*

Before proving the theorem, we wish to point out three subtleties. First, our assumption of accuracy in the limit imposes a condition on the rule as the number of votes *goes to infinity*. This has to be translated into a condition on all *finite profiles*; we do this by leveraging the structure of generalized scoring rules.

Second, if there are several rankings that appear the same number of times, a monotone-robust rule can actually output any ranking with impunity, because in the limit this event happens with probability zero.

Third, every noise model G that is monotone with respect to some distance metric satisfies $\Pr_G[\sigma^*; \sigma^*] > \Pr_G[\sigma; \sigma^*]$ for all pairs of different rankings $\sigma, \sigma^* \in \mathcal{L}(A)$. It seems intuitive that the converse holds, i.e., if a noise model satisfies $\Pr_G[\sigma^*; \sigma^*] > \Pr_G[\sigma; \sigma^*]$ for all $\sigma \neq \sigma^*$ then there exists a distance metric d such that G is monotone with respect to d — but this is false. Hence, our condition asks for accuracy in the limit for noise models that are monotone with respect to some metric, instead of just assuming accuracy in the limit with respect to *all* noise models where the ground truth is the unique mode.

Proof of Theorem 1. Let r be a (possibly) randomized generalized scoring rule with no holes. Using Lemma 1, we represent r as a hyperplane rule $(\mathcal{H}, g)$ where $\mathcal{H} = \{H_i\}_{i=1}^{l}$ is the set of hyperplanes.

First, we show the simpler forward direction. Let r output the most frequent ranking with probability 1 on every profile where it is unique. We want to show that r is monotone-robust with respect to all distance metrics. Take a distance metric d, a d-monotonic noise model G, and a true ranking σ^*. We need to show that r outputs σ^* with probability 1 given infinitely many samples from $G(\sigma^*)$.

Note that d satisfies $d(\sigma^*, \sigma^*) = 0 < d(\sigma, \sigma^*)$ for all $\sigma \neq \sigma^*$. Hence, G must satisfy $\Pr_G[\sigma^*; \sigma^*] > \Pr_G[\sigma; \sigma^*]$ for all $\sigma \neq \sigma^*$. Now, given infinite samples from $G(\sigma^*)$, σ^* becomes the unique most frequent ranking with probability 1. Thus, r outputs σ^* with probability 1 in the limit, as required.

For the reverse direction, let r be d-monotone-robust for all distance metrics d. Take a profile π^* with a unique most frequent ranking σ^*. Recall that $x_\sigma^{\pi^*}$ denotes the fraction of times σ appears in π^* and note that $x_{\sigma^*}^{\pi^*} > x_\sigma^{\pi^*}$ for all $\sigma \neq \sigma^*$. We also denote by $X_\sigma^{\pi^*}$ the number of times σ appears in π^*.

The rest of the proof is organized in three steps. First, we define a distance metric d, a d-monotonic noise model G, and a true ranking. Second, we show that given samples from $G(\sigma^*)$, in the limit r outputs σ^* with probability 1 in every interior region adjacent to x^{π^*}. Finally, we use the no holes property of r to argue that $\Pr[r(\pi^*) = \sigma^*] = 1$.

Step 1: We define d as

$$d(\sigma, \sigma') = \begin{cases} \max(1, |X_\sigma^{\pi^*} - X_{\sigma'}^{\pi^*}|) & \text{if } \sigma \neq \sigma', \\ 0 & \text{otherwise.} \end{cases}$$

We claim that d is a distance metric. Indeed, the first two axioms are easy to verify. The triangle inequality $d(\sigma, \sigma') \leq d(\sigma, \sigma'') + d(\sigma'', \sigma')$ holds trivially if any two of the three rankings are equal. When all three rankings are distinct,

$$\begin{aligned} d(\sigma, \sigma'') + d(\sigma'', \sigma') &= \max(1, |X_\sigma^{\pi^*} - X_{\sigma''}^{\pi^*}|) + \max(1, |X_{\sigma''}^{\pi^*} - X_{\sigma'}^{\pi^*}|) \\ &\geq \max(1 + 1, |X_\sigma^{\pi^*} - X_{\sigma''}^{\pi^*}| + |X_{\sigma''}^{\pi^*} - X_{\sigma'}^{\pi^*}|) \\ &\geq \max(1, |X_\sigma^{\pi^*} - X_{\sigma'}^{\pi^*}|) = d(\sigma, \sigma'). \end{aligned}$$

Now, define the noise model G where

$$\Pr_G[\sigma;\sigma'] = \frac{1/(1+d(\sigma,\sigma'))}{\sum_{\tau\in\mathcal{L}(A)} 1/(1+d(\tau,\sigma'))} \quad \text{for } \sigma' \neq \sigma^*.$$

and $\Pr_G[\sigma;\sigma^*] = x_\sigma^{\pi^*}$. Note that G is trivially d-monotone for true rankings other than σ^*. Denoting the number of votes in π^* by n^*, since σ^* is the unique most frequent ranking, we have that $d(\sigma,\sigma^*) = n^*(x_{\sigma^*}^{\pi^*} - x_\sigma^{\pi^*})$ for all $\sigma \neq \sigma^*$. Hence, $\Pr_G[\sigma_1;\sigma^*] \geq \Pr_G[\sigma_2;\sigma^*]$ if and only if $d(\sigma_1,\sigma^*) \leq d(\sigma_2,\sigma^*)$ and G is also d-monotone for the true ranking σ^*. We conclude that G is a d-monotonic noise model.

Step 2: Let π_n denote a profile consisting of n i.i.d. samples from $G(\sigma^*)$. Since r is monotone-robust for every distance metric, we have

$$\lim_{n\to\infty} \Pr[r(\pi_n) = \sigma^*] = 1. \tag{1}$$

If π^* has only one ranking, then only that ranking will ever be sampled. Hence, we will have $\Pr[x^{\pi_n} = x^{\pi^*}] = 1$, and Equation (1) would imply that the rule must output σ^* with probability 1 on π^*.

Assume that π^* has at least two distinct votes. We want to show that r outputs σ^* with probability 1 in every interior region adjacent to x^{π^*}. As $n \to \infty$, the distribution of x^{π_n} tends to a Gaussian with mean x^{π^*} and concentrated on the hyperplane

$$\sum_{\sigma\in\mathcal{L}(A)|x_\sigma^{\pi^*}>0} x_\sigma^{\pi_n} = 1.$$

This follows from the multivariate central limit theorem; see (Mossel et al., 2013) for a detailed explanation. Note that the sum ranges only over the rankings that appear in π^* because in the distribution $G(\sigma^*)$, the probability of sampling a ranking σ that does not appear in π^* is zero.

Since the Gaussian lies in the subspace $\mathcal{S}(x^{\pi^*})$, we set the coordinates corresponding to rankings that do not appear in π^* to zero in all the hyperplanes, and remove the hyperplanes that become trivial. Hereinafter we only consider the rest of the hyperplanes, and the regions they form around x^{π^*}, all in the subspace $\mathcal{S}(x^{\pi^*})$.

If none of the hyperplanes pass through x^{π^*}, then there is a unique interior region K which actually contains x^{π^*} as its interior point. In this case, the limiting probability of x^{π_n} falling in K will be 1, as the Gaussian becomes concentrated around x^{π^*}. Thus, Equation (1) implies that r outputs σ^* with probability 1 in K, and therefore on π^*.

If there exists a hyperplane passing through x^{π^*}, then each interior region K adjacent to x^{π^*} is the intersection of finitely many halfspaces whose hyperplanes pass through x^{π^*}. Let $\overline{K}$ and $\overline{\mathcal{S}(x^{\pi^*})}$ denote the closures of K and $\mathcal{S}(x^{\pi^*})$ in $\mathbb{R}^{m!}$, respectively.[3] Thus, $\overline{K}$ is a pointed convex cone with its apex at x^{π^*}, and must subtend a positive solid angle (in $\overline{\mathcal{S}(x^{\pi^*})}$) at its apex since the hyperplanes are distinct. By definition, the solid angle that $\overline{K}$ forms at x^{π^*} is the fraction of volume (the Lebesgue measure in $\overline{\mathcal{S}(x^{\pi^*})}$) covered by $\overline{K}$ in a ball of radius ρ (again in $\overline{\mathcal{S}(x^{\pi^*})}$) centered at x^{π^*}, as $\rho \to 0$ (see, e.g., Section 2 in Desario & Robins, 2011).

3. We remark that considering the closures is necessary since $\Delta^{m!}$ contains only points with rational coordinates; hence it (as well as any subset of it) has measure zero.

Since the Gaussian is symmetric in $\overline{\mathcal{S}(x^{\pi^*})}$ around x^{π^*}, and the limiting distribution of x^{π_n} converges to the Gaussian, the limiting probability of x^{π_n} lying in K is positive. This holds for every interior region K adjacent to x^{π^*}. Thus, Equation (1) again implies that r outputs σ^* with probability 1 in every interior region adjacent to x^{π^*}.

Step 3: Finally, since r has no holes and it outputs σ^* with probability 1 in every interior region adjacent to x^{π^*}, we conclude that r must also output σ^* with probability 1 on π^*. ■ (Proof of Theorem 1)

4. How Restrictive is the No Holes Property?

To complete the picture, we wish to show that the no holes condition that Theorem 1 imposes on GSRs is indeed unrestrictive, by establishing that many prominent voting rules (in the sense of receiving attention in the computational social choice literature) are GSRs with no holes. One issue that must be formally addressed is that the definitions of prominent voting rules (see Section 2) typically do not address how ties are broken. For example, the plurality rule ranks the alternatives by their number of voters who rank them first; but what should we do in case of a tie? A common practice is to adopt uniformly random tie-breaking; this is almost always used in political elections (e.g., by throwing dice or drawing cards in small municipal elections where ties are not unlikely to occur). From a theoretical point of view, randomized tie-breaking is necessary in order to achieve neutrality with respect to the alternatives (Moulin, 1983).

We show that prominent voting rules are GSRs with no holes under a wide family of randomized tie-breaking schemes, which we call *inclusive tie-breaking*.

Definition 3 (Inclusive tie-breaking)**.** *A tie-breaking scheme is called* inclusive *if it assigns a positive probability to each of the tied decisions at every stage.*

Such tied decisions could vary for different rules. For rules that assign scores to alternatives and order them according to their scores, the decisions correspond to choosing the order of alternatives with identical scores. For rules that assign scores to rankings and choose the ranking with the highest score, the decisions correspond to breaking ties among rankings with identical highest scores. While most voting rules use tie-breaking only once (either initially or in the end), multi-stage protocols such as single transferable vote (STV) use tie-breaking in each stage.

We note that uniformly random tie-breaking, which assigns equal probability to every decision, is a special case of inclusive tie-breaking. Admittedly, inclusive tie-breaking does not capture deterministic tie-breaking schemes (such as lexicographic tie-breaking). However, we strongly believe that prominent voting rules other than the modal ranking rule are not monotone-robust with respect to all distance metrics even if a deterministic tie-breaking scheme were used.

Before we demonstrate that prominent voting rules are GSRs with no holes, we show that the no holes condition is implied by a property well-known in the social choice literature as *consistency*. This yields a way of leveraging known results from the literature to easily establish that all positional scoring rules, the Kemeny rule, and single transferable vote (STV) are GSRs with no holes. Intuitively, consistency means that if the output of a rule is identical on two profiles, then taking their union should not change the output. For deterministic SWFs, consistency was first studied by Young and Levenglick (1978), who observed that it is incomparable to, but usually much weaker than, consistency of winning alternative in the case of SCFs. Later, Conitzer and Sandholm (2005) showed that consistency (whether in rankings or in winning alternatives) is a necessary condition for a voting rule to be a maximum likelihood estimator under i.i.d. votes. We formalize a related, but weaker notion of consistency for randomized SWFs.

Definition 4 (Weak consistency for rankings)**.** *A randomized SWF r is said to satisfy weak consistency for rankings if* $\Pr[r(\pi_1) = \sigma] = 1$ *and* $\Pr[r(\pi_2) = \sigma] = 1$ *implies* $\Pr[r(\pi_1 \cup \pi_2) = \sigma] = 1$ *for all profiles π_1 and π_2, and all rankings $\sigma \in \mathcal{L}(A)$. Here, $\pi_1 \cup \pi_2$ denotes the profile representing the union of π_1 and π_2.*

For hyperplane rules (generalized scoring rules), weak consistency for rankings is equivalent to convexity of the region where the rule outputs σ with probability 1, for every $\sigma \in \mathcal{L}(A)$. Now, we are ready to prove the following implication.

Lemma 2. *Any generalized scoring rule satisfying weak consistency for rankings has no holes.*

Proof. Take a GSR r that satisfies weak consistency for rankings. Suppose for contradiction that r has a hole at x^π for some profile π. Let r output σ with probability 1 in all interior regions adjacent to x^π, but not on π. Let k be the number of distinct rankings that appear in π. Hence, $\mathcal{S}(x^\pi)$ is a k-dimensional subspace of $\Delta^{m!}$.

If $k = 1$, then π has only one distinct ranking σ. Thus, $x^\pi_\sigma = 1$ and $x^\pi_{\sigma'} = 0$ for all $\sigma' \neq \sigma$. By the definition of $\mathcal{S}(x^\pi)$, for every $y \in \mathcal{S}(x^\pi)$ we have $y_{\sigma'} = 0$ for all $\sigma' \neq \sigma$. Thus, $y_\sigma = 1$, implying that $\mathcal{S}(x^\pi) = \{x^\pi\}$. Hence, trivially, x^π cannot be a hole.

Let $k > 1$. Define

$$V = \left\{ v \in \{-1, 0, 1\}^{m!} \;\middle|\; \forall \sigma \in \mathcal{L}(A),\ v_\sigma = 0 \Leftrightarrow x^\pi_\sigma = 0 \right.$$
$$\left. \wedge\ \exists \sigma \in \mathcal{L}(A),\ v_\sigma = 1\ \wedge\ \exists \sigma \in \mathcal{L}(A),\ v_\sigma = -1 \right\}.$$

Now, for every $v \in V$, define the orthant

$$O^v = \left\{ y \in \mathcal{S}(x^\pi) \;\middle|\; \forall \sigma \in \mathcal{L}(A),\ (v_\sigma = 1 \Rightarrow y_\sigma > x^\pi_\sigma) \wedge (v_\sigma = -1 \Rightarrow y_\sigma < x^\pi_\sigma) \right\}.$$

Note that we do not consider the orthants where all the k coordinates are higher (respectively, lower) than those of x^π because such orthants do not have any points in $\mathcal{S}(x^\pi)$ as the sum of those k coordinates must be equal to 1. The rest $3^k - 2$ orthants have non-empty intersection with $\mathcal{S}(x^\pi)$. Further, since the interior regions adjacent to x^π surround it in the space $\mathcal{S}(x^\pi)$ and so do the $3^k - 2$ orthants, each orthant O^v must have a point x^v in some interior region adjacent to x^π, where the output is σ. Now, a convexity lemma (Lemma 6 in the appendix) shows that we can get x^π as a convex combination of points in $\{x^v | v \in V\}$,[4] on all of which r outputs σ with probability 1. Hence, due to weak consistency for rankings, r must also output σ with probability 1 on x^π, a contradiction. Hence, r has no holes. ■ (Proof of Lemma 2)

Finally, we are ready to prove that prominent voting rules are GSRs with no holes under all inclusive tie-breaking schemes (which contain uniformly random tie-breaking).

Theorem 2. *Under any inclusive tie-breaking scheme, all positional scoring rules, the Kemeny rule, STV, Copeland's method, Bucklin's rule, the maximin rule, Slater's rule, and the ranked pairs method are generalized scoring rules with no holes.*

4. In Lemma 6, take $FIX = \{\sigma \in \mathcal{L}(A) | x^\pi_\sigma = 0\}$.

Proof. It can easily be checked that the f functions of the GSR constructions given by Xia and Conitzer (2008) and the hyperplanes for the hyperplane rule constructions given by Mossel et al. (2013) encode enough information (including all the ties) in their input to the g functions such that it is possible to change the output of g from a winning alternative (for deterministic SCFs) to any desired distribution over rankings (for randomized SWFs) for the rules mentioned in the statement of the theorem. In particular, these rules can be implemented with any inclusive tie-breaking scheme as GSRs.

It is well-known and easy to check that all positional scoring rules are consistent for rankings (see Conitzer & Sandholm, 2005; Conitzer et al., 2009). Young and Levenglick (1978) showed that the Kemeny rule is also consistent for rankings. Conitzer and Sandholm (2005) showed that STV is consistent for rankings; later it was shown that STV is consistent for rankings only in the absence of tie-breaking (Conitzer et al., 2009). It can be checked that under inclusive tie-breaking, STV returns a single ranking with probability 1 if and only if there are no ties. Hence, consistency in the absence of tie-breaking implies weak consistency for rankings. Thus, we have the following.

Lemma 3. *Under any inclusive tie-breaking scheme, all positional scoring rules, the Kemeny rule, and single transferable vote (STV) satisfy weak consistency for rankings.*

Hence, the no holes property of all positional scoring rules, the Kemeny rule, and single transferable vote (STV) follows by Lemmas 2 and 3. Conitzer and Sandholm (2005) showed that other rules such as Bucklin's rule, Copeland's method, the maximin rule, and the ranked pairs method are not consistent for rankings even in the absence of ties. Hence, these rules do not satisfy weak consistency for rankings. Still, we will show that they satisfy the no holes property as well, albeit using a different (and significantly more involved) approach.

Take a hyperplane rule r. We want to show that r does not have a hole at a profile π. Let $k = |\{\sigma \in \mathcal{L}(A) | x^{\pi}_{\sigma} > 0\}|$ be the number of distinct rankings in π. If $k = 1$, then as shown in the proof of Lemma 2, $\mathcal{S}(x^{\pi}) = \{x^{\pi}\}$, and there cannot be a hole at π. Assume $k \geq 2$. For a set of profiles P, let $x^P = \{x^{\pi'} | \pi' \in P\}$.

Let $dim(\cdot)$ denote the Hausdorff dimension of a given subset of $\mathbb{R}^{m!}$. For any set $C \subseteq \Delta^{m!}$, let $\overline{C}$ denote its closure in $\mathbb{R}^{m!}$. Hence, we have that $dim(\overline{\mathcal{S}(x^{\pi})}) = k-1$, because $\mathcal{S}(x^{\pi})$ has $k-1$ free variables.

Lemma 4. *Let T denote the set of points in $\mathcal{S}(x^{\pi})$ that lie on at least one of the hyperplanes of r. Then, $dim(\overline{T}) \leq k-2$.*

Proof. Take a hyperplane $\sum_{\sigma \in \mathcal{L}(A)} w_{\sigma} x_{\sigma} = 0$ of r. Consider its intersection with $\overline{\mathcal{S}(x^{\pi})}$. First, we notice that all but k of the x_{σ}'s must be set to zero. Among the remaining k, if we substitute values for $k-2$ of the variables, we get two equations in two variables, which can be seen to be independent since the one obtained from the hyperplane has the RHS zero, while the one obtained from $\overline{\mathcal{S}(x^{\pi})}$ has the RHS one. Hence, there is at most one solution of the pair of equations.

That is, every combination of values of $k-2$ free variables lead to at most one solution for the remaining variables. Thus, the dimension of the intersection of the hyperplane with $\overline{\mathcal{S}(x^{\pi})}$ is at most $k-2$. Taking union over finitely many hyperplanes does not increase the Hausdorff dimension. Hence, we have $dim(\overline{T}) \leq k-2$. ■ (Proof of Lemma 4)

Next, we describe an outline that we follow in order to prove that the no holes property is satisfied by many prominent voting rules. We consider rules that assign a score to every alternative, and then order them in a non-increasing or non-decreasing order of their scores, breaking ties among alternatives with identical scores. This applies to Copeland's method, the maximin rule, and Bucklin's rule.[5] Such rules return a single ranking with probability 1 if and only if the scores of the alternatives are strictly ordered according to that ranking. Let us denote the score of alternative c in profile π by $SC^\pi(c)$.

1. For the sake of contradiction, we assume that the rule under consideration, say r, has a hole at a profile π. Hence, r outputs a ranking τ with probability 1 in every interior region adjacent to x^π, but there exists a ranking $\tau' \neq \tau$ such that $\Pr[r(\pi) = \tau'] > 0$.

2. Since $\tau' \neq \tau$, there must exist alternatives a and b such that $a \succ_\tau b$, but $b \succ_{\tau'} a$. Due to the inclusive tie-breaking scheme, we must have that
   - $SC^\pi(b) \geq SC^\pi(a)$, and
   - $SC(a) > SC(b)$ in every interior region adjacent to x^π.

3. Finally, we find a neighborhood of π where we also have $SC(b) \geq SC(a)$. Formally, we find a set of profiles P such that
   - for every profile $\pi' \in P$, $SC^{\pi'}(b) \geq SC^{\pi'}(a)$, and $x^{\pi'}$ either lies in an interior region adjacent to x^π or on one of the hyperplanes of r, and
   - $dim(\overline{x^P}) = k - 1$.

Given this, we argue that a contradiction can be reached. Recall that T is the intersection of the hyperplanes of r with $\mathcal{S}(x^\pi)$. Suppose that $x^P \subseteq T$. Then, $\overline{x^P} \subseteq \overline{T}$, which is impossible because $dim(\overline{x^P}) > dim(\overline{T})$ (Lemma 4). Hence, there must exist a profile π' such that $x^{\pi'} \in x^P \setminus T$ lies in an interior region adjacent to x^π. However $SC^{\pi'}(b) \geq SC^{\pi'}(a)$, which is the desired contradiction.

Note that the first two steps are common to all voting rules. All we need to do is to find a set of profiles P satisfying the stated conditions. For many of the voting rules, P is obtained by increasing $x^\pi_{\sigma^*}$ for some $\sigma^* \in \pi$, and decreasing x^π_σ for all $\sigma \neq \sigma^*$ that appear in π. Formally,

$$P = \left\{ \pi' \;\middle|\; \forall \sigma \in \mathcal{L}(A), \right.$$

$$\left. x^{\pi'}_\sigma = \begin{cases} 0, & \text{if } x^\pi_\sigma = 0, \\ x^\pi_{\sigma^*} + \delta, & \text{if } \sigma = \sigma^*, \\ x^\pi_\sigma - \delta_\sigma, & \text{otherwise,} \end{cases} \quad \text{where } 0 < \delta \leq \delta_{\max} \wedge \sum_{\sigma \neq \sigma^*, x^\pi_\sigma > 0} \delta_\sigma = \delta \right\},$$

By the construction, for every profile $\pi' \in P$, the weights of the edges of the weighted PM graph of π increase in the direction of σ^*, and decrease in the direction opposite to σ^* (except the edges with weights 1 and 0 do not change). If $\delta_{\max}$ is chosen to be small enough, this change does not alter the direction of any existing edge in the unweighted PM graph, but breaks all existing ties between pairs of alternatives in one direction or the other. Clearly, $dim(\overline{x^P}) = k - 1$ since decreasing the fractions of all rankings $\sigma \neq \sigma^*$ with $x^\pi_\sigma > 0$ so that the decrements sum up to δ

5. We will see that Slater's rule, which assigns a score to every ranking, can also be handled in this outline.

gives $k-2$ degrees of freedom choosing δ gives another degree of freedom. This observation is very crucial to the proofs for many of the voting rules.

Below, we describe appropriate choices of σ^* and $\delta_{\max}$ for various prominent voting rules, namely for Copeland's method, Bucklin's rule, the maximin rule, and Slater's rule. We also provide a proof for the ranked pairs method in which we use completely different arguments.

(a) Copeland's method.

Recall that for Copeland's method, $SC^\pi(c)$ is the number of outgoing edges from c in the unweighted PM graph of π. If there are no ties in the unweighted PM graph of π, then choosing any $\sigma^* \in \pi$ and a small enough $\delta_{\max}$ ensures that the set P obtained fits the requirements of step 3 of the outline and preserves all the edges in the unweighted PM graph. Hence, $SC(b) \geq SC(a)$ is preserved, as required.

In case of ties, let $TIE^\pi(c)$ be the set of alternatives with which c is tied in the unweighted PM graph of π.[6] For $\sigma \in \mathcal{L}(A)$, let

$$s(\sigma) = \sum_{c \in TIE^\pi(b)} \mathbb{I}[b \succ_\sigma c] - \sum_{c \in TIE^\pi(b)} \mathbb{I}[c \succ_\sigma b] - \sum_{c \in TIE^\pi(a)} \mathbb{I}[a \succ_\sigma c] + \sum_{c \in TIE^\pi(a)} \mathbb{I}[c \succ_\sigma a].$$

Let $n^\pi_{x \succ y}$ denote the number of rankings that prefer alternative x to alternative y in π. Summing over all rankings in π and changing the order of the summation in each term, we get

$$\sum_{i=1}^{n} s(\sigma_i) = \sum_{c \in TIE^\pi(b)} \left(n^\pi_{b \succ c} - n^\pi_{c \succ b} \right) - \sum_{c \in TIE^\pi(a)} \left(n^\pi_{a \succ c} - n^\pi_{c \succ a} \right) = 0,$$

where the last equality holds due to the definitions of $TIE^\pi(b)$ and $TIE^\pi(a)$. Also, note that the sum evaluates to zero even if either $TIE^\pi(b)$ or $TIE^\pi(a)$ or both are empty sets.

Hence, there exists a ranking $\sigma^* \in \pi$ such that $s(\sigma^*) \geq 0$. There exists a $\delta_{\max} > 0$ such that increasing $x^\pi_{\sigma^*}$ by at most $\delta_{\max}$ and decreasing the fractions of other rankings that appear in π would not change the non-tied edges of the PM graph, and among the ties, b would defeat at least as many previously tied alternatives as a does. Hence, such a change preserves $SC(b) \geq SC(a)$. Further, $\delta_{\max}$ is chosen to be small enough so that for the new profile π', $x^{\pi'}$ does not fall in an interior region that is not adjacent to x^π, i.e., it either lies in an interior region adjacent to x^π or on one of the hyperplanes of Copeland's method. Thus, the set of profiles P obtained in this way fits the requirements of step 3 of the outline.

(b) Bucklin's rule.

Let $SC^\pi(a) = k$. We know that $SC^\pi(b) \leq SC^\pi(a) = k$.[7] Let $T^\pi(j, c)$ denote the fraction of rankings where c is ranked in the first j positions. Then, by the definition of the Bucklin score,

$$T^\pi(k, b) > 1/2 \quad \text{and} \quad T^\pi(k-1, a) \leq 1/2. \tag{2}$$

If we find σ^* such that the set P defined in the outline preserves the two inequalities in Equation (2), then we will have $SC(a) \geq k$ and $SC(b) \leq k$, i.e., $SC(b) \leq SC(a)$ will be preserved.

Let $T^\pi(k, b) = 1/2 + \gamma$. Then, it is easy to check that if the fractions of all the rankings in π are altered by less than $\gamma/m!$, then we would still have $T(k, b) > 1/2$. Now, we simply observe that

6. We add zero to the Copeland score of an alternative for its tied edges; this is also known as $Copeland^0$.
7. Recall that the Bucklin score is to be *minimized*.

since $T^\pi(k-1, a) \leq 1/2$, more than half of the rankings in π, in particular, at least one ranking ranks a not in the first $k-1$ positions. Choosing this as σ^* and taking $\delta_{\max} < \gamma/m!$ (and also small enough so that the new profile does not lie in an interior region not adjacent to x^π) would preserve both inequalities in Equation (2).

(c) The maximin rule.

Here, $SC^\pi(c)$ is the minimum of the weights of the outgoing edges from c in the weighted PM graph of π. Let $MINW^\pi(c)$ denote the set of alternatives to which c has an outgoing edge with the minimum weight in the weighted PM graph of π. Now, take an alternative $c \in MINW^\pi(a)$. Let w be the weight of the edge from a to c. First, we note that $w \neq 1$, because $w = 1$ would imply that a has an outgoing edge with weight 1 to every other alternative, i.e., a is ranked first in all votes in π. This would contradict $SC^\pi(b) \geq SC^\pi(a)$. Next, if $w = 0$, then c beats a in every vote in π. Now, all profiles in $\mathcal{S}(x^\pi)$ have the same set of rankings as π, and hence have zero maximin score of a. Thus, $SC(b) \geq SC(a)$ is trivially satisfied in any point of $\mathcal{S}(x^\pi)$ and, subsequently, we can define P so that x^P is the union of the interior regions adjacent to x^π.

Let us assume $w \in (0, 1)$. Let $R_{c \succ a}(\pi)$ be the set of rankings in π where $c \succ a$, and define $R_{a \succ c}$ to be the set of rankings in π where $a \succ c$. Since $w \in (0, 1)$, $R_{a \succ c} \neq \emptyset$ and $R_{c \succ a} \neq \emptyset$. To obtain P, we do not choose one $\sigma^* \in \pi$, increase its fraction and decrease the fractions of the rest of the rankings in π. Rather, we increase the fractions of all rankings in $R_{c \succ a}$ by a total of δ, and decrease the fractions of all rankings in $R_{a \succ c}$ by a total of δ, where $0 < \delta \leq \delta_{\max}$. Once again, we choose $\delta_{\max} > 0$ small enough so that x^P does not intersect with interior regions not adjacent to x^π. Increasing the fractions of all rankings $R_{c \succ a}$ so that the increments add up to δ gives $|R_{c \succ a}| - 1$ degrees of freedom. Similarly, decreasing the fractions of all rankings in $R_{a \succ c}$ so that the decrements add up to δ gives another $|R_{a \succ c}| - 1$ degrees of freedom. Finally, choosing δ itself gives one degree of freedom. Hence, the set of profiles P obtained satisfy $dim(x^P) = k - 1$.

Further, note that by construction, the weight of the edge from a to c drops by δ. Hence, the maximin score of a also drops by at least (in fact, by exactly) δ. To show that the rest of the proof follows from the outline, we need to show that the maximin score of b drops by at most δ. For each $d \in A \setminus \{b\}$, the weight of the edge from b to d is the sum of fractions of a subset R_d of rankings in π. Now, the collective weight of rankings in $R_d \cap R_{a \succ c}$ drops by at most δ, and the collective weight of rankings in $R_d \cap R_{c \succ a}$ can only increase. Hence, the weight of each outgoing edge from b drops by at most δ, which means that the maximin score of b also drops by at most δ, as required.

(d) Slater's rule.

Recall that Slater's rule associates a score to every ranking, and then chooses the ranking with the lowest Slater score,[8] breaking ties to choose among all rankings with the lowest Slater score. Even though Slater's rule does not associate scores to alternatives, we show that it fits our framework with a little modification. First, if there are no ties in the unweighted PM graph of a profile π, then similarly to Bucklin's rule, its unweighted PM graph and therefore the Slater scores of all rankings can be preserved in a small enough neighborhood of π, eliminating the possibility of π being a hole. In the general case, we slightly abuse the notation, and use $SC^\pi(\sigma)$ to denote the Slater score of ranking σ in profile π.

As in the step 1 of the outline, assume that π is a hole for Slater's rule; the rule returns τ with probability 1 in all interior regions adjacent to x^π, but returns a different ranking τ' with a positive probability on π. Then, due to all-inclusivity of the tie-breaking scheme, we must have

8. Recall that Slater's score is the disagreement of a ranking from a profile, which must be minimized.

$SC^\pi(\tau') \le SC^\pi(\tau)$.[9] We again need to find a σ^* and its associated set of profiles P. P must satisfy all the conditions in the third step of the outline, except we replace the inequality in the scores of alternatives by the inequality in the scores of rankings, namely $SC(\tau') \le SC(\tau)$.

Since $SC^\pi(\sigma)$ counts the number of pairwise disagreements of σ with the unweighted PM graph of π, and since small deviations in the fractions x^π_σ would not change the edges that are not tied, we concentrate on the edges of the PM graph of π that are tied. Formally, let $TIE(\pi)$ denote the set of ordered pairs of alternatives that are tied in the PM graph of π. For $\sigma \in \mathcal{L}(A)$, define

$$s(\sigma) = \sum_{\substack{(c,d)\in TIE(\pi) \\ \text{s.t. } c \succ_{\tau'} d}} \mathbb{I}[c \succ_\sigma d] - \sum_{\substack{(c,d)\in TIE(\pi) \\ \text{s.t. } c \succ_{\tau} d}} \mathbb{I}[c \succ_\sigma d].$$

It is clear that taking $\sigma^* \in \pi$ such that $s(\sigma) \ge 0$ would ensure that in every profile in P, at least as much will be added to the Slater score of τ as to the Slater score of τ' compared to π, ensuring $SC(\tau') \le SC(\tau)$. To see why such a ranking exists, we sum $s(\sigma_i)$ over all votes σ_i in π and, by interchanging the order of summations, we get

$$\begin{aligned} \sum_{i=1}^{n} s(\sigma_i) &= \sum_{\substack{(c,d)\in TIE(\pi) \\ \text{s.t. } c \succ_{\tau'} d}} n^\pi_{c\succ d} - \sum_{\substack{(c,d)\in TIE(\pi) \\ \text{s.t. } c \succ_{\tau} d}} n^\pi_{c\succ d} \\ &= \frac{n}{2} \cdot \Big(\big|\{(c,d) \in TIE(\pi) \text{ s.t. } c \succ_{\tau'} d\}\big| - \big|\{(c,d) \in TIE(\pi) \text{ s.t. } c \succ_{\tau} d\}\big| \Big) \qquad (3) \\ &= 0, \end{aligned}$$

where the last step follows since both terms inside the brackets in Equation (3) are the number of unordered pairs of alternatives that are tied in the PM graph of π. Hence, there exists a ranking $\sigma^* \in \pi$ with $s(\sigma^*) \ge 0$, as required. Finally, $\delta_{\max}$ is chosen so that the non-tied pairs in the PM graph stay non-tied, and the new profile does not fall in an interior region that is not adjacent to x^π.

(e) The ranked pairs method.

This proof does not follow the general outline given above. For an ordered pair of alternatives (c,d), let $w^\pi(c,d)$ denote the weight of the edge from c to d in the weighted PM graph of π. Suppose r outputs a ranking τ with probability 1 in every interior region adjacent to x^π, but does not output τ with probability 1 on π.

Let L denote the list in the ranked pairs process in π where ordered pairs of alternatives are sorted by their weight. Let Δ denote the minimum positive difference between the weights of any two pairs in L. Let (a,b) be the first pair in the list that is chosen with a positive probability and is inconsistent with τ (such a pair exists because r does not output τ with probability 1 on π).

Lemma 5. *Let PRE denote the set of pairs in L that have weight strictly greater than the weight of (a,b). Then, each pair in PRE must be chosen with probability 1 or 0 in the ranked pairs process on π, and the subset that is chosen with probability 1 must be consistent with τ.*

Proof. Let L^p be the largest prefix of L such that such that every pair in L^p is chosen with probability 1 or 0 in the ranked pairs process under an inclusive tie-breaking scheme.[10] Let $C^p \subseteq L^p$ be the set of pairs in L^p that are chosen with probability 1.

9. As with Bucklin's rule, the sign of the inequality is reversed because the Slater ranking *minimizes* the Slater score.
10. Note that if L^p has a group of pairs with equal weight, they will all be chosen with probability 1 or all be chosen with probability 0 irrespective of the tie-breaking.

First, we argue that all pairs in C^p are consistent with τ. Let P denote the space of profiles obtained by changing the fractions of all the rankings by at most $\Delta/(2m!)$. Note that this may only break ties in L, but cannot invert the order of two pairs that were strictly ordered by their weight in L. Similarly to the general outline, P has Hausdorff dimension $k-1$, and hence contains a point in an interior region adjacent to x^π. Further, since ties do not matter for pairs in P, all pairs in P chosen with probability 1 in π would also be chosen with probability 1 in all profiles in P. Hence, all pairs in C^p must be consistent with τ.

Since r does not output τ with probability 1 on π, $L^p \neq L$. Consider the group G of pairs with equal weight that follows L^p. First, G cannot be consistent with C^p, otherwise it would have been part of L^p. Therefore, there must exist a pair $p \in G$ that is chosen with a probability strictly in $(0,1)$ (i.e., not equal to 0 or 1). Thus, there must exist a feasible subset of G such that when it is chosen in the ranked pairs process along with C^p to produce a partial order l, l is inconsistent with p. If l is consistent with τ, then p must be inconsistent with τ. If l is inconsistent with τ, then since C^p is consistent with τ, there must exist a pair in G that is inconsistent with τ.

In either case, all pairs in C^p are consistent with τ, and the group G of pairs with equal weight that follows C^p has a pair that is inconsistent with τ. Thus, $(a,b) \in G$, and $PRE = L^p$. ■ (Proof of Lemma 5)

Next, we argue that $0 < w^\pi(a,b) < 1$. If $w^\pi(a,b) = 0$, then $w^\pi(b,a) = 1$. An ordered pair with weight 1 is consistent with all rankings in the profile. Hence, the set of ordered pairs in π with weight 1 do not contain a cycle. Thus, they are all selected with probability 1 in the ranked pairs process, which is a contradiction as we assumed (a,b) is chosen with a positive probability on π.

On the other hand, if $w^\pi(a,b) = 1$, then all rankings in π must prefer a to b. However, all profiles in $\mathcal{S}(x^\pi)$ have the same set of rankings as π. Hence, the weight of (a,b) is 1 everywhere in $\mathcal{S}(x^\pi)$. Due to the argument presented in the previous paragraph, this implies that in an interior region K adjacent to x^π, a is preferred to b with probability 1. This is a contradiction because r outputs τ with probability 1 in K that prefers b to a.

Hence, indeed $0 < w^\pi(a,b) < 1$. Let $R_{a \succ b}$ be the set of rankings in π that prefer a to b, and let $R_{b \succ a}$ be the set of rankings in π that prefer b to a. Since $0 < w^\pi(a,b) < 1$, we have $R_{a \succ b} \neq \emptyset$ and $R_{b \succ a} \neq \emptyset$.

Recall that Δ is the minimum positive difference between the weights of any two pairs in L. Choose $\delta_{\max} = \Delta/2$. Let P denote the set of profiles obtained by increasing the fractions of rankings in $R_{a \succ b}$ by a total of δ and decreasing the fractions of the rankings in $R_{b \succ a}$ by a total of δ, for $0 < \delta < \delta_{\max}$. This increases the weight of (a,b) by exactly δ and changes the weight of every other pair by at most δ. Due to the choice of $\delta_{\max}$, it is clear that the set of pairs with weight greater than that of (a,b) must be PRE for every profile in P.

Further, the changes in the fractions can only break ties among pairs in PRE, but cannot invert the order of two pairs with different weight in π. Since ties do not matter for pairs in PRE,[11] we see that the same subset of pairs in PRE are chosen in every profile in P. This would imply that under an inclusive tie-breaking scheme, (a,b) has a positive probability of being selected in each profile in P. However, P has Hausdorff dimension $k-1$, and therefore must contain a point in an interior region K adjacent to x^π. This contradicts the fact that r outputs τ that prefers b to a with probability 1 in K. Hence, π cannot be a hole. ■ (Proof of Theorem 2)

11. The pairs in P that were chosen with probability 1 and 0 in π would still be chosen with probability 1 and 0, respectively.

The comprehensive list of GSRs with no holes includes all prominent rules that are known to be GSRs (Xia & Conitzer, 2008; Mossel et al., 2013) — suggesting that the no holes property does not impose a significant restriction beyond the assumption that the rule is a GSR. One prominent rule is conspicuously missing — the fascinating but peculiar Dodgson rule (Dodgson, 1876), which is indeed not a GSR (Xia & Conitzer, 2008).

5. Impossibility for PM-c and PD-c Rules

Theorem 1 establishes the uniqueness of the modal ranking rule within a large family of voting rules (GSRs with no holes). Next we further expand this result by showing that no PM-c or PD-c rule is monotone-robust with respect to all distance metrics. Thus, the modal ranking rule is the unique rule that is monotone-robust with respect to all distance metrics in the union of GSRs with no holes, PM-c rules, and PD-c rules. Crucially, as shown in Figure 1, the families of PM-c and PD-c rules are disjoint, and neither one is a strict subset of GSRs.

Theorem 3. *For $m \geq 3$ alternatives, no PM-c rule or PD-c rule is monotone-robust with respect to all distance metrics.*

Proof. In both parts of this proof (for PM-c rules and PD-c rules), we use an intuitive, but somewhat technical lemma, which is given as Lemma 7 in the appendix. Let $A = \{a_1, \ldots, a_m\}$ be the set of alternatives. We use $a_{4\ldots m}$ as shorthand for $a_4 \succ \ldots \succ a_m$. Fix $\tau = a_1 \succ \ldots \succ a_m$, and $\sigma^* = a_2 \succ a_1 \succ a_3 \succ a_{4\ldots m}$.

First, we prove that no PM-c rule is monotone-robust with respect to all distance metrics. In particular, using Lemma 7, we will construct a distance metric d and a d-monotonic noise model G such that no PM-c rule is accurate in the limit for G.

Consider the distribution D over $\mathcal{L}(A)$ defined as follows:

$$\begin{aligned}
&\Pr_D[a_2 \succ a_1 \succ a_3 \succ a_{4\ldots m}] = \tfrac{4}{9},\\
&\Pr_D[a_1 \succ a_2 \succ a_3 \succ a_{4\ldots m}] = \tfrac{3}{9},\\
&\Pr_D[a_1 \succ a_3 \succ a_2 \succ a_{4\ldots m}] = \tfrac{2}{9},\\
&\Pr_D[\sigma] = 0, \text{ for all } \sigma \text{ not covered above.}
\end{aligned}$$

By Lemma 7, we know that there exist a distance metric d and a d-monotonic noise model G such that $\Pr_G[\sigma; \sigma^*] = \Pr_D[\sigma]$ for every $\sigma \in \mathcal{L}(A)$.

Given infinite samples from $G(\sigma^*)$, a $5/9$ fraction — a majority — of the votes have a_1 in the top position. A $7/9$ fraction of the votes prefer a_2 to a_3, while all votes prefer a_2 and a_3 to any other alternative besides a_1. Clearly, a_i is preferred to a_{i+1} for $i \geq 4$. Hence, in the PM graph, the alternatives are ordered according to $\tau = a_1 \succ a_2 \succ a_3 \succ a_{4\ldots m}$. Thus, every PM-c rule outputs τ in the limit, which is not the ground truth. Thus, no PM-c rule is accurate in the limit for G.

The construction for PD-c rules is more complex. Here, we will show that there is a noise model such that, given infinite samples for a specific ground truth, the PD graph of the profile induces a ranking that is different from the ground truth. The distribution D above is not sufficient for our purposes since there are pairs of alternatives (e.g., a_2 and a_3) that have the same probability of appearing in the first three positions of the outcome; hence, the PD graph of profiles with infinite samples may not be complete. Instead, we will use a distribution D' so that all probability values of this kind are different.

Let $0 = \delta_1 < \delta_2 < ... < \delta_m$ so that $\sum_{i=1}^{m} \delta_i = 1$. Define the probability distribution D'' as follows. Pick one out of the m alternatives so that alternative a_i is picked with probability δ_i. Rank alternative a_i last and complete the ranking by a uniformly random permutation of the alternatives in $\mathcal{L}(A) \setminus \{a_i\}$. Now, the distribution D' is defined as follows: With probability $9/10$ (resp., $1/10$), the output ranking is sampled from the distribution D (resp., D'').

The important property of distribution D'' is that for every $k \in [m-1]$, the probability that alternative a_i is ranked in the first k positions is exactly $\frac{(1-\delta_i)k}{m-1}$, i.e., strictly decreasing in i. On the other hand, distribution D has the property that for every $k \in [m-1]$, the probability that alternative a_i is ranked in the first k positions is non-increasing in i. Hence, their linear combination D' has the property that for every $k \in [m-1]$, the probability that alternative a_i is ranked in the first k positions is strictly decreasing in i. Additionally,

$$\arg\max_{\tau \in \mathcal{L}(A)} \Pr_{D'}[\tau] = \{\sigma^*\}.$$

Hence, we can apply Lemma 7 to obtain a distance metric d' and a d'-monotonic noise model G' so that an infinite number of samples from $G'(\sigma^*)$ induce a complete PD graph corresponding to the ranking $\tau = a_1 \succ a_2 \succ a_3 \succ a_{4...m}$, which is different from the ground truth σ^*. Thus, no PD-c rule is accurate in the limit for G'.

We conclude that no PM-c rule or PD-c rule is monotone-robust with respect to all distance metrics. ■ (Proof of Theorem 3)

The restriction on the number of alternatives in Theorem 3 is indeed necessary. For two alternatives, $\mathcal{L}(A)$ contains only two rankings, and all reasonable voting rules coincide with the majority rule that outputs the more frequent of the two rankings. It can be shown that, in this case, the majority rule is monotone-robust with respect to all distance metrics.

It is known that the union of PM-c and PD-c rules includes all positional scoring rules, Bucklin's rule, the Kemeny rule, ranked pairs, Copeland's method, and Slater's rule (Caragiannis et al., 2013). Two prominent SWFs that are neither PM-c nor PD-c are the maximin rule and STV. In the example given in the proof of Theorem 3, the maximin rule and STV would also rank the wrong alternative (a_1) in the first position with probability 1 in the limit. Thus, Theorem 3 gives another proof that prominent voting rules are not monotone-robust with respect to all distance metrics.

6. Discussion

Perhaps our main conceptual contribution is the realization that the modal ranking rule — a natural voting rule that was previously disregarded — can be exceptionally useful in crowdsourcing settings. Interestingly, from a classic social choice viewpoint the modal ranking rule would appear to be a poor choice. It does satisfy some axiomatic properties, such as Pareto efficiency — if all voters rank x above y, the output ranking places x above y (indeed, the rule always outputs one of the input rankings). But the modal ranking rule fails to satisfy many other basic desiderata, such as monotonicity — if a voter pushes an alternative upwards, and everything else stays the same, that alternative's position in the output should only improve. So our uniqueness result implies an impossibility: a voting rule that is monotone-robust with respect to any distance metric d and is a GSR with no holes, PD-c rule, or PM-c rule, cannot satisfy the monotonicity property. A similar statement is true for any social choice axiom not satisfied by the modal ranking rule. That said, social choice axioms like monotonicity were designed with subjective opinions and notions of social justice in mind. These axioms are incompatible with the settings that motivate our work on a conceptual level, and — as our results show — on a technical level.

Acknowledgments

Procaccia and Shah were partially supported by the NSF under grants NSF IIS-1350598 and NSF CCF-1215883. Caragiannis was partially supported by COST Action IC1205 on Computational Social Choice and by Caratheodory research grant E.114 from the University of Patras.

Appendix

Lemma 6 (Convexity Lemma)**.** *Consider a point $x \in \Delta^{m!}$. Let $FIX \subseteq \mathcal{L}(A)$, and $VARY = \mathcal{L}(A) \setminus FIX$. Further, assume that $\{\sigma \in \mathcal{L}(A) | x_\sigma = 0\} \subseteq FIX$, and let $k = |VARY| \geq 2$. Define*

$$V = \Big\{ v \in \{-1,0,1\}^{m!} \,\Big|\, \forall \sigma \in \mathcal{L}(A),\ v_\sigma = 0 \Leftrightarrow \sigma \in FIX \wedge \exists \sigma \in \mathcal{L}(A),\ v_\sigma = 1 \wedge \exists \sigma \in \mathcal{L}(A),\ v_\sigma = -1 \Big\}.$$

For every $v \in V$, define the orthant

$$O^v = \Big\{ y \in \Delta^{m!} \,\Big|\, \forall \sigma \in \mathcal{L}(A), (v_\sigma = 0 \Rightarrow y_\sigma = x_\sigma) \wedge (v_\sigma = 1 \Rightarrow y_\sigma > x_\sigma) \wedge (v_\sigma = -1 \Rightarrow y_\sigma < x_\sigma) \Big\}.$$

Given points $x^v \in O^v$ for all $v \in V$, $x \in co\{x^v | v \in V\}$, where co denotes the convex hull.

Proof. We prove this by induction on k.

For $k = 2$, let $VARY = \{\sigma_1, \sigma_2\}$. Thus, O^v contains two orthants: one consisting of y's where $y_{\sigma_1} < x_{\sigma_1}$ and $y_{\sigma_2} > x_{\sigma_2}$, and another consisting of y's where $y_{\sigma_1} > x_{\sigma_1}$ and $y_{\sigma_2} < x_{\sigma_2}$. We are given a point x^1 in the former orthant and a point x^2 in the latter orthant. For both points, the values of coordinates other than σ_1 and σ_2 match those for x. Hence, it is easy to check that $x = \lambda x^1 + (1-\lambda) x^2$, where $\lambda = (x_{\sigma_1} - x^2_{\sigma_1})/(x^1_{\sigma_1} - x^2_{\sigma_1})$. It is further easy to check that $0 < \lambda < 1$. Hence, $x \in \mathrm{co}\{x^1, x^2\}$.

Suppose that the theorem holds for all $FIX, VARY$ with $k = |VARY| = d - 1$, for some $d \leq m!$. Let us consider $FIX, VARY$ with $k = |VARY| = d$. Define V and O^v for every $v \in V$ from FIX. Take any $\tau \in VARY$, construct $\widehat{FIX} = FIX \cup \{\tau\}$, and define $\widehat{V}$ and $\widehat{O^v}$ for every $v \in \widehat{V}$ according to $\widehat{FIX}$.

If we can find a point $\widehat{x^v} \in \widehat{O^v}$ for each $v \in \widehat{V}$ which is also in $\mathrm{co}\{x^v | v \in V\}$, then by the induction hypothesis, we have $x \in \mathrm{co}\{\widehat{x^v} | v \in \widehat{V}\} \subseteq \mathrm{co}\{x^v | v \in V\}$. Take any $v \in \widehat{V}$. We construct $v^{+1}, v^{-1} \in V$ as follows: $v^+_\tau = +$, $v^-_\tau = -$, and $v^+_\sigma v^-_\sigma = v_\sigma$ for all $\sigma \neq \tau$. We show that we can find $\widehat{x^v} \in \widehat{O^v}$ as a convex combination of x^{v^+} and x^{v^-}. It is easy to check that taking $\widehat{x^v} = \lambda x^{v^+} + (1-\lambda) x^{v^-}$ works, where $\lambda = (x_\tau - x^{v^-}_\tau)/(x^{v^+}_\tau - x^{v^-}_\tau)$. It is easy to check that $0 < \lambda < 1$. Further, we have $\widehat{x^v}_\tau = x_\tau$ by construction, which is desired because $\tau \in \widehat{FIX}$. For every $\sigma \neq \tau$, $v^+_\sigma v^-_\sigma = v_\sigma$. Hence,

$$v_\sigma = + \Rightarrow \left(x^{v^+}_\sigma > x_\sigma \wedge x^{v^-}_\sigma > x_\sigma \right) \Rightarrow \widehat{x^v}_\sigma > x_\sigma,$$

and

$$v_\sigma = - \Rightarrow \left(x_\sigma^{v^+} < x_\sigma \wedge x_\sigma^{v^-} < x_\sigma \right) \Rightarrow \widehat{x^v}_\sigma < x_\sigma.$$

For arbitrary $v \in \widehat{V}$, we found $\widehat{x^v} \in \widehat{O^v}$, which is also in $\text{co}\{x^v | v \in V\}$ as desired. Thus, $x \in \text{co}\{x^v | v \in V\}$. ■ (Proof of Lemma 6)

Lemma 7. *Given a specific ranking $\sigma^* \in \mathcal{L}(A)$ and a probability distribution D over the rankings in $\mathcal{L}(A)$ such that*

$$\arg\max_{\tau \in \mathcal{L}(A)} \Pr_D[\tau] = \{\sigma^*\},$$

there exists a distance metric d over $\mathcal{L}(A)$ and a d-monotonic noise model G with $\Pr_G[\sigma; \sigma^*] = \Pr_D[\sigma]$ for every $\sigma \in \mathcal{L}(A)$.

Proof. First, let $V = \{\Pr_D[\sigma] | \sigma \in \mathcal{L}(A)\}$ be the set of distinct probability values in D. Now, we construct the distance metric d as follows. For all $\sigma \in \mathcal{L}(A)$, set $d(\sigma, \sigma^*) = d(\sigma^*, \sigma) = |\{v \in V | v > \Pr_D[\sigma]\}|$ for every $\sigma \in \mathcal{L}(A)$. For every pair of rankings σ, σ' different than σ^*, we set $d(\sigma, \sigma') = 0$ if $\sigma = \sigma'$ and $d(\sigma, \sigma') = d(\sigma, \sigma^*) + d(\sigma', \sigma^*)$.

We can easily show that the function d is indeed a distance metric. The first two properties are preserved by definition. For the triangle inequality, we wish to prove that $d(\sigma, \sigma') + d(\sigma', \sigma'') \geq d(\sigma, \sigma'')$ for all $\sigma, \sigma', \sigma'' \in \mathcal{L}(A)$. The inequality clearly holds when any two of the three rankings are identical. If all three rankings are distinct, we take two cases.

1. Suppose either $\sigma = \sigma^*$ or $\sigma'' = \sigma^*$. Without loss of generality, let us assume $\sigma = \sigma^*$. Then, the above inequality is obvious since, by the definition of d, $d(\sigma', \sigma'') \geq d(\sigma^*, \sigma'') = d(\sigma, \sigma'')$.

2. Suppose that neither σ nor σ'' is equal to σ^*. Then, again by the definition of d, the LHS of the above inequality becomes

$$d(\sigma^*, \sigma) + 2d(\sigma^*, \sigma') + d(\sigma^*, \sigma'') \geq d(\sigma^*, \sigma) + d(\sigma^*, \sigma'') = d(\sigma, \sigma'').$$

We now define the noise model G as $\Pr_G[\sigma; \sigma^*] = \Pr_D[\sigma]$ for every $\sigma \in \mathcal{L}(A)$ and

$$\Pr_G[\sigma; \sigma'] = \frac{1/(1 + d(\sigma, \sigma'))}{\sum_{\tau \in \mathcal{L}(A)} 1/(1 + d(\tau, \sigma'))}$$

for $\sigma' \neq \sigma^*$. The property $\Pr_G[\sigma; \sigma'] \geq \Pr[\sigma''; \sigma']$ iff $d(\sigma; \sigma') \leq d(\sigma'', \sigma')$ is obvious if $\sigma' \neq \sigma^*$. Otherwise, recall that $\Pr_G[\sigma; \sigma^*] = \Pr_D[\sigma]$ and, clearly, $\Pr_G[\sigma; \sigma^*] \geq \Pr_G[\sigma'; \sigma^*]$ iff $|\{v \in V | v > \Pr_D[\sigma]\}| \leq |\{v \in V | v > \Pr_D[\sigma']\}|$, i.e., $d(\sigma, \sigma^*) \leq d(\sigma', \sigma^*)$. ■ (Proof of Lemma 7)

References

Caragiannis, I., Procaccia, A. D., & Shah, N. (2013). When do noisy votes reveal the truth?. In *Proceedings of the 14th ACM Conference on Electronic Commerce (EC)*, pp. 143–160.

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

Conitzer, V., Rognlie, M., & Xia, L. (2009). Preference functions that score rankings and maximum likelihood estimation. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 109–115.

Conitzer, V., & Sandholm, T. (2005). Common voting rules as maximum likelihood estimators. In *Proceedings of the 21st Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 145–152.

Desario, D., & Robins, S. (2011). Generalized solid-angle theory for real polytopes. *The Quarterly Journal of Mathematics*, *62*(4), 1003–1015.

Dodgson, C. L. (1876). *A Method for Taking Votes on More than Two Issues*. Clarendon Press.

Elkind, E., Faliszewski, P., & Slinko, A. (2010). Good rationalizations of voting rules. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI)*, pp. 774–779.

Endriss, U., & Grandi, U. (2014). Binary aggregation by selection of the most representative voters. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI)*, pp. 668–674.

Law, E., & von Ahn, L. (2011). *Human Computation*. Morgan & Claypool.

Little, G., Chilton, L. B., Goldman, M., & Miller, R. C. (2010). Turkit: Human computation algorithms on Mechanical Turk. In *Proceedings of the 23rd Annual ACM symposium on User interface software and technology (UIST)*, pp. 57–66.

Lu, T., & Boutilier, C. (2011). Learning Mallows models with pairwise preferences. In *Proceedings of the 28th International Conference on Machine Learning (ICML)*, pp. 145–152.

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

Mao, A., Procaccia, A. D., & Chen, Y. (2013). Better human computation through principled voting. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence (AAAI)*, pp. 1142–1148.

Mossel, E., Procaccia, A. D., & Rácz, M. Z. (2013). A smooth transition from powerlessness to absolute power. *Journal of Artificial Intelligence Research*, *48*, 923–951.

Moulin, H. (1983). *The Strategy of Social Choice*, Vol. 18 of *Advanced Textbooks in Economics*. North-Holland.

Obraztsova, S., Elkind, E., Faliszewski, P., & Slinko, A. (2013). On swap-distance geometry of voting rules. In *Proceedings of the 12th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pp. 383–390.

Procaccia, A. D., Reddi, S. J., & Shah, N. (2012). A maximum likelihood approach for selecting sets of alternatives. In *Proceedings of the 28th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 695–704.

Saari, D. G. (1995). *Basic geometry of voting*. Springer.

Saari, D. G. (2008). Complexity and the geometry of voting. *Mathematical and Computer Modelling*, *48*(9–10), 1335–1356.

Xia, L., & Conitzer, V. (2008). Generalized scoring rules and the frequency of coalitional manipulability. In *Proceedings of the 9th ACM Conference on Electronic Commerce (EC)*, pp. 109–118.

Xia, L., & Conitzer, V. (2009). Finite local consistency characterizes generalized scoring rules. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 336–341.

Xia, L., & Conitzer, V. (2011). A maximum likelihood approach towards aggregating partial orders. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 446–451.

Xia, L., Conitzer, V., & Lang, J. (2010). Aggregating preferences in multi-issue domains by using maximum likelihood estimators. In *Proceedings of the 9th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pp. 399–408.

Xia, L. (2013). Generalized scoring rules: a framework that reconciles Borda and Condorcet. *SIGecom Exchanges*, *12*(1), 42–48.

Young, H. P. (1975). Social choice scoring functions. *SIAM Journal of Applied Mathematics*, *28*(4), 824–838.

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

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