Strategyproof Voting under Correlated Beliefs

Daniel Halpern
Harvard University
dhalpern@g.harvard.edu

Rachel Li
Harvard University
rachelli@college.harvard.edu

Ariel D. Procaccia
Harvard University
arielpro@seas.harvard.edu

Abstract

In voting theory, when voters have ranked preferences over candidates, the celebrated *Gibbard-Satterthwaite Theorem* essentially rules out the existence of reasonable strategyproof methods for picking a winner. What if we weaken strategyproofness to only hold for Bayesian voters with beliefs over others' preferences? When voters believe other participants' rankings are drawn independently from a fixed distribution, the impossibility persists. However, it is quite reasonable for a voter to believe that other votes are correlated, either to each other or to their own ranking. We consider such beliefs induced by classic probabilistic models in social choice such as the *Mallows*, *Placket-Luce*, and *Thurstone-Mosteller* models. We single out the plurality rule (choosing the candidate ranked first most often) as a particularly promising choice as it is strategyproof for a large class of beliefs containing the specific ones we introduce. Further, we show that plurality is unique among positional scoring rules in having this property: no other scoring rule is strategyproof for beliefs induced by the Mallows model when there are a sufficient number of voters. Finally, we give examples of prominent non-scoring voting rules failing to be strategyproof on beliefs in this class, further bolstering the case for plurality.

1 Introduction

One of the most celebrated results in voting theory is the *Gibbard-Satterthwaite Theorem* [8, 22]. It states that when voters express ordinal preferences over at least 3 candidates, there is no "reasonable" aggregation rule that is *strategy-proof*: there will always exist instances where voters will be incentivized to manipulate and lie about their preferences to achieve a better outcome.

However, one caveat about this strong negative result is that, a priori, a voter may need perfect information about how others vote to manipulate successfully. Perhaps, if the voter is slightly uncertain, no manipulation helps consistently enough to be worthwhile. Majumdar and Sen [13] analyzed exactly this question when voters have *independent beliefs*. That is, when a voter is considering whether or not to manipulate, they assume all others have rankings drawn independently from a fixed distribution. The classic notion of strategyproofness no longer makes sense in this probabilistic Bayesian setting, so they instead use the natural extension known as *ordinally Bayesian incentive compatible (OBIC)*, essentially that the rules are strategyproof in expectation no matter what underlying cardinal values voters have. Their results, unfortunately, are widely negative. They show that for a "large" set of distributions, Gibbard-Satterthwaite still holds. There do exist distributions where many rules are OBIC, e.g., the uniform distribution over all rankings. Still, these positive examples are extremely brittle: even a slight perturbation leads back to the impossibility.

Preprint. Under review.

But independent beliefs are quite restrictive. They cannot capture several kinds of beliefs that would likely occur in practice. For one, when the number of voters is large, the uncertainty essentially vanishes. Suppose a distribution places probability $1/4$ on other voters having the ranking $a \succ b \succ c$. In that case, when the number of voters is large, it is extremely unlikely that the proportion of voters with this ranking is anything other than $1/4 \pm \varepsilon$. In a real presidential election, a voter may quite plausibly believe that a candidate will receive anywhere between 45% and 55% of the votes, but this situation simply cannot be captured by a single independent ranking distribution. Second, one's own ranking may influence the probability placed on others. Suppose a voter, after much research, discovers that they prefer one proposal to another; they may reasonably believe others are clever enough to have reached a similar conclusion. In terms of their beliefs, they may place a slightly higher probability on others voting more similarly to them than not, no matter what their realized preferences are.

This has led follow-up work to consider the same question under *correlated beliefs* [1, 16, 14, 2]. However, besides some impossibilities, the work so far has largely been of the following form: for any reasonable voting rule, *there exists* a set of beliefs where the rule is OBIC. But perhaps the more natural direction is the converse: under a natural set of beliefs, is there a reasonable voting rule that is consistently OBIC? Can this property help us distinguish between voting rules, showing that under some reasonable beliefs, certain rules are *not* OBIC, thereby bolstering the case for the provably incentive-compatible ones? These are the questions we tackle.

Our contributions. We begin by presenting various classes of beliefs induced by classic probabilistic social choice models such as the *Mallows* [15], *Thurstone-Mosteller* [23, 18], and *Placket-Luce* [20, 12] models. In essence, these are the beliefs a voter would have if they assume that voter preferences were generated by such a model. Inspired by these models, we present a novel class of mildly correlated beliefs that includes all of them. We show that, under this class of beliefs, the *plurality* rule is OBIC.

Next, we provide a negative result: Among positional scoring rules (where each voter assigns a fixed score to each position in their ranking), plurality is unique in being OBIC when voters have Mallows beliefs. All other rules will become not OBIC when there are three candidates, at least when there are a sufficient number of voters. In addition, we provide some robustness checks on this negative result. A popular positional scoring rule known as *Borda Count* fails for any number of voters. By contrast, we identify other positional scoring rules that are OBIC with two voters, meaning our result could not be strengthened by relaxing the sufficient number of voters requirement.

Finally, we complement this more sweeping classification with examples of other prominent rules, such as *Copeland* and *maximin*, which fail to be OBIC with specific Mallows beliefs and few voters. This further bolsters the case for plurality as an unusually attractive rule when viewed through the lens of ordinal Bayesian incentive compatibility under correlated beliefs.

Related work. As mentioned above, the analysis of OBIC voting rules began with Majumdar and Sen [13] essentially providing the final word on independent beliefs; their notion of OBIC dates back to work on committee selection [5].

Since then, there have been a few lines of work on correlated beliefs with slightly different goals. The most closely related is that of Majumdar and Sen [14]. They define a large class of positively correlated beliefs based on the Kemeny metric and then show in a similar fashion to the Gibbard-Satterthwaite Theorem that any voting rule that is OBIC with respect to these beliefs, along with being Pareto efficient, is necessarily dictatorial. They do present one voting rule that is both OBIC with respect to these beliefs and nondictatorial (while not being Pareto efficient), but it is a clearly impractical rule that is designed to make a technical point.[1] Note that all the rules we consider are Pareto efficient.

Another line of work considers *local* OBIC. A voting rule is locally OBIC with respect to a class of beliefs if there *exists* a belief in the class such that any belief in a neighborhood of the original is OBIC. This means the rule remains OBIC even after a slight perturbation to the underlying belief. Bhargava et al. [1] and Bose and Roy [2] attempt to classify the set of locally OBIC voting rules with

[1] Their rule is called *Unanimity with Status Quo* . There is one default candidate x. If there is a candidate y which every single voter places as their top choice, then y is elected, but in any other case, x wins.

respect to a large class of correlated beliefs and show that under minimal conditions, this requirement can be satisfied.

Mandal and Parkes [16] consider a different notion of incentive compatibility which, rather than requiring that no manipulation can lead to a utility gain in expectation, bounds the probability under which there is a utility gain. They again do this with respect to several different classes of beliefs, including one that we consider based on the Mallows Model.

Further afield, there are extensive lines of research on circumventing the Gibbard-Satterthwaite Theorem. We provide examples of three here, although there are many others. One line considers the complexity, showing that some rules, while in principle susceptible to manipulation, have instances where it is hard (in the worst case) to find such a manipulation [7, 3]. Another considers strategyproofness under restricted domains, where a voter's set of possible rankings is limited [7, 3]. A third considers the likelihood of an individual arriving in an instance where they are able to manipulate at all [17, 25].

Finally, without considering strategyproofness, there has been much work on probabilistic social choice, making use of the models on which our results are based, especially in learning preferences from data [11, 24, 10, 19].

2 Model

We begin by introducing the classic social model and then later describe relevant definitions for social choice under uncertainty.

Classic social choice model. Let $N = \{1, \dots, n\}$ be a set of n *voters*, and let $\mathcal{A} = \{a_1, \dots, a_m\}$ be a set of m *alternatives*. Let $\mathcal{L}$ be the set of rankings over $\mathcal{A}$, where for $\sigma \in \mathcal{L}$, $\sigma(j)$ is the j'th candidate in ranking σ and $\sigma^{-1}(a)$ is the ranking index of candidate a. We use the notation $a \succ_\sigma b$ to denote that $\sigma^{-1}(a) < \sigma^{-1}(b)$ and $a \succeq_\sigma b$ to denote $\sigma^{-1}(a) \leq \sigma^{-1}(b)$, i.e., a is strictly (or weakly) preferred to b under σ. Additionally, instead of writing $\sigma = a \succ b \succ c$, when it is clear from context, we will sometimes shorten this to $\sigma = abc$. Each voter i has a ranking $\sigma_i \in \mathcal{L}$ and the tuple of these rankings $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_n) \in \mathcal{L}^n$ is called the *preference profile* . We let $\boldsymbol{\sigma}_{-i} \in \mathcal{L}^{n-1}$ denote the profile without voter i, and for a ranking $\sigma_i' \in \mathcal{L}$, we let $(\boldsymbol{\sigma}_{-i}, \sigma_i')$ be the profile with σ_i replaced with σ_i'.

A *voting rule* is a function f that, given a profile $\boldsymbol{\sigma}$, outputs a distribution over winning alternatives. We define several voting rules of interest here. Our theoretical results will primarily focus on *positional scoring rules* [26]. A positional scoring rule f is parameterized by a vector of $(s_1, \dots, s_m)$ where each $s_j \in \mathbb{Z}_{\geq 0}$ with $s_1 \geq \cdots \geq s_m$ and $s_1 > s_m$. On a profile $\boldsymbol{\sigma}$, for each voter i, their j'th candidate $\sigma_i(j)$ is given s_j points. The points are added up over all voters, and the winning candidate is the one with the most points. More formally, for a ranking $\sigma \in \mathcal{L}$ and candidate $c \in \mathcal{A}$, we write $\text{SC}_c^f(\sigma) = s_{\sigma^{-1}(c)}$ for the points (or score) given to c by σ. For a profile $\boldsymbol{\sigma}$, we write $\text{SC}_c^f(\boldsymbol{\sigma}) = \sum_i \text{SC}_c^f(\sigma_i)$ to be the total points. When f is clear from context, we may drop it from the notation. In deterministic settings, when there is a tie, a tie-breaking rule needs to be given (i.e., tie-break in favor of lower index candidates). Since we will be working in a probabilistic setting, it will be more convenient to assume *uniform random tie-breaking*, so that if there is a tie among k candidates, each wins with probability $1/k$. However, our results would continue to hold even with arbitrary deterministic choices. Two rules of particular interest are *plurality*, parameterized by the vector $(1, 0, \dots, 0)$, and *Borda count*, parameterized by the vector $(m-1, m-2, \dots, 1, 0)$.

We consider two additional rules beyond positional scoring rules, *Copeland* and *maximin*. To define them, for a profile $\boldsymbol{\sigma}$ we define the pairwise margin for two candidates a and b, $N_{ab}(\boldsymbol{\sigma}) = |\{i | a \succ_i b\}|$, i.e., the number of voters that prefer candidate a to candidate b.

For Copeland, we define the Copeland score for a candidate a as $\sum_{b \neq a} \mathbf{I}[N_{ab}(\boldsymbol{\sigma}) > n/2] + (1/2)\mathbf{I}[N_{ab}(\boldsymbol{\sigma}) = n/2]$. In words, the candidate gets one point for every other candidate they pairwise beat and a half point for every other candidate they pairwise tie. The Copeland winners are those with the highest Copeland scores (with uniform tie-breaking). For maximin, we define the maximin score for a candidate a as $\min_{b \neq a} N_{ab}(\boldsymbol{\sigma})$, i.e., the smallest margin by which a beats another candidate. Again, the maximin winners are those with the highest maximin scores (with uniform tie-breaking).

A voting rule f is called *strategy-proof* if for all profiles $\boldsymbol{\sigma}$, all voters i, and all alternative manipulations $\sigma_i' \in \mathcal{L}$, $f(\boldsymbol{\sigma}) \succeq_{\sigma_i} f(\boldsymbol{\sigma}_{-i}, \sigma_i')$. That is, no voter can ever improve the outcome of the vote by misreporting. A rule is called onto if for all candidates $a \in \mathcal{A}$, there is a profile $\boldsymbol{\sigma}$ where $f(\boldsymbol{\sigma}) = a$, and is called dictatorial if there is a voter i for which $f(\boldsymbol{\sigma}) = \sigma_i(1)$, i.e., voter i always gets their top choice. The Gibbard-Satterthwaite Theorem states that any rule for $m \geq 3$ candidates that is strategy-proof and onto is necessarily dictatorial. Since "reasonable" rules must be onto and nondictatorial, this eliminates the possibility of any being strategy-proof.

These notions can also be extended to randomized rules. A *utility function* u is a mapping from candidates to real numbers. We say that u is consistent with a ranking σ if $u(x) > u(y) \Leftrightarrow x \succ_\sigma y$. A (randomized) voting rule f is called SD-strategy-proof if for all profiles $\boldsymbol{\sigma}$, all voters i, all manipulations σ_i', and all utility functions u consistent with σ_i, $\mathbb{E}[u(f(\boldsymbol{\sigma}))] \geq \mathbb{E}[u(f(\boldsymbol{\sigma}_{-i}, \sigma_i'))]$. This says that no matter what underlying utilities an agent has, as long as they are consistent with their ranking, they cannot improve their expected utility by manipulation. An equivalent definition can be given with respect to stochastic dominance (hence the SD in the name). For $k \leq m$, let $B_k(\sigma) = \{\sigma(1), \ldots, \sigma(k)\}$ be the set of the k best alternatives according to σ. Then, SD-strategy-proofness can be rephrased as requiring that for all profiles $\boldsymbol{\sigma}$, all voters i, all manipulations σ_i', and all $k \leq m$, $\Pr[f(\boldsymbol{\sigma}) \in B_k(\sigma_i)] \geq \Pr[f(\boldsymbol{\sigma}_{-i}, \sigma_i') \in B_k(\boldsymbol{\sigma}_i)]$.

We say a rule f is unilateral if it depends only on a single voter, i.e., there is a voter i such that for all $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$, if $\sigma_i = \sigma_i'$, then $f(\boldsymbol{\sigma}) = f(\boldsymbol{\sigma}')$. A rule f is called a duple if its range is two candidates, i.e., there is a pair of candidates a and b such that for all $\boldsymbol{\sigma}$, $f(\boldsymbol{\sigma}) \in \{a, b\}$. Gibbard [9] extended the Gibbard-Satterthwaite theorem to randomized rules as follows: Any randomized rule f that is strategyproof is a mixture over unilateral and duple rules. Since unilateral and duple rules are seen as undesirable, this implies that finding a reasonable, strategy-proof voting rule, even allowing randomization, is a hopeless endeavor.

Social choice under uncertainty. A *belief* for voter i is a probability measure $\mathbb{P}_i$ over the set of profiles $\mathcal{L}^n$ (when the i is clear from context we will drop it from the notation). This describes i's prior probability over profiles before considering their own ranking. After observing their own ranking $\hat{\sigma}_i$, the voter can update their posterior using the conditional distribution $\mathbb{P}[\cdot \mid \sigma_i = \hat{\sigma}_i]$. For notational convenience, we will often shorten this to $\mathbb{P}[\cdot \mid \hat{\sigma}_i]$.

In this model, we need a slightly different notion of strategyproofness. A voting rule is called *ordinally Bayesian incentive compatible (OBIC)* with respect to a beliefs $(\mathbb{P}_1, \ldots, \mathbb{P}_n)$ if for all voters i, all rankings $\hat{\sigma}_i$, all manipulations σ_i', and all utility functions u consistent with $\hat{\sigma}_i$, $\mathbb{E}[u(f(\boldsymbol{\sigma}_{-i}, \hat{\sigma}_i)) \mid \hat{\sigma}_i] \geq \mathbb{E}[u(f(\boldsymbol{\sigma}_{-i}, \sigma_i') \mid \hat{\sigma}_i]$. This is the natural generalization of SD-strategyproofness to a Bayesian setting. Just as with SD-strategyproofness, an equivalent definition is for all voters i, all rankings $\hat{\sigma}_i$, all manipulations σ_i', and all $k \leq m$, $\mathbb{P}[f(\boldsymbol{\sigma}_{-i}, \hat{\sigma}_i) \in B_k(\hat{\sigma}_i) \mid \hat{\sigma}_i] \geq \mathbb{P}[f(\boldsymbol{\sigma}_{-i}, \sigma_i') \in B_k(\hat{\sigma}_i) \mid \hat{\sigma}_i]$.

We now present a few possible choices of "reasonable" priors based on well-known probabilistic models of social choice. The first is based off of a *Mallows Model* [15]. This model is parameterized by a *ground truth* ranking $\tau \in \mathcal{L}$ and a dispersion quantity φ. We define the Kendall tau distance between rankings $d(\sigma_1, \sigma_2) = |\{(a, b) \in \mathcal{A}^2 | a \succ_{\sigma_1} b \wedge b \succ_{\sigma_2} a\}|$, i.e., the number of pairs of candidates on which σ_1 and σ_2 disagree. In a Mallows Model, each voter's ranking is drawn independently with probability proportional to $\varphi^{d(\sigma,\tau)}$. More formally, the probability that a specific ranking σ is drawn is equal to $\frac{\varphi^{d(\sigma,\tau)}}{Z}$ where $Z = \sum_{\sigma \in \mathcal{L}} \varphi^{d(\sigma,\tau)}$ is the normalizing constant. One can easily check that if we extend the notion of Kendall tau distance to operate on a profile and a ranking, with $d(\boldsymbol{\sigma}, \tau) = \sum_i d(\sigma_i, \tau)$, then the probability of sampling a profile $\boldsymbol{\sigma}$ is proportional to $\varphi^{d(\boldsymbol{\sigma},\tau)}$ (this time with a Z^n normalizing constant).

We convert this model into a prior in two ways. The first we call a *confident Mallows prior* parameterized by φ. The agent assumes a ground truth τ is first drawn from some (arbitrary) distribution, then, given this ground truth, $\sigma_i = \tau$ with probability 1 and the remainder of the profile $\boldsymbol{\sigma}_{-i}$ is drawn from a Mallows Model with a fixed φ using τ. Essentially, the agent believes that they correctly know the ground truth, but all others only approximate this truth using a Mallows Model. The conditional distribution over the remainder of the profile $\boldsymbol{\sigma}_{-i}$ given $\hat{\sigma}_i$ then follows a standard Mallows model with the ground truth equal to $\hat{\sigma}_i$, so $\mathbb{P}[\boldsymbol{\sigma}_{-i} \mid \hat{\sigma}_i] \propto \varphi^{d(\boldsymbol{\sigma}_{-i}, \hat{\sigma}_i)}$.[2]

[2]This is equivalent to the *Conditional Mallows Model* of Mandal and Parkes [16].

The second we call an *unconfident Mallows prior*. Here, the agent believes that the ground truth τ is drawn uniformly at random, and then the entire profile (including their own ranking) is drawn from a Mallows Model. Therefore, $\mathbb{P}[\boldsymbol{\sigma}] = \frac{1}{m!}\sum_{\tau\in\mathcal{L}} \frac{\varphi^{d_{kt}(\boldsymbol{\sigma},\tau)}}{Z^n}$. Since for any $\hat{\sigma}_i$, by symmetry $\Pr[\sigma_i = \hat{\sigma}_i] = \frac{1}{m!}$, we can write the conditional probability as

$$\mathbb{P}[\boldsymbol{\sigma}_{-i} \mid \hat{\sigma}_i] = \sum_{\tau\in\mathcal{L}} \frac{\varphi^{d((\boldsymbol{\sigma}_{-i},\hat{\sigma}_i),\tau)}}{Z^n} = \sum_{\tau\in\mathcal{L}} \frac{\varphi^{d(\boldsymbol{\sigma}_{-i},\tau)}}{Z^{n-1}} \cdot \frac{\varphi^{d(\hat{\sigma}_i,\tau)}}{Z}.$$

We will abuse notation slightly and write $\mathbb{P}[\tau \mid \hat{\sigma}_i] = \frac{\varphi^{d(\hat{\sigma}_i,\tau)}}{Z}$ and $\mathbb{P}[\boldsymbol{\sigma}_{-i} \mid \tau] = \frac{\varphi^{d(\boldsymbol{\sigma}_{-i},\tau)}}{Z^{n-1}}$, so that

$$\mathbb{P}[\boldsymbol{\sigma}_{-i} \mid \hat{\sigma}_i] = \sum_{\tau\in\mathcal{L}} \mathbb{P}[\boldsymbol{\sigma}_{-i} \mid \tau] \cdot \mathbb{P}[\tau \mid \hat{\sigma}_i].$$

We can interpret this as saying the voter has a posterior over ground truths, $\mathbb{P}[\tau \mid \hat{\sigma}_i]$, and is using this posterior to infer the probability of the rest of the profile. Intuitively, the agent is uncertain over ground truths but, due to the observation of their ranking, places higher weight on ground truths that are closer to their own ranking. This decomposition is possible because the rest of the profile $\boldsymbol{\sigma}_{-i}$ is conditionally independent of $\hat{\sigma}_i$ given the ground truth τ.

The *Thurstone-Mosteller* model is defined with respect to underlying means μ_c for each candidate $c \in \mathcal{A}$. To sample a ranking, a value $X_c \sim \mathcal{N}(\mu_c, 1)$ is drawn independently for each candidate c from a normal distribution with variance 1 around the mean. The resulting ranking is the order of the X_c values from highest to lowest. The *Placket-Luce* model is defined with respect to underlying weights $w_c > 0$ for each candidate $c \in \mathcal{A}$. To sample a ranking, we iteratively select a candidate c from the remaining unchosen candidates P with probability $\frac{w_c}{\sum_{c'\in P} w'_c}$.

To convert these models to beliefs, we assume that the voter believes there are underlying distinct means $\mu_1 > \cdots > \mu_m$ (resp. weights $w_1 > \cdots > w_m$) but is uncertain about which candidate has which mean (resp. weight). To relate this to the Mallows belief, we will call this order τ, the ground truth. In the confident version, the voter believes that their ranking is always equal to τ, but all other votes are drawn from the corresponding model. In the unconfident version, the voter believes a priori that τ is drawn uniformly at random, and then all voter rankings, including their own, are drawn from the corresponding model. As with the Mallows beliefs, the voter can do a Bayesian update to compute a posterior about which candidate was assigned to which weight. We can again decompose

$$\mathbb{P}[\boldsymbol{\sigma}_{-i} \mid \hat{\sigma}_i] = \sum_{\tau} \mathbb{P}[\boldsymbol{\sigma}_{-i} \mid \tau] \cdot \mathbb{P}[\tau \mid \hat{\sigma}_i],$$

where, by Bayes' rule, $\mathbb{P}[\tau \mid \hat{\sigma}_i] \propto \mathbb{P}[\hat{\sigma}_i|\tau]$, the probability of generating $\hat{\sigma}_i$ under the model with ground truth τ.

Note that to make this more general, it would also make sense for the voter to believe there is a distribution over means or weights; our results continue to hold with this more general class; however, for ease of presentation, we focus on the more restricted form.

3 Plurality is OBIC

We start by defining a class of beliefs that we call *top-choice correlated*. The class is similar in spirit (although incomparable) to the class of top-set correlated beliefs introduced by Bhargava et al. [1].

To define the class, given a profile $\boldsymbol{\sigma}$ and a candidate $c \in \mathcal{A}$, we let $\text{PLU}_c(\boldsymbol{\sigma}) = |\{i \mid \sigma_i(1) = c\}|$ be the *plurality score of* c, i.e., the number of voters that rank c first. Further, we let $\text{PLU}(\boldsymbol{\sigma})$ be the vector of plurality scores indexed by the candidates. A belief $\mathbb{P}_i$ is top-choice correlated if the following holds. Fix a ranking $\hat{\sigma}_i$ and let $a = \hat{\sigma}_i(1)$. Then, for all candidates $b \neq a$ and all pairs of plurality vectors $\mathbf{r}$ and $\mathbf{r}'$ such that $r_c = r'_c$ for all $c \neq a, b$, $r_a = r'_b$, $r_b = r'_a$, and $r_a > r_b$, $\mathbb{P}[\text{PLU}(\boldsymbol{\sigma}_{-i}) = \mathbf{r} \mid \hat{\sigma}_i] \geq \mathbb{P}[\text{PLU}(\boldsymbol{\sigma}_{-i}) = \mathbf{r}' \mid \hat{\sigma}_i]$. This says that if the voter is told the remaining plurality scores of all other candidates except a and b, as well as possible scores for a and b, they would think it is more likely that a (their top choice) has the higher score. In other words, all else being equal, the voter's top choice is more likely to perform better than other candidates.

We now claim that all of the specific beliefs we have introduced are top-choice correlated, suggesting that this condition is quite weak.

Lemma 1. *The confident and unconfident versions of Mallows, Thurstone-Mosteller, and Placket-Luce beliefs under any parameter settings are top-choice correlated.*

The proof of Lemma 1 can be found in Appendix A. For confident versions of these beliefs, this is relatively straightforward as the models directly place higher mass on the voter's top choice being chosen. For the unconfident versions, slightly more intricate analysis is necessary to show that more mass is placed on ground truth rankings where the voter's top choice is higher, and from this, we can reach the same conclusion.

Despite the breadth of the class of top-choice correlated beliefs, it turns out that plurality is OBIC for all beliefs in this class.

Lemma 2. *Under any top-choice correlated beliefs, plurality is OBIC.*

Proof. Let f be the plurality voting rule, i be an agent, and $\mathbb{P}$ be their top-choice correlated belief. Suppose i observes $\hat{\sigma}_i$, and let u be an arbitrary utility function that is consistent with $\hat{\sigma}_i$. Let $a = \hat{\sigma}_i(1)$ be their top-ranked alternative.

Let σ_i' be a possible manipulation for voter i, and let $b = \sigma_i'(1)$ be the top-ranked alternative. Notice that if $a = b$, then the outcome under plurality is identical, and this manipulation cannot be an improvement. Hence, from now on, we assume that $b \neq a$.

For $\boldsymbol{\sigma}_{-i}$, let $\textsc{ug}(\boldsymbol{\sigma}_{-i}) = \mathbb{E}[u(f(\boldsymbol{\sigma}_{-i}, \sigma_i'))] - \mathbb{E}[u(f(\boldsymbol{\sigma}_{-i}, \hat{\sigma}_i))]$ be the expected utility gain of switching from $\hat{\sigma}_i$ to σ_i' when others report $\boldsymbol{\sigma}_{-i}$. We wish to show that $\mathbb{E}[\textsc{ug}(\boldsymbol{\sigma}_{-i})|\hat{\sigma}_i] \leq 0$, where the expectation is over the belief $\mathbb{P}$. To simplify notation, we will allow the utility function u to operate on (nonempty) sets of candidates S as $u(S) = \frac{1}{|S|}\sum_{c \in S} u(c)$. Note that when the set of plurality winners on a profile is S, the expected utility is $u(S)$.

We now partition the possible $\boldsymbol{\sigma}_{-i}$ based on their utility gain. Let $C \subseteq \mathcal{A} \setminus \{a, b\}$ be a (nonempty) set candidates not including a and b. For each set C, we define eight sets of profiles $\boldsymbol{\sigma}_{-i}$ depending on the winners under $(\boldsymbol{\sigma}_{-i}, \hat{\sigma}_i)$ and $(\boldsymbol{\sigma}_{-i}, \sigma_i')$, $E_1(C)^a, E_1(C)^b, E_2(C)^a, E_2(C)^b, E_3(C)^a, E_3(C)^b, E_4(C), E_5(C)$. In each $E(C)$ set, C will be the set of candidates excluding a and b with the highest plurality score. We abuse notation slightly and write $\textsc{plu}(C)$ for the (tied) plurality score of each of these candidates and $\textsc{plu}(a)$ and $\textsc{plu}(b)$ for the plurality scores of a and b, respectively. The sets are otherwise defined by the set of plurality winners in $(\boldsymbol{\sigma}_{-i}, \hat{\sigma}_i)$ and $(\boldsymbol{\sigma}_{-i}, \sigma_i')$. The definitions can be found in Table 1. One can check that these (disjoint) sets collectively cover all possible $\boldsymbol{\sigma}_{-i}$ where $\textsc{ug}(\boldsymbol{\sigma}_{-i})$ is nonzero.

We can now rewrite the expected utility gain in terms of these sets. For each set $E(C)$, we write $\textsc{ug}(E(C))$ for the expected utility gain for profiles $\boldsymbol{\sigma}_{-i} \in E(C)$ (which will always be the same for all $\boldsymbol{\sigma}_{-i} \in E(C)$). From this, we have

$$\mathbb{E}[\textsc{ug}(\boldsymbol{\sigma}_{-i})|\hat{\sigma}_i] = \sum_{\substack{C \subseteq \mathcal{A} \setminus \{a,b\} \\ C \neq \emptyset}} \Bigg(\sum_{j=1}^{3} \left(\mathbb{P}[E_j(C)^a|\hat{\sigma}_i]\textsc{ug}(E_j(C)^a) + \mathbb{P}[E_j(C)^b|\hat{\sigma}_i]\textsc{ug}(E_j(C)^b) \right) + \mathbb{P}[E_4(C)|\hat{\sigma}_i]\textsc{ug}(E_4(C)) + \mathbb{P}[E_5(C)|\hat{\sigma}_i]\textsc{ug}(E_5(C)) \Bigg)$$

Our goal, again, is to show that this expression is at most 0. Notice that for each C, $\textsc{ug}(E_4(C)) = u(b) - u(a) < 0$ and $\textsc{ug}(E_5(C)) = \frac{1}{|C|}(u(b) - u(a)) < 0$ because $u(a) > u(c)$ for all other candidates c. In what remains, we show that for all C and each $j \leq 3$,

$$\mathbb{P}[E_j(C)^a]\textsc{ug}(E_j(C)^a) + \mathbb{P}[E_j(C)^b]\textsc{ug}(E_j(C)^b) \leq 0. \tag{1}$$

Fix an arbitrary C. To do this, we show that for each j, $\textsc{ug}(E_j(C)^a) \leq 0$, $-\textsc{ug}(E_j(C)^a) \geq \textsc{ug}(E_j(C)^b)$, and $\mathbb{P}[E_j(C)^a] \geq \mathbb{P}[E_j(C)^b]$. Together, these imply (1).

We analyze the case of $j = 1$; the arguments for $j = 2$ and $j = 3$ are very similar. Notice that

$$\textsc{ug}(E_1(C)^a) = u(C \cup \{a\}) - u(a) = \frac{|C|}{|C| + 1}(u(C) - u(a)),$$

Set	Condition	$(\boldsymbol{\sigma}_{-i}, \hat{\sigma}_i)$ Winners	$(\boldsymbol{\sigma}_{-i}, \sigma_i')$ Winners
$E_1(C)^a$	$\text{PLU}(a) = \text{PLU}(C) > \text{PLU}(b) + 1$	a	$C \cup \{a\}$
$E_1(C)^b$	$\text{PLU}(b) = \text{PLU}(C) > \text{PLU}(a) + 1$	$C \cup \{b\}$	b
$E_2(C)^a$	$\text{PLU}(a) = \text{PLU}(C) = \text{PLU}(b) + 1$	a	$C \cup \{a, b\}$
$E_2(C)^b$	$\text{PLU}(b) = \text{PLU}(C) = \text{PLU}(a) + 1$	$C \cup \{a, b\}$	b
$E_3(C)^a$	$\text{PLU}(C) = \text{PLU}(a) + 1 > \text{PLU}(b) + 1$	$C \cup \{a\}$	C
$E_3(C)^b$	$\text{PLU}(C) = \text{PLU}(b) + 1 > \text{PLU}(a) + 1$	C	$C \cup \{b\}$
$E_4(C)$	$\text{PLU}(a) = \text{PLU}(b) \geq \text{PLU}(C)$	a	b
$E_5(C)$	$\text{PLU}(C) = \text{PLU}(a) + 1 = \text{PLU}(b) + 1$	$C \cup \{a\}$	$C \cup \{b\}$

Table 1: Definition of the sets $E_1(C)^a, E_1(C)^b, E_2(C)^a, E_2(C)^b, E_3(C)^a, E_3(C)^b$, $E_4(C)$, and $E_5(C)$. They contain all $\boldsymbol{\sigma}_{-i}$ that satisfy the corresponding condition. In each set, the contained $\boldsymbol{\sigma}_{-i}$ all have the same winners in both $(\boldsymbol{\sigma}_{-i}, \hat{\sigma}_i)$ and $(\boldsymbol{\sigma}_{-i}, \sigma_i')$ as seen in the corresponding columns.

and symmetrically

$$\text{UG}(E_1(C)^b) = u(b) - u(C \cup \{b\}) = \frac{|C|}{|C|+1}(u(b) - u(C)).$$

Since $u(a)$ is maximal, $\text{UG}(E_1(C)^a) \leq 0$. Further,

$$-\text{UG}(E_1(C)^a) - \text{UG}(E_1(C)^b) = \frac{|C|}{|C|+1}(u(a) - u(b)) \geq 0.$$

Finally, to show $\mathbb{P}[E_1(C)^a] \geq \mathbb{P}[E_1(C)^b]$, we consider the vectors of plurality scores (indexed by candidates) $\mathbf{r}^a$ and $\mathbf{r}^b$ that lead a profile $\boldsymbol{\sigma}_{-i}$ to end in up in $E_1(C)^a$ and $E_1(C)^b$, respectively. Due to the symmetry, there is a natural bijection between these two sets of vectors obtained by swapping the a and b components. Further, by the definition of $E_1(C)^a$ and $E_1(C)^b$, the a component of $\mathbf{r}^a$ is always strictly larger than the b component. Since $\mathbb{P}$ is top-choice correlated, for any two vectors that differ by swapping the a and b components, $\mathbb{P}[\cdot|\hat{\sigma}_i]$ always places higher mass on the vector in $\mathbf{r}^a$. Therefore, $\mathbb{P}[E_1(C)^a|\hat{\sigma}_i] \geq \mathbb{P}[E_1(C)^b|\hat{\sigma}_i]$, as needed. □

From these two lemmas, we immediately derive our main positive result.

Theorem 1. *When voters have beliefs that are any of the confident or unconfident versions of Mallows, Thurstone-Mosteller, or Placket-Luce under any parameter settings, plurality is OBIC.*

4 Other Voting Rules Are Not OBIC

From the positive result about plurality, one might wonder whether satisfying OBIC with respect to these beliefs is a relatively weak condition. If several rules satisfy it, this property is not useful for a mechanism designer who is comparing between rules to implement. In this section, we show this is not the case. Specifically, we focus on both the confident and unconfident variants of Mallows beliefs. Our main theoretical negative result is that plurality is uniquely OBIC among positional scoring rules in certain regimes of Mallows beliefs.

Theorem 2. *Let f be a non-plurality positional scoring rule on three candidates. If a voter has unconfident or confident Mallows beliefs with $\varphi \leq .988$, for a sufficiently large n, f is not OBIC.*

Below we provide a detailed proof sketch. However, we shunt some unwieldy technical derivations into two lemmas relegated to the appendix.

Proof sketch of Theorem 2. Fix a non-plurality scoring vector (s_1, s_2, s_3). Without loss of generality, we can translate and scale the vector such that $s_3 = 0$ and s_1 and s_2 are relatively prime integers with $s_2 > 0$. Fix a voter i with unconfident Mallows beliefs $\mathbb{P}$ with parameter $\varphi \leq .988$; we will describe later how to extend it to a confident Mallow's belief. Suppose they observe ranking $\hat{\sigma}_i = a \succ b \succ c$. We will show that for sufficiently large n, it will be beneficial to switch to $\sigma_i' = a \succ c \succ b$. More specifically, we will show that this manipulation increases the probability

that candidate a wins, which means the rule cannot be OBIC with respect to these beliefs. Formally, we will show that,

$$\Pr[f(\boldsymbol{\sigma}_{-i}, \sigma_i') = a \mid \hat{\sigma}_i] > \Pr[f(\boldsymbol{\sigma}_{-i}, \hat{\sigma}_i) = a \mid \hat{\sigma}_i],$$

or equivalently,

$$\Pr[f(\boldsymbol{\sigma}_{-i}, \sigma_i') = a \mid \hat{\sigma}_i] - \Pr[f(\boldsymbol{\sigma}_{-i}, \hat{\sigma}_i) = a \mid \hat{\sigma}_i] > 0.$$

Notice that since we are looking at the difference in probabilities of two events, we can ignore their intersection when both reports lead to a as the winner. That is, the left-hand side is equal to[3]

$$\Pr[f(\boldsymbol{\sigma}_{-i}, \sigma_i') = a \wedge f(\boldsymbol{\sigma}_{-i}, \hat{\sigma}_i) \neq a \mid \hat{\sigma}_i] - \Pr[f(\boldsymbol{\sigma}_{-i}, \hat{\sigma}_i) = a \wedge f(\boldsymbol{\sigma}_{-i}, \sigma_i') \neq a \mid \hat{\sigma}_i].$$

To shorten notation, we will label the events of interest as $\mathcal{E}^{cb} = \{f(\boldsymbol{\sigma}_{-i}, \sigma_i') = a \wedge f(\boldsymbol{\sigma}_{-i}, \hat{\sigma}_i) \neq a \mid \hat{\sigma}_i]\}$ (i.e., ranking c above b causes a to win) and $\mathcal{E}^{bc} = \{f(\boldsymbol{\sigma}_{-i}, \hat{\sigma}_i) = a \wedge f(\boldsymbol{\sigma}_{-i}, \sigma_i') \neq a \mid \hat{\sigma}_i]\}$ (i.e., ranking b above c causes a to win), so we wish to show that

$$\mathbb{P}[\mathcal{E}^{cb} | \hat{\sigma}_i] - \mathbb{P}[\mathcal{E}^{bc} | \hat{\sigma}_i] > 0.$$

Using the definition of the Mallows belief model, we can expand the left-hand side using ground truths to

$$\sum_{\tau \in \mathcal{L}} \left(\mathbb{P}[\mathcal{E}^{cb} \mid \tau] - \mathbb{P}[\mathcal{E}^{bc} \mid \tau]\right) \mathbb{P}[\tau \mid \hat{\sigma}_i]. \tag{2}$$

Consider the $\tau = a \succ c \succ b$ term. Notice that, by symmetry $\mathbb{P}[\mathcal{E}^{cb} \mid a \succ c \succ b] = \mathbb{P}[\mathcal{E}^{bc} \mid a \succ b \succ c]$ and $\mathbb{P}[\mathcal{E}^{bc} \mid a \succ c \succ b] = \mathbb{P}[\mathcal{E}^{cb} \mid a \succ b \succ c]$. Further $\mathbb{P}[a \succ c \succ b \mid \hat{\sigma}_i] = \varphi \cdot \mathbb{P}[a \succ b \succ c \mid \hat{\sigma}_i]$ as $d(a \succ c \succ b, \hat{\sigma}_i) = d(a \succ b \succ c, \hat{\sigma}_i) + 1$. Hence,

$$\begin{aligned}
&\left(\mathbb{P}[\mathcal{E}^{cb} \mid \tau = a \succ c \succ b] - \mathbb{P}[\mathcal{E}^{bc} \mid \tau = a \succ c \succ b]\right) \mathbb{P}[a \succ c \succ b \mid \hat{\sigma}_i] \\
= -\varphi \cdot &\left(\left(\mathbb{P}[\mathcal{E}^{cb} \mid \tau = a \succ b \succ c] - \mathbb{P}[\mathcal{E}^{bc} \mid \tau = a \succ b \succ c]\right) \mathbb{P}[a \succ b \succ c \mid \hat{\sigma}_i]\right),
\end{aligned}$$

or in words, the $\tau = a \succ c \succ b$ term is exactly equal to $-\varphi$ times the $\tau = a \succ b \succ c$ term. In fact, this same property holds for the other two pairs of ground truth rankings where a remains in the same position and b is swapped with c, so $b \succ a \succ c$ with $c \succ a \succ b$ and $b \succ c \succ a$ with $c \succ b \succ a$. Hence, we can write the entire expression (2) as

$$\begin{aligned}
(1-\varphi) \cdot \Bigg(&\left(\mathbb{P}[\mathcal{E}^{cb} \mid \tau = a \succ b \succ c] - \mathbb{P}[\mathcal{E}^{bc} \mid \tau = a \succ b \succ c]\right) \mathbb{P}[a \succ b \succ c \mid \hat{\sigma}_i] \\
&+ \left(\mathbb{P}[\mathcal{E}^{cb} \mid \tau = b \succ a \succ c] - \mathbb{P}[\mathcal{E}^{bc} \mid \tau = b \succ a \succ c]\right) \mathbb{P}[b \succ a \succ c \mid \hat{\sigma}_i] \\
&+ \left(\mathbb{P}[\mathcal{E}^{cb} \mid \tau = b \succ c \succ a] - \mathbb{P}[\mathcal{E}^{bc} \mid \tau = b \succ c \succ a]\right) \mathbb{P}[b \succ c \succ a \mid \hat{\sigma}_i]\Bigg).
\end{aligned}$$

Notice that since we wish to show this is strictly larger than 0 and $1 - \varphi > 0$, we show only that the sum of the probability terms is positive. Additionally, subbing in the values of $\mathbb{P}[\tau \mid \hat{\sigma}_i]$ with the corresponding Kendall tau distances, the above simplifies to

$$\begin{aligned}
&\left(\mathbb{P}[\mathcal{E}^{cb} \mid \tau = a \succ b \succ c] - \mathbb{P}[\mathcal{E}^{bc} \mid \tau = a \succ b \succ c]\right) \\
&+ \varphi \left(\mathbb{P}[\mathcal{E}^{cb} \mid \tau = b \succ a \succ c] - \mathbb{P}[\mathcal{E}^{bc} \mid \tau = b \succ a \succ c]\right) \\
&+ \varphi^2 \left(\mathbb{P}[\mathcal{E}^{cb} \mid \tau = b \succ c \succ a] - \mathbb{P}[\mathcal{E}^{bc} \mid \tau = b \succ c \succ a]\right).
\end{aligned} \tag{3}$$

We will now show that for some $c_1 > c_2$ to be chosen later, the first positive term $\mathbb{P}[\mathcal{E}^{cb} \mid \tau = a \succ b \succ c] \in \Omega(c_1^n)$ and each negative term $\mathbb{P}[\mathcal{E}^{bc} \mid \tau] \in O(c_2^n)$, which implies that for sufficiently large n, the entire sum is positive, as needed. In addition, for the result to hold with confident Mallows rather than unconfident, it is only required that the first difference be positive, i.e.,

$$\mathbb{P}[\mathcal{E}^{cb} \mid \tau = a \succ b \succ c] - \mathbb{P}[\mathcal{E}^{bc} \mid \tau = a \succ b \succ c] > 0.$$

This is also directly implied by showing the above bounds.

We relegate these arguments to the following two lemmas, established in Appendices B and C, respectively.

[3]Note that since f is randomized, to make this precise, we would need to specify the joint distribution of its outputs on different inputs. However, the remainder of the proof will not rely on how this is done, so the joint distribution can be arbitrary.

Lemma 3. $\mathbb{P}[\mathcal{E}^{bc} \mid \tau] \in O(c_2^n)$ *for* $c_2 = e^{\frac{1-\varphi^2}{2(1+\varphi+\varphi^2)}}\sqrt{\frac{\varphi(1+2\varphi)}{1+\varphi+\varphi^2}}$.

Lemma 4. $\mathbb{P}[\mathcal{E}^{cb} \mid \tau = a \succ b \succ c] \in \Omega(c_1^n)$ *for* $c_1 > e^{\frac{1-\varphi^2}{2(1+\varphi+\varphi^2)}}\sqrt{\frac{\varphi(1+2\varphi)}{1+\varphi+\varphi^2}}$.

Together, the two lemmas imply the desired result. □

We now complement this result with some additional robustness checks. First, even though the result is asymptotic, in special cases of interest, this is, in fact, not necessary.

Theorem 3. *With three candidates, when a voter has unconfident or confident Mallows beliefs with* $\varphi < 1$*, Borda Count is not OBIC for* any $n \geq 2$.

Notice that $n = 1$ is a degenerate case with no other voters, so $n \geq 2$ is the strongest we can hope for. The proof of Theorem 3 can be found in Appendix D. The beginning is nearly identical to the proof sketch of Theorem 2, but they diverge after this point. While Lemmas 3 and 4 are asymptotic in nature, the corresponding portion for Theorem 3 requires careful counting of the number of profiles satisfying different conditions to ensure that for any fixed n, the inequalities hold.

In light of Theorem 3, it may seem plausible that Theorem 2 could be strengthened to hold for all n rather than just asymptotically. However, we can give examples where this is not the case, suggesting that a "sufficiently large n" requirement may be necessary. In particular, there are scoring rules that, while not being plurality, are "close" to plurality in the sense that s_2 is so tiny it only matters when there is a tie among the plurality winners. For example, say we have the scoring rule $(4, 1, 0)$ with $n = 3$ voters. If any candidate receives two first-place votes, they immediately have 8 out of the 15 available points, so they necessarily win. Only when each candidate is ranked first by one voter is there any difference. We show in Appendix E that such close-to-plurality rules, at least for $n = 3$, are OBIC for confident Mallows beliefs with any $\varphi < 1$.

Finally, we consider other prominent non-scoring rules, namely Copeland and maximin. Note that for explicitly-defined beliefs and a number of voters n, we can determine whether or not a rule is OBIC by computing the probabilities of winners under all possible manipulations. We do so for the aforementioned rules under both confident and unconfident Mallows beliefs with $\varphi = 0.25, 0.5, 0.75$, $n = 2, \ldots, 50$, and $m = 3$. The results can be found in Table 2. Although slightly mixed in the sense that in a few specific cases, OBIC holds, the key takeaway is that none of the rules considered are as consistently OBIC as plurality.

	Copeland	maximin with $\varphi = 0.25, 0.5$	maximin with $\varphi = 0.75$
confident Mallows	even n	$n \neq 3$	$n \neq 3$
unconfident Mallows	all	all	$n \neq 6$

Table 2: Scenarios where the Copeland and maximin rules *fail* to be OBIC with respect to Confident and Unconfident Mallows beliefs with $\varphi = 0.25, 0.50, 0.75$, $n = 2, \ldots 50$, and $m = 3$.

5 Discussion

In summary, we have considered the problem of strategic voting when voters have certain correlated beliefs over others. We have singled out plurality as an auspicious choice, being incentive compatible for a large class of beliefs, and have complemented this with negative results showing other prominent voting rules do not satisfy this property. However, our work is certainly not the final word on this topic. The current negative results are only for three candidates, and although we believe they should extend to a larger number, the technical work in showing this seems to get quite messy. Further, although we have checked many prominent voting rules, we have not ruled out the existence of other "reasonable" rules that perform as well as plurality while simultaneously satisfying other desiderata.

Finally, taking a more practical viewpoint, although OBIC is a theoretically compelling condition, it is susceptible to common criticisms of models of voter behavior. As with many models of this form, the utility difference of misreporting, or of choosing any vote for that matter, can be very low, and it is debatable whether this is the driving force in how voters make decisions. It raises questions

similar in spirit to the so-called *Paradox of Voting* [6]: why would any rational agent choose to vote if the cost almost certainly outweighs the expected benefits? Despite these challenges, we do believe that the exploration and refinement of models such as OBIC can lead to an improved understanding of voter behavior and, ultimately, to the development of more effective voting systems.

References

[1] M. Bhargava, D. Majumdar, and A. Sen. Incentive-compatible voting rules with positively correlated beliefs. *Theoretical Economics*, 10(3):867–885, 2015.

[2] A. Bose and S. Roy. Correction to "Incentive-compatible voting rules with positively correlated beliefs". *Theoretical Economics*, 17(2):929–942, 2022.

[3] V. Conitzer and T. Walsh. Barriers to manipulation in voting. In F. Brandt, V. Conitzer, U. Endress, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 6. Cambridge University Press, 2016.

[4] T. M. Cover and J. A. Thomas. Information theory and statistics. *Elements of Information Theory*, 1(1):279–335, 1991.

[5] C. d'Aspremont and B. Peleg. Ordinal Bayesian incentive compatible representations of committees. *Social Choice and Welfare*, 5(4):261–279, 1988.

[6] A. Downs. An economic theory of political action in a democracy. *Journal of Political Economy*, 65(2):135–150, 1957.

[7] P. Faliszewski and A. D. Procaccia. AI's war on manipulation: Are we winning? *AI Magazine*, 31(4):53–64, 2010.

[8] A. Gibbard. Manipulation of voting schemes. *Econometrica*, 41:587–602, 1973.

[9] A. Gibbard. Manipulation of schemes that mix voting with chance. *Econometrica*, 45:665–681, 1977.

[10] A. Kahng, M. K. Lee, R. Noothigattu, A. D. Procaccia, and C.-A. Psomas. Statistical foundations of virtual democracy. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pages 3173–3182, 2019.

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

[12] R. D. Luce. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, 1959.

[13] D. Majumdar and A. Sen. Ordinally Bayesian incentive compatible voting rules. *Econometrica*, 72(2):523–540, 2004.

[14] D. Majumdar and A. Sen. Robust incentive compatibility of voting rules with positively correlated beliefs. *Social Choice and Welfare*, pages 1–33, 2021.

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

[16] D. Mandal and D. C. Parkes. Correlated voting. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 366–372, 2016.

[17] E. Mossel, A. D. Procaccia, and M. Z. Rácz. A smooth transition from powerlessness to absolute power. *Journal of Artificial Intelligence Research*, 48:923–951, 2013.

[18] F. Mosteller. Remarks on the method of paired comparisons: I. the least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, 16(1):3–9, 1951.

[19] R. Noothigattu, D. Peters, and A. D Procaccia. Axioms for learning from pairwise comparisons. In *Proceedings of the 33rd Annual Conference on Neural Information Processing Systems (NeurIPS)*, pages 17745–17754, 2020.

[20] R. Plackett. The analysis of permutations. *Applied Statistics*, 24:193–202, 1975.

[21] I. N. Sanov. *On the probability of large deviations of random variables*. United States Air Force, Office of Scientific Research, 1958.

[22] M. Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10: 187–217, 1975.

[23] L. L. Thurstone. A law of comparative judgment. *Psychological Review*, 34:273–286, 1927.

[24] V. Vitelli, Ø. Sørensen, M. Crispino, A. Frigessi Di Rattalma, and E. Arjas. Probabilistic preference learning with the Mallows rank model. *Journal of Machine Learning Research*, 18 (158):1–49, 2018.

[25] L. Xia. How likely a coalition of voters can influence a large election? arXiv:2202.06411, 2022.

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

Appendix

A Proof of Lemma 1

Fix a voter i, ranking $\hat{\sigma}_i$, and let $a = \hat{\sigma}_i(1)$ be their top choice. Fix an alternative $b \neq a$ and let $j^b = \hat{\sigma}_i^{-1}(b)$ be b's position in $\hat{\sigma}_i$. Let $\mathbf{r}$ be a vector of plurality scores with $r_a > r_b$ and let $\mathbf{r}'$ be the same vector but with the a and b components swapped.

Let $\mathbb{P}$ be an unconfident Mallows, Placket-Luce, or Thurstone-Mosteller belief. We show later how this proof directly implies it for the confident version as well. We abuse notation slightly, and write $\mathbb{P}[\mathbf{r} \mid \hat{\sigma}_i]$ to denote the probability of the event that $\boldsymbol{\sigma}_{-i}$ has plurality vector $\mathbf{r}$, and $\mathbb{P}[\mathbf{r} \mid \tau]$ for the same under ground truth τ. We wish to show that $\sum_\tau \mathbb{P}[\mathbf{r} \mid \tau] \cdot \mathbb{P}[\tau \mid \hat{\sigma}_i] \geq \sum_\tau \mathbb{P}[\mathbf{r}' \mid \tau] \cdot \mathbb{P}[\tau \mid \hat{\sigma}_i]$, or equivalently,

$$\sum_\tau (\mathbb{P}[\mathbf{r} \mid \tau] - \mathbb{P}[\mathbf{r}' \mid \tau]) \cdot \mathbb{P}[\tau \mid \hat{\sigma}_i] \geq 0.$$

Let τ be an arbitrary ground truth ranking with $a \succ_\tau b$ and let τ' be the same ranking, but with a and b switched. Notice that by symmetry, $\mathbb{P}[\mathbf{r} \mid \tau] = \mathbb{P}[\mathbf{r}' \mid \tau']$ and $\mathbb{P}[\mathbf{r}' \mid \tau] = \mathbb{P}[\mathbf{r} \mid \tau']$. Hence, in the above sum we can combine these two terms to be $(\mathbb{P}[\mathbf{r} \mid \tau] - \mathbb{P}[\mathbf{r}' \mid \tau])(\mathbb{P}[\tau \mid \hat{\sigma}_i] - \mathbb{P}[\tau' \mid \hat{\sigma}_i])$. We prove for all such τ with $a \succ_\tau b$, both of these terms are positive. Note that this immediately implies that this also holds for the confident version, as for that, we simply need to show $\mathbb{P}[\mathbf{r} \mid \tau] - \mathbb{P}[\mathbf{r}' \mid \tau]$ for $\tau = \hat{\sigma}_i$, and we have $a \succ_{\hat{\sigma}_i} b$.

We begin by showing $\mathbb{P}[\mathbf{r} \mid \tau] - \mathbb{P}[\mathbf{r}' \mid \tau] \geq 0$ for all τ with $a \succ_\tau b$. Fix such a τ. Notice that, conditioned on τ, all other rankings are drawn independently from the same distribution, namely, the corresponding model with ground truth τ. Let p_c be the probability that a ranking drawn from the corresponding model has top choice c. We can directly compute $\mathbb{P}[\mathbf{r} \mid \tau] = \binom{n-1}{\mathbf{r}} \prod_{c \in \mathcal{A}} p_c^{r_c}$ and $\mathbb{P}[\mathbf{r}' \mid \tau] = \binom{n-1}{\mathbf{r}'} \prod_{c \in \mathcal{A}} p_c^{r'_c}$, where $\binom{n-1}{\mathbf{r}}$ and $\binom{n-1}{\mathbf{r}'}$ are the multinomial coefficients, i.e., $\frac{(n-1)!}{\prod_{c \in \mathcal{A}} r_c!}$. To show the $\mathbb{P}[\mathbf{r} \mid \tau] \geq \mathbb{P}[\mathbf{r}' \mid \tau]$, observe that the two multinomial coefficients are equal as $\mathbf{r}$ and $\mathbf{r}'$ are the same up to swapping components. Further, since $r_c = r'_c$ for all $c \neq a, b$, the terms other than a and b are equal. Hence, all we need to show is that $p_a^{r_a} p_b^{r_b} \geq p_a^{r'_a} p_b^{r'_b}$. This will be directly implied by $p_a \geq p_b$.

For Mallow's, it is known that if $c = \tau(j)$, then the probability c is the highest rank is proportional to φ^j. Hence, since $\tau^{-1}(a) < \tau^{-1}(b)$, $p_a > p_b$. For Placket-Luce, observe that each p_c is proportional to w_c. Hence, $p_a > p_c$.

For Thurstone-Mosteller, things are more technical. Let $\mu_a > \mu_b$ be the corresponding means. We condition on arbitrary samples x_c for $c \neq a, b$, and show that even conditioned on this, the probability X_a is largest is greater than the probability that X_b is largest. Since the conditioning was arbitrary, the law of total probability tells us that this is true in general.

Let $x_c^{max} = \max_{c \neq a,b} x_c$. Then, integrating over the standard normal PDF, the probability that X_a is the largest is exactly

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{1}{2}(x-\mu_a)^2} e^{-\frac{1}{2}(y-\mu_b)^2} \mathbb{I}[x > \max(y, x_c^{max})] \, dx \, dy.$$

We can break up this integral depending on whether $X_b \geq x_c^{max}$ or not, to get that this is equal to

$$\frac{1}{2\pi} \Bigg(\int_{-\infty}^{x_c^{max}} \int_{x_c^{max}}^{\infty} e^{-\frac{1}{2}(x-\mu_a)^2} e^{-\frac{1}{2}(y-\mu_b)^2} \, dx \, dy + \int_{x_c^{max}}^{\infty} \int_{x_c^{max}}^{\infty} e^{-\frac{1}{2}(x-\mu_a)^2} e^{-\frac{1}{2}(y-\mu_b)^2} \mathbb{I}[x > y] \, dx \, dy \Bigg).$$

The same can be done symmetrically for X_b. To show the probability is larger for X_a, we show that each of the terms is bigger, i.e.,

$$\int_{-\infty}^{x_c^{max}} \int_{x_c^{max}}^{\infty} e^{-\frac{1}{2}(x-\mu_a)^2} e^{-\frac{1}{2}(y-\mu_b)^2} \, dx \, dy \geq \int_{-\infty}^{x_c^{max}} \int_{x_c^{max}}^{\infty} e^{-\frac{1}{2}(x-\mu_b)^2} e^{-\frac{1}{2}(y-\mu_a)^2} \, dx \, dy$$

and

$$\int_{x_c^{max}}^{\infty}\int_{x_c^{max}}^{\infty} e^{-\frac{1}{2}(x-\mu_a)^2}e^{-\frac{1}{2}(y-\mu_b)^2}\mathbb{I}[x>y]\,dx\,dy$$
$$\geq \int_{x_c^{max}}^{\infty}\int_{x_c^{max}}^{\infty} e^{-\frac{1}{2}(x-\mu_b)^2}e^{-\frac{1}{2}(y-\mu_a)^2}\mathbb{I}[x>y]\,dx\,dy.$$

Both of these inequalities are implied by the fact that for all fixed $x > y$,

$$e^{-\frac{1}{2}(x-\mu_a)^2}e^{-\frac{1}{2}(y-\mu_b)^2} > e^{-\frac{1}{2}(x-\mu_b)^2}e^{-\frac{1}{2}(y-\mu_b)^2}.$$

Note that this is equivalent to showing

$$-\frac{1}{2}(x-\mu_a)^2 + -\frac{1}{2}(y-\mu_b)^2 \geq -\frac{1}{2}(x-\mu_b)^2 + -\frac{1}{2}(y-\mu_b)^2.$$

Indeed, we have that

$$-\frac{1}{2}((x-\mu_a)^2+(y-\mu_b)^2) + \frac{1}{2}((x-\mu_b)^2+(y-\mu_b)^2) = x\mu_a + y\mu_b - x\mu_b - y\mu_a$$
$$= (x-y)(\mu_a-\mu_b).$$

Since $x > y$ and $\mu_a > \mu_b$, this is positive, as needed.

Next, we wish to show $\mathbb{P}[\tau \mid \hat{\sigma}_i] \geq \mathbb{P}[\tau' \mid \hat{\sigma}_i]$. Recall that by Baye's rule, these are each proportional to $\mathbb{P}[\hat{\sigma}_i \mid \tau]$ and $\mathbb{P}[\hat{\sigma}_i \mid \tau']$, where these are the probabilities of drawing $\hat{\sigma}_i$ from the corresponding model with ground truth τ and τ'. In the Mallows model, note $d(\hat{\sigma}_i, \tau) < d(\hat{\sigma}_i, \tau')$ because $a \succ_{\hat{\sigma}_i} b$, so swapping them can only increase the distance. Hence, $\mathbb{P}[\hat{\sigma}_i \mid \tau] \geq \mathbb{P}[\hat{\sigma}_i \mid \tau']$. For Placket-Luce, observe that the probability of generating a ranking σ is

$$\prod_{j=1}^{m} \frac{w_{\sigma(j)}}{\sum_{j'\geq j} w_{\sigma(j')}}.$$

Notice that even reordering the weights w, the product of the numerators is always $\prod_{c\in\mathcal{A}} w_c$. However, the denominators can change. Let w_a and w_b be the weights of a and b under τ, so $w_a > w_b$. The only difference between the denominators are those in terms with $j = 2, \ldots j^b$. Under τ these denominators include w_b while under τ', this is replaced with w_a. Hence, under τ', all the denominators are at least as large, and hence the overall probability is less.

Finally, we handle Thurstone-Mosteller. Notice that under both ground truths τ and τ', X_c for $c \neq a, b$ follow the same distributions. Hence, as before, we condition on values x_c for $c \neq a, b$. If we show conditioned on any values, it is more likely to generate $\hat{\sigma}_i$ under τ then τ', then we are done. We first restrict to x_c such that their order matches $\hat{\sigma}_i$, as otherwise, the probability of generating $\hat{\sigma}_i$ is 0. Again, let $x_c^{max} = \max_c x_c$. We now split into two cases based on if $j^b = 2$ or if $j^b > 2$. If $j^b = 2$. Then, the probability of generating $\hat{\sigma}_i$ under τ is

$$\int_{x_c^{max}}^{\infty}\int_{x_c^{max}}^{\infty} e^{-\frac{1}{2}(x-\mu_a)^2}e^{-\frac{1}{2}(y-\mu_b)^2}\mathbb{I}[x>y]\,dx\,dy.$$

Under τ', it is the same with μ_a and μ_b swapped. Similarly, when $j^b > 2$, then let $c^u = \hat{\sigma}_i(j^b - 1)$ and let $c^\ell = \hat{\sigma}_i(j^b + 1)$ be the candidates appearing directly before and after b in $\hat{\sigma}_i$. The probability here of generating $\hat{\sigma}_i$ under τ is

$$\int_{x_{c^\ell}}^{x_{c^u}}\int_{x_c^{max}}^{\infty} e^{-\frac{1}{2}(x-\mu_a)^2}e^{-\frac{1}{2}(y-\mu_b)^2}\mathbb{I}[x>y]\,dx\,dy.$$

Under τ', it is again the same with μ_a and μ_b swapped. The proof that the τ versions are larger than the τ' follow the identical argument to the earlier ones showing $p_a > p_b$. □

B Proof of Lemma 3

Recall that $\mathcal{E}^{bc}$ is the event that a wins in $(\boldsymbol{\sigma}_{-i}, \hat{\sigma}_i)$ but not $(\boldsymbol{\sigma}_{-i}, \sigma_i')$ where $\hat{\sigma}_i = abc$ and $\sigma_i' = acb$. Notice that in terms of scores, the only change when swapping from σ_i' to $\hat{\sigma}_i$ is that b has increased

by r_1 while c has decreased by r_1. We claim that a necessary condition on $\boldsymbol{\sigma}_{-i}$ such that the probability of a winning increases under this switch is that both $\text{SC}_c(\boldsymbol{\sigma}_{-i}, \sigma_i') \geq \text{SC}_a(\boldsymbol{\sigma}_{-i}, \sigma_i')$ and $\text{SC}_a(\boldsymbol{\sigma}_{-i}, \hat{\sigma}_i) \geq \text{SC}_b(\boldsymbol{\sigma}_{-i}, \hat{\sigma}_i)$. Indeed, if $\text{SC}_c(\boldsymbol{\sigma}_{-i}, \sigma_i') < \text{SC}_c(\boldsymbol{\sigma}_{-i}, \sigma_i')$, then even before the switch, c was not a winning candidate, so decreasing their score and increasing b's cannot improve a's chances. Further, if $\text{SC}_a(\boldsymbol{\sigma}_{-i}, \hat{\sigma}_i) < \text{SC}_b(\boldsymbol{\sigma}_{-i}, \hat{\sigma}_i)$, then a is winning with probability 0 on $(\boldsymbol{\sigma}_{-i}, \hat{\sigma}_i)$, so this cannot be an increase. Writing this only as a function of $\boldsymbol{\sigma}_{-i}$, we have that a necessary condition is that $\text{SC}_c(\boldsymbol{\sigma}_{-i}) + r_2 \geq \text{SC}_a(\boldsymbol{\sigma}_{-i}) + r_1$, $\text{SC}_a(\boldsymbol{\sigma}_{-i}) + r_1 \geq \text{SC}_b(\boldsymbol{\sigma}_{-i}) + r_2$, and (transitively from the previous two) $\text{SC}_c(\boldsymbol{\sigma}_{-i}) \geq \text{SC}_b(\boldsymbol{\sigma}_{-i})$.

We will show that for any $\tau = xyz$, $\mathbb{P}[\text{SC}_z(\boldsymbol{\sigma}_{-i}) \geq \text{SC}_x(\boldsymbol{\sigma}_{-i}) \mid \tau] \in O(c_2^n)$ (for c_2 to be chosen later). The above necessary conditions imply that this upper bounds each $\mathbb{P}[\mathcal{E}^{bc} \mid \tau]$ term.

To upperbound $\mathbb{P}[\text{SC}_z(\boldsymbol{\sigma}_{-i}) \geq \text{SC}_x(\boldsymbol{\sigma}_{-i}) \mid \tau]$, we will use a Chernoff bound. We begin by rewriting it as

$$\mathbb{P}[\text{SC}_z(\boldsymbol{\sigma}_{-i}) - \text{SC}_x(\boldsymbol{\sigma}_{-i}) \geq 0 \mid \tau] = \mathbb{P}\left[\sum_{j \neq i} \text{SC}_z(\sigma_j) - \text{SC}_x(\sigma_j) \geq 0 \;\middle|\; \tau\right].$$

Notice that conditioned on a ground truth τ, each σ_j (for $j \neq i$) is sampled independently from a Mallow's distribution around τ. Hence, if we write $X_j = \text{SC}_z(\sigma_j) - sc_x(\sigma_j)$, this is now the sum of independent random variables. To apply Chernoff, we will need that these are bounded between 0 and 1. As they are currently bounded in $[-r_1, r_1]$, we define $Y_j = \frac{X_j}{2r_1} + 1/2$, which is now bounded in $[0, 1]$. Hence, we wish to upperbound

$$\mathbb{P}\left[\frac{1}{n-1}\sum_{j \neq i} Y_j \geq 1/2 \;\middle|\; \tau\right]$$

To compute $\mathbb{E}[Y_j]$, we first compute $\mathbb{E}[X_j]$:

$$\begin{aligned}
\mathbb{E}[X_j] &= \sum_{\sigma \in \{xyz, xzy, yxz, zxy, yzx, zyx\}} (\text{SC}_z(\sigma) - \text{SC}_x(\sigma))\varphi^{d(\sigma,\tau)} \\
&= \frac{(0-r_1)\cdot 1 + (r_2 - r_1)\cdot\varphi + (0-r_2)\cdot\varphi + (r_1-r_2)\cdot\varphi^2 + (r_2-0)\cdot\varphi^2 + (r_1-0)\cdot\varphi^3}{1 + 2\varphi + 2\varphi^2 + \varphi^3} \\
&= \frac{(-1-\varphi+\varphi^2+\varphi^3)r_1 + (\varphi - \varphi - \varphi^2 + \varphi^2)r_2}{1 + 2\varphi + 2\varphi^2 + \varphi^3} \\
&= r_1 \cdot \frac{(1+\varphi)(\varphi^2-1)}{(1+\varphi)(1+\varphi+\varphi^2)} = r_1 \cdot \frac{\varphi^2 - 1}{1+\varphi+\varphi^2}.
\end{aligned}$$

From this we have that

$$\mathbb{E}[Y_j] = \frac{1}{2r_1}\mathbb{E}[X_j] + 1/2 = \frac{\varphi^2 - 1 + (1+\varphi+\varphi^2)}{2(1+\varphi+\varphi^2)} = \frac{\varphi(1+2\varphi)}{2(1+\varphi+\varphi^2)}.$$

We will use the form of the Chernoff bound that states that if each $W_1, \ldots W_k$ is i.i.d. drawn from a distribution supported on $[0, 1]$ with $\mathbb{E}[W_j] = \mu$, then

$$\Pr\left[\frac{1}{k}\sum_j W_j \geq (1+\delta)\mu\right] \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{k\mu} = \left(e^{(1+\delta)\mu - \mu}\left(\frac{\mu}{(1+\delta)\mu}\right)^{(1+\delta)\mu}\right)^k. \quad (4)$$

Notice that in our case, $k = n-1$, $\mu = \frac{\varphi(1+\varphi)}{2(1+\varphi+\varphi^2)}$, and $(1+\delta)\mu = 1/2$. Hence, plugging in our values, we get that this is at most

$$\left(e^{\frac{1-\varphi^2}{2(1+\varphi+\varphi^2)}}\sqrt{\frac{\varphi(1+2\varphi)}{1+\varphi+\varphi^2}}\right)^{n-1}.$$

Therefore, this quantity is $O(c_2^n)$ for $c_2 = e^{\frac{1-\varphi^2}{2(1+\varphi+\varphi^2)}}\sqrt{\frac{\varphi(1+2\varphi)}{1+\varphi+\varphi^2}}$. □

C Proof of Lemma 4

To lower bound $\mathbb{P}[\mathcal{E}^{cb} \mid \tau = abc]$, our strategy will be the following. First, we call a vector $\mathbf{h} = (h_\sigma)_{\sigma \in L}$ of integers indexed by $\mathcal{L}$ a *histogram*, and we will say that a profile $\boldsymbol{\sigma}$ has histogram $\mathbf{h}$ if $|\{i \mid \sigma_i = \sigma\}| = h_\sigma$. For all sufficiently large n, we will find histograms $(h_\sigma)_{\sigma \in L}$ with $\sum_{\sigma \in L} h_\sigma = n - 1$ such that on profiles $(\boldsymbol{\sigma}_{-i}, \sigma_i)$ with histogram $\mathbf{h}$, a is tied with b for the largest score, while on $(\boldsymbol{\sigma}_{-i}, \sigma_i')$, a is the unique winner. This implies that the probability a wins for such profiles increases by at least $1/2$. We will then show that the probability that $\boldsymbol{\sigma}_{-i}$ has the corresponding histogram h_σ is lower bounded by $\Omega(c_1^n)$.

To do this, we first must understand how likely it is to sample a profile with specific histogram $\mathbf{h}$. Let $p_\sigma = \varphi^{d(\sigma, a \succ b \succ c)}/Z$ be the probability of sampling σ from the Mallow's distribution. Notice that sampling $\boldsymbol{\sigma}_{-i}$ and considering the counts $|\{i \in \boldsymbol{\sigma}_{-i} \mid \sigma_i = \sigma\}|$ is equivalent to drawing from a multinomial distribution over the alphabet $\mathcal{L}$ with probabilities $(p_\sigma)_{\sigma \in \mathcal{L}}$ of size $n-1$. If we write $q_\sigma = h_\sigma/(n-1)$ as the proportion of voters with σ, it is known that the probability of observing $(h_\sigma)_{\sigma \in \mathcal{L}}$ (with each $h_\sigma > 0$) is at least $\left(\prod_\sigma \left(\frac{p_\sigma}{q_\sigma}\right)^{q_\sigma} - o(1)\right)^{n-1}$. Note that $\prod_\sigma \left(\frac{p_\sigma}{q_\sigma}\right)^{q_\sigma} = 1/e^{D_{KL}(\mathbf{p} \| \mathbf{q})}$ where D_{KL} is the KL-divergence and $\mathbf{p}$ and $\mathbf{q}$ are treated as probability distributions over $\mathcal{L}$. This is essentially (without uniform convergence) an immediate consequence of the tightness of Sanov's theorem [21], although it can easily be derived by known bounds on multinomial coefficients [4].

With this property in hand, we now wish to find profiles satisfying the tie conditions such that $\left(\frac{p_\sigma}{q_\sigma}\right)^{q_\sigma}$ is bounded away from $e^{\frac{1-\varphi^2}{2(1+\varphi+\varphi^2)}}\sqrt{\frac{\varphi(1+2\varphi)}{1+\varphi+\varphi^2}}$. To that end, we now show the following:

Lemma 5. *For all $\varphi \leq .988$ and positional scoring rules $(r_1, r_2, 0)$, there exists real numbers $(q_\sigma)_{\sigma \in L}$ such that:*

1. *They are valid proportions: $\sum_\sigma q_\sigma = 1$ and each $q_\sigma > 0$.*
2. *Candidates a and b are tied in score: $\sum_\sigma \mathrm{SC}_a(\sigma) q_\sigma = \sum_\sigma sc_b(\sigma) q_\sigma$.*
3. *Candidate c is not beating a and b: $\sum_\sigma \mathrm{SC}_a(\sigma) q_\sigma \geq \sum_\sigma sc_c(\sigma) q_\sigma$.*
4. *The objective of these q's are large $\prod_\sigma \left(\frac{p_\sigma}{q_\sigma}\right)^{q_\sigma} > e^{\frac{1-\varphi^2}{2(1+\varphi+\varphi^2)}}\sqrt{\frac{\varphi(1+2\varphi)}{1+\varphi+\varphi^2}}$.*

Proof. We first handle the case where $\varphi \leq 0.1$. Under this assumption of φ, we can explicitly choose q_σ as follows.

$$q_{abc} = q_{bac} = \frac{\sqrt{p_{abc}p_{bac}}}{2(\sqrt{p_{abc}p_{bac}} + \sqrt{p_{acb}p_{bca}} + \sqrt{p_{cab}p_{cba}})}$$
$$q_{acb} = q_{bca} = \frac{\sqrt{p_{acb}p_{bca}}}{2(\sqrt{p_{abc}p_{bac}} + \sqrt{p_{acb}p_{bca}} + \sqrt{p_{cab}p_{cba}})}$$
$$q_{cab} = q_{cba} = \frac{\sqrt{p_{cab}p_{cba}}}{2(\sqrt{p_{abc}p_{bac}} + \sqrt{p_{acb}p_{bca}} + \sqrt{p_{cab}p_{cba}})}.$$

Since all p_σ are positive, each q_σ is positive. Further, they are explicitly chosen to add up to one. In addition, due to the symmetry between a and b (they appear in each position at the same frequency), their corresponding scores are equal. Finally, since $p_{abc} > p_{bac} \geq p_{acb} > p_{bca} \geq p_{cab} > p_{cba}$, it follows that $q_{abc} = q_{bac} > q_{acb} = q_{bca} > q_{cab} = q_{cba}$, so the score of c is strictly less than the score of a. It remains to be shown that $\prod_\sigma \left(\frac{p_\sigma}{q_\sigma}\right)^{q_\sigma} > e^{\frac{1-\varphi^2}{2(1+\varphi+\varphi^2)}}\sqrt{\frac{\varphi(1+2\varphi)}{1+\varphi+\varphi^2}}$. Let $d = 2(\sqrt{p_{abc}p_{bac}} + \sqrt{p_{acb}p_{bca}} + \sqrt{p_{cab}p_{cba}})$ be the denominator in each of the q values. Let us consider the contribution to the product of the abc and bac terms. We have,

$$\left(\frac{p_{abc}}{q_{abc}}\right)^{q_{abc}} \cdot \left(\frac{p_{bac}}{q_{bac}}\right)^{q_{bac}} = \left(\frac{dp_{abc}}{\sqrt{p_{abc}p_{bca}}}\right)^{q_{abc}} \cdot \left(\frac{dp_{bac}}{\sqrt{p_{abc}p_{bca}}}\right)^{q_{bac}}$$
$$= d^{q_{abc}+q_{bac}} \cdot \left(\frac{p_{abc}}{\sqrt{p_{abc}p_{bca}}} \cdot \frac{p_{bac}}{\sqrt{p_{abc}p_{bca}}}\right)^{q_{abc}}$$

$$= d^{q_{abc}+q_{bca}} \cdot 1$$

The same argument holds for the other two pairs, which implies that

$$\prod_{\sigma} \left(\frac{p_\sigma}{q_\sigma}\right)^{q_\sigma} = d^{\sum_\sigma q_\sigma} = d.$$

Expanding the value of d,

$$\begin{aligned} 2\left(\sqrt{\frac{\varphi^0 \cdot \varphi^1}{Z^2}} + \sqrt{\frac{\varphi^1 \cdot \varphi^2}{Z^2}} + \sqrt{\frac{\varphi^2 \cdot \varphi^3}{Z^2}}\right) &= \frac{2\sqrt{\varphi}(1+\varphi+\varphi^2)}{Z} \\ &= \frac{2\sqrt{\varphi}(1+\varphi+\varphi^2)}{1+2\varphi+2\varphi^2+\varphi^3} \\ &= \frac{2\sqrt{\varphi}}{1+\varphi}. \end{aligned}$$

Finally, using the assumption that $\varphi \leq .1$, we have

$$\begin{aligned} \frac{2\sqrt{\varphi}}{1+\varphi} &\geq \frac{2\sqrt{\varphi}}{1.1} \\ &= \sqrt{e \cdot 1.2} \cdot \sqrt{\varphi} \\ &\geq e^{1/2} \cdot \sqrt{\varphi(1+2\varphi)} \\ &> (e^{1/2})^{\frac{1-\varphi^2}{1+\varphi+\varphi^2}} \cdot \frac{1}{\sqrt{1+\varphi+\varphi^2}} \cdot \sqrt{\varphi(1+2\varphi)} \\ &= e^{\frac{1-\varphi^2}{2(1+\varphi+\varphi^2)}} \sqrt{\frac{\varphi(1+2\varphi)}{1+\varphi+\varphi^2}}, \end{aligned}$$

where the second inequality uses the fact that $2/1.1 \approx 1.82 > \sqrt{1.2e} \approx 1.81$.

Next, we consider $\varphi > 0.1$. We formalize finding valid qs in the following form. Notice first that we can rescale the scoring vector to be of the form $(1, \alpha, 0)$ where $\alpha = r_2/r_1 \in [0,1]$. We will use $\mathrm{sc}_x^\alpha(\sigma)$ to denote the score of candidate x on ranking σ with the positional scoring rule $(1,\alpha,0)$. Let Q_α be the set of vectors $\mathbf{q}$ (indexed by $\mathcal{L}$), which satisfy the constraints for a specific α. Expanding the objective in terms of φ, let $d_\sigma = d(\sigma, a \succ b \succ c)$, $f(\varphi, \mathbf{q}) = \frac{1}{1+2\varphi+2\varphi^2+\varphi^3} \prod_\sigma \left(\frac{\varphi^{d_\sigma}}{q_\sigma}\right)^{q_\sigma}$, and $\ell(\varphi) = e^{\frac{1-\varphi^2}{2(1+\varphi+\varphi^2)}} \sqrt{\frac{\varphi(1+2\varphi)}{1+\varphi+\varphi^2}}$. Let $g(\varphi, \mathbf{q}) = f(\varphi, \mathbf{q}) - \ell(\varphi)$. Our goal is to show that for all $\varphi \in (.1, .99]$ and for all $\alpha \in [0,1]$, there is a $\mathbf{q} \in Q_\alpha$ such that $g(\varphi, \mathbf{q}) > 0$. When $\mathbf{q}$ satisfies this, we will say that $\mathbf{q}$ is a *solution* for φ and α.

To that end, we will first show using the smoothness of g and the Q_α sets that as long as a solution $\mathbf{q}$ for a specific φ and α satisfies reasonable conditions, then that will imply the existence of solutions for nearby φ and α. We will then present several solutions found using a computational search that cover the α and φ region, implying the existence of solutions for all necessary values.

Fix α, φ, and suppose we have a corresponding solution $\mathbf{q}$. Fix some $\varepsilon > 0$, we now find sufficient conditions such that for all $\varphi' \in [\varphi - \varepsilon, \varphi + \varepsilon]$ and $\alpha' \in [\alpha - \varepsilon, \alpha + \varepsilon]$, there exists a solution $\mathbf{q}'$ for φ' and ε'. We begin by extending it to the same α, but for $\varphi' \in [\varphi - \varepsilon, \varphi + \varepsilon]$. We first show that ℓ is an increasing function on $[0,1]$ which implies (as long as $\varphi + \varepsilon \leq 1$), on $[\varphi - \varepsilon, \varphi + \varepsilon]$, it is upper bounded by $\ell(\varphi + \varepsilon)$.

Indeed, notice that μ (from (4)) is equal to $\frac{\varphi+2\varphi^2}{2(1+\varphi+\varphi^2)} = 1/2 - \frac{1-\varphi^2}{2(1+\varphi+\varphi^2)}$and its derivative with respect to φ is $\frac{\varphi^2+4\varphi+1}{(\varphi^2+\varphi+1)^2} > 0$. Therefore, it is an increasing function of φ bounded in $[0, 1/2]$. As a function of μ, $\ell(\varphi)$ is equal to $(e/2)^{1/2} e^{-\mu}\sqrt{\mu}$. The derivative of this with respect to μ is $(e/2)^{1/2}e^{-\mu}(1-2\mu)/(2\sqrt{\mu})$, positive for $\mu \in [0, 1/2)$. Therefore, as the composition of two increasing functions, $\ell(\varphi)$ is increasing on $[0,1]$.

Next, we wish to lower bound $f(\varphi', \mathbf{q})$. To do this, suppose the derivative $\frac{\partial f}{\partial \varphi}(\varphi', \mathbf{q})$ for $\varphi' \in [\varphi - \varepsilon, \varphi + \varepsilon]$ lower bounded by $B \leq 0$. Notice that $-B$ upper bounds the rate at which f can

decrease, so we get that $f(\varphi', \mathbf{q}) \geq f(\varphi - \varepsilon, \mathbf{q}) + 2\varepsilon B$. To compute such a B, we first compute $\frac{\partial f}{\partial \varphi}(\varphi', \mathbf{q})$. We will use the fact that $\frac{\partial f}{\partial \varphi}(\varphi', \mathbf{q}) = \frac{\partial \log(f)}{\partial \varphi}(\varphi', \mathbf{q}) \cdot f(\varphi', \mathbf{q})$. Since $\log(f(\varphi', \mathbf{q})) = \sum_\sigma q_\sigma(d_\sigma \log(\varphi') - q\sigma) - \log(1 + 2\varphi' + 2\varphi'^2 + \varphi'^3)$, we have that

$$\frac{\partial f}{\partial \varphi}(\varphi', \mathbf{q}) = \left(\frac{\sum_\sigma q_\sigma d_\sigma}{\varphi'} - \frac{2 + 4\varphi' + 3\varphi'^2}{1 + 2\varphi' + 2\varphi'^2 + \varphi'^3} \right) \cdot f(\varphi', \mathbf{q}).$$

Notice that $\frac{\sum_\sigma q_\sigma d_\sigma}{\varphi'}$ is decreasing in φ'. Further, we can also show that $\frac{2+4\varphi'+3\varphi'^2}{1+2\varphi'+2\varphi'^2+\varphi'^3}$ is decreasing, as its derivative is

$$-\frac{\varphi'(3\varphi'^3 + 8\varphi'^2 + 8\varphi' + 2)}{(\varphi'^3 + 2\varphi'^2 + 2\varphi' + 1)^2},$$

negative for all positive values of φ'. Finally, notice that f is defined as $1/e^{D_{KL}(\mathbf{q}\|\mathbf{p})}$ and D_{KL} is nonnegative, f is upperbounded by 1. Hence, for all $\varphi' \in [\varphi - \varepsilon, \varphi + \varepsilon]$,

$$\frac{\partial f}{\partial \varphi}(\varphi', \mathbf{q}) \geq \min\left(\frac{\sum_\sigma q_\sigma d_\sigma}{\varphi + \varepsilon} - \frac{2 + 4(\varphi - \varepsilon) + 3(\varphi - \varepsilon)^2}{1 + 2(\varphi - \varepsilon) + 2(\varphi - \varepsilon)^2 + (\varphi - \varepsilon)^3}, 0 \right).$$

Let $B(\varphi, \mathbf{q}, \varepsilon) = \min\left(\frac{\sum_\sigma q_\sigma d_\sigma}{\varphi+\varepsilon} - \frac{2+4(\varphi-\varepsilon)+3(\varphi-\varepsilon)^2}{1+2(\varphi-\varepsilon)+2(\varphi-\varepsilon)^2+(\varphi-\varepsilon)^3}, 0 \right)$. We then have that for all $\varphi' \in [\varphi - \varepsilon, \varphi + \varepsilon]$, $g(\varphi', q) \geq f(\varphi - \varepsilon, \mathbf{q}) + 2\varepsilon B(\varphi, \mathbf{q}, \varepsilon) - \ell(\varphi + \varepsilon, \mathbf{q})$.

Next, we consider modifying α to $\alpha' \in [\alpha - \varepsilon, \alpha + \varepsilon]$. Let $\beta = \alpha' - \alpha$. Notice that the current $\mathbf{q}$ may not be an element of $Q_{\alpha'}$. Although $\sum_\sigma q_\sigma = 1$, and each $q_\sigma > 0$, it may not be the case $\sum_\sigma \text{SC}_b^{\alpha'}(\sigma) q_\sigma = \sum_\sigma \text{SC}_a^{\alpha'}(\sigma) q_\sigma$. Instead, we have that $\sum_\sigma \text{SC}_b^{\alpha'}(\sigma) q_\sigma + \beta(q_{abc} + q_{cba}) = \sum_\sigma \text{SC}_a^{\alpha'}(\sigma) q_\sigma + \beta(q_{bac} + q_{cab})$. Let $r = q_{abc} + q_{cba} - q_{bac} - q_{cab}$; this is the current amount b is beating a by (it may be negative). Notice that we can find a $\mathbf{q}' \in Q_{\alpha'}$ by simply shifting $r/2 \cdot \beta$ mass from q_{acb} to q_{bca}. This will result in a valid $\mathbf{q}'$ as long as $q_{acb} > r/2 \cdot \beta$ when $r/2 \cdot \beta$ is positive or $q_{bca} > -r/2 \cdot \beta$ when it is negative. A sufficient condition for this is that both $q_{acb} > |r\varepsilon/2|$ and $q_{bca} > |r\varepsilon/2|$. Under this assumption, we now consider the effect on the solution value $g(\varphi, \mathbf{q}')$. To do this, we can consider the directional derivative of g with respect to increasing q_{acb} and decreasing q_{bca}. We have that for each σ,

$$\frac{\partial g}{\partial q_\sigma} = f(\varphi, \mathbf{q}) \cdot \left(\log\left(\frac{\varphi^{d_\sigma}}{q_\sigma} \right) - 1 \right).$$

Therefore, the derivative with respect the the vector of increasing q_{acb} and dcreasing q_{bca} is

$$\frac{\partial g}{\partial q_{acb}} - \frac{\partial g}{\partial q_{bca}} = f(\varphi, \mathbf{q}) \cdot \left(\log(\varphi^{d_{acb} - d_{bca}}) + \log\left(\frac{q_{bca}}{q_{acb}} \right) \right) = f(\varphi, \mathbf{q}) \cdot \left(\log\left(\frac{q_{bca}}{q_{acb}} \right) - \log(\varphi) \right).$$

We will now upperbound the magnitude of this. Recall that f is upper bounded by 1. Further, for any $\varphi' \in [\varphi + \varepsilon, \varphi - \varepsilon]$ and $\mathbf{q}'$ constructed by shifting at most $r\varepsilon/2$ mass between q_{acb} and q_{bca},

$$\log\left(\frac{q_{bca} - |r\varepsilon/2|}{q_{acb} + |r\varepsilon/2|} \right) - \log(\varphi + \varepsilon) \leq \log\left(\frac{q_{bca}}{q_{acb}} \right) - \log(\varphi) \leq \log\left(\frac{q_{bca} + |r\varepsilon/2|}{q_{acb} - |r\varepsilon/2|} \right) - \log(\varphi - \varepsilon).$$

Hence, the magnitude of the derivative is always at most:

$$\max\left(\left| \log\left(\frac{q_{bca} - |r\varepsilon/2|}{q_{acb} + |r\varepsilon/2|} \right) - \log(\varphi + \varepsilon) \right|, \left| \log\left(\frac{q_{bca} + |r\varepsilon/2|}{q_{acb} - |r\varepsilon/2|} \right) - \log(\varphi - \varepsilon) \right| \right).$$

Let $m(\varphi, \mathbf{q}, \varepsilon)$ be this value. Then, from shifting the at most $r\varepsilon/2$ mass between q_{acb} and q_{bca}, this decreases $g(\varphi, \mathbf{q})$ by at most $r\varepsilon/2 \cdot m(\varphi, \mathbf{q}, \varepsilon)$. Hence, putting this all together, we have that for any vector $\mathbf{q}$, as long as both $q_{acb}, q_{bca} > r\varepsilon/2$, and as long as

$$f(\varphi - \varepsilon, \mathbf{q}) + 2\varepsilon B(\varphi, \mathbf{q}, \varepsilon) - \ell(\varphi + \varepsilon, \mathbf{q}) - \frac{r\varepsilon}{2} m(\varphi, \mathbf{q}, \varepsilon) > 0,$$

then this implies that for all $\varphi' \in [\varphi - \varepsilon, \varphi + \varepsilon]$ and $\alpha' \in [\alpha - \varepsilon, \alpha + \varepsilon]$, there exists a solution $\mathbf{q}'$.

Finally, for all $0.1 \leq \varphi \leq .988$ and $0 \leq \alpha \leq 1$ that are multiples of $1/1000$, we compute corresponding $\mathbf{q}$ that satisfy the above conditions with $\varepsilon = 1/2000$. Together, these cover the space of φ and α, which implies that the lemma holds. This can be done (approximately enough) using a convex program to find $\mathbf{q}$ that maximizes f given φ and α. The computed values can be found in the supplementary material. □

Notice that the solutions $(q_\sigma)_{\sigma\in\mathcal{L}}$ from Lemma 5 need not be rational which would be necessary for a valid profile with corresponding $(h_\sigma)_{\sigma\in\mathcal{L}}$ to be sampled. However, we claim that given a non-rational solution, we can always find a rational one, so it is without loss of generality to assume they are. Notice that since the strict inequalities are all continuous functions of q, so there must be an $\varepsilon > 0$ such that all q vectors in an ε-ball around these qs (in $\mathbb{R}^6$) satisfy the strict inequalities. In addition, the linear equalities form an affine subspace. Since all coefficients are rational, all-rational vectors are dense within this subspace. Hence, there are rational $(q'_\sigma)_{\sigma\in\mathcal{L}}$ within ε of $(q_\sigma)_{\sigma\in L}$ that satisfies the equalities and is, therefore, a rational solution to the four properties.

Using rational $\mathbf{q}$, we can find a corresponding integral $\mathbf{h}$ such that on profiles with ranking counts equal to $\mathbf{h}$, a and b are tied for winning. Let $s = \sum_\sigma h_\sigma$ be the number of rankings in $\mathbf{h}$. For a ranking σ, let $\mathbf{e}_\sigma$ be the unit vector with 1 in the σ coordinate and 0 elsewhere. Notice that if $n - 1 = ks + 1$ for some integer k and $\boldsymbol{\sigma}_{-i}$ has ranking counts equal to $k\mathbf{h} + \mathbf{e}_{bac}$, then it is indeed the case that on $(\boldsymbol{\sigma}_{-i}, a \succ b \succ c)$, a is tied with b, while on $(\boldsymbol{\sigma}_{-i}, a \succ c \succ b)$, a is the unique winner.

To handle cases where $n - 2$ is not a multiple of s, suppose we write $n - 1 = k \cdot h + 1 + r$ where $2 \le r \le s + 1$. If r is odd, we can first add a cycle $\mathbf{e}_{abc} + \mathbf{e}_{bca} + \mathbf{e}_{cab}$ which does not affect relative scores. After doing this, we can add $r/2$ (or $(r-3)/2$ if r was odd) copies of $\mathbf{e}_{abc} + \mathbf{e}cab$ which again keeps a and b at the same relative scores and only pushes c down. By doing this, we can get a histogram of arbitrary size where $(\boldsymbol{\sigma}_{-i}, a \succ b \succ c)$ has a tied with b and $(\boldsymbol{\sigma}_{-i}, a \succ c \succ b)$ has a as a unique winner. Finally, notice that as n grows large, the proportion of this histogram approaches $\mathbf{q}$. Hence, for sufficiently large n, the probability of sampling this histogram will be $\Omega(c_1^n)$ for any $c_1 < \prod_\sigma \left(\frac{p_\sigma}{q_\sigma}\right)^{q_\sigma}$. Since $\prod_\sigma \left(\frac{p_\sigma}{q_\sigma}\right)^{q_\sigma} > c_2$, we can choose $c_1 > c_2$. This completes the proof. □

D Proof of Theorem 3

Consider the Borda scoring rule $(2, 1, 0)$, and a voter i with unconfident Mallows belief $\mathbb{P}$ with $\varphi < 1$. As usual, we will describe how to extend it to confident Mallows later. The proof begins identically to Theorem 2, up to the point of needing to show (3) is nonnegative. We restate (3) here for convenience.

$$\begin{aligned}
&\left(\mathbb{P}[\mathcal{E}^{cb} \mid \tau = a \succ b \succ c] - \mathbb{P}[\mathcal{E}^{bc} \mid \tau = a \succ b \succ c]\right) \\
&+ \varphi\left(\mathbb{P}[\mathcal{E}^{cb} \mid \tau = b \succ a \succ c] - \mathbb{P}[\mathcal{E}^{bc} \mid \tau = b \succ a \succ c]\right) \\
&+ \varphi^2\left(\mathbb{P}[\mathcal{E}^{cb} \mid \tau = b \succ c \succ a] - \mathbb{P}[\mathcal{E}^{bc} \mid \tau = b \succ c \succ a]\right).
\end{aligned}$$

Here, we show each of the probability differences are nonnegative, and the first is strictly positive. This also implies that the result holds for confident Mallows where only the first strict inequality is necessary.

To do this, we provide an equivalent way of computing $\mathbb{P}[\mathcal{E}^{cb} \mid \tau] - \mathbb{P}[\mathcal{E}^{bc} \mid \tau]$. Let us consider the profiles $\boldsymbol{\sigma}_{-i}$ where swapping from $a \succ b \succ c$ to $a \succ c \succ b$ leads to an increase in the probability a wins. Notice that the swap decreases the score of b by 1 and increases the score of c by 1. For this to help a win, b must have been one of the winners before. Therefore, one of the following must hold.

1. On $(\boldsymbol{\sigma}_{-i}, \sigma_i)$, a was tied with b with c being at least two behind them. Then, after the swap, a wins outright, an increase in winning probability of $1/2$.
2. On $(\boldsymbol{\sigma}_{-i}, \sigma_i)$, b was winning outright, a was one point behind, and c was more than one point behind a, then, after the swap, a and b are tied winners, an increase in winning probability of $1/2$.
3. On $(\boldsymbol{\sigma}_{-i}, \sigma_i)$, b was winning outright, a was one point behind, and c was one point behind a, then, after the swap, all three are tied, an increase of winning probability of $1/3$.

We define sets A_1, A_2, A_3 of profiles $\boldsymbol{\sigma}_{-i}$ that correspond to these three events. More formally,

$$\begin{aligned}
A_1 &= \{\boldsymbol{\sigma}_{-i} \mid \textsc{sc}_b(\boldsymbol{\sigma}_{-i}) = \textsc{sc}_a(\boldsymbol{\sigma}_{-i}) + 1 \ge \textsc{sc}_c(\boldsymbol{\sigma}_{-i}) + 1\}, \\
A_2 &= \{\boldsymbol{\sigma}_{-i} \mid \textsc{sc}_b(\boldsymbol{\sigma}_{-i}) = \textsc{sc}_a(\boldsymbol{\sigma}_{-i}) + 2 \ge \textsc{sc}_c(\boldsymbol{\sigma}_{-i}) + 2\},
\end{aligned}$$

$$A_3 = \{\boldsymbol{\sigma}_{-i} \mid \text{SC}_b(\boldsymbol{\sigma}_{-i}) = \text{SC}_a(\boldsymbol{\sigma}_{-i}) + 2 = \text{SC}_c(\boldsymbol{\sigma}_{-i}) + 1\}.$$

We can analogously define B_1, B_2, and B_3 with b and c swapped, which correspond to profiles where swapping causes the probability of a winning to decrease. From this, we get that

$$\begin{aligned} P[\mathcal{E}^{cb} \mid \tau] - \mathbb{P}[\mathcal{E}^{bc} \mid \tau] =& \frac{1}{2}\mathbb{P}[A_1 \mid \tau] + \frac{1}{2}\mathbb{P}[A_2 \mid \tau] + \frac{1}{3}\mathbb{P}[A_3 \mid \tau] \\ & - \left(\frac{1}{2}\mathbb{P}[B_1 \mid \tau] + \frac{1}{2}\mathbb{P}[B_2 \mid \tau] + \frac{1}{3}\mathbb{P}[B_3 \mid \tau]\right) \end{aligned}$$

Further, notice that there is a natural bijection π between the sets of profiles A_k and B_k for $k \leq 3$, namely, swapping every occurrence of b with c and vice-versa.

To prove a weak inequality, we will show that for each k and each τ, $\mathbb{P}[A_k \mid \tau] \geq \mathbb{P}[B_k \mid \tau]$. Notice that this is simply a statement about draws of profiles from a Mallows model; voter i and their report do not have an impact. To make calculations less messy we will simply refer to these profiles without i as $\boldsymbol{\sigma}$ instead of $\boldsymbol{\sigma}_{-i}$ and refer to the set of voters V as the ones without i, and the size of these profiles as n (even though this is technically $n-1$). This means that our assumption now is that $n \geq 1$.

Next, as a simplifying step, fix an arbitrary partition of the voters K_1, K_2, K_3, i.e., $V = K_1 \sqcup K_2 \sqcup K_3$. Let

$$A_k^{K_1,K_2,K_3} = \{\boldsymbol{\sigma}_{-i} \in A_k \mid \sigma_j(1) = a, \forall j \in K_1 \wedge \sigma_j(2) = a, \forall j \in K_2 \wedge \sigma_j(3) = a \forall j \in K_3\}.$$

In words $A_k^{K_1,K_2,K_3}$ is the subset of A_k such that the voters in K_1 rank a first, voters in K_2 rank a second, and voters in K_3 rank a third. We define this analogously for the B_k sets. We will show for all partitions K_1, K_2, K_3, $\mathbb{P}[A_k^{K_1,K_2,K_3} \mid \tau] \geq \mathbb{P}[B_k^{K_1,K_2,K_3} \mid \tau]$ which implies it holds for the original sets.

Fix an arbitrary K_1, K_2, K_3 and $k \leq 3$. Writing this out more explicitly and using the π bijection, we see that it suffices to show for each τ,

$$\sum_{\boldsymbol{\sigma} \in A_k^{K_1,K_2,K_3}} (\varphi^{d(\boldsymbol{\sigma},\tau)} - \varphi^{d(\pi(\boldsymbol{\sigma}),\tau)}) \geq 0. \tag{5}$$

We assume now that $A_k^{K_1,K_2,K_3} \neq \emptyset$ as otherwise this inequality trivially holds.

Notice that for all $\boldsymbol{\sigma} \in A_k^{K_1,K_2,K_3}$,

$$\text{SC}_a(\boldsymbol{\sigma}) = 2|K_1| + |K_2| \tag{6}$$

In other words, the score of a on all profiles in $A_k^{K_1,K_2,K_3}$ is constant. From this, we can derive the scores of the other candidates.

$$\text{SC}_b(\boldsymbol{\sigma}) = \text{SC}_a(\boldsymbol{\sigma}) + \kappa = 2|K_1| + |K_2| + \kappa. \tag{7}$$

where $\kappa = 1, 2$ depending on whether $k = 1$ or $k \in \{2, 3\}$. Finally, for all $\boldsymbol{\sigma}$, $\text{SC}_a(\boldsymbol{\sigma}) + \text{SC}_b(\boldsymbol{\sigma}) + \text{SC}_c(\boldsymbol{\sigma}) = 3n$. Therefore,

$$\text{SC}_a(\boldsymbol{\sigma}) = 3n - \text{SC}(a) - sc(b) = 3n - 4|K_1| - 2|K_2| - \kappa. \tag{8}$$

Further, these equations are an equivalent condition for defining $A_k^{K_1,K_2,K_3}$, a profile $\boldsymbol{\sigma} \in A_k^{K_1,K_2,K_3}$ if and only if the voters in each of K_1, K_2, and K_3 rank a accordingly and Equations (6) to (8) are all satisfied.

Additionally, we have that for any $\boldsymbol{\sigma} \in A_1 \cup A_2 \cup A_3$, $\text{SC}_c(\boldsymbol{\sigma}) \leq \text{SC}_b(\boldsymbol{\sigma}) - 1$. Therefore, by the assumption that $A_k^{K_1,K_2,K_3}$ was nonempty, we can derive some constraints on $|K_1|, |K_2|$, and $|K_3|$. Namely, for any $\boldsymbol{\sigma} \in A_k^{K_1,K_2,K_3}$,

$$\begin{aligned} 3(|K_1| + |K_2| + |K_3|) &= 3n \\ &= \text{SC}_a(\boldsymbol{\sigma}) + \text{SC}_b(\boldsymbol{\sigma}) + \text{SC}_c(\boldsymbol{\sigma}) \end{aligned}$$

$$= \text{SC}_a(\boldsymbol{\sigma}) + 2\text{SC}_b(\boldsymbol{\sigma}) - 1$$
$$= 6|K_1| + 3|K_2| + 2\kappa - 1.$$

Therefore,

$$|K_1| \geq |K_3| - \frac{2\kappa - 1}{3}. \tag{9}$$

Recall that for a pair of candidates x and y, $N_{xy}(\boldsymbol{\sigma}) = |\{i | x \succ_i y\}|$. Note that $N_{bc}(\boldsymbol{\sigma}) = n - N_{bc}(\pi(\boldsymbol{\sigma}))$ since all occurrences of b and c are swapped. It can be shown that for Borda scores,

$$\text{SC}_x(\boldsymbol{\sigma}) = \sum_{y \neq x} N_{xy}(\boldsymbol{\sigma}) \tag{10}$$

In addition, by the definition of the Kendall tau distance,

$$d(\boldsymbol{\sigma}, xyz) = N_{yx}(\boldsymbol{\sigma}) + N_{zx}(\boldsymbol{\sigma}) + N_{zy}(\boldsymbol{\sigma}), \tag{11}$$

as this counts the total number of swapped pairs.

We now handle the cases of each $\tau \in \{abc, bac, cba\}$ separately.

Case 1: $\tau = abc$**.** By Equations (10) and (11), we have that

$$\begin{aligned} d(\boldsymbol{\sigma}, abc) &= N_{ba}(\boldsymbol{\sigma}) + N_{ca}(\boldsymbol{\sigma}) + N_{cb}(\boldsymbol{\sigma}) \\ &= n - N_{ab}(\boldsymbol{\sigma}) + n - N_{ac}(\boldsymbol{\sigma}) + n - N_{bc}(\boldsymbol{\sigma}) \\ &= 2n - \text{SC}_a(\boldsymbol{\sigma}) + (n - N_{bc}(\boldsymbol{\sigma})) \end{aligned}$$

and

$$\begin{aligned} d(\pi(\boldsymbol{\sigma}), abc) &= N_{ba}(\pi(\boldsymbol{\sigma})) + N_{ca}(\pi(\boldsymbol{\sigma})) + N_{cb}(\pi(\boldsymbol{\sigma})) \\ &= 2n - \text{SC}_a(\pi(\boldsymbol{\sigma})) + (n - N_{bc}(\pi(\boldsymbol{\sigma})) \\ &= 2n - \text{SC}_a(\boldsymbol{\sigma}) + N_{bc}(\boldsymbol{\sigma}). \end{aligned}$$

Substituting this into the left-hand side of (5), we have

$$\begin{aligned} &\sum_{\boldsymbol{\sigma} \in A_k^{K_1,K_2,K_3}} \varphi^{d(\boldsymbol{\sigma},abc)} - \varphi^{d(\pi(\boldsymbol{\sigma}),abc)} \\ = &\sum_{\boldsymbol{\sigma} \in A_k^{K_1,K_2,K_3}} \varphi^{2n - \text{SC}_a(\boldsymbol{\sigma}) + n - N_{bc}(\boldsymbol{\sigma})} - \varphi^{2n - \text{SC}_a(\boldsymbol{\sigma}) + N_{bc}(\boldsymbol{\sigma})} \\ = &\sum_{\boldsymbol{\sigma} \in A_k^{K_1,K_2,K_3}} \varphi^{2n - 2|K_1| - |K_2|} \left(\varphi^{n - N_{bc}(\boldsymbol{\sigma})} - \varphi^{N_{bc}(\boldsymbol{\sigma})}\right) \end{aligned}$$

Note that the term in front is always nonnegative and constant for fixed K_1, K_2, K_3, so it is sufficient to show

$$\sum_{\boldsymbol{\sigma} \in A_k^{K_1,K_2,K_3}} \left(\varphi^{n - N_{bc}(\boldsymbol{\sigma})} - \varphi^{N_{bc}(\boldsymbol{\sigma})}\right) \geq 0. \tag{12}$$

Notice that these terms depend only on $N_{bc}(\boldsymbol{\sigma})$ which must take on a value in $\{0, \ldots, n\}$. Hence, we can instead consider counting the number of profiles $\boldsymbol{\sigma} \in A_k^{K_1,K_2,K_3}$ with a specific $N_{bc}(\boldsymbol{\sigma})$. More formally, let $Q_j = |\{\boldsymbol{\sigma} \in A_k^{K_1,K_2,K_3} | N_{bc}(\boldsymbol{\sigma}) = j\}|$ for $j \in \{0, \ldots, n\}$. We can now write

$$\sum_{\boldsymbol{\sigma} \in A_k^{K_1,K_2,K_3}} \left(\varphi^{n - N_{bc}(\boldsymbol{\sigma})} - \varphi^{N_{bc}(\boldsymbol{\sigma})}\right) = \sum_{j=0}^{n} Q_j \left(\varphi^{n-j} - \varphi^j\right).$$

Notice that for each $Q_j(\varphi^{n-j} - \varphi^j)$ term in the sum, there is a corresponding term $Q_{n-j}(\varphi^j - \varphi^{n-j})$. Pairing up these opposite terms, we can rewrite the sum as

$$\sum_{j=0}^{\lfloor (n-1)/2 \rfloor} (Q_{n-j} - Q_j)\left(\varphi^j - \varphi^{n-j}\right)$$

The $\lfloor (n-1)/2 \rfloor$ expression is simply the largest integer strictly less than $n/2$ (we exclude the $j = n/2$ term since this is 0 if it exists). Note that $(\varphi^j - \varphi^{n-j}) > 0$ for $j < n/2$, so we have reduced the problem to counting the number of profiles $\boldsymbol{\sigma}$ with a specific value of $N_{bc}(\boldsymbol{\sigma})$. More formally, Inequality (12) to show for $j < n/2$,

$$Q_j \leq Q_{n-j}. \tag{13}$$

Fix a $j < n/2$. For a profile $\boldsymbol{\sigma} \in A_k^{K_1,K_2,K_3}$, define

$$t(\boldsymbol{\sigma}) = \{i \in K_2 | b \succ_i c\}$$
$$o(\boldsymbol{\sigma}) = \{i \in K_1 \cup K_3 | b \succ_i c\},$$

In words, $t(\boldsymbol{\sigma})$ is the number of voters in K_2 that prefer b to c and $o(\boldsymbol{\sigma})$ is the number of voters in $K_1 \cup K_3$ that prefer b to c. This is useful for us because these values allow us to calculate $\text{SC}_b(\boldsymbol{\sigma})$. Voters in $|K_3|$ give a minimum of one point to b. For all voters counted in $o(\boldsymbol{\sigma})$, an additional one point is given versus those not counted. For all voters counted in $t(\boldsymbol{\sigma})$ an additional two points are given versus those not counted. Hence,

$$\text{SC}_b(\boldsymbol{\sigma}) = |K_3| + 2t(\boldsymbol{\sigma}) + o(\boldsymbol{\sigma}).$$

When $\boldsymbol{\sigma} \in A_k^{K_1,K_2,K_3}$, we know that $\text{SC}_b(\boldsymbol{\sigma}) = \text{SC}_a(\boldsymbol{\sigma}) + \kappa$, so we have that

$$2t(\boldsymbol{\sigma}) + o(\boldsymbol{\sigma}) = 2|K_2| + |K_1| + \kappa - |K_3| \tag{14}$$

Further,

$$t(\boldsymbol{\sigma}) + o(\boldsymbol{\sigma}) = N_{bc}(\boldsymbol{\sigma})$$

as it is simply a different way of counting the number of voters with $b \succ c$.

Observe that if $\boldsymbol{\sigma}$ is counted toward Q_j, both Equation (14) must hold and $t(\boldsymbol{\sigma}) + o(\boldsymbol{\sigma}) = j$. These are two independent linear equations on $t(\boldsymbol{\sigma})$ and $o(\boldsymbol{\sigma})$ and hence there is exactly one solution for $t(\boldsymbol{\sigma})$ and $o(\boldsymbol{\sigma})$ that satisfies them. Further, notice that this is a necessary and sufficient condition: $\boldsymbol{\sigma} \in A_k^{K_1,K_2,K_3}$ is counted toward Q_j if and only if it satisfies both Equation (14) and $t(\boldsymbol{\sigma}) + o(\boldsymbol{\sigma}) = j$ (along with the K_1, K_2, K_3 constraint).

Let t and o be the solutions satisfying the above equations for Q_j with $0 \leq t \leq |K_2|$ and $0 \leq o \leq |K_1| + |K_3|$. Note that if t or o are not integers or do not satisfy the inequalities then $Q_j = 0$, so $Q_j \leq Q_{n-j}$ as Q_{n-j} is necessarily nonnegative. For t and o satisfying the inequalities, we have that

$$Q_j = \binom{|K_2|}{t}\binom{|K_1| + |K_3|}{o}$$

since we choose t voters in $|K_2|$ and o voters in $|K_1 \cup |K_3|$ to rank $b \succ c$. We first claim that $t' := t - n + 2j$ and $o' := o + 2n - 4j$ are solutions for Q_{n-j} since

$$2t' + o' = 2(t - n + 2j) + (o + 2n - 4j) = 2t + o = 2|K_2| + |K_1| + \kappa - |K_3|$$
$$t' + o' = t - n + 2j + o + 2n - 4j = t + o + n - 2j = j + n - 2j = n - j$$

Since t and o were integers, so are t' and o'. We want to show

$$Q_j = \binom{|K_2|}{t}\binom{|K_1| + |K_3|}{o} \leq \binom{|K_2|}{t'}\binom{|K_1| + |K_3|}{o'} = Q_{n-j}$$

We will show individually that $\binom{|K_2|}{t'} \geq \binom{|K_2|}{t}$ and $\binom{|K_1|+|K_3|}{o'} \geq \binom{|K_1|+|K_3|}{o}$. Notice that $t' \leq t$ and $o' \geq o$, so this is implied by showing that $t' \geq |K_2| - t$ and $o' \leq |K_1| + |K_3| - o$. Both rely on the inequality $2(2t + o) > n + |K_2|$, which follows from

$$\begin{aligned}
2(2t + o) &= 2(2|K_1| + |K_2| + \kappa - |K_3|) \\
&= 4|K_1| + 2|K_2| - 2|K_3| + 2\kappa \\
&\geq |K_1| + 2|K_2| + 3|K_3| - 2|K_3| + 2\kappa - (2\kappa - 1) && (|K_1| \geq |K_3| - \tfrac{2\kappa-1}{3}) \\
&\geq n + |K_2| && (|K_1| + |K_2| + |K_3| = n)
\end{aligned}$$

Using the derived inequality, we have

$$t' = t - n + 2j$$

$$
\begin{aligned}
&= t - n + 2(t + o) \\
&= 3t + 2o - n \\
&= 2(2t + o) - n - t \\
&\geq n + |K_2| - n - t \\
&= |K_2| - t
\end{aligned}
$$

and

$$
\begin{aligned}
o' &= o + 2n - 4N \\
&= o + 2n - 4(o + t) \\
&= 2n - 3o - 4t \\
&= 2n - 2(o + 2t) - o \\
&\leq 2n - (n - |K_2|) - o \\
&= n - |K_2| - o \\
&= |K_1| + |K_3| - o, && (|K_1| + |K_2| + |K_3| = n)
\end{aligned}
$$

Therefore, Inequality (13) holds, as needed.

Case 2: $\tau = bca$**.** Again, by Equations (10) and (11), we have that

$$
\begin{aligned}
d(\boldsymbol{\sigma}, bca) &= N_{ab}(\boldsymbol{\sigma}) + N_{ac}(\boldsymbol{\sigma}) + N_{cb}(\boldsymbol{\sigma}) \\
&= \mathrm{SC}_a(\boldsymbol{\sigma}) + n - N_{bc}(\boldsymbol{\sigma})
\end{aligned}
$$

and

$$
\begin{aligned}
d(\pi(\boldsymbol{\sigma}), bca) &= N_{ab}(\pi(\boldsymbol{\sigma})) + N_{ac}(\pi(\boldsymbol{\sigma})) + N_{cb}(\pi(\boldsymbol{\sigma})) \\
&= \mathrm{SC}_a(\pi(\boldsymbol{\sigma})) + (n - N_{bc}(\pi(\boldsymbol{\sigma})) \\
&= \mathrm{SC}_a(\boldsymbol{\sigma}) + N_{bc}(\boldsymbol{\sigma}).
\end{aligned}
$$

Substituting this into the left-hand side of (5), we have,

$$
\begin{aligned}
\sum_{\boldsymbol{\sigma} \in A_k^{K_1,K_2,K_3}} \varphi^{d(\boldsymbol{\sigma},bca)} - \varphi^{d(\pi(\boldsymbol{\sigma}),bca)} &= \sum_{\boldsymbol{\sigma} \in A_k^{K_1,K_2,K_3}} \varphi^{\mathrm{SC}_a(\boldsymbol{\sigma})+(n-N_{bc}(\boldsymbol{\sigma}))} - \varphi^{\mathrm{SC}_a(\boldsymbol{\sigma})+N_{bc}(\boldsymbol{\sigma})} \\
&= \sum_{\boldsymbol{\sigma} \in A_k^{K_1,K_2,K_3}} \varphi^{2|K_1|+|K_2|} \left(\varphi^{n-N_{bc}(\boldsymbol{\sigma})} - \varphi^{N_{bc}(\boldsymbol{\sigma})} \right).
\end{aligned}
$$

Again we notice that the term in front is always nonnegative and constant for fixed K_1, K_2, K_3, so, it is sufficient to show

$$
\sum_{\boldsymbol{\sigma} \in A_k^{K_1,K_2,K_3}} \left(\varphi^{n-N_{bc}(\boldsymbol{\sigma})} - \varphi^{N_{bc}(\boldsymbol{\sigma})} \right) \geq 0,
$$

which we already proved in the last case.

Case 3: $\tau = bac$**.** Using Equations (10) and (11), we have

$$
\begin{aligned}
d(\boldsymbol{\sigma}, bac) &= N_{ab}(\boldsymbol{\sigma}) + N_{ca}(\boldsymbol{\sigma}) + N_{cb}(\boldsymbol{\sigma}) \\
&= (n - N_{ba}(\boldsymbol{\sigma})) + (n - N_{ac}(\boldsymbol{\sigma})) + (n - N_{bc}(\boldsymbol{\sigma})) \\
&\quad + \underbrace{(n - N_{ab}(\boldsymbol{\sigma}) - N_{ba}(\boldsymbol{\sigma}))}_{0} + \underbrace{(N_{bc}(\boldsymbol{\sigma}) - N_{bc}(\boldsymbol{\sigma})}_{0} \\
&= 4n - \mathrm{SC}_a(\boldsymbol{\sigma}) - 2\mathrm{SC}_b(\boldsymbol{\sigma}) + N_{bc}(\boldsymbol{\sigma}) \\
&= 4n - 6|K_1| - 3|K_2| - 2\kappa + N_{bc}(\boldsymbol{\sigma}) \\
&= (-2|K_1| + |K_2| + 4|K_3| - 2\kappa) + N_{bc}(\boldsymbol{\sigma})
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
d(\pi(\boldsymbol{\sigma}), bac) &= N_{ab}(\pi(\boldsymbol{\sigma})) + N_{ca}(\pi(\boldsymbol{\sigma})) + N_{cb}(\pi(\boldsymbol{\sigma})) \\
&= (n - N_{ab}(\boldsymbol{\sigma})) + (n - N_{ca}(\boldsymbol{\sigma})) + (n - N_{cb}(\boldsymbol{\sigma}))
\end{aligned}
$$

$$= 3n - d(\boldsymbol{\sigma}, bac)$$
$$= 3n - (-2|K_1| + |K_2| + 4|K_3| - 2\kappa) - N_{bc}(\boldsymbol{\sigma})$$

Let $C = (-2|K_1| + |K_2| + 4|K_3| - 2\kappa)$. Substituting this into the left-hand side of (5), we have .

$$\sum_{\boldsymbol{\sigma} \in A_k^{K_1,K_2,K_3}} \varphi^{d(\boldsymbol{\sigma},bac)} - \varphi^{d(\pi(\boldsymbol{\sigma}),cab)} = \sum_{\boldsymbol{\sigma} \in A_k^{K_1,K_2,K_3}} \varphi^{C+N_{bc}(\boldsymbol{\sigma})} - \varphi^{3n-C-N_{bc}(\boldsymbol{\sigma})}$$
$$= \sum_{j=0}^{n} Q_j \left(\varphi^{C+j} - \varphi^{3n-C-j}\right).$$

Observe that $\left(\varphi^{C+j} - \varphi^{3n-C-j}\right)$ is negative only for $j > \frac{3n}{2} - C$. Additionally, for each of these terms, there is a corresponding positive term in the sum for $j' = 3n - 2C - j < \frac{3n}{2} - C$, where

$$\left(\varphi^{C+j'} - \varphi^{3n-C-j'}\right) = \left(\varphi^{C+3n-2C-j} - \varphi^{3n-C-3n+2C+j}\right) = -\left(\varphi^{C+j} - \varphi^{3n-C-j}\right).$$

Thus, it suffices to show for $j > \frac{3n}{2} - C$ that $Q_j \leq Q_{j'}$ where $j' = 3n - 2C - j$.

As before, let t and o be solutions for Q_j. Then we have the following solutions for $Q_{j'}$

$$t' = t + (j - j')$$
$$o' = o - 2(j - j')$$

since

$$t' + o' = t + (j - j') + o - 2(j - j') = j'$$
$$2t' + o' = 2t + 2(j + j')o - 2(j - j') = 2t + o.$$

Recall that

$$Q_j = \binom{|K_2|}{t}\binom{|K_1| + |K_3|}{o} \text{ and } Q_{j'} = \binom{|K_2|}{t'}\binom{|K_1| + |K_3|}{o'}.$$

We will show that $\binom{|K_2|}{t} \leq \binom{|K_2|}{t'}$ and $\binom{|K_1|+|K_3|}{o} \leq \binom{|K_1|+|K_3|}{o'}$. Note that $t' \geq t$ and $o' \leq o$, so it suffices to show that $t' \leq |K_2| - t$ and $o' \geq |K_1| + |K_3| - o$. Let us directly consider

$$\begin{aligned}
t' &= t + (j - j') \\
&= t + (2j - 3n + 2C) \\
&= t + (2(t + o) - 3n + 2C) \\
&= 2(2t + o) - 3n + 2C - t \\
&= 2(2|K_1| + |K_2| + \kappa - |K_3|) - 3n + 2(-2|K_1| + |K_2| + 4|K_3| - 2\kappa) - t \\
&= -3n + 4|K_2| + 6|K_3| - 2\kappa - t \\
&= -3|K_1| + 3|K_3| - 2\kappa + |K_2| - t \\
&< |K_2| - t.
\end{aligned}$$

We also have that

$$\begin{aligned}
o' &= o - 2(j - j') \\
&= o - 2(2j - 3n + 2C) \\
&= o - 2(2(t + o) - 3n + 2C) \\
&= -2(2t + o) + 6n - 4C - o \\
&= -2(2|K_1| + |K_2| + \kappa - |K_3|) + 6n - 4(-2|K_1| + |K_2| + 4|K_3| - 2\kappa) - o \\
&= 6n + 4|K_1| - 6|K_2| - 14|K_3| + 6\kappa - o \\
&= 10|K_1| - 8|K_3| + 6\kappa - o \\
&= 9|K_1| - 9|K_3| + 6\kappa + |K_1| + |K_3| - o \\
&> |K_1| + |K_3| - o.
\end{aligned}$$

Report	Probability a wins
abc	$1\varphi^0 + 4\varphi^1 + 7\varphi^2 + 8\varphi^3 + 8/3\varphi^4 + 0\varphi^5 + 0\varphi^6$
bac	$1\varphi^0 + 2\varphi^1 + 3\varphi^2 + 2\varphi^3 + 2/3\varphi^4 + 0\varphi^5 + 0\varphi^6$
acb	$1\varphi^0 + 4\varphi^1 + 7\varphi^2 + 8\varphi^3 + 8/3\varphi^4 + 0\varphi^5 + 0\varphi^6$
bca	$1\varphi^0 + 2\varphi^1 + 5/3\varphi^2 + 0\varphi^3 + 0\varphi^4 + 0\varphi^5 + 0\varphi^6$
cab	$1\varphi^0 + 4\varphi^1 + 11/3\varphi^2 + 0\varphi^3 + 0\varphi^4 + 0\varphi^5 + 0\varphi^6$
cba	$1\varphi^0 + 2\varphi^1 + 5/3\varphi^2 + 0\varphi^3 + 0\varphi^4 + 0\varphi^5 + 0\varphi^6$

Report	Probability c wins
abc	$0\varphi^0 + 0\varphi^1 + 0\varphi^2 + 0\varphi^3 + 5/3\varphi^4 + 2\varphi^5 + 1\varphi^6$
bac	$0\varphi^0 + 0\varphi^1 + 0\varphi^2 + 0\varphi^3 + 5/3\varphi^4 + 2\varphi^5 + 1\varphi^6$
acb	$0\varphi^0 + 0\varphi^1 + 0\varphi^2 + 0\varphi^3 + 11/3\varphi^4 + 4\varphi^5 + 1\varphi^6$
bca	$0\varphi^0 + 0\varphi^1 + 2/3\varphi^2 + 2\varphi^3 + 3\varphi^4 + 2\varphi^5 + 1\varphi^6$
cab	$0\varphi^0 + 0\varphi^1 + 8/3\varphi^2 + 8\varphi^3 + 7\varphi^4 + 4\varphi^5 + 1\varphi^6$
cba	$0\varphi^0 + 0\varphi^1 + 8/3\varphi^2 + 8\varphi^3 + 7\varphi^4 + 4\varphi^5 + 1\varphi^6$

Table 3: Probability that a and c each win under different reports for voter i. This assumes their observed ranking was abc.

This completes the proof for the weak inequality.

To show that the first inequality is strict, observe that in case 1, all of the inequalities about $Q_j \leq Q_{n-j}$ can be shown to be strict. Hence, all we need to show is that there is some k such that A_k is nonempty. Fix some arbitrary n. If n is even, we can take a profile $\boldsymbol{\sigma}$ where $n/2+1$ voters have the ranking bac and $n/2-1$ have abc. Such a profile is always an element of A_2 since $\text{SC}_b(\boldsymbol{\sigma}) = \text{SC}_a(\boldsymbol{\sigma}) + 2$ and $\text{SC}_a(\boldsymbol{\sigma}) \geq \text{SC}_c(\boldsymbol{\sigma})$. Similarly, if n is odd, we can take a profile $\boldsymbol{\sigma}$ where $\lceil n/2 \rfloor$ voters have the ranking bac and $\lfloor n/2 \rfloor$ have the ranking acb. Such a profile is always an element of A_1 since $\text{SC}_b(\boldsymbol{\sigma}) = \text{SC}_a(\boldsymbol{\sigma}) + 1$ and $\text{SC}_a(\boldsymbol{\sigma}) \geq \text{SC}_c(\boldsymbol{\sigma})$. □

E OBIC Positional Scoring Rule Example

Let f be a scoring rule $(r_1, r_2, 0)$ such that $r_2/r_1 < 1/3$. Note that on three voters, all these rules coincide. Indeed, if there is a strict plurality winner, that candidate is necessarily the winner. If not, this means each candidate appeared first exactly once. I some candidate appears in second twice, then that candidate is the winner. Finally, if no candidate appears in second twice, they all appear in second once, and therefore all appear in third once, so there is a three-way tie. Under such rules with a confident Mallows prior with a fixed φ, we can explicitly compute the probability that each candidate wins as a function of φ. This assumes, without loss of generality, that the voter's observed ranking is abc. The probability that a and c each win under possible reports are shown in Table 3. One can check that reporting anything other than abc neither increases the probability that a wins or decreases the probability that c wins, which means the rule is OBIC.