

---

# Is Approval Voting Optimal Given Approval Votes?

---

**Ariel D. Procaccia**

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

**Nisarg Shah**

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

## Abstract

Some crowdsourcing platforms ask workers to express their opinions by approving a set of  $k$  good alternatives. It seems that the only reasonable way to aggregate these  $k$ -approval votes is the *approval voting rule*, which simply counts the number of times each alternative was approved. We challenge this assertion by proposing a probabilistic framework of noisy voting, and asking whether approval voting yields an alternative that is most likely to be the best alternative, given  $k$ -approval votes. While the answer is generally positive, our theoretical and empirical results call attention to situations where approval voting is suboptimal.

## 1 Introduction

It is surely no surprise to the reader that modern machine learning algorithms thrive on large amounts of data — preferably labeled. Online labor markets, such as *Amazon Mechanical Turk* ([www.mturk.com](http://www.mturk.com)), have become a popular way to obtain labeled data, as they harness the power of a large number of human workers, and offer significantly lower costs compared to expert opinions. But this low-cost, large-scale data may require compromising quality: the workers are often unqualified or unwilling to make an effort, leading to a high level of noise in their submitted labels.

To overcome this issue, it is common to hire multiple workers for the same task, and aggregate their noisy opinions to find more accurate labels. For example, TurKit [17] is a toolkit for creating and managing crowdsourcing tasks on Mechanical Turk. For our purposes its most important aspect is that it implements plurality voting: among available alternatives (e.g., possible labels), workers report the best alternative in their opinion, and the alternative that receives the most votes is selected.

More generally, workers may be asked to report the  $k$  best alternatives in their opinion; such a vote is known as a  $k$ -approval vote. This has an advantage over plurality (1-approval) in noisy situations where a worker may not be able to pinpoint the best alternative accurately, but can recognize that it is among the top  $k$  alternatives [23].<sup>1</sup> At the same time,  $k$ -approval votes, even for  $k > 1$ , are much easier to elicit than, say, rankings of the alternatives, not to mention full utility functions. For example, EteRNA [16] — a citizen science game whose goal is to design RNA molecules that fold into stable structures — uses 8-approval voting on submitted designs, that is, each player approves up to 8 favorite designs; the designs that received the largest number of approval votes are selected for synthesis in the lab.

So, the elicitation of  $k$ -approval votes is common practice and has significant advantages. And it may seem that the only reasonable way to aggregate these votes, once collected, is via the *approval voting rule*, that is, tally the number of approvals for each alternative, and select the most approved one.<sup>2</sup> But is it? In other words, do the  $k$ -approval votes contain useful information that can lead to

---

<sup>1</sup> $k$ -approval is also used for picking  $k$  winners, e.g., various cities in the US such as San Francisco, Chicago, and New York use it in their so-called “participatory budgeting” process [15].

<sup>2</sup>There is a subtle distinction, which we will not belabor, between  $k$ -approval voting, which is the focus of this paper, and approval voting [8], which allows voters to approve as many alternatives as they wish. The latter

significantly better outcomes, and is ignored by approval voting? Or is approval voting an (almost) optimal method for aggregating  $k$ -approval votes?

**Our Approach.** We study the foregoing questions within the *maximum likelihood estimation* (MLE) framework of social choice theory, which posits the existence of an underlying ground truth that provides an objective comparison of the alternatives. From this viewpoint, the votes are noisy estimates of the ground truth. The optimal rule then selects the alternative that is most likely to be the best alternative given the votes. This framework has recently received attention from the machine learning community [18, 3, 2, 4, 21], in part due to its applications to crowdsourcing domains [20, 21, 9], where, indeed, there is a ground truth, and individual votes are objective.

In more detail, in our model there exists a ground truth ranking over the alternatives, and each *voter* holds an opinion, which is another ranking that is a noisy estimate of the ground truth ranking. The opinions are drawn i.i.d. from the popular Mallows model [19], which is parametrized by the ground truth ranking, a noise parameter  $\varphi \in [0, 1]$ , and a distance metric  $d$  over the space of rankings. We use five well-studied distance metrics: the Kendall tau (KT) distance, the (Spearman) footrule distance, the maximum displacement distance, the Cayley distance, and the Hamming distance.

When required to submit a  $k$ -approval vote, a voter simply approves the top  $k$  alternatives in his opinion. Given the votes, an alternative  $a$  is the *maximum likelihood estimate* (MLE) for the best alternative if the votes are most likely generated by a ranking that puts  $a$  first.

We can now reformulate our question in slightly more technical terms:

*Is approval voting (almost) a maximum likelihood estimator for the best alternative, given votes drawn from the Mallows model? How does the answer depend on the noise parameter  $\phi$  and the distance metric  $d$ ?*

**Our results.** Our first result (Theorem 1) shows that under the Mallows model, the set of winners according to approval voting coincides with the set of MLE best alternatives under the Kendall tau distance, but under the other four distances there may exist approval winners that are not MLE best alternatives. Our next result (Theorem 2) confirms the intuition that the suboptimality of approval voting stems from the information that is being discarded: when only a single alternative is approved or disapproved in each vote, approval voting — which now utilizes all the information that can be gleaned from the anonymous votes — is optimal under mild conditions.

Going back to the general case of  $k$ -approval votes, we show (Theorem 3) that even under the four distances for which approval voting is suboptimal, a weaker statement holds: in cases with very high or very low noise, every MLE best alternative is an approval winner (but some approval winners may not be MLE best alternatives). And our experiments, using real data, show that the accuracy of approval voting is usually quite close to that of the MLE in pinpointing the best alternative.

We conclude that approval voting is a good way of aggregating  $k$ -approval votes in most situations. But our work demonstrates that, perhaps surprisingly, approval voting may be suboptimal, and, in situations where a high degree of accuracy is required, exact computation of the MLE best alternative is an option worth considering. We discuss our conclusions in more detail in Section 6.

## 2 Model

Let  $[t] \triangleq \{1, \dots, t\}$ . Denote the set of *alternatives* by  $A$ , and let  $|A| = m$ . We use  $\mathcal{L}(A)$  to denote the set of rankings (total orders) of the alternatives in  $A$ . For a ranking  $\sigma \in \mathcal{L}(A)$ , let  $\sigma(i)$  denote the alternative occupying position  $i$  in  $\sigma$ , and let  $\sigma^{-1}(a)$  denote the rank (position) of alternative  $a$  in  $\sigma$ . With a slight abuse of notation, let  $\sigma([t]) \triangleq \{a \in A \mid \sigma^{-1}(a) \in [t]\}$ . We use  $\sigma_{a \leftrightarrow b}$  to denote the ranking obtained by swapping the positions of alternatives  $a$  and  $b$  in  $\sigma$ . We assume that there exists an unknown *true ranking* of the alternatives (the ground truth), denoted  $\sigma^* \in \mathcal{L}(A)$ . We also make the standard assumption of a uniform prior over the true ranking.

---

framework of approval voting has been studied extensively, both from the axiomatic point of view [7, 8, 13, 22, 1], and the game-theoretic point of view [14, 12, 6]. However, even under this framework it is a standard assumption that votes are tallied by counting the number of times each alternative is approved, which is why we simply refer to the aggregation rule under consideration as approval voting.

Let  $N = \{1, \dots, n\}$  denote the set of *voters*. Each voter  $i$  has an *opinion*, denoted  $\pi_i \in \mathcal{L}(A)$ , which is a noisy estimate of the true ranking  $\sigma^*$ ; the collection of opinions — the (opinion) *profile* — is denoted  $\pi$ . Fix  $k \in [m]$ . A  $k$ -*approval vote* is a collection of  $k$  alternatives approved by a voter. When asked to submit a  $k$ -approval vote, voter  $i$  simply submits the *vote*  $V_i = \pi_i([k])$ , which is the set of alternatives at the top  $k$  positions in his opinion. The collection of all votes is called the *vote profile*, and denoted  $V = \{V_i\}_{i \in [n]}$ . For a ranking  $\sigma$  and a  $k$ -approval vote  $v$ , we say that  $v$  is generated from  $\sigma$ , denoted  $\sigma \rightarrow_k v$  (or  $\sigma \rightarrow v$  when the value of  $k$  is clear from the context), if  $v = \sigma([k])$ . More generally, for an opinion profile  $\pi$  and a vote profile  $V$ , we say  $\pi \rightarrow_k V$  (or  $\pi \rightarrow V$ ) if  $\pi_i \rightarrow_k V_i$  for every  $i \in [n]$ .

Let  $\mathcal{A}_k = \{A_k \subseteq A \mid |A_k| = k\}$  denote the set of all subsets of  $A$  of size  $k$ . A *voting rule* operating on  $k$ -approval votes is a function  $(\mathcal{A}_k)^n \rightarrow A$  that returns a *winning alternative* given the votes.<sup>3</sup> In particular, let us define the approval score of an alternative  $a$ , denoted  $\text{SC}^{\text{APP}}(a)$ , as the number of voters that approve  $a$ . Then, *approval voting* simply chooses an alternative with the greatest approval score. Note that we do not break ties. Instead, we talk about the *set of approval winners*.

Following the standard social choice literature, we model the opinion of each voter as being drawn i.i.d. from an underlying *noise model*. A noise model describes the probability of drawing an opinion  $\sigma$  given the true ranking  $\sigma^*$ , denoted  $\Pr[\sigma \mid \sigma^*]$ . We say that a noise model is *neutral* if the labels of the alternatives do not matter, i.e., renaming alternatives in the true ranking  $\sigma$  and in the opinion  $\sigma^*$ , in the same fashion, keeps  $\Pr[\sigma \mid \sigma^*]$  intact. A popular noise model is the *Mallows model* [19], under which  $\Pr[\sigma \mid \sigma^*] = \varphi^{d(\sigma, \sigma^*)} / Z_\varphi^m$ . Here,  $d$  is a distance metric over the space of rankings. Parameter  $\varphi \in [0, 1]$  governs the noise level;  $\varphi = 0$  implies that the true ranking is generated with probability 1, and  $\varphi = 1$  implies the uniform distribution.  $Z_\varphi^m$  is the normalization constant, which is independent of the true ranking  $\sigma^*$  given that distance  $d$  is *neutral*, i.e., renaming alternatives in the same fashion in two rankings does not change the distance between them. Below, we review five popular distances used in the social choice literature; they are all neutral.

- *The Kendall tau (KT) distance*, denoted  $d_{KT}$ , measures the number of pairs of alternatives over which two rankings disagree. Equivalently, it is the number of swaps required by bubble sort to convert one ranking into another.
- *The (Spearman) footrule (FR) distance*, denoted  $d_{FR}$ , measures the total displacement (absolute difference between positions) of all alternatives in two rankings.
- *The Maximum Displacement (MD) distance*, denoted  $d_{MD}$ , measures the maximum of the displacements of all alternatives between two rankings.
- *The Cayley (CY) distance*, denoted  $d_{CY}$ , measures the minimum number of swaps (not necessarily of adjacent alternatives) required to convert one ranking into another.
- *The Hamming (HM) distance*, denoted  $d_{HM}$ , measures the number of positions in which two rankings place different alternatives.

Since opinions are drawn independently, the probability of a profile  $\pi$  given the true ranking  $\sigma^*$  is  $\Pr[\pi \mid \sigma^*] = \prod_{i=1}^n \Pr[\pi_i \mid \sigma^*] \propto \varphi^{d(\pi, \sigma^*)}$ , where  $d(\pi, \sigma^*) = \sum_{i=1}^n d(\pi_i, \sigma^*)$ . Once we fix the noise model, for a fixed  $k$  we can derive the probability of observing a given  $k$ -approval vote  $v$ :  $\Pr[v \mid \sigma^*] = \sum_{\sigma \in \mathcal{L}(A): \sigma \rightarrow v} \Pr[\sigma \mid \sigma^*]$ . Then, the probability of drawing a given vote profile  $V$  is  $\Pr[V \mid \sigma^*] = \prod_{i=1}^n \Pr[V_i \mid \sigma^*]$ . Alternatively, this can also be expressed as  $\Pr[V \mid \sigma^*] = \sum_{\pi \in \mathcal{L}(A)^n: \pi \rightarrow V} \Pr[\pi \mid \sigma^*]$ . Hereinafter, we omit the domains  $\mathcal{L}(A)^n$  for  $\pi$  and  $\mathcal{L}(A)$  for  $\sigma^*$  when they are clear from the context.

Finally, given the vote profile  $V$  the likelihood of an alternative  $a$  being the best alternative in the true ranking  $\sigma^*$  is proportional to (via Bayes' rule)  $\Pr[V \mid \sigma^*(1) = a] = \sum_{\sigma^*: \sigma^*(1) = a} \Pr[V \mid \sigma^*]$ . Using the two expressions derived earlier for  $\Pr[V \mid \sigma^*]$ , and ignoring the normalization constant  $Z_\varphi^m$  from the probabilities, we define the likelihood function of  $a$  given votes  $V$  as

$$L(V, a) \triangleq \sum_{\sigma^*: \sigma^*(1) = a} \sum_{\pi: \pi \rightarrow V} \varphi^{d(\pi, \sigma^*)} = \sum_{\sigma^*: \sigma^*(1) = a} \prod_{i=1}^n \left[ \sum_{\pi_i: \pi_i \rightarrow V_i} \varphi^{d(\pi_i, \sigma^*)} \right]. \quad (1)$$

The *maximum likelihood estimate* (MLE) for the best alternative is given by  $\arg \max_{a \in A} L(V, a)$ . Again, we do not break ties; we study the *set of MLE best alternatives*.

<sup>3</sup>Technically, this is a social choice function; a social welfare function returns a ranking of the alternatives.

### 3 Optimal Voting Rules

At first glance, it seems natural to use approval voting (that is, returning the alternative that is approved by the largest number of voters) given  $k$ -approval votes. However, consider the following example with 4 alternatives ( $A = \{a, b, c, d\}$ ) and 5 voters providing 2-approval votes:

$$V_1 = \{b, c\}, V_2 = \{b, c\}, V_3 = \{a, d\}, V_4 = \{a, b\}, V_5 = \{a, c\}. \quad (2)$$

Notice that alternatives  $a$ ,  $b$ , and  $c$  receive 3 approvals each, while alternative  $d$  receives only a single approval. Approval voting may return any alternative other than alternative  $d$ . But is that always optimal? In particular, while alternatives  $b$  and  $c$  are symmetric, alternative  $a$  is qualitatively different due to different alternatives being approved along with  $a$ . This indicates that under certain conditions, it is possible that not all three alternatives are MLE for the best alternative. Our first result shows that this is indeed the case under three of the distance functions listed above, and a similar example works for a fourth. However, surprisingly, under the Kendall tau distance the MLE best alternatives are exactly the approval winners, and hence are polynomial-time computable, which stands in sharp contrast to the NP-hardness of computing them given *rankings* [5].

**Theorem 1.** *The following statements hold for aggregating  $k$ -approval votes using approval voting.*

1. *Under the Mallows model with a fixed distance  $d \in \{d_{MD}, d_{CY}, d_{HM}, d_{FR}\}$ , there exist a vote profile  $V$  with at most six 2-approval votes over at most five alternatives, and a choice for the Mallows parameter  $\varphi$ , such that not all approval winners are MLE best alternatives.*
2. *Under the Mallows model with the distance  $d = d_{KT}$ , the set of MLE best alternatives coincides with the set of approval winners, for all vote profiles  $V$  and all values of the Mallows parameter  $\varphi \in (0, 1)$ .*

*Proof.* For the Mallows model with  $d \in \{d_{MD}, d_{CY}, d_{HM}\}$  and any  $\varphi \in (0, 1)$ , the profile from Equation (2) is a counterexample: alternatives  $b$  and  $c$  are MLE best alternatives, but  $a$  is not. For the Mallows model with  $d = d_{FR}$ , we could not find a counter example with 4 alternatives; computer-based simulations generated the following counterexample with 5 alternatives that works for any  $\varphi \in (0, 1)$ :  $V_1 = V_2 = \{a, b\}$ ,  $V_3 = V_4 = \{c, d\}$ ,  $V_5 = \{a, e\}$ , and  $V_6 = \{b, c\}$ . Here, alternatives  $a$ ,  $b$ , and  $c$  have the highest approval score of 3. However, alternative  $b$  has a strictly lower likelihood of being the best alternative than alternative  $a$ , and hence is not an MLE best alternative. The calculation verifying these counterexamples is presented in the online appendix (specifically, Appendix A).

In contrast, for the Kendall tau distance, we show that all approval winners are MLE best alternatives, and vice-versa. We begin by simplifying the likelihood function  $L(V, a)$  from Equation (1) for the special case of the Mallows model with the Kendall tau distance. In this case, it is well known that the normalization constant satisfies  $Z_\varphi^m = \prod_{j=1}^m T_\varphi^j$ , where  $T_\varphi^j = \sum_{i=0}^{j-1} \varphi^i$ . Consider a ranking  $\pi_i$  such that  $\pi_i \rightarrow V_i$ . We can decompose  $d_{KT}(\pi_i, \sigma^*)$  into three types of pairwise mismatches: i)  $d_1(\pi_i, \sigma^*)$ : The mismatches over pairs  $(b, c)$  where  $b \in V_i$  and  $c \in A \setminus V_i$ , or vice-versa; ii)  $d_2(\pi_i, \sigma^*)$ : The mismatches over pairs  $(b, c)$  where  $b, c \in V_i$ ; and iii)  $d_3(\pi_i, \sigma^*)$ : The mismatches over pairs  $(b, c)$  where  $b, c \in A \setminus V_i$ .

Note that every ranking  $\pi_i$  that satisfies  $\pi_i \rightarrow V_i$  has identical mismatches of type 1. Let us denote the number of such mismatches by  $d_{KT}(V_i, \sigma^*)$ . Also, notice that  $d_2(\pi_i, \sigma^*) = d_{KT}(\pi_i|_{V_i}, \sigma^*|_{V_i})$ , where  $\sigma|_S$  denotes the ranking of alternatives in  $S \subseteq A$  dictated by  $\sigma$ . Similarly,  $d_3(\pi_i, \sigma^*) = d_{KT}(\pi_i|_{A \setminus V_i}, \sigma^*|_{A \setminus V_i})$ . Now, in the expression for the likelihood function  $L(V, a)$ ,

$$\begin{aligned} L(V, a) &= \sum_{\sigma^*: \sigma^*(1)=a} \prod_{i=1}^n \sum_{\pi_i: \pi_i \rightarrow V_i} \varphi^{d_{KT}(V_i, \sigma^*) + d_{KT}(\pi_i|_{V_i}, \sigma^*|_{V_i}) + d_{KT}(\pi_i|_{A \setminus V_i}, \sigma^*|_{A \setminus V_i})} \\ &= \sum_{\sigma^*: \sigma^*(1)=a} \prod_{i=1}^n \varphi^{d_{KT}(V_i, \sigma^*)} \left[ \sum_{\pi_i^1 \in \mathcal{L}(V_i)} \varphi^{d_{KT}(\pi_i^1, \sigma^*|_{V_i})} \right] \cdot \left[ \sum_{\pi_i^2 \in \mathcal{L}(A \setminus V_i)} \varphi^{d_{KT}(\pi_i^2, \sigma^*|_{A \setminus V_i})} \right] \\ &= \sum_{\sigma^*: \sigma^*(1)=a} \prod_{i=1}^n \varphi^{d_{KT}(V_i, \sigma^*)} \cdot Z_\varphi^k \cdot Z_\varphi^{m-k} \propto \sum_{\sigma^*: \sigma^*(1)=a} \varphi^{d_{KT}(V, \sigma^*)} \triangleq \widehat{L}(V, a). \end{aligned}$$

The second equality follows because every ranking  $\pi_i$  that satisfies  $\pi_i \rightarrow V$  can be generated by picking rankings  $\pi_i^1 \in \mathcal{L}(V_i)$  and  $\pi_i^2 \in \mathcal{L}(A \setminus V_i)$ , and concatenating them. The third equality follows from the definition of the normalization constant in the Mallows model. Finally, we denote  $d_{KT}(V, \sigma^*) \triangleq \sum_{i=1}^n d_{KT}(V_i, \sigma^*)$ . It follows that maximizing  $L(V, a)$  amounts to maximizing  $\widehat{L}(V, a)$ . Note that  $d_{KT}(V, \sigma^*)$  counts the number of times alternative  $a$  is approved while alternative  $b$  is not for all  $a, b \in A$  with  $b \succ_{\sigma^*} a$ . That is, let  $n^V(a, -b) \triangleq |\{i \in [n] | a \in V_i \wedge b \notin V_i\}|$ . Then,  $d_{KT}(V, \sigma^*) = \sum_{a, b \in A: b \succ_{\sigma^*} a} n^V(a, -b)$ . Also, note that for alternatives  $c, d \in A$ , we have  $\text{SC}^{\text{APP}}(c) - \text{SC}^{\text{APP}}(d) = n^V(c, -d) - n^V(d, -c)$ .

Next, we show that  $\widehat{L}(V, a)$  is a monotonically increasing function of  $\text{SC}^{\text{APP}}(a)$ . Equivalently,  $\widehat{L}(V, a) \geq \widehat{L}(V, b)$  if and only if  $\text{SC}^{\text{APP}}(a) \geq \text{SC}^{\text{APP}}(b)$ . Fix  $a, b \in A$ . Consider the bijection between the sets of rankings placing  $a$  and  $b$  first, which simply swaps  $a$  and  $b$  ( $\sigma \leftrightarrow \sigma_{a \leftrightarrow b}$ ). Then,

$$\widehat{L}(V, a) - \widehat{L}(V, b) = \sum_{\sigma^*: \sigma^*(1)=a} \varphi^{d_{KT}(V, \sigma^*)} - \varphi^{d_{KT}(V, \sigma_{a \leftrightarrow b}^*)}. \quad (3)$$

Fix  $\sigma^*$  such that  $\sigma^*(1) = a$ . Note that  $\sigma_{a \leftrightarrow b}^*(1) = b$ . Let  $C$  denote the set of alternatives positioned between  $a$  and  $b$  in  $\sigma^*$  (equivalently, in  $\sigma_{a \leftrightarrow b}^*$ ). Now,  $\sigma^*$  and  $\sigma_{a \leftrightarrow b}^*$  have identical disagreements with  $V$  on a pair of alternatives  $(x, y)$  unless i) one of  $x$  and  $y$  belongs to  $\{a, b\}$ , and ii) the other belongs to  $C \cup \{a, b\}$ . Thus, the difference of disagreements of  $\sigma^*$  and  $\sigma_{a \leftrightarrow b}^*$  with  $V$  on such pairs is

$$\begin{aligned} & d_{KT}(V, \sigma^*) - d_{KT}(V, \sigma_{a \leftrightarrow b}^*) \\ &= \left[ n^V(b, -a) - n^V(a, -b) \right] + \sum_{c \in C} [n^V(c, -a) + n^V(b, -c) - n^V(c, -b) - n^V(a, -c)] \\ &= (|C| + 1) \cdot (\text{SC}^{\text{APP}}(b) - \text{SC}^{\text{APP}}(a)). \end{aligned}$$

Thus,  $\text{SC}^{\text{APP}}(a) = \text{SC}^{\text{APP}}(b)$  implies  $d_{KT}(V, \sigma^*) = d_{KT}(V, \sigma_{a \leftrightarrow b}^*)$  (and thus,  $\widehat{L}(V, a) = \widehat{L}(V, b)$ ), and  $\text{SC}^{\text{APP}}(a) > \text{SC}^{\text{APP}}(b)$  implies  $d_{KT}(V, \sigma^*) < d_{KT}(V, \sigma_{a \leftrightarrow b}^*)$  (and thus,  $\widehat{L}(V, a) > \widehat{L}(V, b)$ ). ■

Suboptimality of approval voting for distances other than the KT distance stems from the fact that in counting the number of approvals for a given alternative, one discards information regarding other alternatives approved along with the given alternative in various votes. However, no such information is discarded when only one alternative is approved (or not approved) in each vote. That is, given plurality ( $k = 1$ ) or veto ( $k = m - 1$ ) votes, approval voting should be optimal, not only for the Mallows model but for any reasonable noise model. The next result formalizes this intuition.

**Theorem 2.** *Under a neutral noise model, the set of MLE best alternatives coincides with the set of approval winners*

1. *given plurality votes, if  $p_1 > p_i > 0, \forall i \in \{2, \dots, m\}$ , where  $p_i$  is the probability of the alternative in position  $i$  in the true ranking appearing in the first position in a sample, or*
2. *given veto votes, if  $0 < q_1 < q_i, \forall i \in \{2, \dots, m\}$ , where  $q_i$  is the probability of the alternative in position  $i$  in the true ranking appearing in the last position in a sample.*

*Proof.* We show the proof for plurality votes. The case of veto votes is symmetric: in every vote, instead of a single approved alternative, we have a single alternative that is not approved. Note that the probability  $p_i$  is independent of the true ranking  $\sigma^*$  due to the neutrality of the noise model.

Consider a plurality vote profile  $V$  and an alternative  $a$ . Let  $T = \{\sigma^* \in \mathcal{L}(A) | \sigma^*(1) = a\}$ . The likelihood function for  $a$  is given by  $L(V, a) = \sum_{\sigma^* \in T} \Pr[V | \sigma^*]$ . Under every  $\sigma^* \in T$ , the contribution of the  $\text{SC}^{\text{APP}}(a)$  plurality votes for  $a$  to the product  $\Pr[V | \sigma^*] = \prod_{i=1}^n \Pr[V_i | \sigma^*]$  is  $(p_1)^{\text{SC}^{\text{APP}}(a)}$ . Note that the alternatives in  $A \setminus \{a\}$  are distributed among positions in  $\{2, \dots, m\}$  in all possible ways by the rankings in  $T$ . Let  $i_b$  denote the position of alternative  $b \in A \setminus \{a\}$ . Then,

$$\begin{aligned} L(V, a) &= p_1^{\text{SC}^{\text{APP}}(a)} \cdot \sum_{\{i_b\}_{b \in A \setminus \{a\}} = \{2, \dots, m\}} \prod_{b \in A \setminus \{a\}} p_{i_b}^{\text{SC}^{\text{APP}}(b)} \\ &= (p_1)^{n \cdot k} \cdot \sum_{\{i_b\}_{b \in A \setminus \{a\}} = \{2, \dots, m\}} \prod_{b \in A \setminus \{a\}} \left( \frac{p_{i_b}}{p_1} \right)^{\text{SC}^{\text{APP}}(b)}. \end{aligned}$$

The second transition holds because  $\text{SC}^{\text{APP}}(a) = n \cdot k - \sum_{b \in A \setminus \{a\}} \text{SC}^{\text{APP}}(b)$ . Our assumption in the theorem statement implies  $0 < p_{i_b}/p_1 < 1$  for  $i_b \in \{2, \dots, m\}$ . Now, it can be checked that for  $a, b \in A$ , we have  $\widehat{L}(V, a)/\widehat{L}(V, b) = \sum_{i \in \{2, \dots, m\}} (p_i/p_1)^{\text{SC}^{\text{APP}}(b) - \text{SC}^{\text{APP}}(a)}$ . Thus,  $\text{SC}^{\text{APP}}(a) \geq \text{SC}^{\text{APP}}(b)$  if and only if  $\widehat{L}(V, a) \geq \widehat{L}(V, b)$ , as required. ■

Note that the conditions of Theorem 2 are very mild. In particular, the condition for plurality votes is satisfied under the Mallows model with all five distances we consider, and the condition for veto votes is satisfied under the Mallows model with the Kendall tau, the footrule, and the maximum displacement distances. This is presented as Theorem 4 in the online appendix (Appendix B).

## 4 High Noise and Low Noise

While Theorem 1 shows that there are situations where at least some of the approval winners may not be MLE best alternatives, it does not paint the complete picture. In particular, in both profiles used as counterexamples in the proof of Theorem 1, it holds that every MLE best alternative is an approval winner. That is, the optimal rule choosing an MLE best alternative works as if a tie-breaking scheme is imposed on top of approval voting. Does this hold true for all profiles? Part 2 of Theorem 1 gives a positive answer for the Kendall tau distance. In this section, we answer the foregoing question (largely) in the positive under the other four distance functions, with respect to the two ends of the Mallows spectrum: the case of low noise ( $\varphi \rightarrow 0$ ), and the case of high noise ( $\varphi \rightarrow 1$ ). The case of high noise is especially compelling (because that is when it becomes hard to pinpoint the ground truth), but both extreme cases have received special attention in the literature [24, 21, 11]. In contrast to previous results, which have almost always yielded different answers in the two cases, we show that every MLE best alternative is an approval winner in *both cases*, in almost every situation.

We begin with the likelihood function for alternative  $a$ :  $L(V, a) = \sum_{\sigma^*: \sigma^*(1)=a} \sum_{\pi: \pi \rightarrow V} \varphi^{d(\pi, \sigma^*)}$ . When  $\varphi \rightarrow 0$ , maximizing  $L(V, a)$  requires minimizing the minimum exponent. Ties, if any, are broken using the number of terms achieving the minimum exponent, then the second smallest exponent, and so on. At the other extreme, let  $\varphi = 1 - \epsilon$  with  $\epsilon \rightarrow 0$ . Using the first-order approximation  $(1 - \epsilon)^{d(\pi, \sigma^*)} \approx 1 - \epsilon \cdot d(\pi, \sigma^*)$ , maximizing  $L(V, a)$  requires minimizing the sum of  $d(\pi, \sigma^*)$  over all  $\sigma^*, \pi$  with  $\sigma^*(1) = a$  and  $\pi \rightarrow V$ . Ties are broken using higher-order approximations. Let

$$L_0(V, a) = \min_{\sigma^*: \sigma^*(1)=a} \min_{\pi: \pi \rightarrow V} d(\pi, \sigma^*) \quad L_1(V, a) = \sum_{\sigma^*: \sigma^*(1)=a} \sum_{\pi: \pi \rightarrow V} d(\pi, \sigma^*).$$

We are interested in minimizing  $L_0(V, a)$  and  $L_1(V, a)$ ; this leads to novel combinatorial problems that require detailed analysis. We are now ready for the main result of this section.

**Theorem 3.** *The following statements hold for using approval voting to aggregate  $k$ -approval votes drawn from the Mallows model.*

1. *Under the Mallows model with  $d \in \{d_{FR}, d_{CY}, d_{HM}\}$  and  $\varphi \rightarrow 0$ , and under the Mallows model with  $d \in \{d_{FR}, d_{CY}, d_{HM}, d_{MD}\}$  and  $\varphi \rightarrow 1$ , it holds that for every  $k \in [m - 1]$ , and every profile with  $k$ -approval votes, every MLE best alternative is an approval winner.*
2. *Under the Mallows model with  $d = d_{MD}$  and  $\varphi \rightarrow 0$ , there exists a profile with seven 2-approval votes over 5 alternatives such that no MLE best alternative is an approval winner.*

Before we proceed to the proof, we remark that in part 1 of the theorem, by  $\varphi \rightarrow 0$  and  $\varphi \rightarrow 1$ , we mean that there exist  $0 < \varphi_0^*, \varphi_1^* < 1$  such that the result holds for all  $\varphi \leq \varphi_0^*$  and  $\varphi \geq \varphi_1^*$ , respectively. In part 2 of the theorem, we mean that for every  $\varphi^* > 0$ , there exists a  $\varphi < \varphi^*$  for which the negative result holds. Due to space constraints, we only present the proof for the Mallows model with  $d = d_{FR}$  and  $\varphi \rightarrow 0$ ; the full proof appears in the online appendix (Appendix C).

*Proof of Theorem 3 (only for  $d = d_{FR}$ ,  $\varphi \rightarrow 0$ ).* Let  $\varphi \rightarrow 0$  in the Mallows model with the footrule distance. To analyze  $L_0(V, \cdot)$ , we first analyze  $\min_{\pi: \pi \rightarrow V} d_{FR}(\sigma^*, \pi)$  for a fixed  $\sigma^* \in \mathcal{L}(A)$ . Then, we minimize it over  $\sigma^*$ , and show that the set of alternatives that appear first in the minimizers (i.e., the set of alternatives minimizing  $L_0(V, a)$ ) is exactly the set of approval winners. Since every MLE best alternative in the  $\varphi \rightarrow 0$  case must minimize  $L_0(V, \cdot)$ , the result follows.

Fix  $\sigma^* \in \mathcal{L}(A)$ . Imagine a boundary between positions  $k$  and  $k + 1$  in all rankings, i.e., between the approved and the non-approved alternatives. Now, given a profile  $\pi$  such that  $\pi \rightarrow V$ , we first apply the following operation repeatedly. For  $i \in [n]$ , let an alternative  $a \in A$  be in positions  $t$  and  $t'$  in  $\sigma^*$  and  $\pi_i$ , respectively. If  $t$  and  $t'$  are on the same side of the boundary (i.e., either both are at most  $k$  or both are greater than  $k$ ) and  $t \neq t'$ , then swap alternatives  $\pi_i(t)$  and  $\pi_i(t') = a$  in  $\pi_i$ . Note that this decreases the displacement of  $a$  in  $\pi_i$  with respect to  $\sigma^*$  by  $|t - t'|$ , and increases the displacement of  $\pi_i(t)$  by at most  $|t - t'|$ . Hence, the operation *cannot increase*  $d_{FR}(\pi, \sigma^*)$ . Let  $\pi^*$  denote the profile that we converge to. Note that  $\pi^*$  satisfies  $\pi^* \rightarrow V$  (because we only swap alternatives on the same side of the boundary),  $d_{FR}(\pi^*, \sigma^*) \leq d_{FR}(\pi, \sigma^*)$ , and the following condition:

*Condition X:* for  $i \in [n]$ , every alternative that is on the same side of the boundary in  $\sigma^*$  and  $\pi_i^*$  is in the same position in both rankings.

Because we started from an arbitrary profile  $\pi$  (subject to  $\pi \rightarrow V$ ), it follows that it is sufficient to minimize  $d_{FR}(\pi^*, \sigma^*)$  over all  $\pi^*$  with  $\pi^* \rightarrow V$  satisfying condition X. However, we show that subject to  $\pi^* \rightarrow V$  and condition X,  $d_{FR}(\pi^*, \sigma^*)$  is actually a constant.

Note that for  $i \in [n]$ , every alternative that is in different positions in  $\pi_i^*$  and  $\sigma^*$  must be on different sides of the boundary in the two rankings. It is easy to see that in every  $\pi_i^*$ , there is an equal number of alternatives on both sides of the boundary that are not in the same position as they are in  $\sigma^*$ . Now, we can divide the total footrule distance  $d_{FR}(\pi^*, \sigma^*)$  into four parts:

1. Let  $i \in [n]$  and  $t \in [k]$  such that  $\sigma^*(t) \neq \pi_i^*(t)$ . Let  $a = \sigma^*(t)$  and  $(\pi_i^*)^{-1}(a) = t' > k$ . Then, the displacement  $t' - t$  of  $a$  is broken into two parts: (i)  $t' - k$ , and (ii)  $k - t$ .
2. Let  $i \in [n]$  and  $t \in [m] \setminus [k]$  such that  $\sigma^*(t) \neq \pi_i^*(t)$ . Let  $a = \sigma^*(t)$  and  $(\pi_i^*)^{-1}(a) = t' \leq k$ . Then, the displacement  $t - t'$  of  $a$  is broken into two parts: (i)  $k - t'$ , and (ii)  $t - k$ .

Because the number of alternatives of type 1 and 2 is equal for every  $\pi_i^*$ , we can see that the total displacements of types 1(i) and 2(ii) are equal, and so are the total displacements of types 1(ii) and 2(i). By observing that there are exactly  $n - \text{SC}^{\text{APP}}(\sigma^*(t))$  instances of type 1 for a given value of  $t \leq k$ , and  $\text{SC}^{\text{APP}}(\sigma^*(t))$  instances of type 2 for a given value of  $t > k$ , we conclude that

$$d_{FR}(\pi^*, \sigma^*) = 2 \cdot \left[ \sum_{t=1}^k (n - \text{SC}^{\text{APP}}(\sigma^*(t))) \cdot (k - t) + \sum_{t=k+1}^m \text{SC}^{\text{APP}}(\sigma^*(t)) \cdot (t - k) \right].$$

Minimizing this over  $\sigma^*$  reduces to minimizing  $\sum_{t=1}^m \text{SC}^{\text{APP}}(\sigma^*(t)) \cdot (t - k)$ . By the rearrangement inequality, this is minimized when alternatives are ordered in a non-increasing order of their approval scores. Note that exactly the set of approval winners appear first in such rankings. ■

Theorem 3 shows that under the Mallows model with  $d \in \{d_{FR}, d_{CY}, d_{HM}\}$ , every MLE best alternative is an approval winner for both  $\varphi \rightarrow 0$  and  $\varphi \rightarrow 1$ . We believe that the same statement holds for all values of  $\varphi$ , as we were unable to find a counterexample despite extensive simulations.

**Conjecture 1.** *Under the Mallows model with distance  $d \in \{d_{FR}, d_{CY}, d_{HM}\}$ , every MLE best alternative is an approval winner for every  $\varphi \in (0, 1)$ .*

## 5 Experiments

We perform experiments with two real-world datasets — Dots and Puzzle [20] — to compare the performance of approval voting against that of the rule that is MLE *for the empirically observed distribution of  $k$ -approval votes* (and not for the Mallows model). Mao et al. [20] collected these datasets by asking workers on Amazon Mechanical Turk to rank either four images by the number of dots they contain (Dots), or four states of an 8-puzzle by their distance to the goal state (Puzzle). Hence, these datasets contain ranked votes over 4 alternatives in a setting where a true ranking of the alternatives indeed exists. Each dataset has four different noise levels; higher noise was created by increasing the task difficulty [20]. For Dots, ranking images with a smaller difference in the number of dots leads to high noise, and for Puzzle, ranking states farther away from the goal state leads to high noise. Each noise level of each dataset contains 40 profiles with approximately 20 votes each.

In our experiments, we extract 2-approval votes from the ranked votes by taking the top 2 alternatives in each vote. Given these 2-approval votes, approval voting returns an alternative with the largest number of approvals. To apply the MLE rule, however, we need to learn the underlying distribution of 2-approval votes. To that end, we partition the set of profiles in each noise level of each dataset into training (90%) and test (10%) sets. We use a high fraction of the profiles for training in order to examine the maximum advantage that the MLE rule may have over approval voting.

Given the training profiles (which approval voting simply ignores), the MLE rule learns the probabilities of observing each of the 6 possible 2-subsets of the alternatives given a fixed true ranking.<sup>4</sup> On the test data, the MLE rule first computes the likelihood of each ranking given the votes. Then, it computes the likelihood of each alternative being the best by adding the likelihoods of all rankings that put the alternative first. It finally returns an alternative with the highest likelihood.

We measure the accuracy of both methods by their frequency of being able to pinpoint the correct best alternative. For each noise level in each dataset, the accuracy is averaged over 1000 simulations with random partitioning of the profiles into training and test sets.

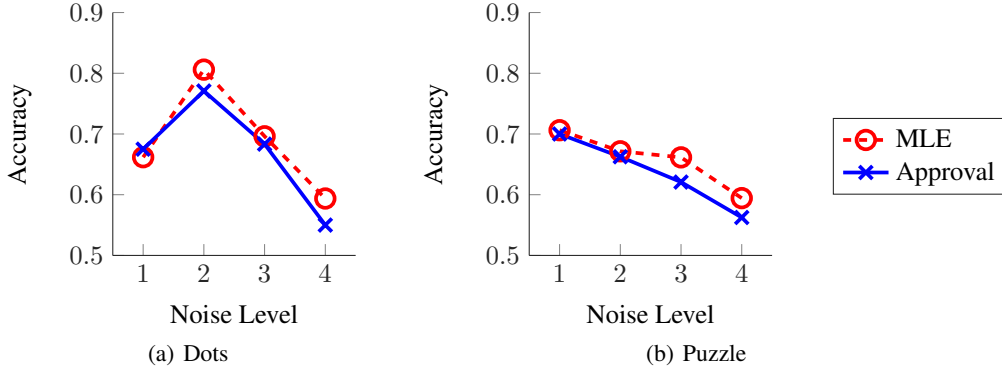


Fig. 1: The MLE rule (trained on 90% of the profiles) and approval voting for 2-approval votes.

Figures 1(a) and 1(b) show that in general the MLE rule does achieve greater accuracy than approval voting. However, the increase is at most 4.5%, which may not be significant in some contexts.

## 6 Discussion

Our main conclusion from the theoretical and empirical results is that approval voting is typically close to optimal for aggregating  $k$ -approval votes. However, the situation is much subtler than it appears at first glance. Moreover, our theoretical analysis is restricted by the assumption that the votes are drawn from the Mallows model. A recent line of work in social choice theory [9, 10] has focused on designing voting rules that perform well — simultaneously — under a wide variety of noise models. It seems intuitive that approval voting would work well for aggregating  $k$ -approval votes under any reasonable noise model; an analysis extending to a wide family of realistic noise models would provide a stronger theoretical justification for using approval voting.

On the practical front, it should be emphasized that approval voting is *not always optimal*. When maximum accuracy matters, one may wish to switch to the MLE rule. However, learning and applying the MLE rule is much more demanding. In our experiments we learn the entire distribution over  $k$ -approval votes given the true ranking. While for 2-approval or 3-approval votes over 4 alternatives we only need to learn 6 probability values, in general for  $k$ -approval votes over  $m$  alternatives one would need to learn  $\binom{m}{k}$  probability values, and the training data may not be sufficient for this purpose. This calls for the design of estimators for the best alternative that achieve greater statistical efficiency by avoiding the need to learn the entire underlying distribution over votes.

<sup>4</sup>Technically, we learn a neutral noise model where the probability of a subset of alternatives being observed only depends on the positions of the alternatives in the true ranking.



## References

- [1] C. Alós-Ferrer. A simple characterization of approval voting. *Social Choice and Welfare*, 27(3):621–625, 2006.
- [2] H. Azari Soufiani, W. Z. Chen, D. C. Parkes, and L. Xia. Generalized method-of-moments for rank aggregation. In *Proc. of 27th NIPS*, pages 2706–2714, 2013.
- [3] H. Azari Soufiani, D. C. Parkes, and L. Xia. Random utility theory for social choice. In *Proc. of 26th NIPS*, pages 126–134, 2012.
- [4] H. Azari Soufiani, D. C. Parkes, and L. Xia. Computing parametric ranking models via rank-breaking. In *Proc. of 31st ICML*, pages 360–368, 2014.
- [5] J. Bartholdi, C. A. Tovey, and M. A. Trick. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare*, 6:157–165, 1989.
- [6] D. Baumeister, G. Erdélyi, E. Hemaspaandra, L. A. Hemaspaandra, and J. Rothe. Computational aspects of approval voting. In *Handbook on Approval Voting*, pages 199–251. Springer, 2010.
- [7] S. J. Brams. *Mathematics and democracy: Designing better voting and fair-division procedures*. Princeton University Press, 2007.
- [8] S. J. Brams and P. C. Fishburn. *Approval Voting*. Springer, 2nd edition, 2007.
- [9] I. Caragiannis, A. D. Procaccia, and N. Shah. When do noisy votes reveal the truth? In *Proc. of 14th EC*, pages 143–160, 2013.
- [10] I. Caragiannis, A. D. Procaccia, and N. Shah. Modal ranking: A uniquely robust voting rule. In *Proc. of 28th AAAI*, pages 616–622, 2014.
- [11] E. Elkind and N. Shah. Electing the most probable without eliminating the irrational: Voting over intransitive domains. In *Proc. of 30th UAI*, pages 182–191, 2014.
- [12] G. Erdélyi, M. Nowak, and J. Rothe. Sincere-strategy preference-based approval voting fully resists constructive control and broadly resists destructive control. *Math. Log. Q.*, 55(4):425–443, 2009.
- [13] P. C. Fishburn. Axioms for approval voting: Direct proof. *Journal of Economic Theory*, 19(1):180–185, 1978.
- [14] P. C. Fishburn and S. J. Brams. Approval voting, condorcet’s principle, and runoff elections. *Public Choice*, 36(1):89–114, 1981.
- [15] A. Goel, A. K. Krishnaswamy, S. Sakshuwong, and T. Aitamurto. Knapsack voting. In *Proc. of Collective Intelligence*, 2015.
- [16] J. Lee, W. Kladwang, M. Lee, D. Cantu, M. Azizyan, H. Kim, A. Limpaecher, S. Yoon, A. Treuille, and R. Das. RNA design rules from a massive open laboratory. *Proceedings of the National Academy of Sciences*, 111(6):2122–2127, 2014.
- [17] G. Little, L. B. Chilton, M. Goldman, and R. C. Miller. TurKit: Human computation algorithms on Mechanical Turk. In *Proc. of 23rd UIST*, pages 57–66, 2010.
- [18] T. Lu and C. Boutilier. Learning Mallows models with pairwise preferences. In *Proc. of 28th ICML*, pages 145–152, 2011.
- [19] C. L. Mallows. Non-null ranking models. *Biometrika*, 44:114–130, 1957.
- [20] A. Mao, A. D. Procaccia, and Y. Chen. Better human computation through principled voting. In *Proc. of 27th AAAI*, pages 1142–1148, 2013.
- [21] A. D. Procaccia, S. J. Reddi, and N. Shah. A maximum likelihood approach for selecting sets of alternatives. In *Proc. of 28th UAI*, pages 695–704, 2012.
- [22] M. R. Sertel. Characterizing approval voting. *Journal of Economic Theory*, 45(1):207–211, 1988.
- [23] N. Shah, D. Zhou, and Y. Peres. Approval voting and incentives in crowdsourcing. In *Proc. of 32nd ICML*, pages 10–19, 2015.
- [24] H. P. Young. Condorcet’s theory of voting. *The American Political Science Review*, 82(4):1231–1244, 1988.

## Online Appendix To: Is Approval Voting Optimal Given Approval Votes?

### A Calculations for the Counterexamples in Theorem 1

Recall from Equation (1) that the likelihood of an alternative  $a$  given  $n$  ( $k$ -approval) votes  $V$  is proportional to

$$L(V, a) \triangleq \sum_{\sigma^*: \sigma^*(1)=a} \sum_{\pi: \pi \rightarrow V} \varphi^{d(\pi, \sigma^*)} = \sum_{\sigma^*: \sigma^*(1)=a} \prod_{i=1}^n \left[ \sum_{\pi_i: \pi_i \rightarrow V_i} \varphi^{d(\pi_i, \sigma^*)} \right].$$

In this section, we illustrate how to use this formula to verify the counterexamples used in the proof of Theorem 1. We show the detailed procedure in the case of the counterexample used for distances  $d_{MD}$ ,  $d_{CY}$ , and  $d_{HM}$ . The counterexample for  $d_{FR}$  can be verified similarly.

Recall the counterexample for  $d \in \{d_{MD}, d_{CY}, d_{HM}\}$ . We have 5 voters providing 2-approval votes over the set of 4 alternatives  $A = \{a, b, c, d\}$ , where

$$V_1 = V_2 = \{b, c\}, V_3 = \{a, d\}, V_4 = \{a, b\}, V_5 = \{a, c\}.$$

We want to verify that under the Mallows model with  $d \in \{d_{MD}, d_{CY}, d_{HM}\}$ , an approval winner (in particular,  $a$ ) is not an MLE best alternative for any value of  $\varphi \in (0, 1)$  by showing that a different alternative  $b$  has greater likelihood. Let  $T_a^*$  (resp.  $T_b^*$ ) denote the set of rankings of  $A$  that place  $a$  (resp.  $b$ ) in the first position, i.e.,

$$\begin{aligned} T_a^* &= \{a \succ b \succ c \succ d, a \succ b \succ d \succ c, \\ &\quad a \succ c \succ b \succ d, a \succ c \succ d \succ b, \\ &\quad a \succ d \succ b \succ c, a \succ d \succ c \succ b\}, \\ T_b^* &= \{b \succ a \succ c \succ d, b \succ a \succ d \succ c, \\ &\quad b \succ c \succ a \succ d, b \succ c \succ d \succ a, \\ &\quad b \succ d \succ a \succ c, b \succ d \succ c \succ a\}. \end{aligned}$$

For  $i \in [5]$ , let  $T_i = \{\pi_i \in \mathcal{L}(A) \mid \pi_i \rightarrow V_i\}$  denote the set of rankings that could generate the  $k$ -approval vote  $V_i$ , i.e.,

$$\begin{aligned} T_1 = T_2 &= \{b \succ c \succ a \succ d, b \succ c \succ d \succ a, c \succ b \succ a \succ d, c \succ b \succ d \succ a\}, \\ T_3 &= \{a \succ d \succ b \succ c, a \succ d \succ c \succ b, d \succ a \succ b \succ c, d \succ a \succ c \succ b\}, \\ T_4 &= \{a \succ b \succ d \succ c, a \succ b \succ c \succ d, b \succ a \succ d \succ c, b \succ a \succ c \succ d\}, \\ T_5 &= \{a \succ c \succ d \succ b, a \succ c \succ b \succ d, c \succ a \succ d \succ b, c \succ a \succ b \succ d\}. \end{aligned}$$

Using these, we can now evaluate the likelihoods

$$L(V, a) = \sum_{\sigma^* \in T_a^*} \prod_{i=1}^5 \left[ \sum_{\pi_i \in T_i} \varphi^{d(\pi_i, \sigma^*)} \right], \quad L(V, b) = \sum_{\sigma^* \in T_b^*} \prod_{i=1}^5 \left[ \sum_{\pi_i \in T_i} \varphi^{d(\pi_i, \sigma^*)} \right].$$

Using a computer program, it can be verified that for  $d \in \{d_{MD}, d_{HM}, d_{CY}\}$ , the difference in likelihoods  $L(V, b) - L(V, a)$ , which is a polynomial in  $\varphi$ , has no roots in  $(0, 1)$ , and is positive at  $\varphi = 0.5$ . Hence, due to the intermediate value theorem, it must be positive in the entire interval. Hence,  $a$  is not an MLE best alternative for any value of  $\varphi \in (0, 1)$ , as required.

### B Mildness of the Conditions in Theorem 2

In this section, we want to prove that the conditions in Theorem 2 are mild, by showing that they hold for popular distance metrics. Let  $T_{i,j}(\sigma^*) = \{\sigma \in \mathcal{L}(A) \mid \sigma^{-1}(\sigma^*(i)) \leq j\}$  be the set of rankings where alternative in position  $i$  in  $\sigma^*$  appears among the top  $j$  positions. Caragiannis et al. [9] study three of our distance metrics: the Kendall tau distance, the footrule distance, and the maximum displacement distance. Combining their Lemma 5.5 and Theorem 5.9, one gets the following lemma.

**Lemma 1** ([9]). For  $d \in \{d_{KT}, d_{FR}, d_{MD}\}$ ,  $\sigma^* \in \mathcal{L}(A)$ ,  $i, i' \in [m]$  with  $i < i'$ , and  $j \in [m - 1]$ , there exists a bijection  $f : T_{i,j}(\sigma^*) \rightarrow T_{i',j}(\sigma^*)$  such that  $d(f(\sigma), \sigma^*) \geq d(\sigma, \sigma^*)$  for all  $\sigma \in T_{i,j}(\sigma^*)$ , and  $d(f(\sigma), \sigma^*) > d(\sigma, \sigma^*)$  for at least one  $\sigma \in T_{i,j}(\sigma^*)$ .

That is, rankings placing a less preferred alternative among the top positions are generally farther away from  $\sigma^*$ ; this is highly intuitive. We use this property to establish our required conditions for the three distance metrics  $\{d_{KT}, d_{FR}, d_{MD}\}$ . For the Cayley and the Hamming distances, unfortunately, Caragiannis et al. [9] give an example demonstrating that they violate the property in Lemma 1. For these metrics, we give a different result that is sufficient for our purpose. First, we need the following characterization (see, e.g., [9]) of the Cayley distance.

**The Cayley distance:** Given two rankings  $\sigma_1$  and  $\sigma_2$ , suppose we want to convert  $\sigma_1$  to  $\sigma_2$  by pairwise swaps of alternatives. We create a directed graph (let us call it the conversion graph) over the alternatives where each alternative  $a$  points to the alternative  $\sigma_1(\sigma_2^{-1}(a))$ , that is, the alternative whose position in  $\sigma_1$  matches the position of  $a$  in  $\sigma_2$ . This indicates that we need to move  $a$  to the position of this alternative in  $\sigma_1$ . It can be checked that this graph is a collection of disjoint directed cycles, and the Cayley distance between the two rankings is  $m$  minus the number of cycles.

We are now ready to prove two useful results.

**Lemma 2.** Let  $d \in \{d_{CY}, d_{HM}\}$ ,  $\sigma_1, \sigma_2 \in \mathcal{L}(A)$ , and  $i, j \in [m]$ . For  $t \in [2]$ , let  $\widehat{\sigma}_t = (\sigma_t)_{\sigma_t(i) \leftrightarrow \sigma_t(j)}$  denote the ranking obtained by swapping alternatives at positions  $i$  and  $j$  in  $\sigma_t$ . Then,  $d(\sigma_1, \sigma_2) = d(\widehat{\sigma}_1, \widehat{\sigma}_2)$ .

*Proof.* For the Hamming distance, this follows immediately from the definition. For the Cayley distance, the result follows directly from the definition of the conversion graph. Indeed, when we swap alternatives in positions  $i$  and  $j$  in two rankings  $\sigma_1$  and  $\sigma_2$ , for every alternative  $a$  the alternative in position  $\sigma_2^{-1}(a)$  in  $\sigma_1$  is the same, before or after the swap. In other words, the function  $\sigma_1(\sigma_2^{-1}(\cdot))$ , which induces the conversion graph, remains intact. Hence, the Cayley distance remains intact, as required. ■ (Proof of Lemma 2)

Note that swapping alternatives in two fixed positions in two rankings is different from swapping two fixed alternatives. Hence, Lemma 2 is not satisfied by all neutral distance metrics.

**Lemma 3.** For  $d \in \{d_{CY}, d_{HM}\}$ ,  $\sigma \in \mathcal{L}(A)$ , and  $i \in [m - 1]$ , there exists a bijection  $f : T_{i,1}(\sigma^*) \rightarrow T_{i+1,1}(\sigma^*)$  such that

1. for  $i \geq 2$ ,  $d(f(\sigma), \sigma^*) = d(\sigma, \sigma^*)$  for all  $\sigma \in T_{i,1}(\sigma^*)$ , and
2. for  $i = 1$ ,  $d(f(\sigma), \sigma^*) > d(\sigma, \sigma^*)$  for all  $\sigma \in T_{1,1}(\sigma^*)$ .

*Proof. The Hamming distance,  $i \geq 2$ :* Let  $d = d_{HM}$ . Fix rankings  $\sigma^* \in \mathcal{L}(A)$  and  $\sigma \in T_{i,1}(\sigma^*)$ . Let  $a_i = \sigma^*(i) = \sigma(1)$  and  $a_{i+1} = \sigma^*(i + 1)$ . Next, let  $\widehat{\sigma}$  and  $\widehat{\sigma}^*$  denote the rankings obtained by swapping the alternatives in positions  $i$  and  $i + 1$  in  $\sigma$  and  $\sigma^*$ , respectively. Lemma 2 implies that  $d(\sigma, \sigma^*) = d(\widehat{\sigma}, \widehat{\sigma}^*)$ . Further,  $i \geq 2$  implies that  $\widehat{\sigma}(1) = \sigma(1) = a_i$ . Also,  $\widehat{\sigma}^*(i + 1) = a_i$  and  $\widehat{\sigma}^*(i) = a_{i+1}$ . Finally, we exchange the labels of alternatives  $a_i$  and  $a_{i+1}$ . Applying this operation on  $\widehat{\sigma}^*$  yields  $\sigma^*$  back. Let  $\tilde{\sigma}$  denote the ranking obtained by applying this operation to  $\widehat{\sigma}$ . Then, clearly,  $\tilde{\sigma}(1) = a_{i+1} = \sigma^*(i + 1)$ . Hence,  $\tilde{\sigma} \in T_{i+1,1}(\sigma^*)$ . Due to the neutrality of the Hamming distance,  $d(\tilde{\sigma}, \sigma^*) = d(\widehat{\sigma}, \widehat{\sigma}^*) = d(\sigma, \sigma^*)$ . The proof is complete by assigning  $f(\sigma) = \tilde{\sigma}$  for every  $\sigma \in T_{i,1}(\sigma^*)$ . The fact that  $f$  is a bijection follows from the observation that its two parts — swapping alternatives in columns  $i$  and  $i + 1$ , and exchanging the labels of alternatives  $a_i$  and  $a_{i+1}$  — are bijections themselves.

**The Cayley distance,  $i \geq 2$ :** Let  $d = d_{CY}$ . For the Cayley distance, we use the same bijection that we used for the Hamming distance. Once again, the first operation (swapping alternatives in positions  $i$  and  $i + 1$ ) keeps the Cayley distance intact due to Lemma 2, and the second operation (exchanging labels of alternatives  $a_i$  and  $a_{i+1}$ ) keeps the Cayley distance intact due to the neutrality of the metric. Hence, we get that  $d(f(\sigma), \sigma^*) = d(\sigma, \sigma^*)$ , as required.

**The Hamming distance,  $i = 1$ :** Let  $d = d_{HM}$ . Let  $\sigma^* \in \mathcal{L}(A)$  and  $\sigma \in T_{1,1}(\sigma^*)$ . Again, let  $a_1 = \sigma^*(1) = \sigma(1)$  and  $a_2 = \sigma^*(2)$ . Let  $t = \sigma^{-1}(a_2)$ . Let  $f(\sigma) = \sigma_{a_1 \leftrightarrow a_2}$ . Observe that not only

$f(\sigma) \in T_{2,1}(\sigma^*)$ , but  $f$  is a bijection from  $T_{1,1}(\sigma^*)$  to  $T_{2,1}(\sigma^*)$ . Next, note that only positions 1 and  $t$  are affected when we swap alternatives  $a_1$  and  $a_2$  in  $\sigma$ . Hence, to show  $d(f(\sigma), \sigma^*) > d(\sigma, \sigma^*)$ , we only need to establish that

$$\mathbb{I}[f(\sigma)(1) \neq \sigma^*(1)] + \mathbb{I}[f(\sigma)(t) \neq \sigma^*(t)] > \mathbb{I}[\sigma(1) \neq \sigma^*(1)] + \mathbb{I}[\sigma(t) \neq \sigma^*(t)]. \quad (4)$$

However, in the LHS, we have  $\mathbb{I}[f(\sigma)(1) \neq \sigma^*(1)] = \mathbb{I}[a_2 \neq a_1] = 1$  and  $\mathbb{I}[f(\sigma)(t) \neq \sigma^*(t)] = \mathbb{I}[a_1 \neq \sigma^*(t)] = 1$ , whereas in the RHS, we have  $\mathbb{I}[\sigma(1) \neq \sigma^*(1)] = \mathbb{I}[a_1 \neq a_1] = 0$ . Hence, Equation (4) holds, as required.

**The Cayley distance,  $i = 1$ :** Let  $d = d_{CY}$ . Let  $\sigma^* \in \mathcal{L}(A)$  and  $\sigma \in T_{1,1}(\sigma^*)$ . Again, let  $a_1 = \sigma^*(1) = \sigma(1)$  and  $a_2 = \sigma^*(2)$ . Define  $f(\sigma) = \sigma_{a_1 \leftrightarrow a_2}$ . Once again,  $f$  is a bijection from  $T_{1,1}(\sigma^*)$  to  $T_{2,1}(\sigma^*)$ . Next, let us compare two conversion graphs: graph  $G_1$  for converting  $\sigma$  to  $\sigma^*$ , and graph  $G_2$  for converting  $f(\sigma)$  to  $\sigma^*$ . It can be checked that  $G_2$  is identical to  $G_1$  except that, instead of having a self-loop at  $a_1$  as in  $G_1$ ,  $a_1$  has an incoming edge that was originally incoming to  $a_2$  in  $G_1$ , and  $a_1$  has an outgoing edge to  $a_2$ . In other words,  $G_1$  has a self-loop at  $a_1$ , which, in  $G_2$ , is absorbed into the loop that contains  $a_2$ . Thus,  $G_2$  has one less cycle than  $G_1$ , i.e.,  $d(f(\sigma), \sigma^*) > d(\sigma, \sigma^*)$ , as required. Note that this is also true in the special case where  $\sigma(2) = a_2$ . In that case,  $G_1$  has self-loops at both  $a_1$  and  $a_2$ , whereas  $G_2$  has a 2-cycle between  $a_1$  and  $a_2$ . ■ (Proof of Lemma 3)

We are now ready to analyze the conditions for plurality and veto votes in Theorem 2.

**Theorem 4.** For  $i \in [m]$ , let  $p_i$  and  $q_i$  denote the probabilities of the alternative in position  $i$  in the true ranking  $\sigma^*$  appearing in the first and the last positions, respectively, under the Mallows model with distance metric  $d$ . Then,

1. For  $d \in \{d_{KT}, d_{FR}, d_{MD}\}$ , we have  $p_1 > p_i > 0$  and  $0 < q_1 < q_i$  for  $i \in \{2, \dots, m\}$ .
2. For  $d \in \{d_{CY}, d_{HM}\}$ , we have  $p_1 > p_i > 0$  for  $i \in \{2, \dots, m\}$ .

*Proof.* Note that  $p_i, q_i > 0$  holds trivially for all  $i \in [m]$  because under the Mallows model (with any distance metric and  $\varphi > 0$ ), every ranking has a positive probability. For  $d \in \{d_{KT}, d_{FR}, d_{MD}\}$ , we directly leverage Lemma 1. Fix  $\sigma^* \in \mathcal{L}(A)$ . Let  $f$  denote the bijection established in Lemma 1 from  $T_{i,1}(\sigma^*)$  to  $T_{i+1,1}(\sigma^*)$ . For  $i \in [m-1]$ , under the Mallows model we have

$$\begin{aligned} p_i - p_{i+1} &= \frac{1}{Z_\varphi^m} \cdot \left[ \sum_{\sigma \in T_{i,1}(\sigma^*)} \varphi^{d(\sigma, \sigma^*)} - \sum_{\sigma \in T_{i+1,1}(\sigma^*)} \varphi^{d(\sigma, \sigma^*)} \right] \\ &= \frac{1}{Z_\varphi^m} \cdot \sum_{\sigma \in T_{i,1}(\sigma^*)} \left[ \varphi^{d(\sigma, \sigma^*)} - \varphi^{d(f(\sigma), \sigma^*)} \right] \\ &> 0, \end{aligned}$$

where the last inequality follows because Lemma 1 ensures  $d(f(\sigma, \sigma^*)) \geq d(\sigma, \sigma^*)$  for every  $\sigma \in T_{i,1}(\sigma^*)$ , and  $d(f(\sigma, \sigma^*)) > d(\sigma, \sigma^*)$  for at least one  $\sigma \in T_{i,1}(\sigma^*)$ . Hence,  $p_i > p_{i+1}$  for all  $i \in [m-1]$ , which directly implies  $p_1 > \max_{i \in \{2, \dots, m\}} p_i$ .

Using a similar argument, we can also see that for  $i \in [m-1]$ ,

$$(1 - q_i) - (1 - q_{i+1}) = \frac{1}{Z_\varphi^m} \cdot \left[ \sum_{\sigma \in T_{i,m-1}(\sigma^*)} \varphi^{d(\sigma, \sigma^*)} - \sum_{\sigma \in T_{i+1,m-1}(\sigma^*)} \varphi^{d(\sigma, \sigma^*)} \right] > 0.$$

Hence, we have that  $q_i < q_{i+1}$  for all  $i \in [m-1]$ ; it follows that  $q_1 < \min_{i \in \{2, \dots, m\}} q_i$ , as required.

The argument for the Hamming and the Cayley distances, i.e., for  $d \in \{d_{HM}, d_{CY}\}$ , is similar, but we use Lemma 3 instead of Lemma 1. It can be checked that this leads to  $p_1 > p_2$  and  $p_i = p_{i+1}$  for  $i \in \{2, \dots, m-1\}$ . This still implies  $p_1 > \max_{i \in \{2, \dots, m\}} p_i$ , as required. ■ (Proof of Theorem 4)

Caragiannis et al. [9] show (in Appendix D.4 in the full version of their paper) that in the case of the Cayley distance (resp., the Hamming distance) with  $m = 3$ , both  $\sigma^*(1)$  and  $\sigma^*(2)$  appear last

in exactly two rankings: one at distance 1 (resp., 3) from  $\sigma^*$ , and another at distance 2 from  $\sigma^*$ . Hence, under any noise model where the probability of a ranking is a function of its distance from  $\sigma^*$  (e.g., under the Mallows model), we have  $q_1 = q_2$ , which violates the condition for veto votes in Theorem 2.

## C Continued Proof of Theorem 3

While for the footrule distance we could show that only approval winners minimize  $L_0(V, a)$ , for the Cayley and the Hamming distances it can be shown that even alternatives with suboptimal approval score might sometimes minimize  $L_0(V, a)$ . In many cases, several higher orders of tie-breaking do not help distinguish approval winners from other alternatives. Hence, we devise a way to completely avoid the analysis of multiple levels of tie-breaking.

First, we say that a set  $S_1$  is lexicographically smaller than another set  $S_2$  if, after sorting their elements in non-decreasing order and comparing from the lowest to the highest, when they differ for the first time  $S_1$  has a smaller element than  $S_2$ . Recall that in the  $\varphi \rightarrow 0$  case, we can determine the MLE best alternative as follows. We consider the set of distances  $D_a = \{d(\pi, \sigma^*)\}_{\sigma^*: \sigma^*(1)=a, \pi: \pi \rightarrow V}$  for each alternative  $a$ . Then, the alternative whose set is lexicographically smallest is the MLE best alternative. There may be multiple tied MLE best alternatives whose sets are identical. We are now ready to prove the following simple and useful lemma.

**Lemma 4.** *Let distance metric  $d \in \{d_{CY}, d_{HM}\}$ . For alternatives  $a, b \in A$ , define*

$$m_{a,-b} = \min_{\substack{\sigma^*: \sigma^*(1)=a \wedge (\sigma^*)^{-1}(b) > k \\ \pi: \pi \rightarrow V}} d(\pi, \sigma^*).$$

*Then, there exists  $\varphi^* > 0$  such that for all  $0 < \varphi \leq \varphi^*$ , alternative  $a$  has higher likelihood than alternative  $b$  if  $m_{a,-b} < m_{b,-a}$ .*

*Proof.* Let  $\mathcal{L}_{a,b}(A) = \{\sigma^* \in \mathcal{L}(A) \mid \sigma^*(1) = a \wedge (\sigma^*)^{-1}(b) \leq k$  (i.e., the set of rankings that put  $a$  first and  $b$  among the top  $k$  positions), and  $\pi = \{\pi \mid \pi \rightarrow V\}$ . For  $\sigma^* \in \mathcal{L}_{a,b}(A)$  and  $\pi \in \pi$ , let  $\hat{\sigma}^*$  and  $\hat{\pi}$  denote the ranking and the profile obtained by swapping alternatives at positions 1 and  $(\sigma^*)^{-1}(b)$  in  $\sigma^*$  and in every  $\pi_i \in \pi$ , respectively. Then, it can be checked that  $(\sigma^*, \pi) \leftrightarrow (\hat{\sigma}^*, \hat{\pi})$  is a bijection from  $\mathcal{L}_{a,b}(A) \times \pi$  to  $\mathcal{L}_{b,a}(A) \times \pi$ . Further, Lemma 2 implies that  $d(\pi, \sigma^*) = d(\hat{\pi}, \hat{\sigma}^*)$ . Hence, when comparing  $D_a$  to  $D_b$ , rankings that put  $a$  first and  $b$  among the top  $k$  positions cancel out with rankings that put  $b$  first and  $a$  among the top  $k$  positions. The result now follows immediately. ■ (Proof of Lemma 4)

We now use Lemma 4 to analyze the Cayley and the Hamming distances.

**$\varphi \rightarrow 0$ , the Cayley and the Hamming distances:** We show that if alternative  $a$  is an approval winner while alternative  $b$  is not, then there exists  $\varphi^* > 0$  such that for all  $0 < \varphi \leq \varphi^*$ , alternative  $a$  has a higher likelihood of being the best alternative than  $b$ . Due to Lemma 4, it is sufficient to show that  $m_{a,-b} < m_{b,-a}$ .

Fix a ranking  $\sigma^* \in \mathcal{L}(A)$ . We can show that the profile  $\pi$  with  $\pi \rightarrow V$  that minimizes the distance from  $\sigma^*$  must also satisfy condition  $X$  in the proof for the footrule distance. For  $i \in [n]$ , define  $A_i = (V_i \cap \sigma^*([k])) \cup ((A \setminus V_i) \cap \sigma^*([m] \setminus [k]))$ . In words,  $A_i$  is the set of alternatives that we can place in the same position in  $\pi_i$  as in  $\sigma^*$  given the restriction  $\pi_i \rightarrow V_i$ . Condition  $X$  says that  $\pi_i$  and  $\sigma^*$  would indeed agree on the positions of the alternatives in  $A_i$ , for each  $i \in [n]$ .

For the Hamming distance, observe that the restriction  $\pi \rightarrow V$  ensures that  $d_{HM}(\pi_i, \sigma^*) \geq |A \setminus A_i|$ . The lower bound is achieved by satisfying condition  $X$ , i.e., putting all alternatives in  $A_i$  in the same positions in  $\pi_i$  as they are in  $\sigma^*$ .

For the Cayley distance, note that

$$d_{CY}(\pi_i, \sigma^*) \geq \frac{1}{2} \cdot d_{HM}(\pi_i, \sigma^*) \geq \frac{1}{2} \cdot |A \setminus A_i|.$$

Once again, to achieve this lower bound, we must align  $\pi_i$  with  $\sigma^*$  on the alternatives in  $A_i$ , and pair up alternatives from  $A \setminus A_i$  in  $\pi_i$  such that for every  $a \in A \setminus A_i$ , there exists a unique  $b \in A \setminus A_i$

with  $\pi_i((\sigma^*)^{-1}(a)) = b$  and  $\pi_i((\sigma^*)^{-1}(b)) = a$ . The pairing is possible because the number of alternatives from  $A \setminus A_i$  on each side of the boundary in  $\pi_i$  must be equal. Therefore, the lower bound on the distance can indeed be achieved by swapping these paired alternatives.

Define

$$P(\sigma^*) = \sum_{i=1}^n |A \setminus A_i| = \sum_{t=1}^k (n - \text{SC}^{\text{APP}}(\sigma^*(t))) + \sum_{t=k+1}^m \text{SC}^{\text{APP}}(\sigma^*(t)).$$

Then, we have established that for the Hamming distance,  $\min_{\pi: \pi \rightarrow V} d_{HM}(\sigma^*, \pi) = P(\sigma^*)$ , and for the Cayley distance,  $\min_{\pi: \pi \rightarrow V} d_{CY}(\sigma^*, \pi) = (1/2) \cdot P(\sigma^*)$ . Further, note that  $P(\sigma^*) = C - \sum_{t=1}^k \text{SC}^{\text{APP}}(\sigma^*(t))$ , where  $C$  is independent of  $\sigma^*$ . Hence, minimizing  $\min_{\pi: \pi \rightarrow V} d(\sigma^*, \pi)$  reduces to, in both cases, maximizing  $\sum_{t=1}^k \text{SC}^{\text{APP}}(\sigma^*(t))$ .

We are not done yet, because a ranking can maximize the sum of approval scores of its top  $k$  alternatives while not placing the alternative with the maximum approval score at the top. However, Lemma 4 comes to our rescue. Let

$$w_{a,-b} = \max_{\sigma^*: \sigma^*(1)=a \wedge (\sigma^*)^{-1}(b) > k} \sum_{t=1}^k \text{SC}^{\text{APP}}(\sigma^*(t)).$$

Then, we only need to show that  $w_{a,-b} > w_{b,-a}$  for any two alternatives  $a, b \in A$  such that  $a$  is an approval winner, while  $b$  is not.

By the choice of  $a$  and  $b$ ,  $\text{SC}^{\text{APP}}(a) > \text{SC}^{\text{APP}}(b)$ . For  $i \in [m]$ , let  $S_i$  denote the sum of the  $i$  highest approval scores. If the approval score of  $b$  is among the  $k$  highest approval scores, then clearly  $w_{a,-b} = S_{k+1} - \text{SC}^{\text{APP}}(b)$  and  $w_{b,-a} = S_{k+1} - \text{SC}^{\text{APP}}(a)$ . Hence,  $w_{a,-b} > w_{b,-a}$ , as required.

If the approval score of  $b$  is not among the  $k$  highest approval scores, then there are two cases.

1. *There are at most  $k$  approval winners.* In this case, we can see that  $w_{a,-b} = S_k$  while  $w_{b,-a} < S_k$  (because in the latter case, we need to find a ranking that does not put  $a$  among the top  $k$  positions).
2. *There are  $t$  approval winners, where  $t > k$ .* In this case, while  $w_{a,-b} = w_{b,-a} = S_k$ , and thus the exponents  $m_{a,-b}$  and  $m_{b,-a}$  in the likelihood expression — Equation (1) — are equal, the number of terms achieving this exponent in the likelihood function of  $b$  is proportional to  $\binom{t}{k}$ , while the number of terms achieving this exponent in the likelihood function of  $a$  is proportional to  $\binom{t-1}{k}$  (because in the latter case, an approval winner —  $a$  — is not allowed to be among the top  $k$  positions). Hence, the tie is broken in favor of  $a$ .

**$\varphi \rightarrow 0$ , the maximum displacement distance:**

Consider the following vote profile  $V$  over the set of alternatives  $A = \{a, b, c, d, e\}$ .

$$V_1 = \{a, b\} \quad V_2 = \{a, c\} \quad V_3 = \{b, c\} \quad V_4 = \{c, d\} \quad V_5 = \{c, e\} \quad V_6 = V_7 = \{d, e\}.$$

Note that alternative  $c$  is the unique approval winner with an approval score of 4. However, it can be checked that  $L_0(V, d) = L_0(V, e) = 11 < 13 = L_0(V, c)$ . Hence, neither of the two MLE best alternatives ( $d$  and  $e$ ) is an approval winner, as required.

**$\varphi \rightarrow 1$ , all five distances:** The case of  $\varphi \rightarrow 1$  is easier because, as we will show, the analysis of  $L_1(V, \cdot)$  is sufficient to differentiate approval winners from other alternatives for all five distance metrics. Let the distance metric be denoted  $d \in \{d_{KT}, d_{FR}, d_{CY}, d_{HM}, d_{MD}\}$ . Note that for an alternative  $a \in A$ ,

$$\begin{aligned} L_1(V, a) &= \sum_{\sigma^*: \sigma^*(1)=a} \sum_{\pi: \pi \rightarrow V} d(\pi, \sigma^*) \\ &= \sum_{\sigma^*: \sigma^*(1)=a} \sum_{i=1}^n \sum_{\pi_i: \pi_i \rightarrow V_i} d(\pi_i, \sigma^*) \\ &= \sum_{i=1}^n \sum_{\pi_i: \pi_i \rightarrow V_i} \left( \sum_{\sigma^*: \sigma^*(1)=a} d(\pi_i, \sigma^*) \right). \end{aligned} \tag{5}$$

Now, define  $S_t(\pi_i) = \sum_{\sigma^* \in \mathcal{L}(A): \sigma^*(1) = \pi_i(t)} d(\sigma^*, \pi_i)$ . That is,  $S_t(\pi_i)$  is the sum of distances of  $\pi_i$  from all rankings that put its  $t^{\text{th}}$  ranked alternative first. Due to neutrality of the distance metric  $d$ , this quantity is independent of the ranking  $\pi_i$ . Substituting this into Equation (5), we get

$$\begin{aligned}
L_1(V, a) &= \sum_{i=1}^n \sum_{\pi_i: \pi_i \rightarrow V_i} S_{\pi_i^{-1}(a)} \\
&= \sum_{i=1}^n \mathbb{I}[a \in V_i] \cdot \left[ k! \cdot (m-k-1)! \cdot \sum_{j=1}^k S_j \right] + \mathbb{I}[a \notin V_i] \cdot \left[ (k-1)! \cdot (m-k)! \cdot \sum_{j=k+1}^m S_j \right] \\
&= k! \cdot (m-k)! \cdot \left[ \text{SC}^{\text{APP}}(a) \cdot \frac{\sum_{j=1}^k S_j}{k} + (n - \text{SC}^{\text{APP}}(a)) \cdot \frac{\sum_{j=k+1}^m S_j}{m-k} \right] \\
&= k! \cdot (m-k)! \cdot n \cdot \frac{\sum_{j=k+1}^m S_j}{m-k} + k! \cdot (m-k)! \cdot \text{SC}^{\text{APP}}(a) \cdot \left[ \frac{\sum_{j=1}^k S_j}{k} - \frac{\sum_{j=k+1}^m S_j}{m-k} \right].
\end{aligned}$$

To show that this quantity is minimized when  $a$  is an approval winner, we need to show that

$$P_k = \frac{\sum_{j=1}^k S_j}{k} - \frac{\sum_{j=k+1}^m S_j}{m-k} \quad (6)$$

satisfies  $P_k < 0$  for all  $k \in [m-1]$ . To that end, we show the following lemma.

**Lemma 5.** *The following statements hold.*

1. For distance  $d \in \{d_{KT}, d_{FR}, d_{MD}\}$ , we have  $S_i < S_{i+1}$  for all  $i \in [m-1]$ .
2. For distance  $d \in \{d_{CY}, d_{HM}\}$ , we have  $S_1 < S_2$  and  $S_i = S_{i+1}$  for all  $i \in \{2, \dots, m-1\}$ .

*Proof.* Consider a bijection  $f$  from  $T_{i,1}(\sigma^*)$  to  $T_{i+1,1}(\sigma^*)$ . Then, for  $i \in [m-1]$ ,

$$\begin{aligned}
S_{i+1} - S_i &= \sum_{\sigma \in T_{i+1,1}(\sigma^*)} d(\sigma, \sigma^*) - \sum_{\sigma \in T_{i,1}(\sigma^*)} d(\sigma, \sigma^*) \\
&= \sum_{\sigma \in T_{i,1}(\sigma^*)} d(f(\sigma), \sigma^*) - d(\sigma, \sigma^*).
\end{aligned}$$

Now, for  $d \in \{d_{KT}, d_{FR}, d_{MD}\}$ , Lemma 1 ensures the existence of a bijection  $f$  for which  $d(f(\sigma), \sigma^*) \geq d(\sigma, \sigma^*)$  for all  $\sigma \in T_{i,1}(\sigma^*)$ , and  $d(f(\sigma), \sigma^*) > d(\sigma, \sigma^*)$  for at least one  $\sigma \in T_{i,1}(\sigma^*)$ . Hence, it follows that  $S_{i+1} > S_i$  for all  $i \in [m-1]$ , as required.

For  $d \in \{d_{CY}, d_{HM}\}$ , Lemma 3 ensures the existence of a bijection  $f$  such that for all  $\sigma \in T_{i,1}(\sigma^*)$ , we have  $d(f(\sigma), \sigma^*) = d(\sigma, \sigma^*)$  if  $i \geq 2$ , and  $d(f(\sigma), \sigma^*) > d(\sigma, \sigma^*)$  if  $i = 1$ . Hence, it follows that  $S_2 > S_1$  and  $S_{i+1} = S_i$  for all  $i \in \{2, \dots, m-1\}$ . ■ (Proof of Lemma 5)

Using Lemma 5, it is easy to check that  $P_k < 0$  in Equation (6), as required. ■ (Proof of Theorem 3)