# Agreement-based Aggregation of Rankings

Emin, Nihar, Vince

## Contents

**1 Roadmap** **1**

**2 NP-hardness of the perfect agreement problem** **2**

**3 Agreement Helps with Generalization** **6**

**4 Axiomatic analysis** **7**
4.1 Consistency axioms . . . 7
4.2 Reversal symmetry . . . 9
4.3 'Preserved' axioms. . . . 10
4.3.1 Monotonicity . . . 10
4.3.2 Homogenity . . . 11
4.4 Shuffling axioms . . . 12

**5 Questions to think about** **13**

# 1 Roadmap

Suppose you have a collection of items, and evaluators who each rank some subsets of items. Aggregating such data into a total ranking of items is an age-old problem. The most common approach to this problem is to assume a statistical model (e.g., Plackett-Luce, Mallows, i.i.d. flips) for the data generating process. The aggregating ranking is then typically chosen to be the one that maximizes the likelihood of the data under the assumed model (or computationally efficient computations of it if the maximum likelihood estimate is intractable). A second common approach is to adopt a "voting rule" from literature on social choice theory. Such voting rules are typically designed and selected via "axiomatic" approaches that identify a set of necessary conditions for the aggregation method and select aggregation methods satisfying these conditions. However, these axioms usually consider settings in which every evaluator ranks all subsets.

In this work our goal is to obtain such an aggregate ranking. We do not wish to assume a specific statistical model a priori, and the axiomatic approaches in the literature are inapplicable in our setting where each evaluator is given a specific (non random) subset to rank.

We thus take an alternative approach. Our motivation stems from the following question: suppose that the entire process – evaluators independently ranking the items and the algorithm then ranking them – is repeated again. Then which algorithm would lead to the ranking that is most consistent across the two copies of the process? Of course, a useless algorithm which always outputs a fixed ranking is maximally consistent, and in order to avoid such pathological outcomes, we will further impose the basic "anonymity" condition that relabeling the items should result in an identical relabeling of the output. Thus given a (possibly infinite) set of aggregation methods which satisfy the anonymity condition, the goal is to pick the method that maximizes the agreement in outcomes between the two copies of the process. With this idea in place, we now discuss the key challenges associated with realizing this idea, and our approaches to address these challenges.

- *Challenge:* In practice, we have only one copy of the process.
  - We split the evaluators uniformly at random into two groups. We consider each of these two groups as one copy of the process (albeit with only half the number of reviewers).
- *Challenge:* Such a split may result in an uneven split of the samples for any item, for instance, some item may have all its evaluations in only one split. In that case, the outcomes for different items may have uneven reliabilities.
  - We reweigh the loss to account for uneven splits via a principled method. In particular, the weight associated to an item is *exponential* in the evenness of its split.
- *Challenge:* If the number of methods is infinite, computing the best method may be computationally challenging.
  - Here we restrict attention to (the union of) specific classes of methods. One class is the set of all position voting rules for which we present a principled approach. For other classes that have only one or few parameters such as the Plackett-Luce model or the Mallow's model, we opt for a grid search for this parameter's value.

We execute our approach on a proposal reviewing dataset...

# 2 NP-hardness of the perfect agreement problem

Consider the following computational problem, specifically for choosing the best *positional scoring rule*: Given voters $N = [n]$ and candidates $K$ with $|K| = m$, say each voter $i \in N$ ranks a subset $K_i \subseteq K$ with $|K_i| = k$. Say we are also given a split $N = N_1 \sqcup N_2$ with $|N_1| = |N_2| = n/2$ and for each $\beta \in \{1, 2\}, a \in K$ and $i \in [k]$, $M_\beta[a, i]$ indicates the number of voters in $N_\beta$ that ranked $a$ as their $i^{\text{th}}$ candidate. Then, PerfPos asks: is there a *positional scoring rule* that outputs exactly the same ranking when applied to the two sets of rankings, defined by $N_1$ and $N_2$. More formally, is there a vector $\boldsymbol{s} = (s_1, s_2, \ldots, s_k)$ with $1 = s_1 \geq s_2 \geq \ldots s_k = 0$ such that for all $a, b \in K$:

$$(T_1[a] - T_1[b])(T_2[a] - T_2[b]) > 0,$$

where $T_\beta[a] = \sum_{i=1}^{k} M_\beta[a, i]s_i$ for all $a \in K$ and $\beta \in \{1, 2\}$.

**Theorem 1.** PerfPos *is* NP*-complete.*

*Proof.* Membership follows from the fact that $T_\beta$ is easy to compute given $\boldsymbol{s}$. For hardness, we will be reducing from 3SAT. Say $\phi$ is a 3CNF formula with clauses $C_1, C_2, \ldots, C_\ell$ and binary variables $x_1, x_2, \ldots, x_t$. We will construct and instance of PerfPos by first specifying $M_\beta$ for each candidate, and then explicitly designing rankings that is consistent with that $M_\beta$. We start with setting $k = t + 2$ and $\varepsilon = \frac{1}{7(k+2)}$.

For each $i \in [t]$, add two candidates $a_i$ and $b_i$, with

$$M_1[a_i, j] = \begin{cases} 1 + (k+3)(i-1) & \text{if } j = 1 \\ k+2 & \text{if } j = i+1 \\ 0 & \text{otherwise} \end{cases} \qquad M_1[b_i, j] = \begin{cases} (k+3)(i-1) & \text{if } j = 1 \text{ and } i \neq 1 \\ k+2 & \text{if } j = i \\ 1 & \text{if } j = k \\ 0 & \text{otherwise} \end{cases}$$

$$M_2[a_i, j] = \begin{cases} 1 + (1/\varepsilon+1)(i-1) & \text{if } j = 1 \\ 1/\varepsilon & \text{if } j = i+1 \\ 0 & \text{otherwise} \end{cases} \qquad M_2[b_i, j] = \begin{cases} (1/\varepsilon+1)(i-1) & \text{if } j = 1 \text{ and } i \neq 1 \\ 1/\varepsilon & \text{if } j = i \\ 1 & \text{if } j = k \\ 0 & \text{otherwise} \end{cases}.$$

where if any of the two entries For each $i \in [\ell]$, say the clause $C_i$ in $\phi$ consist of variables $\{x_j\}_{j \in V_i}$ for $V_i \subseteq [t]$, with $1 \leq |V_i| \leq 3$. For this clause $C_i$, add two candidates $c_j, d_j$, with

$$M_1[c_i, j] = \begin{cases} t(k+3) + 2i & \text{if } j = 1 \\ 0 & \text{otherwise} \end{cases} \qquad M_1[d_i, j] = \begin{cases} t(k+3) + 2i - 1 & \text{if } j = 1 \\ 1 & \text{if } j = k \\ 0 & \text{otherwise} \end{cases}.$$

For constructing $M_2[c_i, j]$ and $M_2[d_i, j]$, say $z \in \{0, 1, 2, 3\}$ is the number of negated clauses in $C_i$. To build $M_2[c_i, j]$ and $M_2[d_i, j]$ for each $j \in [k]$, start from the following:

$$M_2'[c_i, j] = \begin{cases} 2z + t(1/\varepsilon + 1) + 6(k+3)(i-1) & \text{if } j = 1 \\ 1 & \text{if } j = k \\ 0 & \text{otherwise} \end{cases}$$

$$M_2'[d_i, j] = \begin{cases} 1 + t(1/\varepsilon + 1) + 6(k+3)(i-1) & \text{if } j = 1 \\ 2z & \text{if } j = k \\ 0 & \text{otherwise} \end{cases}.$$

Now for each $j \in V_i$, do the following:

- If $x_j$ appears non-negated in $C_i$, then add $2(k+2)$ to $M_2'[c_i, j]$ and $2(k+2)$ to $M_2'[d_i, j+1]$.
- If $x_j$ appears negated in $C_i$, then add $2(k+2)$ to $M_2'[c_i, j+1]$ and $2(k+2)$ to $M_2'[d_i, j]$.

Finally, set $M_2[c_i, j]$ and $M_2[d_i, j]$ to the resulting $M_2'[c_i, j]$ and $M_2'[d_i, j]$ for each $j \in [k]$.

Say $A = \{a_i\}_{i \in [t]}, B = \{b_i\}_{i \in [t]}, C = \{C_i\}_{i \in [\ell]}$, and $D = \{d_i\}_{i \in [\ell]}$. We now construct a profile that corresponds to the above $M_\beta$. First, add $k$ more candidates $E = \{e_i\}_{i \in [k]}$. Now for each for each $\beta \in \{1, 2\}$, each $f \in A \cup B \cup C \cup D$, and each $i \in [k]$, add $M_\beta[f, i]$ voters to the set $N_\beta$ that rank $f$ in the $i^{\text{th}}$ position, and ranks $e_j$ in the $j^{\text{th}}$ position for all $j \in [k] \setminus \{i\}$. Say we have added $n_1$ and $n_2$ voters to $N_1$ and $N_2$ so far respectively and that $\beta' = \operatorname{argmax}_{\beta \in \{1,2\}} n_\beta$. Add $(14k + 28) n_{\beta'}$ and $(14k + 29) n_{\beta'} - n_{3-\beta'}$ voters to $N_{\beta'}$ and $N_{3-\beta'}$, respectively, all of whom rank $e_i$ in the $i^{\text{th}}$ position for all $i \in [k]$. This also ensures that $|N_1| = |N_2|$. For each $\beta \in \{1, 2\}$ say $\boldsymbol{\sigma}_\beta$ is the final vector of rankings of $N_\beta$, as specified.

By construction, for each $\beta \in \{1, 2\}$, each $f \in A \cup B \cup C \cup D$, and each $i \in [k]$, $\boldsymbol{\sigma}_\beta$ indeed has $M_\beta[f, i]$ voters that rank $f$ in their $i^{\text{th}}$ position. Further, we have a total of $|K| = 2(t + \ell) + k = 3t + 2\ell + 2$ candidates (which polynomial in $t, m$), and $|N_1| = |N_2| = (14k + 29) \max_{\beta \in \{1,2\}} \sum_{f \in A \cup B \cup C \cup D} \sum_{i=1}^{k} M_\beta[f, i] \leq (14k + 29) 2(t + \ell) k \cdot \max_{f \in A \cup B \cup C \cup D, i \in [k]} M_\beta[f, i]$ (which is also polynomial in $t, m$ since all entries of $M_\beta[f, i]$ are).

We now claim that $\phi$ is satisfiable if and only if there exists a positional scoring rule that gives full agreement between $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$.

$(\Leftarrow)$ : Assume there is a positional scoring rule $\boldsymbol{s}$ that gives full agreement for $M_1$ and $M_2$. Say $T_\beta[a] = \sum_{i=1}^{k} M_\beta[a, i] s_i$ for all $a \in K$ and $\beta \in \{1, 2\}$. In particular, we must have agreement between $a_i$ and $b_i$ for each $i \in [t]$. We have:

$$\begin{aligned} T_1[a_i] &= (1 + (k+3)(i-1)) s_1 + (k+2) s_{i+1} & T_1[b_i] &= ((k+3)(i-1)) s_1 + (k+2) s_i + s_k \\ T_2[a_i] &= (1 + (1/\varepsilon + 1)(i-1)) s_1 + s_{i+1}/\varepsilon & T_2[b_i] &= ((1/\varepsilon + 1)(i-1)) s_1 + s_i/\varepsilon + s_k \end{aligned}$$

Agreement implies that we must have:

$$(T_1[a_i] - T_1[b_i])(T_2[a_i] - T_2[b_i]) = (1 - (k+2)(s_i - s_{i+1}))(1 - (s_i - s_{i+1})/\varepsilon) > 0$$

This implies we must either have $s_i - s_{i+1} < \varepsilon$ or $s_i - s_{i+1} > \frac{1}{k+2}$. Set the binary variable $x_i$ to False if $s_i - s_{i+1} < \varepsilon$ and to True if $s_i - s_{i+1} > \frac{1}{k+1}$. We now argue that the resulting $\{x_i\}_{i \in [t]}$ satisfies $\phi$, i.e., satisfies all of its clauses. Fix any $i \in [\ell]$. We will show that $C_i$ is satisfied. By assumption of full agreement,

$$(T_1[c_i] - T_1[d_i])(T_2[c_i] - T_2[d_i]) > 0.$$

Since $T_1[c_i]-T_1[d_i] = (t(k+3)+2i)s_1 - (t(k+3)+2i-1)s_1 - s_k = 1 > 0$, this implies that $(T_2[c_i]-T_2[d_i]) > 0$. Say $\{x_j\}_{j\in V_i}$ are the variables that appear in $C_i$ (for $V_i \subseteq [t]$), and that $Z = \{j \in V_i : x_j \text{ is negated in } C_i\}$ with $|Z| = z$. This implies we have

$$T_2[c_i] = (2z + t(1/\varepsilon+1) + 6(k+3)(i-1))s_1 + s_k + 2(k+2)\left(\sum_{g\in V_i\setminus Z} s_g + \sum_{g\in Z} s_{g+1}\right)$$

$$T_2[d_i] = (1 + t(1/\varepsilon+1) + 6(k+3)(i-1))s_1 + zs_k + 2(k+2)\left(\sum_{g\in V_i\setminus Z} s_{g+1} + \sum_{g\in Z} s_g\right)$$

$$T_2[c_i] - T_2[d_i] = 2z - 1 + 2(k+2)\left(\sum_{g\in V_i\setminus Z}(s_g - s_{g+1}) - \sum_{g\in Z}(s_g - s_{g+1})\right).$$

The only way for $C_i$ to be not satisfied is if $x_g$ is assigned to True (i.e. $s_g - s_{g+1} > \frac{1}{k+2}$) for all $g \in Z$ and $x_h$ is assigned to False (i.e. $s_h - s_{h+1} < \varepsilon$) for all $h \in V_i \setminus Z$. This would imply, however $T_2[c_i] - T_2[d_i] < 2z-1+2(k+2)\left(\sum_{g\in V_i\setminus Z}\varepsilon - \sum_{g\in Z}\frac{1}{k+2}\right) = 2z-1+\frac{2(|V_i|-z)}{7} - 2z < -1+\frac{6}{7} < 0$, which gives a contradiction.

Since assuming $C_i$ is not satisfied gives a contradiction to the assumption that $\boldsymbol{s}$ gives agreement between $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ for $c_i$ and $d_i$, $C_i$ must be satisfied. Since this is true for all $i \in [\ell]$, this implies that $\phi$ is satisfiable.

$(\Rightarrow)$ : Assume $\phi$ is satisfiable for truth assignments $\{x_i^*\}_{i\in[t]}$. For each $i \in [t]$, define $\delta_i = \begin{cases} \frac{1}{k+1} & \text{if } x_i \text{ is True} \\ \frac{\varepsilon}{2} & \text{otherwise.}\end{cases}$.

Define the positional scoring rule $\boldsymbol{s} = \{s_i\}_{i\in[k]}$ as $s_i = \begin{cases} 1 & \text{if } i = 1 \\ 1 - \left(\sum_{j=1}^{i-1}\delta_j\right) & \text{if } k > i > 1 \\ 0 & \text{if } i = k\end{cases}$. Since $s_{k-1} = 1 - \left(\sum_{j=1}^{k-2}\delta_j\right) = \geq 1 - \frac{k-2}{k+1} > 0 = s_k$, monotonicity is satisfied, and $\boldsymbol{s}$ is a valid scoring rule. We will show that $\boldsymbol{s}$ gives perfect agreement between $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$. As always, say $T_\beta[f] = \sum_{i=1}^{k} M_\beta[f,j]$ for all $\beta \in \{1,2\}$ and $f \in K$. Since $1 = s_1 \geq s_2 \geq \ldots s_k = 0$ and since $s_i > 0$ for each $i \in [k-1]$, the total scores for each $i \in [t]$ and $j \in [\ell]$ are

$$T_1[a_i] = (1 + (k+3)(i-1))s_1 + (k+2)s_{i+1} \Rightarrow (k+3)i \geq T_1[a_i] > (k+3)(i-1) \tag{A1}$$
$$T_2[a_i] = (1 + (1/\varepsilon+1)(i-1))s_1 + (1/\varepsilon)s_{i+1} \Rightarrow (1/\varepsilon+1)i \geq T_2[a_i] > (1/\varepsilon+1)(i-1) \tag{A2}$$
$$T_1[b_i] = (k+3)(i-1)s_1 + (k+2)s_i + s_k \Rightarrow (k+3)i > T_1[b_i] > (k+3)(i-1) \tag{B1}$$
$$T_2[b_i] = (1/\varepsilon+1)(i-1)s_1 + (1/\varepsilon)s_i + s_k \Rightarrow (1/\varepsilon+1)i > T_2[b_i] > (1/\varepsilon+1)(i-1) \tag{B2}$$
$$T_1[c_j] = (t(k+3)+2j)s_1 \Rightarrow t(k+3)+2j = T_1[c_j] > t(k+3)+2(j-1) \tag{C1}$$
$$\begin{aligned} T_2[c_j] &= (2|Z_j| + t(1/\varepsilon+1) + 6(k+3)(j-1))s_1 + s_k + 2(k+2)\left(\sum_{g\in V_i\setminus Z} s_g + \sum_{g\in Z} s_{g+1}\right) \\ &\Rightarrow t(1/\varepsilon+1) + 6(k+3)j \geq T_2[c_j] > t(1/\varepsilon+1) + 6(k+3)(j-1)\end{aligned} \tag{C2}$$
$$T_1[d_j] = (t(k+3)+2j-1)s_1 \Rightarrow t(k+3)+2j > T_1[d_j] > t(k+3)+2(j-1) \tag{D1}$$
$$\begin{aligned} T_2[d_j] &= (1 + t(1/\varepsilon+1) + 6(k+3)(j-1))s_1 + 2|Z_j|s_k + 2(k+2)\left(\sum_{g\in V_j\setminus Z_j} s_{g+1} + \sum_{g\in Z_j} s_g\right) \\ &\Rightarrow t(1/\varepsilon+1) + 6(k+3)j > T_2[d_j] > t(1/\varepsilon+1) + 6(k+3)(j-1)\end{aligned} \tag{D2}$$

where $V_j \subseteq [t]$ and $Z_j \subseteq V_j$ indicate the indices of the variables that appear in and that appear negated in the clause $C_j$, respectively. We now show that $\boldsymbol{s}$ gives perfect agreement between $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$, i.e.,

$(T_1[g]-T_1[h])(T_2[g]-T_2[h])>0$ for all distinct pairs of $g,h\in K=A\sqcup B\sqcup C\sqcup D\sqcup E$. We will proceed by a case by case analysis:

**Case 1:** $g\in A\sqcup B$, $h\in C\sqcup D$. By (A1) and (B1), we have $T_1[g]\le t(k+3)$, since $i\le t$. By (C1) and (D1), we have $T_1[h]>t(k+3)$, as $j\ge 1$. This implies $T_1[g]<T_1[h]$. Similarly, by (A2) and (B2) we have $T_2[g]\le t(1/\varepsilon+1)$, and by (C2) and (D2) we have $T_2[h]>t(1/\varepsilon+1)$, as $j\ge 1$. This implies $T_2[g]<T_2[h]$. Hence, $(T_1[g]-T_1[h])(T_2[g]-T_2[h])>0$, as desired.

**Case 2:** $g\in\{a_i,b_i\}$, $h\in\{a_j,b_j\}$ for some $t\ge i>j\ge 1$. By (A1) and (B1), we have $T_1[g]>(k+3)(i-1)\ge(k+3)j$ and $T_1[h]\le(k+3)j$. This implies $T_1[g]>T_1[h]$. Similarly, by (A2) and (B2) we have $T_2[g]>(1/\varepsilon+1)(i-1)\ge(1/\varepsilon+1)j$ and $T_2[h]\le(1/\varepsilon+1)j$. This implies $T_2[g]>T_2[h]$. Hence, $(T_1[g]-T_1[h])(T_2[g]-T_2[h])>0$, as desired.

**Case 3:** $g=a_i$, $h=b_i$ for some $i\in[t]$. In this case $T_1[g]-T_1[h]=1-(k+2)(s_i-s_{i+1})=1-(k+2)\delta_i$ and $T_2[g]-T_2[h]=1-(1/\varepsilon)(s_i-s_{i+1})=1-\delta_i/\varepsilon$. If $\delta_i=\frac{\varepsilon}{2}$ (i.e. $x_i^*$ is False), $(T_1[g]-T_1[h])(T_2[g]-T_2[h])=(1-\frac{1}{14})(1-\frac{1}{2})>0$. If $\delta_i=\frac{1}{k+1}$ (i.e. $x_i^*$ is True), $(T_1[g]-T_1[h])(T_2[g]-T_2[h])=(1-\frac{k+2}{k+1})(1-\frac{7(k+2)}{k+1})=(\frac{-1}{k+1})(\frac{-6k-13}{k})>0$.

**Case 4:** $g\in\{c_i,d_i\}$, $h\in\{c_j,d_j\}$ for some $t\ge i>j\ge 1$. By (C1) and (D1), we have $T_1[g]>t(k+3)+2(i-1)\ge t(k+3)+2j$ and $T_1[h]\le t(k+3)+2j$. This implies $T_1[g]>T_1[h]$. Similarly, by (C2) and (D2) we have $T_2[g]>t(1/\varepsilon+1)+6(k+3)(i-1)\ge t(1/\varepsilon+1)+6(k+3)j$ and $T_1[h]\le t(1/\varepsilon+1)+6(k+3)j$. This implies $T_1[g]>T_2[h]$. Hence, $(T_1[g]-T_1[h])(T_2[g]-T_2[h])>0$, as desired.

**Case 5:** $g=c_i$, $h=d_i$ for some $i\in[\ell]$. In this case $T_1[g]-T_1[h]=1$ and

$$\begin{aligned}T_2[g]-T_2[h]&=2|Z_i|-1+2(k+2)\left(\sum_{g\in V_i\setminus Z_i}(s_g-s_{g+1})-\sum_{h\in Z_i}(s_h-s_{h+1})\right)\\&=2|Z_i|-1+2(k+2)\left(\sum_{g\in V_i\setminus Z_i}\delta_g-\sum_{h\in Z_i}\delta_h\right),\end{aligned}$$

where $V_i\subseteq[t]$ and $Z_i\subseteq V_i$ indicate the indices of the variables that appear in and that appear negated in the clause $C_i$, respectively. Since $\{x_i^*\}_{i\in t}$ is a satisfying assignment by assumption, we must either have $\delta_{g'}=\frac{1}{k+1}$ (i.e. $x_{g'}^*$ is True) for some $g'\in V_i\setminus Z_i$ or $\delta_{h'}=\frac{\varepsilon}{2}$ (i.e. $x_{h'}^*$ is False) for some $h'\in Z_i$. If the former is true ($\delta_{g'}=\frac{1}{k+1}$ for some $g'\in V_i\setminus Z_i$):

$$\begin{aligned}T_2[g]-T_2[h]&=2|Z_i|-1+2(k+2)\left(\frac{1}{k+1}+\sum_{g\in V_i\setminus(Z_i\cup\{g'\})}\delta_g-\sum_{h\in Z_i}\delta_h\right)\\&\ge 2|Z_i|-1+2(k+2)\left(\frac{1}{k+1}+\frac{(|V_i|-|Z_i|-1)\varepsilon}{2}-\frac{|Z_i|}{k+1}\right)\\&\ge 2|Z_i|-1+2(k+2)\left(\frac{1-|Z_i|}{k+1}-\frac{\varepsilon}{2}\right)=-\frac{2|Z_i|}{k+1}+\frac{k+3}{k+1}-\frac{1}{7}=\frac{-2|Z_i|+\frac{6}{7}k+\frac{20}{7}}{k+1}.\end{aligned}$$

Similarly, if the latter is true ($\delta_{h'}=\frac{\varepsilon}{2}$ for some $h'\in Z_i$):

$$\begin{aligned}T_2[g]-T_2[h]&=2|Z_i|-1+2(k+2)\left(\sum_{g\in V_i\setminus Z_i}\delta_g-\frac{\varepsilon}{2}-\sum_{h\in Z_i\setminus\{h'\}}\delta_h\right)\\&\ge 2|Z_i|-1+2(k+2)\left(-\frac{\varepsilon}{2}-\frac{|Z_i|-1}{k+1}\right)=-\frac{2|Z_i|}{k+1}+\frac{k+3}{k+1}-\frac{1}{7}=\frac{-2|Z_i|+\frac{6}{7}k+\frac{20}{7}}{k+1},\end{aligned}$$

which gives the same inequality. Consider two cases: if $t \geq 2$ (i.e., $k = t+2 \geq 4$), we get

$$T_2[g] - T_2[h] \geq \frac{-2|Z_i| + \frac{6}{7}k + \frac{20}{7}}{k+1} \geq \frac{-6 + \frac{24}{7} + \frac{20}{7}}{k+1} = \frac{\frac{2}{7}}{k+1} > 0,$$

since $|Z_i| \leq 3$. If $t = 1$ (i.e., $k = 3$), then $|Z_i| \leq 1$, since $t$ is the number of variables in $\phi$. Then,

$$T_2[g] - T_2[h] \geq \frac{-2|Z_i| + \frac{6}{7}k + \frac{20}{7}}{k+1} \geq \frac{-2 + \frac{18}{7} + \frac{20}{7}}{k+1} = \frac{\frac{24}{7}}{k+1} > 0.$$

In both cases, we have $T_2[g] - T_2[h] > 0$ and hence $(T_1[g] - T_1[h])(T_2[g] - T_2[h]) > 0$, as desired.

**Case 6:** $g = E \setminus \{e_k\}, h \notin E$. Say $g = e_i$ for some $1 \leq i < k$, and that $n_\beta = \sum_{f \in K \setminus E} \sum_{j=1}^{k} M_\beta[f, j]$ for $\beta \in \{1, 2\}$. Due to the last set of voters that we added while constructing $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ (those that rank only elements of $E$), we have

$$M_\beta[e_i, i] \geq (k+1) \max_{\gamma \in \{1,2\}} n_\gamma \geq 14(k+2) \sum_{j=1}^{k} M_\beta[h, j] \geq 14(k+2) \sum_{j=1}^{k} M_\beta[h, j] s_j = 14(k+2) T_\beta[h]$$

Moreover, since $s_i \geq s_{k-1} \geq 1 - \frac{k-2}{k+1} = \frac{3}{k+1} > \frac{1}{k+1}$, we have $T_\beta[g] = M_\beta[e_i, i] s_i > \frac{M_\beta[g,i]}{k+1} > T_\beta[h]$ for each $\beta \in \{1, 2\}$. This gives $(T_1[g] - T_1[h])(T_2[g] - T_2[h]) > 0$, as desired.

**Case 7:** $g = e_k$. By construction $M_\beta[e_k, j] > 0 \Rightarrow j = k$, and hence $T_\beta[g] = T_\beta[e_k] = M_\beta[e_k, k] s_k = 0$ for each $\beta \in \{1, 2\}$. If $h \in K \setminus E$, we have $T_\beta[h] > 0$ for both $\beta \in \{1, 2\}$ by (A1) to (D2). If $h = e_i$ for some $i \in [k-1]$, on the other hand, $M_\beta[h, i] > 0$ by the last set of voters added to $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ (those that only rank the elements of $E$), so once again we have $T_\beta[h] = M_\beta[h, i] s_i > 0$. Hence, we have $(T_1[g] - T_1[h])(T_2[g] - T_2[h]) > 0$, as desired.

**Case 8:** $g = e_i$, $h = e_j$ for some $i < j < k$. Fix some $\beta \in \{1, 2\}$. By construction, $\boldsymbol{\sigma}_\beta$ contains two types of rankings: (1) those that contain one elements in $K \setminus E$ and all remaining elements are those in $E$, of which there $n_\beta = \sum_{f \in K \setminus E} \sum_{l=1}^{k} M_\beta[f, l]$ many, and (2) those that rank only and all elements in $E$, of which there are say $y_\beta$ many, with $y_\beta \geq \max_{\gamma \in \{1,2\}} 14(k+2) n_\gamma \geq 14(k+2) n_\beta$. Hence, for any $l \in [k]$, we have $n_\beta + y_\beta > M[e_l, l] \geq y_\beta$, where the first inequality is strict since $M_\beta[f, \ell] > 0$ for some $f \in K \setminus E$ for all $\ell \in [k-1]$ (see, for example, the definitions for $M_\beta[a_l, l+1]$ and $M_\beta[b_l, l]$ for each $l \in [t]$), so there is at least some voters in the group of $n_\beta$ that do not rank $e_l$. This implies

$$\begin{aligned} T_\beta[g] - T_\beta[h] = T_\beta[e_i] - T_\beta[e_j] = M_\beta[e_i, i] s_i - M_\beta[e_j, j] s_j > y_\beta s_i - (n_\beta + y_\beta) s_j = y_\beta(s_i - s_j) - n_\beta s_j \\ \geq y_\beta(s_i - s_{i+1}) - n_\beta = y_\beta \delta_i - n_\beta \geq 14(k+2) n_\beta \cdot \frac{\varepsilon}{2} - n_\beta = n_\beta - n_\beta = 0 \end{aligned}$$

Hence, we have $(T_1[g] - T_1[h])(T_2[g] - T_2[h]) > 0$, as desired.

We have shown that $(T_1[g] - T_1[h])(T_2[g] - T_2[h]) > 0$ for all $g, h \in K$, proving that $\boldsymbol{s}$ indeed gives full agreement between $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$. □

# 3 Agreement Helps with Generalization

Say you are the program chair of a conference, and you have a number of reviewer rankings over some submitted papers. You are trying to choose what rule to use to decide on your aggregate ranking. One concern is that rejected authors might complain that if they were given more reviews, their paper would have been accepted. Hence, you want to choose a rule (among some specified set of 'reasonable' rules) such that if you were suddenly given access to a previously-unseen (and much larger) set of new reviews, the resulting ranking would not have changed too much.

Let's formalize this: You get $n$ rankings ($\boldsymbol{\sigma} = \{\sigma_i\}_{i\in[n]} \sim D^n$) over candidates $K$ i.i.d. from some hidden distribution ($D$), and you have a set of social preference functions $F$. You would like to pick an $f \in F$ such that if you received $m$ more rankings ($\boldsymbol{\sigma}^* = \{\sigma_i^*\}_{i\in[m]} \sim D^m$) from the same distribution, with $m >> n$ (in fact $m$ is large enough that the frequency with which a ranking $r$ appears can be assumed to be equal to the probability of that ranking in the hidden distribution, $D(r)$). You would like a (possibly randomized) 'rule picking rule' $Z : L(A)^n \to F$ (where $L(A)$ is the linear orderings over $A$) such that

$$\mathbb{E}\left[d_{KT}(Z(\boldsymbol{\sigma})(\boldsymbol{\sigma}), Z(\boldsymbol{\sigma})(\boldsymbol{\sigma}^*)\right] \tag{1}$$

is minimized, where $d_{KT}$ is the Kendall-Tau distance and the expectation is taken over the $\boldsymbol{\sigma}$ (from the hidden distribution) and $Z$.

We would like to argue that our method (pick a random split over the voters in $\boldsymbol{\sigma}$ into two halves $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$, pick the $f \in F$ that minimizes $d_{KT}(f(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2))$ also bounds Emin: /minimizes? (1). Since votes are drawn i.i.d., we have $\boldsymbol{\sigma}_i \sim D^{n/2}$ for $i \in \{1, 2\}$.

Emin: next: what exactly do we want to argue here? that as $n$ gets large, we pick $f \in F$ minimzing (1) with probability almost 1? or perhaps a bound/convergence rate?

Emin: From Vince's point: maybe not about minimization (rules can be infinitesimally different), but that if the disagreement error is at least $\delta$ different between two different rules, can we say something about at least how different their generalization error will be?

Emin: point from Nihar: instead of showing that we minimize (1), we can make a PAC-learning-like argument where we bound how far away we are from optimal - maybe go over Ariel's paper

Emin: one can perhaps these two approaches by restricting our model to rule picking rules that output a score an each rule

Emin: for lower boudning the optimal, maybe restrict to positional scoring rules

# 4 Axiomatic analysis

Here, we list a number of axioms and prove whether they are satisfied by our rule. So far, we moslty have counterexamples. First, we formalize our model.

**Definition 1** (Rule picking rules)**.** Given candidates $K$ and the set of full (resp. partial) rankings over candidates $L(K)$ (resp. $L_p(K)$), say $F$ is a set of social welfare rules of form $f : L_p(K)^* \to L(K)$. Then, a *rule picking rule (RPR)* is a (possibly randomized) function $Z : L_p(K)^* \to F$, that given a multi-set of rankings, picks a rule from $F$. We denote by $f_Z$ the social welfare function induced by $Z$; that is, for all profiles $\boldsymbol{\sigma} \in L_p(K)^*$, we have $f_Z(\boldsymbol{\sigma}) = Z(\boldsymbol{\sigma})(\boldsymbol{\sigma})$.

Emin: Specific details of this definition to consider: does each $f \in F$ return a single (weak) ranking, or a set of strict rankings? does $Z$ returns a single $f \in F$, or a subset of $F$? do we allow stochastic? Also, maybe give $F$ as part of input.

We now formalize our rule:

**Definition 2** (Agreement based aggregation)**.** The *agreement based aggregation (ABA)* rule is an RPR that given a profile $\boldsymbol{\sigma}$, uniformly randomly splits the voters in $\boldsymbol{\sigma}$ to two sets to get $\boldsymbol{\sigma} = \boldsymbol{\sigma}_1 + \boldsymbol{\sigma}_2$ and returns $\text{argmin}_{f\in F} KT_w(f(\boldsymbol{\sigma}_1), f(\boldsymbol{\sigma}_2))$, where $KT_w$ is the weighted Kendall-Tau function.

**Definition 3** (Expected agreement based aggregation)**.** The *expected-agreement based aggregation (eABA)* rule is an RPR that given a profile $\boldsymbol{\sigma}$, returns $\text{argmin}_{f\in F} \mathbb{E}[KT_w(f(\boldsymbol{\sigma}_1), f(\boldsymbol{\sigma}_2))]$, where the expectation is taken over uniform splits $\boldsymbol{\sigma} = \boldsymbol{\sigma}_1 + \boldsymbol{\sigma}_2$

## 4.1 Consistency axioms

Here, we try to formalize an axiom that conceptually represents the idea that if our rule 'similar things' with two sets of reviews, then it should do 'a similar' thing when we bring the two sets together. We will have different attempts at formalizing this:

**Definition 4** (Strong consistency)**.** An RPR $Z$ satisfies *strong conistency (s-consistency)* if given $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$, $Z(\boldsymbol{\sigma}_1) = Z(\boldsymbol{\sigma}_2) = f$ with high probability implies $Z(\boldsymbol{\sigma}_1 + \boldsymbol{\sigma}_2) = f$ with high probability.

Here, by $Z(\boldsymbol{\sigma}) = f$ 'with high probability' we mean that $\boldsymbol{\sigma}$ can be parameterized by some $n$ such that as $n \to \infty$, then $\Pr[Z(\boldsymbol{\sigma}) = f] \to 1$.

**Proposition 5.** *ABA fails s-consistency.*

*Proof.* Say $K = \{a, b, c\}$ and $F = \{f_B, f_K\}$ where $f_B$ is Borda score and $f_K$ is Kemeny. Say $\boldsymbol{\sigma}_1$ consists of $n$ voters that rank $a \succ b \succ c$ and $2n$ voters that rank $b \succ c \succ a$. Similarly, $\boldsymbol{\sigma}_2$ consists of $2n$ voters that rank $a \succ b \succ c$ and $n$ voters that rank $b \succ c \succ a$.

Given a profile of $x$ rankings of $a \succ b \succ c$ and $y$ rankings of $b \succ c \succ a$, Kemeny will pick $a \succ b \succ c$ if $x > y$ and $b \succ c \succ a$ if $y > x$. Borda will pick $a \succ b \succ c$ if $x > 2y$, $b \succ c \succ a$ if $y > 2x$, and $b \succ a \succ c$ if $2y > x > y/2$ (the "buffer zone").

Now, if we pick $n$ to be sufficiently large, a random split of $\boldsymbol{\sigma}_1$ from above will (with high probability) have one half in the buffer zone with $2x > y > x$, and the other half outside the buffer zone with $y > 2x$ (since the total numbers "sit at the boundary"), so ABA will pick Kemeny to minimize disagreement. Similarly, for sufficiently large $n$, a random split in $\boldsymbol{\sigma}_2$ will (with high probability) have one half in the buffer zone with $2y > x > y$, and the other half outside the buffer zone with $x > 2y$, again leading to ABA choosing Kemeny.

However, now consider $\boldsymbol{\sigma} = \boldsymbol{\sigma}_1 + \boldsymbol{\sigma}_2$, which has exactly $3n$ rankings of $a \succ b \succ c$ and $3n$ rankings of $b \succ c \succ a$. For sufficiently large $n$, with high probability, one half of a random split will have more of one ranking, while the other will have more of the other, however, both of them will fall within the "buffer zone", leading to ABA choosing Borda.

□

We now introduce a weaker notion of consistency, inspired by binary classification in machine learning.

**Definition 6** (Weak consistency)**.** An RPR $Z$ satisfies *weak conistency (w-consistency)* if given $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$, whenever $Z(\boldsymbol{\sigma}_1) = Z(\boldsymbol{\sigma}_2) = f$ with high probability AND $f(\boldsymbol{\sigma}_1) = f(\boldsymbol{\sigma}_2)$, then we also have $Z(\boldsymbol{\sigma}_1 + \boldsymbol{\sigma}_2) = f$ with high probability.

Note the only difference from Definition 4 is the additional constraint that $f(\boldsymbol{\sigma}_1) = f(\boldsymbol{\sigma}_2)$ (in the language of binary classification, this would correspond to the best fitting function to both training datasets is from the same function class and is exactly the same function).

**Proposition 7.** *eABA fails w-consistency.*

*Proof.* Say $K = \{a, b, c, d\}$ and $F = \{f_P, f_V\}$ where $f_P$ is the plurality rule (positional scoring rule (1,0,0,0)) and $f_V$ is the veto rule (positional scoring rule (1,1,1,0)). Fix some $n \in \mathbb{Z}^+$ and say $\boldsymbol{\sigma}_1$ consist of $n-2$ voters ranking $a \succ b \succ d \succ c$ and 2 voters ranking $a \succ b \succ c \succ d$. Similarly, say $\boldsymbol{\sigma}_2$ consist of $2n-2$ voters ranking $b \succ a \succ d \succ c$ and 2 voters ranking $b \succ a \succ c \succ d$. Notice that for all splits $\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_i^{(1)} + \boldsymbol{\sigma}_i^{(2)}$ for $i \in \{1, 2\}$, we have $f_P(\boldsymbol{\sigma}_1^{(1)}) = f_P(\boldsymbol{\sigma}_1^{(2)}) = (a \succ b = c = d)$, giving a KT of 1.5 due to ties, and $f_P(\boldsymbol{\sigma}_2^{(1)}) = f_P(\boldsymbol{\sigma}_2^{(2)}) = (b \succ a = c = d)$, also giving a KT of 1.5 due to ties.

For $f_V$, first consider $\boldsymbol{\sigma}_1$. If both of the voters that rank $a \succ b \succ c \succ d$ end up in the same side of the split $\boldsymbol{\sigma}_1^{(j)}$, then $f_V(\boldsymbol{\sigma}_1^{(j)}) = (a = b \succ c \succ d)$ and $f_V(\boldsymbol{\sigma}_1^{(3-j)}) = (a = b = c \succ d)$, giving a disagreement score of 1.5. If these two voters end up in opposite sides of the split however, we have $f_V(\boldsymbol{\sigma}_1^{(1)}) = f_V(\boldsymbol{\sigma}_1^{(2}) = (a = b \succ c \succ d)$, giving a disagreement score of 0.5. For sufficently large $n$, these two cases happen with the same probability, so $f_V$ brings 0.5 disagreement $\sim 0.5$ of the time. The same argument applies for $\boldsymbol{\sigma}_2$. Hence, as ABA chooses the rule with the least disagreement over splits, we have $ABA(\boldsymbol{\sigma}_1) = ABA(\boldsymbol{\sigma}_2) = f_V$. Moreover, $f_V(\boldsymbol{\sigma}_1) = f(\boldsymbol{\sigma}_2) = (a = b \succ d \succ c)$

Now consider $\boldsymbol{\sigma} = \boldsymbol{\sigma}_1 + \boldsymbol{\sigma}_2$. For sufficiently large $n$, any split $\boldsymbol{\sigma} = \boldsymbol{\sigma}^{(1)} + \boldsymbol{\sigma}^{(2)}$ with high probability we will have $f_P(\boldsymbol{\sigma}^{(1)}) = f_P(\boldsymbol{\sigma}^{(2)}) = (b \succ a \succ c = d)$, giving a disagreement score of 0.5. However, if all four special voters end up in the same side (happens with probability $\sim 1/8$ as $n \to \infty$), $f_V$ will still give a disagreement score of 1.5 and 0.5 otherwise, showing it performs worse on average. Hence $ABA(\boldsymbol{\sigma}) = f_P$, violating w-consistency. □

**Proposition 8.** *ABA fails w-consistency.*

*Proof.* Say $A = \{a, b, c, d, e, f\}$ and $F = \{f_t, f_b\}$, where $f_v$ and $f_b$ are the positional scoring rules associated with scoring vectors $(1, 1, 1, 1, 1, 0)$ and $(1, 0.8, 0.6, 0.4, 0.2)$, respectively (i.e., veto and Borda). Fix some $n$ and define the profiles as follows:

$$\boldsymbol{\sigma}_1 = \begin{cases} n+2 \text{ voters: } a \succ b \succ c \succ d \succ e \succ f \\ n \text{ voters: } d \succ c \succ b \succ a \succ e \succ f \end{cases} \qquad \boldsymbol{\sigma}_2 = \begin{cases} n+2 \text{ voters: } a \succ b \succ c \succ d \succ e \succ f \\ n \text{ voters: } a \succ e \succ d \succ c \succ b \succ f \end{cases}$$

Notice that $f_v$ gives $(a = b = c = d = e \succ f)$ for any split of any profile, so the disagreement associated with $f_v$ will always be $\frac{1}{2} \cdot \binom{5}{2} = 5$. For a random split of $\boldsymbol{\sigma}_1$, for large $n$, with probability $\sim 1$ one half (say $\boldsymbol{\sigma}_1^{(j)}$) will have more voters that rank $a \succ b \succ c \succ d \succ e \succ f$ while other half $(\boldsymbol{\sigma}_1^{(3-j)})$ will have more voters that rank $d \succ c \succ b \succ a \succ e \succ f$. In this case, $f_v(\boldsymbol{\sigma}_1^{(j)}) = a \succ b \succ c \succ d \succ e \succ f$ and $f_v(\boldsymbol{\sigma}_1^{(j)}) = d \succ c \succ b \succ a \succ e \succ f$, leading to a disagreement error of $\binom{4}{2} = 6$. By the analogous reasoning, given a random split of $\boldsymbol{\sigma}_2$, for large $n$ with probability $\sim 1$ we will have $f_v(\boldsymbol{\sigma}_2^{(j)}) = a \succ b \succ c \succ d \succ e \succ f$ and $f_v(\boldsymbol{\sigma}_1^{(j)}) = a \succ e \succ d \succ c \succ b \succ f$ for some $j \in \{1, 2\}$, again leading to a disagreement error of $\binom{4}{2} = 6$. Hence, $ABA(\boldsymbol{\sigma}_1) = ABA(\boldsymbol{\sigma}_2) = f_v$ with high probability, and $f_v(\boldsymbol{\sigma}_1) = f_v(\boldsymbol{\sigma}_2) = (a = b = c = d = e \succ f)$.

Now consider $\boldsymbol{\sigma} = \boldsymbol{\sigma}_1 + \boldsymbol{\sigma}_2$. With probability $\sim 1$ for large $n$, a split will have more voters that rank $a \succ b \succ c \succ d \succ e \succ f$ than those that rank $d \succ c \succ b \succ a \succ e \succ f$ or $a \succ e \succ d \succ c \succ b \succ f$, and the other side will have less, implying one side will have $b \succ c \succ d$ and the other will have $d \succ c \succ b$ in their Borda rankings. Moreover, with probability $\sim 1$ for large $n$, both halves will have $a \succ b \succ c \succ d \succ e \succ f$ represented most frequently among the three, implying $a \succ \{b, c, d\}$ and $\{b, c, d\} \succ e$ for both. Hence, the output of $f_b$ for the two halves will be $a \succ b \succ c \succ d \succ e \succ f$ and $a \succ b \succ c \succ d \succ e \succ f$, giving a total disagreement of $\binom{3}{2} = 3$, which is less than the disagreement caused by veto (5). Hence, $ABA(\boldsymbol{\sigma}) = f_b$, violating w-consistency. $\square$

## 4.2 Reversal symmetry

Say $F_s$ are the set of positional scoring rules, where each $f_s \in F_S$ is associated with a positional scoring vector $s = (s_i)_{i \in k}$ with $1 = s_1 \geq s_2 \geq \ldots s_k = 0$. We define the reverse of the scoring rule $f_s$, denoted $\text{rev}(f_s)$, as the scoring rule $f_{s'}$ associated with vector $s' = (s'_i)_{i \in [k]}$, where $s'_i = 1 - s_{k-i}$ for each $i \in [k]$. For instance, the reverse of plurality is veto. For any profile $\boldsymbol{\sigma}$, we say the reverrse of the profile, denoted $\text{rev}(\boldsymbol{\sigma})$, is the same profile with every voter having flipped their ranking (e.g., $a \succ b \succ c$ becomes $c \succ b \succ a$, and so forth).

**Definition 9.** [Reversal symmetry] Say $F \subseteq F_S$. Then an RPR $Z$ satisfies *reversal symmetry* if for all profiles $\boldsymbol{\sigma}$, we have $Z(\boldsymbol{\sigma}) = \text{rev}(Z(\text{rev}(\boldsymbol{\sigma})))$.

Emin: ties need to be addressed

**Proposition 10.** *eABA and ABA both satisfy reversal symmetry.*

*Proof.* The weighted Kendall Tau function, $KT_w$, is symmetric with respect to reversals, i.e.,$KT_w(\sigma_1, \sigma_2) = KT_w(\text{rev}(\sigma_1), \text{rev}(\sigma_2))$. Moreover, for any positional scoring rule we have $f_s(\boldsymbol{\sigma}) = \text{rev}(\text{rev}(f_s)(\text{rev}(\boldsymbol{\sigma})))$, so

$$\begin{aligned} KT_w(f_s(\boldsymbol{\sigma}_1), f_s(\boldsymbol{\sigma}_2)) &= KT_w(\text{rev}(\text{rev}(f_s)(\text{rev}(\boldsymbol{\sigma}_1))), \text{rev}(\text{rev}(f_s)(\text{rev}(\boldsymbol{\sigma}_2)))) \\ &= KT_w(\text{rev}(f_s)(\text{rev}(\boldsymbol{\sigma}_1)), \text{rev}(f_s)(\text{rev}(\boldsymbol{\sigma}_2))). \end{aligned}$$

This implies that if $f_s$ minimizes $KT_w$ for a random split of $\boldsymbol{\sigma}$ with high probability (resp. in expectation), then $\text{rev}(f_s)$ minimizes $KT_w$ for a random split of $\text{rev}(\boldsymbol{\sigma})$ with high probability (resp. in expectation), implying $ABA(\boldsymbol{\sigma}) = \text{rev}(ABA(\text{rev}(\boldsymbol{\sigma})))$ (resp. $eABA(\boldsymbol{\sigma}) = \text{rev}(eABA(\text{rev}(\boldsymbol{\sigma})))$). $\square$

### 4.3 'Preserved' axioms.

Another natural question is, can we impose certain properties on an RPR by restricting $F$ to rules that satisfy that property?

**Definition 11.** We say an RPR $Z$ *preserves* a property $P$ if whenever $P$ is true for all $f \in F$, then $P$ is true for $f_Z$.

It is easy to see that any RPR will automatically preserve unanimity, anonymity, neutrality, Condorcet consistency, and Smith criterion. Conceptually these are axioms/properties that are defined for an individual voting profile (i.e. if the profile is XXX then YYY should happen). However, things get more complicated for axioms/properties defined for pairs of profiles, such as monotonicity.

#### 4.3.1 Monotonicity

Consider the following definition for monotonicity: Emin: this is probably the weakest way in which we can define motonicity

**Definition 12** (Monotonicity)**.** A social welfare function $f$ satisfies *monotonicity* if for any profile $\boldsymbol{\sigma}$ with $a = \text{top}(f(\boldsymbol{\sigma}))$ and $\boldsymbol{\sigma}'$ is the same as $\boldsymbol{\sigma}$ except some voters now rank $a$ higher, then $a = \text{top}(f(\boldsymbol{\sigma}))$

**Proposition 13.** *ABA or eABA does not preserve monotonicity.*

*Proof.* Say $F = \{f_p, f_v\}$, where $f_p$ and $f_v$ are plurality and veto, respectively. Both of these rules are monotonic, as all positional scoring rules are. Fix some $n$ and consider the following profile:

- Group 1: $3n$ people voted $a \succ b \succ c \succ d$
- Group 2: $2n$ people voted $b \succ d \succ c \succ a$
- Group 3: $n$ people voted $c \succ d \succ b \succ a$

If $n$ is sufficiently large, each split will have more Group 1 voters than Group 2 voters, which will be more than Group 3 voters. Hence, plurality gives $a \succ b \succ c \succ d$ for both splits (perfect agreement), whereas veto will (w prob $\sim 1$) give $b = c \succ a \succ d$ in one split and $b = c \succ d \succ a$ in the other (disagreement score 1.5). Hence, both eABA and ABA choose plurality, and $a$ is the winner. Now say $n$ people in Group 2 move $a$ to the top of their votes, so the new profile is:

- Group 1: $3n$ people voted $a \succ b \succ c \succ d$
- Group 2a: $n$ people voted $b \succ d \succ c \succ a$
- Group 2b: $n$ people voted $a \succ b \succ d \succ c$
- Group 3: $n$ people voted $c \succ d \succ b \succ a$

In this case, for a random split, plurality will (w prob $\sim 1$) give $a \succ b \succ c \succ d$ in one and $a \succ c \succ b \succ d$ in the other (disagreement score 1). However, veto will with high probability give $b \succ c \succ a \succ d$ for both. Hence, both ABA and eABA pick veto. By promoting $a$, we have caused it to go down from no1 (the winner) in the ranking to no 3. $\square$

In general, for any axiom our rule fails, it would be good to have some properties impossibility results, such as the following:

**Proposition 14.** *No RPR that satisfies reversal symmetry (Definition 9) can preserve monotonicity.*

*Proof (sketch).* Emin: This proof assumes RPRs can only return a single rule. However, if this is the case, eABA needs a tie-breaking over rules, which would violate reversal symmetry Assume that $Z$ is an RPR that satisfies reversal symmetry and $F = \{f_p, f_v\}$. Fix some $n$ and define the following profile:

- Group 1: $n$ people voted$d \succ b \succ a \succ c$
- Group 2: $n$ people voted$c \succ b \succ a \succ d$

We consider two cases.
**Case 1:** $Z$ chooses veto with profile above. Then $a$ and $b$ are winners (the output ranking is $a = b \succ c = d$). Say everyone promotes $a$ by a spot, to get:

- Group 1: $n$ people voted $d \succ a \succ b \succ c$
- Group 2: $n$ people voted $c \succ a \succ b \succ d$

But this is also the original profile reversed! So by the flipping axiom, $Z$ must now choose plurality, and suddenly d,c are the winners. By promoting one of the winners, (a), we have caused it to stop becoming the winner.
**Case 2:** $Z$ chooses plurality with the original profile. By the flipping axiom $Z$ must choose veto when given the the reverse of the original profile:

- Group 1: $n$ people voted $c \succ a \succ b \succ d$
- Group 2: $n$ people voted $d \succ a \succ b \succ c$

With veto, the winners are $\{a, b\}$ in this reverse profile (the output ranking is $a = b \succ c = d$). In this reverse profile, promote $b$ by one spot for both groups, and you get the original profile, where F picks plurality and $\{d, c\}$ are the winners! Again, we have promoted a winner (b) and caused it to lose. Emin: should probably restrict the proposal statement to RPRs that satisfy anonymity $\square$

#### 4.3.2 Homogenity

It would be useful to have a property about pairs of profiles such that if the property is satisfied, then our rule picking rule (RPR) picks the same rule in both cases. The most "benign" transformation one could do to a profile is to multiply all votes by the same number. If this transformation were to preserve the rule picked by our RPR, then the statement "If all available rules are homogeneous, then so is our rule picking rule" would be true. Unfortunately, this is not always the case, at least for some small profiles.

**Proposition 15.** *eABA does not preserve homogeneity.*

*Proof.* Say we have to pick between the positional scoring rules $(1, 1, 1)$ (the dummy rule that always ties everything) and $(1, 0.6, 0)$ (a slightly-perturbed Borda)[1]. Consider a profile with only two rankings: $a \succ b \succ c$ and $c \succ b \succ a$. There is a single possible split, and $(1, 1, 1)$ gives $a = b = c$ for both sides (disagreement score 1.5), whereas $(1, 0.6.0)$ returns $a \succ b \succ c$ and $c \succ b \succ a$ for the two halves (disagreement score 3). Hence, our rule picks $(1, 1, 1)$ and returns $a = b = c$.

Now, say we double the number of each vote, so our new profile (say $\boldsymbol{\sigma}$) has two $a \succ b \succ c$ votes and two $c \succ b \succ a$ votes. Up to symmetries, there are two possible splits: (1) $S1 = \{a \succ b \succ c, a \succ b \succ c\}, S2 = \{a \succ b \succ c, a \succ b \succ c\}$ and (2) $S1 = S2 = \{a \succ b \succ c, c \succ b \succ a\}$. Note that (2) is twice as likely as (1). With (1), we get the same scenario as above, with $(1, 1, 1)$ and $(1, 0.6, 0)$ giving disagreement scores 1.5 and 3, respectively. With (1) however, $(1, 1, 1)$ again gives a disagreement score of 1.5, whereas [1,0.6,0] returns $b \succ c = a$ for both halves, hence a disagreement score of 0.5. Thus, the expected disagreement score of $(1, 1, 1)$ is $(1.5 + 2 \cdot 1.5)/3 = 1.5$, whereas the expected disagreement score of $(1, 0.6, 0)$ is $(3 + 2 \cdot 0.5)/3 = 1.333$, so our RPR now chooses $(1, 0.6, 0)$, outputting $b \succ c = a$ in the process. Hence, by doubling the votes, we have changed the outcome.

$\square$

[1]The $(1, 1, 1)$ rule violates our usual restriction that the last position always has score 0. This can be circumvented easily by adding a fourth candidate d ranked at the bottom of every ballot, and changing the rules to $(1, 1, 1, 0)$ and $(1, 0.6, 0, 0)$.

### 4.4 Shuffling axioms

We like our rule to pass some sanity checks, as in (informally): if we uniformly 'shuffle' the first an interval of positions in everyone's ranking, then our RPR prefers rules that treat these positions equivalently. One way of shuffling could be simply picking a random permutation with equal probabilities. However, this does not work:

**Example 16.** *Say $F$ consists of the positional scoring rules $(1, 1, 0)$ and $(1, 0.5, 0)$. Consider the profile where $n$ voters rank $a \succ b \succ c$, 1 voter ranks $a \succ c \succ b$, and 1 voter ranks $b \succ c \succ a$.*

*First consider the behavior of $(1, 1, 0)$, for which the flips do not matter. If the two special voters end up in the same side of the split, the rule gives $a = b \succ c$ for both sides, for a disagreement error of $0.5$. If they end up in different sides, however, one side will have $a \succ b \succ c$ while the other will have $b \succ a \succ c$, giving a disagreement error of 1. For large $n$, these two happen with roughly equal probabilities.*

*Now consider the behavior of $(1, 0.5, 0)$. We can simply split and then shuffle the first two positions uniformly. For sufficiently large $n$, the 2 voters will not matter with high probabilitiy, and the winner will be determined (in each half) of whether $a$ or $b$ ended up more often in the top in the top more often. Due to symmetry, they are equally likely to end up in the top. If the both sides have the same candidate end up in the top, we get disagreement error 0, otherwise we get disagreemetn error 1. For high $n$, these happen with roughly equal probability.*

*Hence, eABA would pick $(1, 0.5, 0)$ over $(1, 1, 0)$.*

This is since the while each $a \succ b \succ c$ has an equal probability of staying the same or ending up as $b \succ a \succ c$ post-shuffling, the actual realization of the shuffle will have $\Theta(\sqrt{n})$ more of one of these. So instead, we propose a following shuffling definition:

**Definition 17.** Given a (full) profile $\boldsymbol{\sigma}$ over candidates $K$ with $|K| = m$ and any subset $P \subseteq [n]$, the *$k-$shuffling of $\boldsymbol{\sigma}$ with respect to positions $P$* (denoted $\c{S}^k(\boldsymbol{\sigma}, S)$) is profile obtained as follows: For each vote $\sigma_i$ in $\boldsymbol{\sigma}$,

- Create $k \cdot m!$ copies of $\sigma_i$ to be added to the final profile.
- For each of the $|S|!$ possible permutation of the positions $S$, alter $(k \cdot m!)/(|S|!)$ of these copies so the candidates in the positions $S$ are permutated according to this permutation.
- Add all copies to the final profile.

Since $|S|!$ is always a factor of $m!$, the second step in Definition 17 is well-defined. We now define a consistency axiom for a specific shuffling, that of all but last position.

**Definition 18.** A rule picking rule $Z$ satisfies *veto-shuffling-consistency (VSC)* such that given a finite set of candidate rules $F$ that are all positional scoring rules and includes veto $f_v$ (the positional scoring rule defined by the vector with 0 in its last position, and 1s in all other), (1) for all profiles $\boldsymbol{\sigma}$ and integers $k$ we have $f_v \in Z(\c{S}^k(\boldsymbol{\sigma}, [m-1])$ and (2) there exists a profile $\boldsymbol{\sigma}$ and an integer $k$ (that may depend on $m$ and $F$) such that $Z(\c{S}^k(\boldsymbol{\sigma}, [m-1]) = \{f_v\}$, i.e., the RPR picks *only* $f_v$ when given the shuffled profile.

VSC is a desirable property since it shows that the RPR respects symmetry in the profile. Since the first $m-1$ positions and uniformly shuffled in $Z(\c{S}^k(\boldsymbol{\sigma}, [m-1]) = \{f_v\}$, it is natural that the RPR should pick the positional scoring rule that treats all these positions equally, namely $f_v$. Despite this, a large class of RPRs, which we introduce next, fails VSC.

**Definition 19.** A *welfare-maximizing RPR* is an RPR $Z$ that is associated with a utility function $u : L(K) \times L(K) \to \mathbb{R}$ such that $Z(\boldsymbol{\sigma}) = \text{argmax}_{f \in F}\, u(\boldsymbol{\sigma}, f(\boldsymbol{\sigma}))$, where we overload the notation of $u$ so that $u(\boldsymbol{\sigma}, r) = \sum_{i=1}^{n} u(\boldsymbol{\sigma}_i, r)$.

**Proposition 20.** *Any wellfare-maximizing RPR $Z$ fails VSC for any $|F| \geq 2$.*

*Proof.* Take any $f \in F$. We will show that for any $k \in \mathbb{Z}_+$, we have $f(\S^k(\boldsymbol{\sigma}, [m-1])) = f_v(\boldsymbol{\sigma})$. Say $\{s\}_{i \in [m]}$ is the scoring vector associated with $f$, with $1 = s_1 \geq s_2 \geq \dots s_m = 0$ let $M$ be the matrix such that for each $a \in K$ and $i \in [m]$, $M[a, i]$ indicates the number of voters in $\boldsymbol{\sigma}$ that rank $a$ in the $i^{\text{th}}$ position. By Definition 17, the score that $a$ gets in $\S^k(\boldsymbol{\sigma}, [m-1])$ (say $T[a]$) is:

$$T[a] = \sum_{i=2}^{m} M[a, i] \cdot \sum_{j=2}^{m} \frac{k \cdot m!}{(m-1)!} s_j = km \left( \sum_{j=2}^{n} s_j \right) (n - M[a, m]),$$

which implies that for any $a, b \in K$, we have $T[a] > T[b]$ iff $M[a, m] < M[b, m]$, i.e. $a$ is ranked ahead of $b$ in $f(\S^k(\boldsymbol{\sigma}, [m-1]))$ iff $a$ is ranked of $b$ in $f_v(\boldsymbol{\sigma})$. Since $f$ was arbitrarily chosen, this implies all rules in $F$ return the same output on $\S^k(\boldsymbol{\sigma}, [m-1])$. Hence, for any welfare-maximizing RPR $Z$, we have $Z(\S^k(\boldsymbol{\sigma}, [m-1])) = F$, showing that the rule fails condition (2) from Definition 18, and hence is not VSC. Emin: One should perhaps justify why were care about picking just $f_v \in F$ even if all $f \in F$ return the same result on $\S^k(\boldsymbol{\sigma}, [n-1])$. □

Considering the $f(\S^k(\boldsymbol{\sigma}, [m-1]))$ is the same for all $f \in F$, one might expect the negative result from Proposition 20 to generalize to all RPRs. We now show that this is not the case for $eABA$ and $ABA$.

**Theorem 2.** *$eABA$ and $ABA$ both satisfy VSC.*

*Proof.* If $|F| = 1$ we are done. Otherwise, given any $f \in F$, say $\{s_i^f\}_{i \in [m]}$ is the scoring vector associated with $f$ and that $T(f) = \sum_{i=1}^{m} s_i^f$. Say $k = \max_{f \in F \setminus f_v} T(f)$. Since $|F| \geq 2$, we must have $k < m - 1$. □

# 5 Questions to think about

- Potential experiments (to discuss with Ben):
  - What happens to the scores we converge to when certain positions in all rankings are uniformly mixed?
  - Incorporating multiple splits into the optimization over scoring functions: compute all splits a prioiri and keep them in memory, have your loss function be sum over the disagreement scores for all splits.
- Check pref_lib library for existing ranking data on real life elections perhaps? Running our method on that and seeing what the best rule would have been.
- Incorporate Ranked Pairs and Schulze method, but there are two problems with this:
  - All pairs of candidates are not compared the same number of times. Normalizing by the number of times that they match might not be work (e.g. not work .
  - A LOT of ties.
- The hardness of this: given a split, does there exists a positional scoring rule that gives perfect agreement between the two. Seems easy if everyone reviews three proposals. Might be hard if you have a variable number of proposals.
- Theoretical justifications: say you want to find a rule that would produce the same outcome. Specific setup we discussed: Vince is program chair, Nihar will later be running an experiment. Vince gets finitely many reviews and must aggregate these rankings say through some scoring rule. Then Nihar will collect effectively infinitely many reviews for the same papers. Vince wants his aggregate ranking to disagree as little as possible with the ranking that results from applying his rule to Nihar's profile (consisting of the limit frequencies), to prevent authors from complaining. It seems intuitively perhaps a good idea for Vince to "test" scoring rules (or a subset of them, for better PAC-style generalization) on random splits, to choose which one to use (but is it a good idea?).