Supplementary material for Computational Complexity of Verifying the Group No-show Paradox

A Tie-breaking mechanisms

In this paper, we investigate three tie-breaking mechanisms. Each tie-breaking mechanism is defined by a pre-determined linear order, which indicates the order for ties to be broken.

First, the lexicographic tie-breaking method, denoted by LEX, breaks ties between alternatives alphabetically. So for a tie between 1 and 2, LEX will break the tie in favor of 1.

The *fixed-agent tie-breaking*, denoted by FA_j, is parameterized by $j \leq n$, which uses the j-th agent's vote to break ties. In this paper, we'll use always use agent 1's vote for fixed-agent tie-breaking (denoted as FA) w.l.o.g.

The *most popular singleton ranking tie-breaking* (Xia, 2020), denoted by MPSR, is parameterized by another "backup" tie-breaking mechanism. We define the most popular singleton ranking as follows.

Definition 4 (Most Popular Singleton Ranking)**.** *Given a profile P, we define its most popular singleton ranking as $MPSR(P) = \arg\max_R(P[R] : \nexists W \neq R \text{ s.t } P[R] = P[W])$, where $P[R]$ is the number of votes that linear order R receives in P.*

The MPSR tie-breaking method first tries calculates the MPSR. If such a ranking (called the most popular singleton ranking) exists, then it is used to break ties. If a unique MPSR does not exist for a profile P, then a backup tie-breaking mechanism $\mathcal{T}$ is used. In this paper we mainly consider two MPSR tie-breaking mechanisms: MPSR+LEX and MPSR+FA, where lexicographic and fixed-agent tie-breaking mechanisms are used as the backup, respectively.

Example 2 (Tie breaking mechanisms)**.** Suppose there are 3 alternatives in a vote with the Plurality rule. The profile P consists of 6 votes: $P = \{V_1, V_2, V_3, V_3, V_4, V_5\}$, where $V_1 = [2 \succ 1 \succ 3]$, $V_2 = [2 \succ 3 \succ 1]$, $V_3 = [3 \succ 2 \succ 1]$, $V_4 = [1 \succ 2 \succ 3]$, and $V_5 = [1 \succ 3 \succ 2]$. We notice that there is a three-way tie in which alternatives 1, 2, and 3 are all voted top twice.

- When the tie-breaking mechanism is LEX, the tie is broken in order $1 \rhd 2 \rhd 3$, and alternative 1 is the winner.
- When the tie-breaking mechanism is FA_1 (fixed to be the first agent), the tie is broken in order of V_1, i.e., $2 \rhd 1 \rhd 3$. Then alternative 2 is the winner.
- When the tie-breaking mechanism is MPSR, we first find the most popular singleton ranking in P. Note that V_3 occurs twice while other votes occur only once. Therefore, the tie is broken in order $3 \rhd 2 \rhd 1$, and alternative 3 will be the winner.

Then we consider another profile $P' = \{V_1, V_1, V_3, V_3, V_4, V_4\}$. While the tie-breaking results of LEX and FA_1 are the same as those in P, P' does not have the most popular singleton ranking since every vote appears twice. Therefore, when the tie-breaking mechanism is MPSR, a backup mechanism is applied. For example, MPSR+LEX will choose 1 to be the winner, and MPSR+FA_1 will choose 2.

B Full Proof and supplementary proof of the Theorems

Full proof of Proposition 1

Proposition 1. *For any fixed m and anonymous voting rule r,* GNSP-r *can be solved in polynomial time if the winner of r can be computed in polynomial time.*

Proof. Let $n_1, n_2, \ldots, n_{m!}$ be the number of votes that each linear order receives. Assume that after abstention, the number of votes are $x_1, x_2, \ldots, x_{m!}$. There are $\Pi_{k=1}^{m!}(n_k + 1)$ possible combinations for the values of $x_1, \ldots, x_{m!}$. For any anonymous voting rule, the number of votes for each linear order uniquely determines the winner. Using the arithmetic mean-geometric mean inequality, we get the following upper bound: $\Pi_{k=1}^{m!}(n_k + 1) \leq (\frac{n}{m!} + 1)^{m!}$. Since m is a constant, this upper bound is a polynomial of n. □

B.1 Full proof of Theorem 1 (Copeland)

Theorem 1. *For any $0 \leq \alpha \leq 1$,* GNSP-Cd_α *is NP-complete to compute, where the tie-breaking mechanism is* LEX, FA, MPSR+LEX *or* MPSR+FA.

Proof. It is easy to check that the problem is in NP—given a subset of agents P', we run the voting mechanism with and without the group to check if they have the incentive to abstain from voting. The NP-hardness is proved by a reduction from RXC3. W.l.o.g. we assume that q is an even number. If q is odd, then we can use an instance with duplicate X and $\mathcal{S}$. We also assume that q is sufficiently large (for example, $q \geq 10$).

We first show the hardness of GNSP-Cd_α first for $\alpha < 1$, and then for $\alpha = 1$. Finally we give instructions on how to modify the proof for other tie-breaking mechanisms.

For any RXC3 instance $(X, \mathcal{S})$ with $\alpha < \frac{q-4}{q-3}$, we construct a GNSP-Cd_α instance with $q + 2$ alternatives as follows.

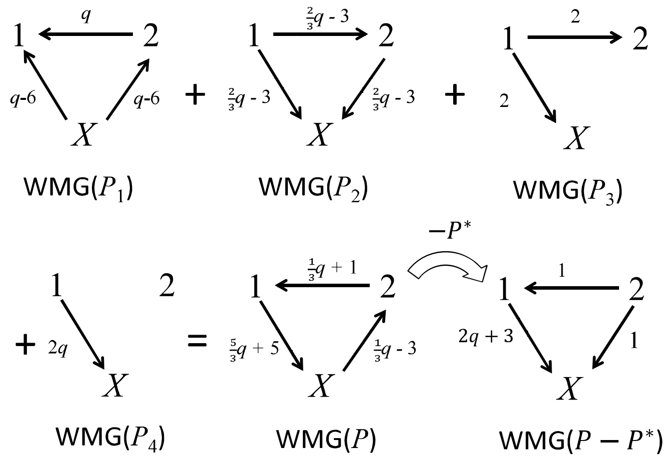

Figure 7: WMG of P for for Cd_α with $\alpha < 1$.

The construction of the GNSP-**Cd**$_\alpha$ **instance.** There are $q + 2$ alternatives $\{1, 2, 3, \dots, q + 2\}$, where for every $3 \leq i \leq q + 2$, alternative i corresponds to x_{i-2}. For convenience, we will use i and x_{i-2} interchangeably and denote alternatives $\{3, 4, \cdots, q + 2\}$ as X.

Let profile $P = P_1 \cup P_2 \cup P_3 \cup P_4$ consist of the following four parts, whose WMGs are illustrated in Figure 7.

- P_1 consists of q votes that correspond to the sets in $\mathcal{S}$: for every $j \leq q$, there is a vote R_{S_j} defined as follows

$$R_{S_j} = (X \setminus S_j) \succ 2 \succ 1 \succ S_j,$$

where alternatives in $(X \setminus S_j)$ and in S_j are ranked alphabetically. More precisely, $P_1 = \{R_S : S \in \mathcal{S}\}$.

- P_2 consists of $\frac{2}{3}q - 3$ copies of $[1 \succ 2 \succ X]$.
- P_3 consists of the following pair of votes

$$\{[1 \succ X \succ 2], [1 \succ 2 \succ X]\}$$

- P_4 consists of q copies of the following pair of votes

$$\{[1 \succ X \succ 2], [2 \succ 1 \succ X]\}.$$

As illustrated in Figure 7, 1 gets a Copeland score of q by beating every alternative in X, 2 gets a score of 1 by beating 1, and an alternative in X gets a score of at most q. Therefore, $\text{Cd}_\alpha(P) = 1$ due to the tie-breaking rule.

Suppose the RXC3 **instance is a yes instance.** There is an exact cover $\mathcal{S}^*$ for X. Then, GNSP-Cd_α is a yes instance. Let the abstention group (denoted by P^*) be those in P_1 that correspond to the 3-sets in $\mathcal{S}^*$, i.e. $P^* = \{R_{S_j} : S_j \in \mathcal{S}^*\}$. After P^* abstain from voting, the whole vote will lose (1) $\frac{q}{3}$ counts of $2 \succ 1$, (2) $\frac{q}{3} - 2$ counts of $X \succ 1$, and (3) $\frac{q}{3} - 2$ counts of $X \succ 2$. Therefore, alternative 2 becomes the Condorcet winner as illustrated in Figure 7. Note that all agents in P_1 prefer 2 to 1. Therefore, this constitutes a group no-show paradox.

Suppose the GNSP-**Cd**$_\alpha$ **instance is a yes instance.** A group of agents, denoted by P^*, have an incentive to abstain from voting. We will show that the RXC3 instance is a yes instance in four steps.

First, $\text{Cd}_\alpha(P - P^*) = 2$. Suppose this is not true, and the new winner is $a \neq 2$. Then $P^* \subseteq P_1$ because only agents in P_1 have an incentive to abstain. However, no matter how many votes in P_1 are removed, 1 beats all alternatives in X in their head-to-head competition and gets a Copeland score of at least q, while a beaten by 1 gets at most q. Therefore, a cannot be the winner because the tie-breaking mechanism favors 1, which is a contradiction.

Second, $|P^*| \leq \frac{q}{3}$. Suppose this is not true. Then at least $\frac{q}{3} + 1$ agents are removed, each of which prefers 2 to 1. Therefore, 2 cannot beat 1 in the head-to-head competition in $P - P^*$. Also, $P^* \subseteq P_1 \cup P_4$ (precisely, the q votes of $[2 \succ 1 \succ X]$ in P_4) because abstaining agents must prefer 2 to 1. Therefore, 1 is not beaten by 2 and beats all alternatives in X in WMG$(P - P^*)$, getting a Copeland score of at least $q + \alpha$, while 2 gets a Copeland score of at most $q + \alpha$. Therefore, 2 cannot be the winner due to the tie-breaking mechanism, which is a contradiction.

Third, $P^* \subseteq P_1$, and the 3-element subsets corresponding to votes in P^* are non-overlapping. We note that alternative 2 cannot lose any alternative $a \in X$ in their head-to-head competition, otherwise 2's Copeland score is not strictly larger than 1's score (which is at least q) and is not the winner. If there exists $a \in X$ such that $2 \succ a$ appears in more than one vote in P^*, 2 is defeated by a in the head-to-head competition as

$$\text{WMG}_{P-P^*}(2 \to a) \leq (|P^*| - 2) - 2 - (\frac{q}{3} - 3) \leq -1,$$

which implies that 2 cannot be the winner. Therefore, all votes in P^* come from P_1 (if P^* contains any vote in P_4, it cannot contain any other vote, which is impossible), and for any two votes R_{S_i} and R_{S_j} in P^*, whose corresponding sets are S_i and S_j, we have $S_i \cap S_j = \emptyset$.

Fourth, $|P^*| = \frac{q}{3}$ and corresponds to an exact cover of X, which implies the yes instance of RXC3. Suppose that $|P^*| \leq \frac{q}{3} - 1$. We show that 2's Copeland score is lower than 1' score (which is at least q), which is a contradiction.

• Case 1: $|P^*| = \frac{q}{3} - 1$. Consider an alternative $a \in X$ such that $a \in S_j$ and $R_{S_j} \in P^*$ for some $j \leq q$. P^* contains 1 count of $2 \succ a$ from S_j (from the third step, we know such j is unique) and $|P^*| - 1$ counts of $a \succ 2$ from other votes. Therefore, 2 is tied with a in the head-to-head competition after P^* abstains from voting. Since S_j is non-overlapping in P^*, there are $3|P^*| = q - 3$ of such alternative a. Therefore, the Copeland score of alternative 2 is at most $4 + \alpha(q-3)$. Recall that we assumed $\alpha < \frac{q-4}{q-3}$. Therefore, 2's score is lower than 1's score.

• Case 2: $|P^*| = \frac{q}{3} - 2$. In this case, 2 is defeated by all $a \in X$ such that $a \in S_j$ and $R_{S_j} \in P^*$ for some $j \leq q$ in the head-to-head competition, and there are $q - 6$ such a. Therefore, the Copeland score of 2 is at most 7. Since we assumed that q is sufficiently large, 2's score is lower than 1's score.

• Case 3: $|P^*| \leq \frac{q}{3} - 3$. In this case, 2 is defeated by or tied with every $a \in X$. Therefore, 2's score is at most $\alpha q + 1 < 4 + \alpha(q-3)$, which is lower than 1's score.

Therefore, we have $|P^*| = q/3$. Let $\mathcal{S}^* = \{S_j : R_{S_j} \in P^*\}$. Since these S_j are non-overlapping, every $x_i \in X$ appears in exactly one $S_j \in \mathcal{S}^*$. Therefore, $\mathcal{S}^*$ is an exact cover of X, and RXC3 is a yes instance.

Proof for $\alpha = 1$. Similarly, we construct a GNSP-Cd$_\alpha$ instance from an RXC3 instance. The profile P is similar to that in the proof for the $\alpha < 1$ except for that (1) 1 and 2 are switched in all votes, and (2) P_2 consist of $\frac{2}{3}q - 2$ copies of the vote, and (3) P_3 consists of $\{[2 \succ X \succ 1], [X \succ 2 \succ 1]\}$. The WMG of each part of the profile is shown in Appendix B.1. Alternative 2 is the winner with the Copeland score q by beating every alternative in X. In order for alternative 1 to win, exactly $\frac{q}{3}$ agents in P_1 that correspond to an exact cover of X need to abstain from voting. In the new profile, 1 is tied with every alternative and becomes the new winner according to the tie-breaking mechanism. Following similar reasoning, we can show that the abstaining group exists if and only if the exact cover exists.

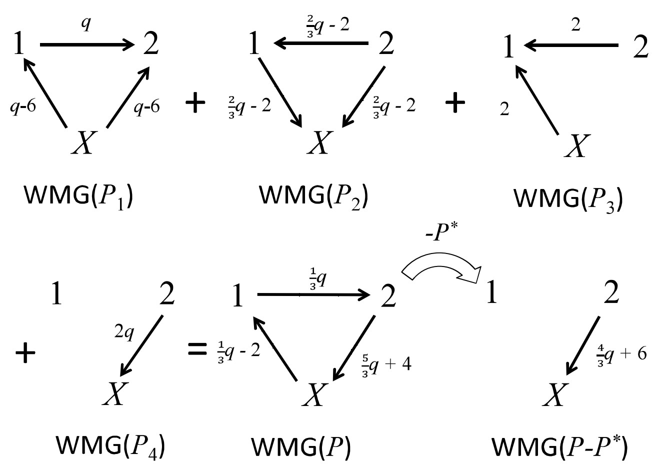

Figure 8: WMG of P for Cd$_\alpha$ with $\alpha = 1$.

Other tie-breaking mechanisms. For other tie-breaking mechanisms, we modify the construction to get a same tie-breaking order as LEX. Our modification for MPSR guarantees to find a most-popular preference, so we write MPSR-LEX and MPSR-FA together as MPSR.

FA. Let every voter of P_2 have the preference of $1 \succ 2 \succ 3 \succ \cdots \succ q+2$. Then switch P_1 and P_2. Then voter 1 becomes the first voter in P_2, with a order same as LEX.

MPSR Let every voter in P_2 has the preference of $1 \succ 2 \succ 3 \succ \cdots \succ q+2$. For other voters, notice that the preference of X has $q!$ different permutations. Therefore, we can assign every other voter a different permutation. (First, we assign each voter in P_1 a permutation of X that in accord with their current preference. Then for P_3 and P_4, we assign each of them a different permutation from the rest.) Therefore $1 \succ 2 \succ 3 \succ \cdots \succ q+2$ is the most popular voting and becomes the tie-breaking order.

Notice that P_2 voters have no incentives to abstain from the voting. Therefore, neither the first voter or the most popular voting will not change, so the tie-breaking order remain still before and after abstaining. □

B.2 Full Proof of Theorem 2 (Maximin)

Theorem 2 (Maximin). GNSP-*MM is NP-complete to compute, where the tie-breaking mechanism is* LEX, FA, MPSR+LEX *or* MPSR+FA.

Proof. It is easy to verify that the problem is in NP. The NP-hardness is proved by a reduction from RXC3. Given a RXC3 instance $X, \mathcal{S}$, we construct the following GNSP-MM with $q + 4$ alternatives as follows.

Alternatives: there are $q + 4$ alternatives $\{1, 2, 3, 4, 5, \ldots, q + 4\}$, where for every $4 \leq i \leq q + 4$, alternative i corresponds to x_{i-4}. For convenience, we will use i and x_{i-4} interchangeably.

Profile: Let profile $P = P_1 \cup P_2 \cup P_3 \cup P_4$ consist of the following four parts, whose WMGs are illustrated in Figure 3.

- **Part 1:** P_1 consists of q votes that correspond to the sets in $\mathcal{S}$: for every $j \leq q$, there is a vote R'_S defined as follows
$$R'_S = S_j \succ 1 \succ 2 \succ (X \setminus S_j) \succ 4 \succ 3,$$
where alternatives in $(X \setminus S_j)$ and in S_j are ranked alphabetically. More precisely, $P_1 = \{R'_S : S \in \mathcal{S}\}$.
- **Part 2:** P_2 consists of the $q - 2$ copies of $[3 \succ 2 \succ 4 \succ X \succ 1]$.
- **Part 3:** P_3 consists of two copies of $[3 \succ 2 \succ 4 \succ 1 \succ X]$.
- **Part 4:** P_4 consists of $\frac{q}{3}$ copies of the following pair of votes
$$\{[1 \succ 4 \succ 3 \succ 2 \succ X], [2 \succ 4 \succ 3 \succ X \succ 1]\}$$

The weights on edges among alternatives in X are not relevant to the proof. It is not hard to verify that MM$(P) = 2$, whose min-score is 0 (via alternatives 1 and 3).

Suppose the RXC3 instance is a yes instance, i.e., it has a solution $\mathcal{S}^*$. Then, we show that GNSP-MM is a yes instance by letting $P' \subset P_1$ denote the votes that correspond to $\mathcal{S}^*$. That is,
$$P' = \{R'_S : S \in \mathcal{S}^*\}$$
Then, we make the following observations about the min-scores of alternatives in $P - P'$, whose WMG is also illustrated in Figure 3.

- The min-scores of $1, 2, 3$ are $-\frac{q}{3}$.
- The min-scores of 4 is $-\frac{5q}{3}$ (via 2).
- For any $x \in X$, the min-scores of x is at most $-\frac{7q}{3} + 6$ (via 2 or alternatives in X).

Therefore, due to the lexicographic tie-breaking, MM$(P - P') = 1$. Notice that all voters in P' prefer 1 to 2. This constitutes a group no-show paradox.

Suppose GNSP-MM is a yes instance and a group of voters, denoted by P^*, have incentive to abstain from voting. We will show that the RXC3 instance is a yes instance in the following steps.

First, MM$(P - P^*) = 1$. Equivalently, we prove that MM$(P - P^*) \notin (\{3, 4\} \cup X)$.

- Suppose for the sake of contradiction that MM$(P - P^*) = 3$. Then, because everyone in P^* prefers 3 to 2, P^* must be contained in $P_2 \cup P_3$ and the $\frac{q}{3}$ copies of $[1 \succ 4 \succ 3 \succ 2 \succ X]$ in P_4. Let $n_1 = |P^* \cap (P_2 \cup P_3)|$ and $n_2 = |P^* \cap P_4|$. It follows that the min-score of 2 in $P - P^*$ is $n_2 - n_1$ (via alternative 1), and the min-score of 3 in $P - P^*$ is at most $n_2 - n_1 - \frac{2q}{3}$ (via alternative 4). This means that 3 cannot be the maximin winner, which is a contradiction.
- Suppose for the sake of contradiction that MM$(P - P^*) = 4$. Then, because everyone in P^* prefers 4 to 2, P^* must be contained in the $\frac{q}{3}$ copies of $[1 \succ 4 \succ 3 \succ 2 \succ X]$ in P_4. Let $n^* = P^* \cap P_4$. It follows that the min-score of 2 in $P - P^*$ is n^* (via 1 and 3), and the min-score of 4 in $P - P^*$ is at most $-n^* - 2q$ (via alternative 2). This means that 4 cannot be the maximin winner, which is a contradiction.
- Suppose for the sake of contradiction that MM$(P - P^*) = x \in X$. Then, because everyone in P^* prefers x to 2, $P^* \subseteq P_1$ and $|P^*| \leq 3$ because x appears in 3 sets in $\mathcal{S}$. Clearly the min-score of 2 is strictly larger than the min-score of x in $P - P^*$, which is a contradiction.

Second, w.l.o.g., $P^* \subseteq P_1$. Because everyone in P^* prefers 1 to 2, P^* is a subset of P_1 and the $\frac{q}{3}$ copies of $[1 \succ 4 \succ 3 \succ 2 \succ X]$ in P_4. Notice that removing a vote of $[1 \succ 4 \succ 3 \succ 2 \succ X]$ always reduces the min-score of 1 by one. Therefore, if the min-score of 1 is not lower than the min-score of 2 in $P - P^*$, then it will also not be lower than that of 2 in $P - (P^* \cap P_4)$, which means that we can assume that $P^* \subset P_1$.

Third, P^* corresponds to a solution to the RXC3 instance. Let $\mathcal{S}^* = \{S : R'_S \in P^*\}$. We first note that $|P^*| \leq \frac{q}{3}$. Otherwise, the min score of 3 is $|P^*| - \frac{2q}{3}$ (via 4), which is strictly larger than $-|P^*|$, which is not smaller than the min-score of 1 (via 2 or alternatives in X). This contradicts the assumption that MM$(P - P^*) = 1$. Now, suppose for the sake of contradiction that $\mathcal{S}^*$ is not a solution to the RXC3 instance, which means that an alternative x is not contained in any set in $\mathcal{S}^*$. Then, the min-score of 1 is $-|P^*| - 2$ (via x), which is strictly smaller than the min-score of 2 (via 3). This contradicts the assumption that MM$(P - P^*) = 1$ and concludes the proof.

Other tie-breaking mechanisms. For Maximin, we cannot find a same order as LEX in the profile. Therefore, we find another order in the profile that satisfies all tie-breaking. Note that the only tie-breaking in the proof is $r(P - P^*)$, where 1 needs to beat 2 and 3 in the tie-breaking. Therefore, we choose $1 \succ 4 \succ 3 \succ 2 \succ X$ in P_4.

FA. Switch P_1 and P_4, let the agent with preference $1 \succ 4 \succ 3 \succ 2 \succ X$ become agent 1.

MPSR For convenience let us denote those in P_4 with preference $1 \succ 4 \succ 3 \succ 2 \succ X$ as P_4^1, others as P_4^2. Let every voter in P_4^1 have preference $1 \succ 4 \succ 3 \succ 2 \succ 5 \succ 6 \succ 7 \succ \cdots \succ q+4$. For every other voter, $q!$ permutations in X can guarantee every voter have a different permutation, thus different preference. Therefore, $1 \succ 4 \succ 3 \succ 2 \succ 5 \succ 6 \succ 7 \succ \cdots \succ q+4$ with $\frac{q}{3}$ votes is the most popular.

In *Second* part we prove that w.l.o.g $P^* \subseteq P_1$. Note that here our goal is to show a yes instance for RXC3. Since we can find one with $P^* \subseteq P_1$, we can recognize that P_4 voters will not abstain from voting. Therefore, the tie-breaking order does not change. □

B.3 Full Proof of Theorem 3(Condorcetified)

Theorem 3 (**Condorcetified positional scoring rules).** *For any Condorcetified integer positional scoring rule $Cond_{\vec{s}}$ and any tie-breaking rule,* GNSP-*$Cond_{\vec{s}}$ is NP-complete to compute.*

Proof. The proof is similar to the proof of Theorem 1. Since both finding a Condorcet winner and running a scoring rule can be done in polynomial time, the problem is obviously in NP. Then, the NP-hardness is proved by the reduction from RXC3. The profile $P = P_1 \cup P_2 \cup P_3 \cup P_4$, where P_1, P_2, and P_3 is the same as the construction in Theorem 1. The difference is P_4. Let σ^i denote the permutation $i \to (i+1) \to \cdots \to (q+2) \to 3 \to 4 \to \cdots \to (i-1)$, i.e., the permutation on X with the position of 1 and 2 unchanged. Let P_4 denote the $2q^2$-profile that consists of q copies of the following $2q$ votes:

$$\{\sigma^i(1 \succ X \succ 2), \sigma^i(2 \succ 1 \succ X) : 1 \leq i \leq q\}.$$

In the rest of the proof, we (1) show that alternative 1 is the winner of profile P under the positional scoring rule, and (2) if RXC3 is a yes instance, there exists a group P^* such that the winner of $P - P^*$ is alternative 2 as a Condorcet winner. The reasoning from a group no-show paradox to an RXC3 yes instance is similar to the proof of Theorem 1.

1 is the winner. First, there is no Condorcet winner under profile P. This is because the WMG of P is very similar to the WMG in Figure 2 except that $1 \succ X$ has a weight of $2q^2 - \frac{1}{3}q + 5$ by the difference in P_4. Therefore, the winner is determined by the positional scoring rule.

We count the number of each rank for each alternative in each part of the profile in Table 3. For example, $q-1 : q$ in the second row and the second column means "alternative 1 is ranked $q-1$ for q times in P_1". Note that the third column is an upper bound of ranks that an alternative $a \in X$ can achieve. The right-bottom entry means "alternative a is *at best* ranked each of $2, 3, \cdots, q+1$ for q times and each of $3, 4, \cdots, q+2$ for q times".

Alternatives	1	2	Upper bound of $a \in X$
P_1(rank: times)	$q-1 : q$	$q-2 : q$	$1 : q$
P_2	$1 : \frac{2q}{3} - 3$	$2 : \frac{2q}{3} - 3$	$3 : \frac{2q}{3} - 3$
P_3	$1 : 2$	$q+2 : 12 : 1$	$2 : 13 : 1$
P_4	$1 : q^2 2 : q^2$	$q+2 : q^2 1 : q^2$	$i : q$ for$i = 2, 3, \cdots, q+1$. $j : q$ for$j = 3, 4, \cdots, q+2$.

Table 3: Number of ranks for alternatives in each part of the profile.

Now we compare the score of each alternative and show that alternative 1 is the winner. Firstly, alternative 1 dominates alternative 2 Denote the scoring vector as $\vec{s} = (s_1, \ldots, s_{q+2}) \in \mathbb{Z}^{q+2}$ with $s_1 \geq s_2 \geq \cdots \geq s_{q+2}$ and $s_1 > s_{q+2}$.

- **1's score is higher than 2's.** First, alternative 1 is ranked higher alternative 2 in P_2, P_3, and $P_1 \cup P_4$. Moreover, in P_3 1 is ranked 1st in both two votes, while 2 is ranked the last $(q+2)$ in one vote. Therefore, 1's score is strictly higher than 2's.

- **1's score is higher than any** $a \in X$**'s.** We'll show that 1's score is higher than the upper bound. Firstly, 1's rank dominates a's rank in P_2 and P_3. Then we look at P_1 and P_4 together. In P_1, a's score exceeds 1's score by $q(s_1 - s_{q-1})$. However, in P_4, 1's score exceeds a's score by

$$\sum_{i=2}^{q+1} q(s_1 - s_i) + \sum_{j=3}^{q+2} q(s_1 - s_j) \geq q(s_1 - s_{q+1} + s_1 - s_{q+2}).$$

Therefore 1's score exceeds a's for at least $s_1 - s_{q+2} > 0$.

Therefore, we prove that 1's integer positional score is strictly higher than any other alternative and is the winner in P.

2 is the Condorcet winner after abstention. For any alternative $a \neq 1$, removing a vote where $a \succ 1$ does not reduce the score difference between 1 and a. Therefore, if the group no-show paradox happens by removing $P^* \subseteq P$, then the winner in $P - P^*$ must be a Condorcet winner. Similarly to Figure 2, after removing a group $P^* \subseteq P_1$ that corresponds to an exact cover in RXC3, alternative 2 becomes the Condorcet winner. $\square$

B.4 Full Proof of Theorem 4(STV)

Theorem 4 (**STV**). *For any* $0 \leq \alpha \leq 1$, GNSP-*STV is NP-complete to compute, where the tie-breaking mechanism is* LEX, FA_1, MPSR+LEX *or* MPSR+FA_1.

Proof. Membership in NP is straightforward. The hardness is proved by a reduction from RXC3 that is similar to the reduction in the hardness proof for the manipulation problem under STV (Bartholdi and Orlin, 1991). For any RXC3 instance $(X, \mathcal{S})$, where $X = \{x_1, \dots, x_q\}$ and $\mathcal{S} = \{S_1, \dots, S_q\}$, we construct the following GNSP-STV instance.

Alternatives: there are $3q + 3$ alternatives $\{w, c\} \cup \{d_0, d_1, \dots, d_q\} \cup \{b_1, \bar{b}_1, \dots, b_q, \bar{b}_q\}$. We assume that b_i has higher priority than $\bar{b}_i$, and $d_1 \succ d_2 \succ \cdots \succ d_q$ in tie-breaking.

Profile: The profile P consists of the following votes, where the top preferences are specified and the remaining alternatives ("others") are ranked arbitrarily.

- $\boldsymbol{P_1}$: There are $12q$ votes of $[c \succ w \succ \text{others}]$.
- $\boldsymbol{P_2}$: There are $12q - 1$ votes of $[w \succ c \succ \text{others}]$.
- $\boldsymbol{P_3}$: There are $10q + 2q/3$ votes of $[d_0 \succ w \succ c \succ \text{others}]$.
- $\boldsymbol{P_4}$: For every $j \in \{1, \dots, q\}$, there are $12q - 2$ votes of $[d_j \succ w \succ c \succ \text{others}]$.
- $\boldsymbol{P_5^1}$: For every $i \in \{1, \dots, q\}$, there are $6q + 4i - 2$ votes of $[b_i \succ \bar{b}_i \succ w \succ c \succ \text{others}]$; and $\boldsymbol{P_5^2}$: for every $i \in \{1, \dots, q\}$, there are two votes of $[b_i \succ d_0 \succ w \succ c \succ \text{others}]$.
- $\boldsymbol{P_6^1}$: For every $i \in \{1, \dots, q\}$, there are $6q + 4i - 6$ votes of $[\bar{b}_i \succ b_i \succ w \succ c \succ \text{others}]$; and $\boldsymbol{P_6^2}$: for every $i \in \{1, \dots, q\}$ and every $j \in S_i$, there are two votes of $[\bar{b}_i \succ d_j \succ w \succ c \succ \text{others}]$.
- $\boldsymbol{P_7}$: For every $i \in \{1, \dots, q\}$, there is a vote $[\bar{b}_i \succ b_i \succ c \succ w \succ \text{others}]$.

We first show STV$(\boldsymbol{P}) = \boldsymbol{w}$. In the first round, the plurality scores of the alternatives are as in the following table.

Rd.	w	c	b_i	$\bar{b}_i$	d_0	d_j
1	$12q - 1$	$12q$	$6q + 4i$	$6q + 4i + 1$	$10q + \frac{2q}{3}$	$12q - 2$
$q + 1$	$12q - 1$	$12q$	Removed	$12q + 8i - 1$	$12q + \frac{2q}{3}$	$12q - 2$

In the first q rounds, the order of elimination is $b_1, b_2, \dots, b_q$ (whose votes are transferred to $\bar{b}_1, \dots, \bar{b}_q$ and d_0). At the beginning of round $q + 1$, the plurality scores of the remaining alternatives are also shown in previous table. Then, d_q is eliminated, whose votes transfer to w. In the remaining rounds w is never eliminated and will become the winner.

Suppose the RXC3 **is a yes instance,** and let $\mathcal{S}^* \subseteq \mathcal{S}$ denote the solution. Let $I = \{i \leq q : S_i \in \mathcal{S}^*\}$. We prove that the GNSP-STV instance is a yes instance by showing that agents in P_7 whose top choices are b_i such that $i \in I$ have incentive to (jointly) abstain from voting. It is not hard to verify that after they abstain from voting, in the first q rounds, for each $i \leq q$, $\bar{b}_i$ is eliminated if and only if $i \in I$, otherwise b_i is eliminated. Notice that if b_i is eliminated, then $6q + 4i - 2$ of its votes (in P_5^1) transfer to $\bar{b}_i$, and two of its votes (in P_5^2) transfer to d_0. If $\bar{b}_i$ is eliminated, then $6q + 4i - 5$ of its votes (in P_6^1) transfer to b_i, and six of its votes (in P_6^2) are distributed evenly among d_j whose indices are the three alternatives in S_i. Therefore, in the beginning of round $q + 1$, the plurality scores of the remaining alternatives are as in the following table.

Rd.	w	c	b_i or $\bar{b}_i$	d_0	d_j
$q + 1$	$12q - 1$	$12q$	$12q + 8i - 1$ or $12q + 8i - 5$	$12q$	$12q$

Therefore, w is eliminated in round $q + 1$, whose votes transfer to c. It is not hard to verify that c will be the winner, and everyone in P_7 (including the agents who abstain from voting) prefers c to w.

Suppose the GNSP-STV **instance is a yes instance.** We prove that the RXC3 instance is a yes instance by proving that the new winner must be c and the absent votes in P_7 constitutes a solution to the RXC3 instance.

First, we prove that the new winner must be c. Suppose for the sake of contradiction this is not true. Notice that c and w are adjacent in all votes. Therefore, if in any round c is eliminated, then all of its votes are transferred to w; and vice versa. Consider the round right after c or w is eliminated. Then, the remaining alternative in $\{c, w\}$ is ranked at the top in at least $24q - 1$ votes (in $P_1 \cup P_2$). Moreover, given that one of $\{c, w\}$ is not eliminated, in any round we have the following upper bounds on the plurality scores of other alternatives (which are no more than the number of votes they ranked higher than c and w).

- For every $i \in \{1, \ldots, q\}$, the plurality score of b_i or $\bar{b}_i$ is at most $21q$ (all votes in $P_5^1 \cup P_5^2 \cup P_6^1 \cup P_6^2 \cup P_7$ where b_i or $\bar{b}_i$ is ranked at the top).
- The plurality score of d_0 is at most $(12 + \frac{2}{3})q$ (all votes in $P_3 \cup P_6^2$).
- For every $j \in \{1, \ldots, q\}$, the plurality score of d_j is at most $12q + 4$ (all votes in $P_4 \cup P_5^2$ where d_j is ranked higher than c).

Notice that all upper bounds are lower than $24q - 1$. Therefore, the winner is c or w, which contradicts the assumption.

Given that the new winner is c, because only agents in P_1 and P_7 rank c above w, only they have incentive to abstain from voting. Let I denote the indices i's of agents abstain from voting in P_7 whose top-ranked preferences are $\bar{b}_i$. It is not hard to verify that at the beginning of round $q + 1$, the plurality score of d_0 is $10q + \frac{2q}{3} + 2(q - |I|)$, and for every $1 \leq j \leq q$, the plurality score of d_j is $12q - 2$ if and only if d_j is not in any S_i with $i \in I$. Suppose for the sake of contradiction that $\mathcal{S}^* = \{S_i : i \in I\}$ is not a solution to the RXC3 instance. Then, in round $q + 1$, d_j for some $j \in \{0, 1, \ldots, q\}$ is removed. Once d_j is removed, all of its votes will transfer to w, and subsequently, w will not lose in any rounds, which contradicts the assumption that the winner is c. Therefore, the RXC3 instance is a yes instance. This completes the proof.

Other tie-breaking mechanisms . For STV, we still use the strategy that constructing agents that have the same tie-breaking order as LEX. We use voters in P_2 to achieve this.

FA$_1$. Let every voter of P_2 have the preference $w \succ c \succ b_1 \succ \bar{b}_1 \succ b_2 \succ \bar{b}_2 \succ \cdots \succ b_q \succ \bar{b}_q \succ d_0 \succ d_1 \succ d_2 \succ \cdots \succ d_q$. Then switch P_1 and P_2 and let agent 1 to be the first agent of P_2.

MPSR Let every voter of P_2 have the preference $w \succ c \succ b_1 \succ \bar{b}_1 \succ b_2 \succ \bar{b}_2 \succ \cdots \succ b_q \succ \bar{b}_q \succ d_0 \succ d_1 \succ d_2 \succ \cdots \succ d_q$. For other groups P_i, let every two voter in the same group have different permutation of their "other" part. This can be easily achieved since no group has more than $12q$ voters, but the 'other' part in a group has at least $q!$ permutations. In this way, the tie-breaking order we want has a population of $12q$ while other orders has at most 7 voters.

For STV, agents in P_2 will not abstain from voting as well. Therefore the tie-breaking order will not change.

□

C Algorithms for Verification of GNSP

C.1 Search-based Algorithms

We first present a brute-force algorithm (Algorithm 1) that solves the GNSP-r problem by doing a breadth first search. For each linear order $R_i \in \mathcal{L}(\mathcal{A})$ for $i = 1$ to $m!$, if voters of that ranking prefer another alternative to the current winner, we check if reducing votes for that ranking can lead to the group no-show paradox. This search enumerates all possible group abstentions, as mentioned in the proof of Proposition 1.

Algorithm 1: BFS algorithm for GNSP-r

Input: A profile P
Output: Minimum k for which GNSP-r occurs if such k exists
Set $a \leftarrow r(P)$
for *every alternative* $b \in A - \{a\}$ **do**
  Initialize $index = 1$
  Initialize a queue $\mathcal{Q}$ with initial state $(P, index)$
  while $\mathcal{Q}$ *is not empty* **do**
    Pop (P', index) from Q
    if $r(P') = b$ **then**
      return $|P| - |P'|$
      # $|P|$ *is the number of voters in profile* P;
    for $i = index$ *to* $m!$ **do**
      if $R_i \in R_{b \succ a}$ **then**
        $P_{new} = ReduceCount(P', R_i)$
        Append (P_{new}, i) to queue $\mathcal{Q}$
return *No solution found*

The ReduceCount function just reduces the number of votes for linear order R_i by 1 if the number of votes for R_i is non-zero. Thus, it returns a new smaller preference profile with abstention.

From Proposition 1, all possible group abstention combinations are bounded by $(\frac{n}{m!})^{m!}$. Thus, the run-time for Algorithm 1 is $O(\frac{n}{m!})^{m!} \cdot$ Run-time(r). Even though, this is a polynomial time for fixed m, the degree of the polynomial is very high indeed. Thus, the BFS algorithm becomes too expensive for large n.

While we are not able to improve the asymptotic run-time for the search, we can propose a couple of possible methods of speeding up the search time.

- **Pruning the search tree.** Consider the following example for Copeland. For a profile P with three alternatives, a, b and c, let the Copeland winner $\text{Cd}_\alpha(P)$ be a. Assume that we want to check if b can be the winner because of a GNSP occurrence. This will happen if some agents with preference $b \succ a$ can abstain from voting to make b the new winner. Rankings with $b \succ a$ are $[b \succ c \succ a], [c \succ b \succ a]$ and $[b \succ a \succ c]$ Between $[b \succ c \succ a]$ and $[c \succ b \succ a]$, if voters with either ranking abstain from voting, it would have the same effect on pairs of alternatives (a, b) and (a, c). The effect would only be different on (b, c), with $[b \succ c \succ a]$ abstentions benefiting c more than b. We remark that if k agents voting $[c \succ b \succ a]$ abstaining does not change the result in favor of b, then k agents voting $[b \succ c \succ a]$ abstaining will not do so either. Thus, $[c \succ b \succ a]$ and $[b \succ c \succ a]$ are similar effect rankings for this scenario. The existence of similar effect rankings allows us to prune certain branches during our breadth-first-search.
- **Parallel BFS.** Breadth-first-search can be parallelized by making the search proceed layer by layer by using modified data structures such as bags instead of regular queues (Leiserson and Schardl, 2010). For our search problem, the diameter can be at most n, the number of agents, while the number of nodes is much larger. This leads us to think that their will be an almost linear improvement with number of processes with a parallel implementation.

Both of the potential improvements can lead to linear improvements of run-time. And particularly for small number of agents, where the search-based algorithm is most effective, these improvements can make the algorithm more useful in practice. We leave such improvements as possible future work. In this work we focus particularly on the ILP-based algorithms as they perform comparatively well for small number of agents, and outperform the search algorithms by a large margin for a larger number of agents.

C.2 Integer Linear Programming (ILP)-based Algorithms

In this section, we present the ILP formulations and ILP-related algorithms for four voting rules. Copeland, maximin, Black's rule and STV. The variable formulation is same for all voting rules, as explained in Section 4. The chosen objective function actually returns the smallest group size, k, for which there exists a group of agents who has incentive to abstain from voting. If there is no feasible solution for any of the formulated ILPs, obviously GNSP does not occur.

Next, we discuss how to construct specific ILPs for each of the voting rules mentioned. For all voting rules, we assume a is the original winner for preference profile P and lexicographic tie-breaking is used. The presented algorithms can very easily adapted to other tie-breaking methods as well.

Copeland

All the alternatives should be enumerated through except for the original Copeland winner for the profile to check for possible no-show paradoxes. So there will be $m - 1$ different ILPs in total. Each of them will be constructed as follows if the alternative b is a possible winner:

Recall that $P_x[c \succ d]$ is the number of occurrences of $c \succ d$ in the profile created by x_i agents voting for each linear order.

$$P_x[c \succ d] = \sum_{R_i \in R_{b \succ a}} x_i \mathbb{1}_{c \succ d}(R_i) + \sum_{R_i \notin R_{b \succ a}} n_i \mathbb{1}_{c \succ d}(R_i)$$

To define the ILP, we use additional auxiliary variables. Let the binary variable q_{cd} indicate that a majority prefers c over d, i.e.,

$$q_{cd} = \begin{cases} 1, & \text{if } P_x[c \succ d] > P_x[d \succ c] \\ 0, & \text{otherwise} \end{cases}$$

And add another binary variable r_{cd} indicating that if there is a majority tie, i.e.,

$$r_{cd} = \begin{cases} 1, & \text{if } P_x[c \succ d] = P_x[d \succ c] \\ 0, & \text{otherwise} \end{cases}$$

The Copeland score of any alternative c in P_x will be

$$\text{CS}_{P_x}(c) = \sum_{d \neq c} q_{cd} + \alpha \sum_{d \neq c} r_{cd} \;\; \forall c$$

Additionally, the constraints on the score and the auxiliary variables will be

$$\mathrm{CS}_{P_x}(b) \geq \mathrm{CS}_{P_x}(c) \;\; \forall c \neq b$$
$$q_{cd} + r_{cd} + q_{dc} = 1 \;\; \forall c \neq d$$

Our objective function would be

$$\text{minimize} \sum_{R_i \in R_{b \succ a}} (n_i - x_i)$$

where we optimize for the minimum number of abstaining agents.

Note that we would have to iterate through all alternatives except for the original Maximin winner for the profile to check for possible no-show paradoxes. So, in total we have $m-1$ different ILPs. It is possible to come up with a single ILP with more conditional constraints covering all of the scenarios. But in practice, that alternate implementation performed worse.

Another formulation that we tried for the ILPs considered a more complex formulation. For example, for Copeland, we would consider all possible unweighted majority graphs (UMG, majority graphs without the edge weights). The Copeland winner is uniquely determined by the UMG. We enumerated all possible UMGs, and constructed separate ILPs for each UMG. While the individual ILPs were much simpler than what we have now, this would lead to an exponential number of ILPs, and in practice performed much less efficiently.[1]

Maximin

Maximin score of any alternative b in P_x, $\mathrm{MS}_{P_x}(b) = \min_{d \in \mathcal{A}} w_{P_x}(b, d)$. An alternative b will be the Maximin winner, if $\mathrm{MS}_{P_x}(b) \geq \mathrm{MS}_{P_x}(c)$ for all $c \neq b$. We can encode this in an ILP by defining auxiliary variables for the maximin scores and the following constraints:

$$\mathrm{MS}_{P_x}(c) = \min_{d \neq c} P_x[c \succ d] \;\; \forall c$$
$$\mathrm{MS}_{P_x}(b) \geq \mathrm{MS}_{P_x}(c) \;\; \forall c \neq b$$

The constraints using the min operator can be used because, by definition, all of the $P_x[c \succ d]$ values are upper bounded by the number of voting agents.

Additionally, the constraints here state that the maximin score of the winner must be greater than or equal to the maximin score of all other alternatives. In reality, based on the tie-breaking rule being used, some of these inequalities need to be strict. For example, if lexicographic tiebreaking is being used, the maximin score of an alternative must be strictly higher than the maximin scores of all alternatives that are lexicographically ahead.

Our objective function would be the same as Copeland, where we optimize for the minimum number of abstaining agents:

$$\text{minimize} \sum_{R_i \in R_{b \succ a}} (n_i - x_i)$$

Again, we would have to iterate through all alternatives except for the original Maximin winner for the profile to check for possible no-show paradoxes. So, in total we have $m-1$ different ILPs.

STV

We introduce some additional notation for STV. For a ranking R, let $R(a)$ be the rank of alternative a, and $R^{-1}(i)$ be the i-th ranked alternative in R. For example, if $R = a \succ b \succ c$, then $R(b) = 2$, and $R^{-1}(2) = b$.

To define the ILP, we use additional auxiliary variables. STV is a multi-round voting rules, where each round a new alternative is eliminated. The STV winner is the alternative that is not eliminated at the end of the $m-1$ rounds. Denote $e_{a,j}$ as a binary variable that is 1 if alternative a is eliminated in round j and 0 otherwise. Assume that n_i is the number of agents with R_i as their ranking.

Now consider the ranking $R = a \succ b \succ c \succ d$. If both a, b have been eliminated, and c has not been eliminated, then in round 3, any agent with this ranking will be counted towards alternative c's score. With this in mind, the score of an alternative b in round r can be calculated as follows.

$$\begin{aligned} \mathrm{STVscore}(b, r) = (1 - \sum_{j=1}^{r} e_{b,j}) \Big[& \sum_{R_i | R_i(b)=1} n_i \\ & + \sum_{k=2}^{r} \sum_{R_i | R_i(b)=k} n_j \big(\prod_{k'=1}^{k-1} (\sum_{j=1}^{r} e_{R^{-1}(k'),j}) \big) \Big] \end{aligned} \tag{1}$$

[1]In fact, we tried such alternate formulations for all of the voting rules discussed below. And in all cases, they were outperformed by the currently presented ILPs.

While this formula may look somewhat formidable, it follows quite simply from the fact that in round r, the score for alternative b would be non-zero only if b has still not been eliminated ($\sum e_{b,j} = 0$). Additionally, the score will get contributions from rankings where all alternatives preferred to b has been already eliminated ($\sum e_{c,j} = 1$ for c when $R(c) < R(b)$).

For a preference profile with abstentions (P_x), STV score will be calculated using x_i instead of n_i for alternatives. With all these, it is now quite simple to present the constraints for the ILP. The winning alternative b will not be eliminated in the $m-1$ rounds, whereas all other alternatives will be. At each round, whichever alternative has $e_{c,j} = 1$, needs to have it's score lesser than (or equal to) all of the thus far non-eliminated alternatives.

Thus we get the following constraints:

$$e_{c,j} = 1 \implies \text{STVscore}_{P_x}(b, r) \leq \min_{c \mid \sum_{j=1}^{r-1} e_{c,j} = 0} \text{STVscore}_{P_x}(c, r)$$

$$\sum_{j=1}^{m-1} e_{c,j} = 1 \;\; \forall c \neq b$$

$$\sum_{j=1}^{m-1} e_{b,j} = 0$$

ILP and MIP (mixed integer programming) solvers, for the most part cannot handle products of more than two (integer or binary) variables. Whereas, our STV score function (Equation 1) has products of binary variables. To solve this problem, we used a dynamic programming approach with auxiliary variables to construct the auxiliary variables and constraints efficiently. In total this leads to this leads to $O(\text{number of unique rankings} \cdot m^2)$ additional variables and constraints. But as long as the number of unique rankings is moderate and m not too large, this is still a reasonable amount of constraints for efficient ILP solvers to solve. The objective function is the same as the other voting rules, optimizing for the minimum number of abstaining agents.

For STV, we also considered a couple of other approaches, also considering the elimination order. The first approach, creates a tree based on the elimination orders, where each leaf indicates a different elimination order and the first level indicating the first alternative that has been eliminated. We added constraints to the ILPs one level of this tree at a time. If a particular alternative cannot be eliminated in round 1 by abstention, then all elimination orders stemming from these can be ignored. While this is somewhat efficient, the downside is that in the worst-case, this requires solving $O(m!)$ different ILPs, leading to bad run-times for high values of m.

The second approach considers binary variables for each elimination order with the sum of these variables being 1. This leads to additional conditional constraints. While this works well for small m, again due to $m!$ different elimination orders possible, this approach also is outperformed by our presented ILP formulation as m starts increasing.

Black's rule

The possible winner b needs to be enumerated as other voting rules. To make sure the Black's winner is b, the Condorcet winner constraints and the Borda winner constraints will be constructed for Black's rule. There will be $m-1$ different ILPs in total. We introduce three additional auxiliary variables. Let the binary variable w_C indicate that b is the Condorcet winner, the binary variable w_B indicate that b is the Borda winner, the binary variable w indicate that b is the Black's winner. So we have,

$$w = \begin{cases} 1, & \text{if } w_C = 1 \text{ or } w_B = 1 \\ 0, & \text{otherwise} \end{cases}$$

The constraints for the Condorcet winner are similar to the Copeland's ones, we still need to compute the Copeland score but the constraints on Copeland's score would be

$$w_C = \begin{cases} 1, & \text{if } \text{CS}_{P_x}(b) = m-1 \\ 0, & \text{otherwise} \end{cases}$$

As for Borda, the Condorcet winner should be b or not exist, so the constraint on the Copeland score and Borda score would be

$$w_B = \begin{cases} 1, & \text{if } \text{CS}_{P_x}(c) < m-1 \text{ and } s_B(b) \geq s_B(c) \;\; \forall c \neq b \\ 0, & \text{otherwise} \end{cases}$$

where $s_B(b)$ is the Borda score for alternative b.

The objective function remains the same:

$$\text{minimize} \sum_{R_i \in R_{b \succ a}} (n_i - x_i)$$

where we optimize for the minimum number of abstaining agents.

Run-time discussion

ILP solving algorithms are also NP-complete. While we can report the worst-case run-time for the different presented ILP formulations, we chose not to because in practice the run-time is often dependent on the particular heuristics used by an ILP solvers. For example, the popular branch and bound method is used by many ILPs to improve run-time, however estimating the size of branch-and-bound trees itself is a computationally hard problem Hendel et al. (2022). Additionally, ILP solvers use parallel computing (by default in most cases), thus also gaining linear run-time with large number of processors.

So, instead of trying to get ILP formulations with good worst-case complexity, we focused on creating ILP formulations where the number of variables and constraints were moderate sized and tested the run-times empirically. For all of our ILP formulations, the number of variables is determined by the number of unique rankings (particularly the size of the set $R_{b \succ a}$ for different b). The number of unique rankings is bound by $min(m!, n)$ and thus in scenarios with either a moderate number of agents or a small number of alternatives, we get reasonably good run-times for our ILP formulations.

D Additional Experiment Details

D.1 Experiment Details

Ranking Models

First we define the different models we use to generante synthetic ranking data.

Mallow's Model (Mallows, 1957) Given $0 < \phi < 1$, a linear order $W \in \mathcal{L}(\mathcal{A})$ and $m \in \mathbb{N}$ alternatives, the Mallows model with fixed dispersion has the following probability for any full linear order V: $\Pr_W(V) = \frac{1}{Z_{m,\phi}}\phi^{KT(V,W)}$, where $Z_{m,\phi}$ is a normalizing factor and $KT(V, W)$ is the Kendall's Tau distance between V, W.

For the Mallow's models, smaller values of ϕ mean more correlation between the votes whereas higher values mean independent votes. $\phi = 1$ coincides with the impartial culture model, which draws all votes in random.

Plackett-Luce (PL) Model (Plackett, 1975; Luce, 1977)

Given the parameters $\Theta = \{\vec{\theta} = \{\theta^j | 1 \leq j \leq m\}\}$, the PL model has the following probability of any full linear order $V = a_{j_1} \succ a_{j_2} \succ \cdots \succ a_{j_m}$: $\Pr_{\text{PL}}(\sigma|\vec{\theta}) = \prod_{p=1}^{m-1} \frac{\exp(\theta^{j_p})}{\sum_{q=p}^{m} \exp(\theta^{j_q})}$.

The parameters θ_j behaves like scores for the models. In our experiments, we use two types of PL models. First, where all of the θ values are sampled uniformly from $[0, 1]$. Second, where we force a tie in the score for the top-2 alternatives. This allows us to look into two types of profiles. The first one, where there is likely high correlation between all agents, the second, where it is possible that there are two groups with the different alternatives as their preference.

Single-peaked preferences (Conitzer, 2009). In single-peaked preferences, all the alternatives are considered on a line, with a peak for one alternative. And then the preference for alternatives decrease as we go further away from the peak.

Euclidean models (Szufa et al., 2020). Euclidean models can be considered a generalization of single-peaked preferences. In this model, all agents and alternatives are randomly placed on a d-dimensional space, and then each agent's preference is measured according to their Euclidean distance to all of the alternatives. For our experiments, we randomly place the agents in two ways. First, we draw uniformly at random from a d-dimensional hypercube, we call this Euclidean (uniform). For the second approach, we first sample a d-dimensional mean for a multivariate Gaussian distribution. Then, we generate agents from a distribution with that mean, a fixed variance in all dimensions, and no covariance. We call this Euclidean (Gaussian). For both approaches, the alternatives are drawn uniformly at random from a hypercube. The Euclidean (Gaussian) will have less randomness compared to the Euclidean (uniform) model.

Computational details.

All experiments were run on the CPU without use of GPU. The CPU configuration is given below.

- Architecture: x86_64
- CPU(s): 16
- RAM: 16GB
- Processor: 3.2 GHz
- Thread(s) per core: 2
- Core(s) per socket: 8
- Socket(s): 1
- Model name: AMD Ryzen 7 5800H with Radeon Graphics

ILP Solver

For all of the ILP formulations, we have the Gurobi ILP solver.

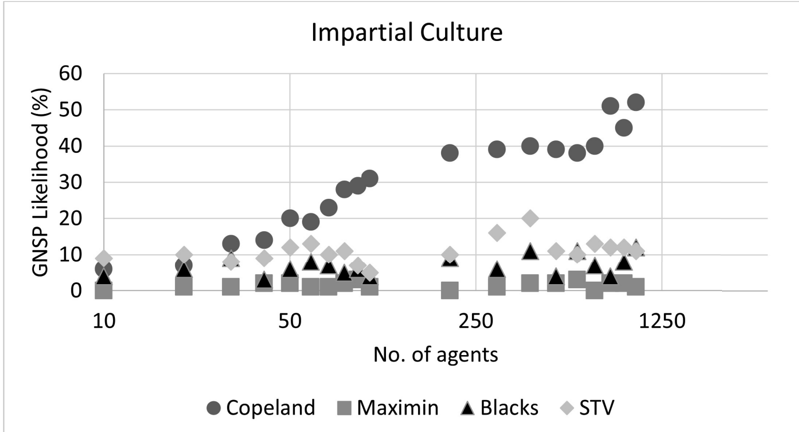

Figure 9: Empirical likelihood of GNSP for different voting rules and different number of agents for $m = 4$ for the IC model

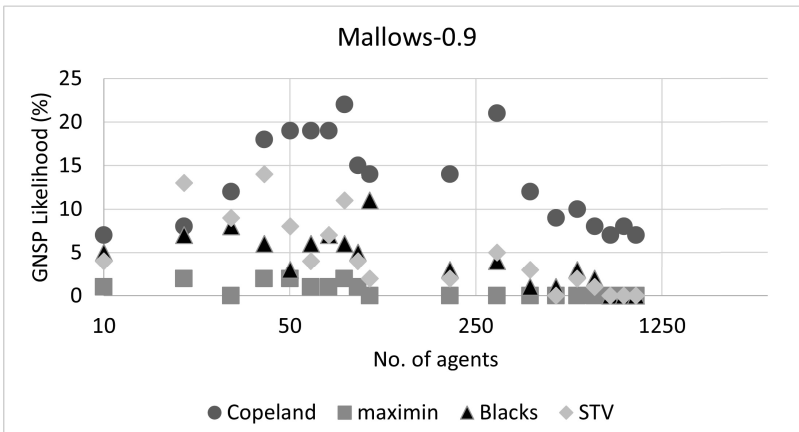

Figure 10: Empirical likelihood of GNSP for different voting rules and different number of agents for $m = 4$ for the Mallows-0.9 model

D.2 Experiment Results

Likelihood of group no-show paradox.

Figures 9-17 shows the empirical results for GNSP likelihood on synthetic data based on various ranking models for a fixed number of alternatives ($m = 4$). We notice some observations here.

- GNSP seems to be more likely for Copeland compared to the other voting rules for most models.
- Maximin seems to be the most robust to GNSP (at least for $m = 4$) for most distributions.
- Just like we saw for Copeland, for the other voting rules too, we see from the IC and the Mallows model plots that as correlation between the voter increases (lower ϕ value), the likelihood for GNSP decreases.
- Similarly, from the PL and PL (tied at top) models, we see that when there are potentially two groups of agents, the likelihood goes up. On the other hand, when the PL parameters are independently drawn, the likelihood is lower for all voting rules. Again, for the two Euclidean models, the one with Gaussian models of agents (correlated agents) have almost 0 likelihood.
- Interestingly, for single-peaked preferences, the likelihood of GNSP for STV is very high. On the other hand, the likelihood is nearly 0 (we don't observe any in our sampples) for the other voting rules. Similarly, for the Euclidean uniform preferences, the likelihood is high for STV. What causes this high likelihood in STV for single-peaked preferences is an important question for the future.

Finally, Figure 18 shows how the likelihood increases with the number of alternatives. We see that likelihood is higher for all of the voting rules with high m, although there is no linear growth here. This result is for the IC model. This matches with what we saw with real data for PrefLib, where the only instances of GNSP we saw came where there were large number of alternatives and unique, random rankings among the agents.

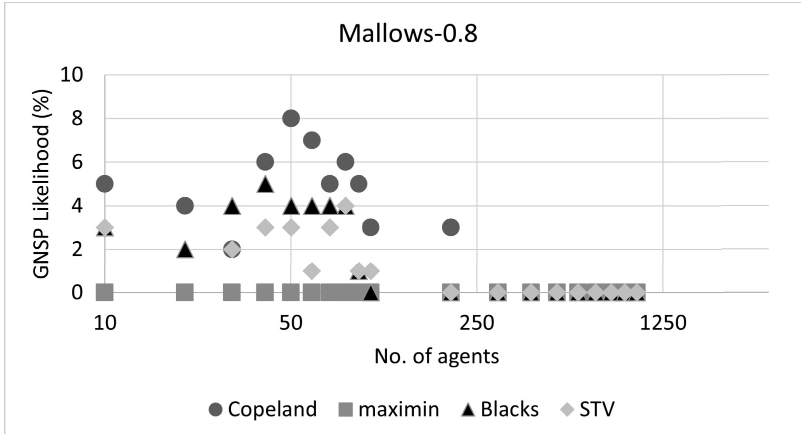

Figure 11: Empirical likelihood of GNSP for different voting rules and different number of agents for $m = 4$ for the Mallows-0.8 model

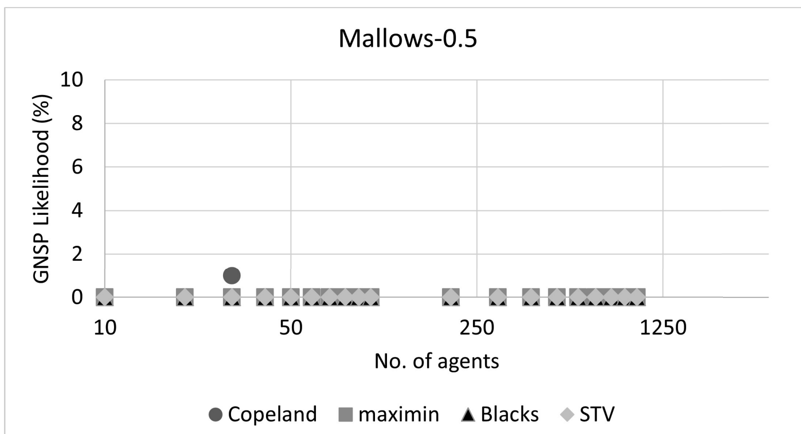

Figure 12: Empirical likelihood of GNSP for different voting rules and different number of agents for $m = 4$ for the Mallows-0.5 model

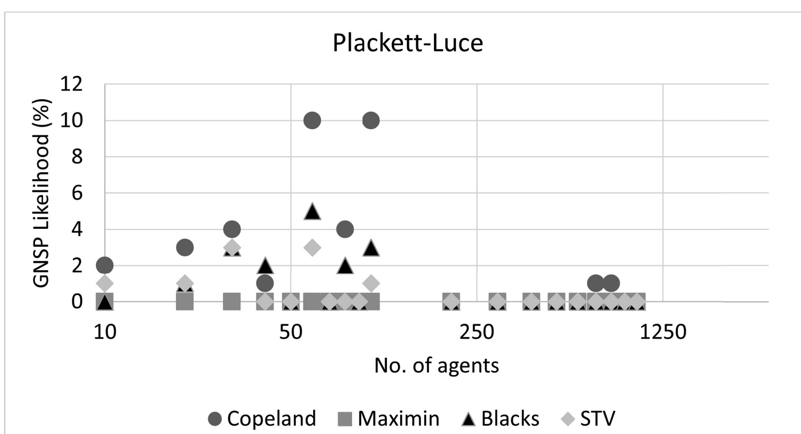

Figure 13: Empirical likelihood of GNSP for different voting rules and different number of agents for $m = 4$ for the PL (regular) model

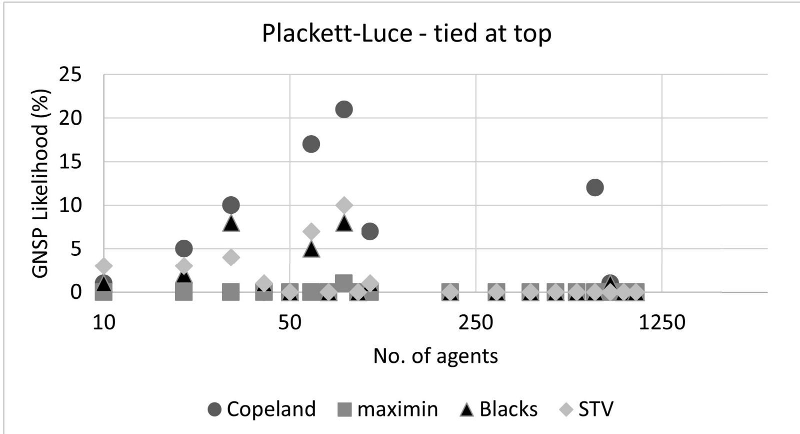

Figure 14: Empirical likelihood of GNSP for different voting rules and different number of agents for $m = 4$ for the PL (tied at top) model

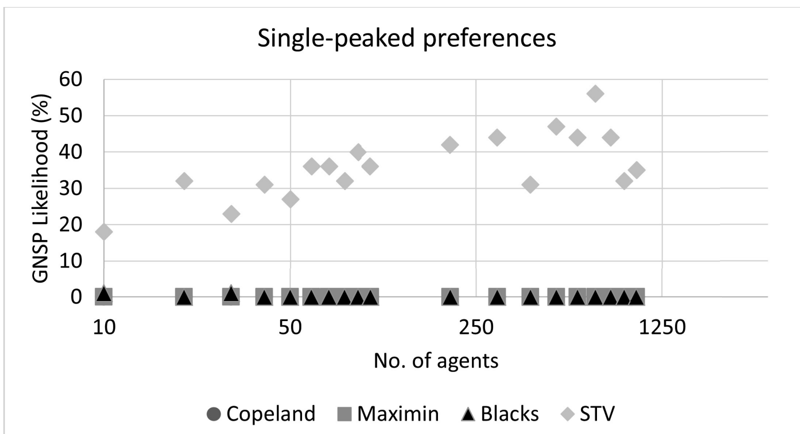

Figure 15: Empirical likelihood of GNSP for different voting rules and different number of agents for $m = 4$ for single-peaked preferences

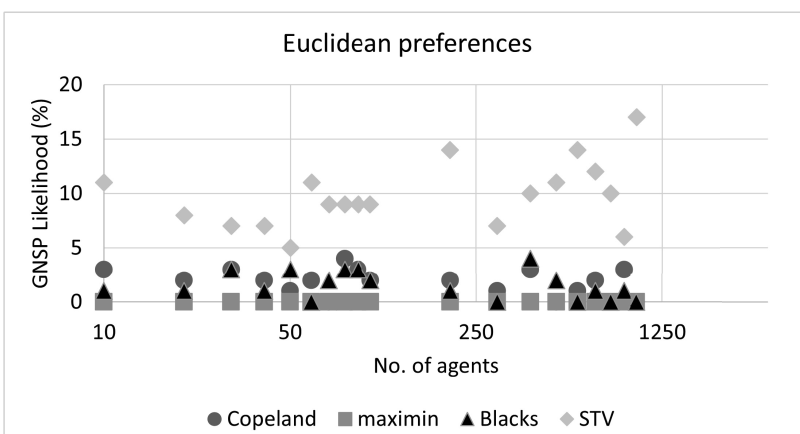

Figure 16: Empirical likelihood of GNSP for different voting rules and different number of agents for $m = 4$ for the Euclidean (uniform) model

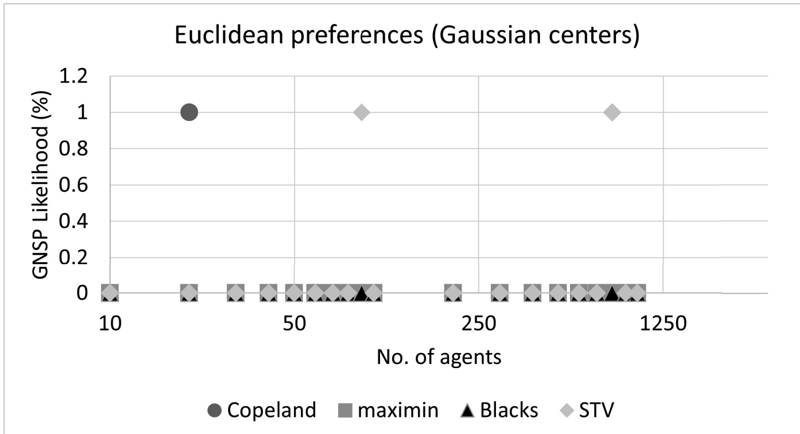

Figure 17: Empirical likelihood of GNSP for different voting rules and different number of agents for $m = 4$ for the Euclidean (Gaussian) model

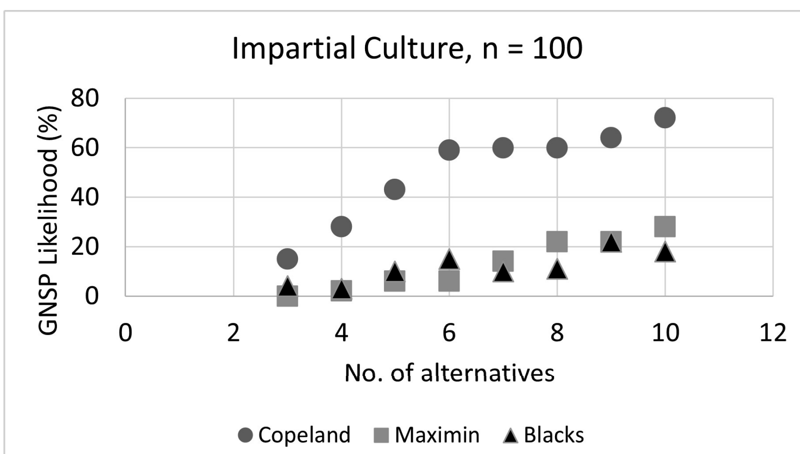

Figure 18: Empirical likelihood of GNSP for different voting rules and different number of alternatives for $n = 100$ for the IC model

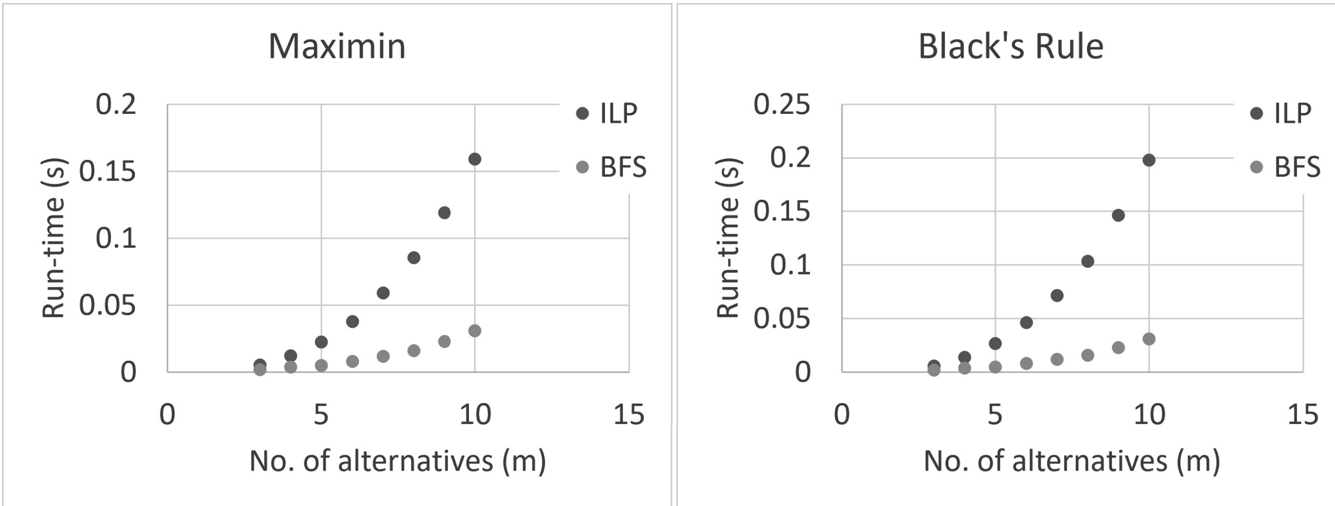

Figure 19: Run-time vs no.of alternatives for $n = 10$ for maximin, Black's and STV voting rule

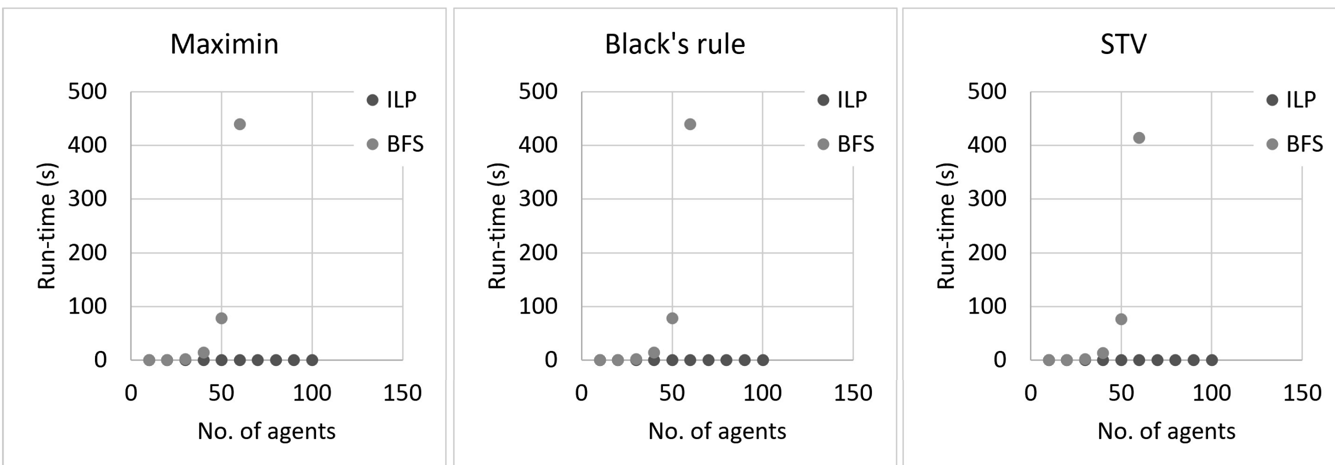

Figure 20: Run-time vs no.of agents for $m = 4$ for maximin, Black's and STV voting rule

Run-time for the BFS and ILP algorithms.

Figures 20 and 19 show the run-time of the two types of algorithms for different scenarios. As expected, for small n, BFS works well for any number of alternatives in the experiment, and even outperforms ILP. But similar to what we saw for Copeland, even for $n = 100$, BFS sees an exponential blow-up of time whereas ILP still works, Additionally, for moderate m, the ILP algorithm works well for all the voting rules.