Computational Complexity of Verifying the Group No-show Paradox

Farhad Mohsin[1] , **Qishen Han**[2] , **Sikai Ruan**[2] , **Pin-Yu Chen**[3] , **Francesca Rossi**[3] and **Lirong Xia**[2]
[1]College of the Holy Cross, Worcester, MA
[2]Rensselaer Polytechnic Institute, Troy, NY
[3]IBM Research, Yorktown Heights, NY

Abstract

The *(group) no-show paradox* refers to the undesirable situation where a group of agents have incentive to abstain from voting to make the winner more favorable to them. To understand whether it is a critical concern in practice, in this paper, we take a computational approach by examining the computational complexity of verifying whether the group no-show paradox exists given agents' preferences and the voting rule. We prove that, unfortunately, the verification problem is NP-hard to compute for some commonly studied voting rules, i.e., Copeland, Maximin, single transferable vote, and all Condorcetified positional scoring rules such as Black's rule. We propose integer linear programming-based algorithms and a search-based algorithm for the verification problem for different voting rules. Experimental results on synthetic data illustrate that the former is efficient when the number of unique rankings in a profile is not too high, and the latter is efficient for a small number of agents. With the help of these algorithms, we observe that group no-show paradoxes rarely occur in real-world data.

1 Introduction

In social choice theory, the *no-show paradox*, first observed by Fishburn and Brams [1983], generally refers to the counter-intuitive event where a group of agents has the incentive to abstain from voting to make the winner more favorable to them. This is undesirable because when it occurs, agents can manipulate the result just by not showing up, which is much easier (thus more threatening) than strategic manipulation [Gibbard, 1977; Satterthwaite, 1975]. The no-show paradox also discourages voters from participating in the election, reducing turnout and undermining democracy.

The significance of the no-show paradox urges researchers to study its existence under different voting rules. Unfortunately, even the *single-voter no-show paradox*,which means that there exists a single voter with the incentive to abstain from voting, always exists under a wide range of voting rules, including all Condorcet rules [Moulin, 1988]. There is an extensive literature on verifying the frequency of various kinds of no-show paradoxes. For example, Ray [1986] studied the likelihood of a *group no-show paradox* for three alternatives under single transferable vote (STV). Lepelley and Merlin [2001] generalized the concept from Ray [1986] and studied various kinds of group no-show paradoxes for scoring run-off methods. Brandt *et al.* [2021] characterized the likelihood of the no-show paradox on Condorcet rules via Ehrhart theory. Xia [2021] characterized the likelihood of the single-voter no-show paradox existing under a semi-random model.

We can verify the existence of a single-voter no-show paradox for many commonly-studied voting rules in polynomial time by simply enumerating the possible absentee. But we have multiple open questions for the group no-show paradox: **How likely is the occurrence of the group no-show paradox under commonly studied voting rules? What is the computational complexity of verifying the paradox?**

The question of computational complexity is interesting from a theoretical perspective and significant from a practical viewpoint. A high complexity of verifying the existence of a paradox can be advantageous for voting rules as that will disallow manipulation from voters. However, a low complexity can be advantageous from a mechanism designer's perspective because it would allow us to verify which voting rules are more robust against group abstention in practice.

Our contributions. We characterize the computational complexity of verifying group no-show paradox (GNSP) under several common voting rules: Copeland, Maximin, STV, and all Condorcetified positional scoring rules, including Black's rule (Section 3). We prove that, unfortunately, the verification problem is NP-complete under all of them (Theorem 1–4). To computationally solve the problem, we propose integer linear program (ILP)-based algorithms and a breadth first search search-based algorithm for verifying GNSP for these voting rules (Section 4). We perform experiments on both synthetic data and PrefLib election data [Mattei and Walsh, 2013] (Section 5). The results on synthetic data illustrate that the search algorithm works well for a small number of agents, outperforming the ILP algorithms only when the number of alternatives is high. On the other hand, the ILP algorithms work well even for a large number of agents, as long the number of unique rankings in a profile is not too large. The results on PrefLib data suggest that no-show paradoxes rarely occur in real-world elections.

1.1 Related Works and Discussion

The no-show paradox. Although it is widely acknowledged that the no-show paradox refers to agent(s) having incentives to abstain from voting, its mathematical definition varies in the literature. Fishburn and Brams [1983], who introduced the paradox, described it as a group of agents having incentive to abstain from voting, but restricted the votes to be "identical" and the full-vote winner being "ranked last". On the other hand, Moulin [1988] restricted the no-show paradox to a single agent while relaxing the "ranked last" constraint, and call the non-existence of such single-agent no-show paradox satisfaction of *participation*. Similarly, Felsenthal and Nurmi [2016] investigated two versions where abstentions of winners being ranked first or other alternatives being ranked last can change the result. Finally, Lepelley and Merlin [2001] investigated the group version as Fishburn and Brams did, while further redefining the paradox to four specific types. In this paper, we adopt a definition that inherits the spirits of all of these papers. We consider the group version of the no-show paradox, but we do not require the votes to be identical or the alternative to be ranked at specific places.

There is a large body of literature on the likelihood of no-show paradox under general assumptions such as impartial culture [Ray, 1986; Lepelley and Merlin, 2001; Plassmann and Tideman, 2014]. Kamwa *et al.* [2021] revisited Lepelley and Merlin [2001]'s setting under a single-peaked preference and found a much lower likelihood under this restriction. Pérez [2001] and Duddy [2014] studied strong versions of no-show paradox's likelihood in Condorcet rules. Brandt *et al.* [2021] also analyzed no-show paradox for Condorcet rules using Ehrhart Thoery. Brandl *et al.* [2015] showed that every Pareto Optimal majoritarian voting rule will suffer from the no-show paradox for some voting profiles. Brandt *et al.* [2017] followed Moulin [1988] and tightened the bound of the number of agents in a no-show paradox in Condorcet-consistent rules. Brandt *et al.* [2022] presented an ILP-based method for finding minimal voting paradoxes, including the no-show paradox to occur. But their work focused on preference profiles limited to 10 alternatives and 20 agents, whereas we explore larger ranges of preference profiles. Also, their ILP formulation depend on individual agents' preferences, which is different from our formulation.

The notion of the no-show paradox has also been extended beyond regular voting rules, e.g., social choice correspondences where the output is a set of alternatives [Jimeno *et al.*, 2009; Pérez *et al.*, 2012; Brandl *et al.*, 2019], and probabilistic or randomized voting rules [Brandl *et al.*, 2017].

Manipulation and control. The no-show paradox is closely related to *manipulation* in voting, particularly "manipulation by abstention" [Brandt *et al.*, 2021], which is a special case of "sincere truncation", where agents partially reveal their truthful preference [Fishburn and Brams, 1984]. Technically, verifying GNSP is similar to the COALITIONAL MANIPULATION (CM) problem in voting [Bartholdi and Orlin, 1991; Conitzer and Walsh, 2016]. where we are asked whether the manipulators can cast votes to make c the winner. One might be tempted to think that verifying GNSP is easier than CM. However, we do not see a formal relationship between the two problems. Because, first, when verifying GNSP, the group of "manipulators" is not fixed; and second, all absentees must prefer the new winner to the old winner. The no-show paradox is also related to *control* in voting, or more specifically, control by deleting agents [Bartholdi *et al.*, 1992; Rothe and Schend, 2013; Faliszewski and Rothe, 2016], where an adversary aims at achieving a goal, e.g., make a favorable alternative win or an unfavorable alternative lose, by deleting agents. The main differences between GNSP and control again stem from the fact that the size of the "deleted" agents is unbounded, and only agents who prefer the new winner to the old winner can be deleted.

2 Preliminaries

For any $m \in \mathbb{N}$, let $\mathcal{A}$ denote the set of $m \geq 3$ *alternatives*. Let $\mathcal{L}(\mathcal{A})$ denote the set of all linear orders or rankings over $\mathcal{A}$. Let $V = \{V_1, \dots, V_n\}$ be the set of agents for $n \in \mathbb{N}$. Each agent uses a linear order $R \in \mathcal{L}(\mathcal{A})$ to represent their preference, called a *vote*, where $a \succ_R b$ means that the agent prefers alternative a to alternative b. The vector of n agents' votes, denoted by P, is called a *(preference) profile*. In this paper, we focus on *resolute voting rules*, which always choose a unique winner. A voting rule is *anonymous* if the winner is insensitive to the identities of agents.

(Un)weighted majority graphs and Condorcet winners. For any profile P and any pair of alternatives a, b, let $P[a \succ b]$ denote the total number of votes in P where a is preferred to b. Let $\text{WMG}(P)$ denote the *weighted majority graph* of P, whose vertices are $\mathcal{A}$ and whose weight on edge $a \to b$ is $w_P(a, b) = P[a \succ b] - P[b \succ a]$. The *Condorcet winner* of a profile P, denoted by $\text{CW}(P)$, is the alternative that only has positive outgoing edges in $\text{WMG}(P)$. Note that a Condorcet winner will not exist for all profiles.

Integer positional scoring rules. An *(integer) positional scoring rule* $r_{\vec{s}}$ is characterized by an integer scoring vector $\vec{s} = (s_1, \dots, s_m) \in \mathbb{Z}^m$ with $s_1 \geq s_2 \geq \dots \geq s_m$ and $s_1 > s_m$. For any alternative a and any linear order $R \in \mathcal{L}(\mathcal{A})$, let $\vec{s}(R, a) = s_i$, where i is the rank of a in R. Given a profile P where n_R agents have R as their vote, the positional scoring rule $r_{\vec{s}}$ chooses the alternative a with maximum $\sum_{R \in P} n_R \cdot \vec{s}(R, a)$, using a tie-breaking mechanism when necessary. For example, the scoring vector is $(1, 0, \dots, 0)$ for *plurality* and $(m-1, m-2, \dots, 0)$ for *Borda*.

STV. The single transferable vote (STV) rule chooses the winner in at most $m-1$ rounds. For each $1 \leq i \leq m-1$, in round i, a loser (an alternative with the lowest score) under plurality is eliminated, applying a tie-breaking mechanism to choose a single loser when necessary.

Copeland. The Copeland rule is parameterized by a number $0 \leq \alpha \leq 1$, denoted by Cd_α. For any profile P, an alternative a gets 1 point for each alternative it beats in head-to-head competitions, and gets α points for each tie. Cd_α chooses the alternative with the highest Copeland score as the winner, using a tie-breaking mechanism when necessary.

Maximin. For each alternative a, its *min-score* is defined to be $\text{MS}_P(a) = \min_{b \in \mathcal{A} - \{a\}} w_P(a, b)$. Maximin, denoted by MM, chooses the alternative with the maximum min-score as the winner, using a tie-breaking mechanism when necessary. **Condorcetified (integer) positional scoring rules.** The rule is defined by an integer scoring vector $\vec{s} \in \mathbb{Z}^m$ and is denoted by $\text{Cond}_{\vec{s}}$, which selects the Condorcet winner when it exits, and otherwise uses $r_{\vec{s}}$ to select the winner. For example, *Black's rule* [Black, 1958] is the Condorcetified Borda rule.

Group no-show Paradox. As discussed in Section 1.1, we adopt the following definition that inherits the spirits of [Fishburn and Brams, 1983] and [Moulin, 1988].

For a profile P with agents V, let $V' \subseteq V$ be a subset of agents and P' the profile when limited to the agents in V'. We further denote $P - P'$ to be the profile of agents in $V - V'$. We now define the group no-show paradox as follows.

Definition 1 (**Group no-show paradox).** *A* group no-show paradox *(*GNSP*) occurs in profile P with agents V under a resolute voting rule r if there exists a subset of agents $V' \subseteq V$ with corresponding profile P', all of whom prefer $r(P - P')$ to $r(P)$, incentivizing them to abstain from voting.*

When the context is clear, we omit V and V' and say that a GNSP occurs in profile P with abstaining profile P'.

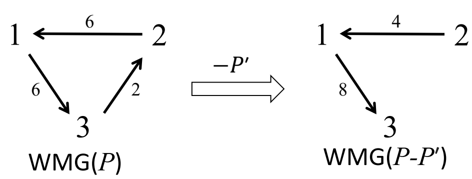

Figure 1: GNSP under $\text{Copeland}_{0.5}$.

Example 1. Let P denote the profile of 14 votes with 6 votes of $[2 \succ 1 \succ 3]$, 4 votes of $[1 \succ 3 \succ 2]$, and 4 votes of $[3 \succ 2 \succ 1]$. As illustrated in Figure 1 (for simplicity, we only show strictly positive edges in a WMG), $\text{Copeland}_{0.5}(P) = 1$ using the following order for tie-breaking $1 \rhd 2 \rhd 3$. For simplicity, only the edges with positive weights are shown in the weighted majority graph. If group P' consisting of 2 votes of $[3 \succ 2 \succ 1]$ abstain from voting, then $\text{Copeland}_{0.5}(P - P') = 2$. Notice that $2 \succ 1$ for both agents in P'. This means that no-show paradox occurs in $\text{Copeland}_{0.5}$ at P. □

3 Complexity of Verifying GNSP

In this section, we investigate the computational complexity of computing the existence of group no-show paradox for Copeland, Maximin, Condocetified positional scoring rules, and STV. No-show paradoxes trivially do not occur for positional scoring rules. The problem is formally defined below:

Definition 2 (GNSP-***r***). *Given a voting rule r,* GNSP-*r is the computational problem that takes a profile P as an input and outputs whether there exists a profile $P' \subseteq P$, each of which prefers $r(P - P')$ to $r(P)$.*

The definition of GNSP-r immediately implies the following easiness result for fixed m.

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.*

The proposition holds because the number of all possible anonymous profiles after abstention is $O\big((\frac{n}{m!})\big)^{m!}$ (proof in Appendix B). So, for constant m, all profiles can be enumerated in polynomial time. Differing from the setting in Proposition 1, in the remainder of this section, we assume a variable m. We investigate the complexity of GNSP-r for voting rules using the following common tie-breaking mechanisms (see Appendix A for formal definitions and examples). The *lexicographic tie-breaking* (LEX) breaks ties alphabetically. Fixed-agent (FA) tie-breaking chooses a fixed agent's preference (w.l.o.g., agent 1) to break ties. Most popular singleton ranking tie-breaking (MPSR) [Xia, 2020] first tries to use the linear order that uniquely occurs most often in the profile, and if such linear order does not exist, a backup tie-breaking mechanism is used. For example, MPSR+LEX uses LEX as the backup. Since the work focuses on resolute voting rules, we must consider tie-breaking mechanisms, and our theoretical results hold under these common mechanisms. Nevertheless, this does not mean that the existence of the group no-show paradox relies on ties. Theorem 3 holds under any tie-breaking mechanism, and its proof contains a group no-show paradox without ties.

We are now ready to present the theoretical results of this paper. Due to the space constraint, we only present the proof or a sketch under LEX, with the full proofs in Appendix B.

Theorem 1 (**Copeland).** *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 sketch. It is easy to check that the problem is in NP—given a profile $P' \in P$, 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, which is a restriction of EXACT 3 COVER that requires every element to be in exactly three sets and is proved to be NP-complete [Gonzalez, 1985].

Definition 3 (RXC3)**.** *RXC3 is a computational problem taking two sets as input: (1) a finite set of elements $X = \{x_1, x_2, \cdots, x_q\}$ (q is divisible by 3) and (2) a set of 3-element subsets of X, $\mathcal{S} = \{S_1, S_2, \cdots, S_q\}$ such that every element x_i appears in exactly three subsets in $\mathcal{S}$. The problem outputs whether $\mathcal{S}$ has an* exact cover *for X, i.e. a subset $\mathcal{S}^* \subseteq \mathcal{S}$ such that every x_i occurs in exactly one subset in $\mathcal{S}^*$.*

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 show the hardness of GNSP-Cd_α here for $\alpha < 1$. The full proof, including the hardness for $\alpha = 1$ is in Appendix B.1. 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.

The construction of the GNSP-**Cd_α instance.** There are $q+2$ alternatives $\{1, 2, 3, \ldots, q+2\}$, 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 2.

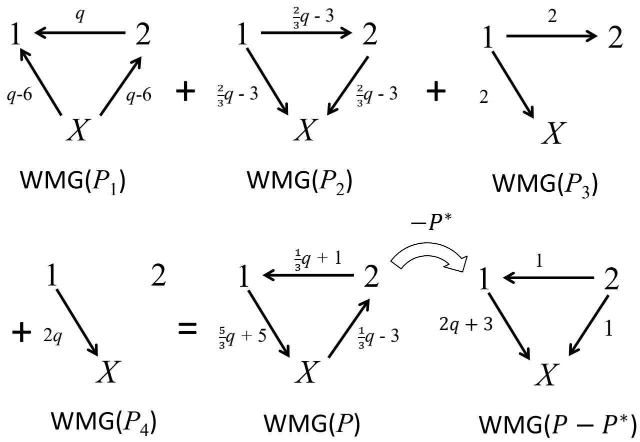

Figure 2: WMG of P for for Cd_α with $\alpha < 1$.

• P_1 consists of q votes that correspond to the sets in $\mathcal{S}$: for every $j \le q$, 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 $[1 \succ 2 \succ X]$; $P_3 = \{[1 \succ X \succ 2], [1 \succ 2 \succ X]\}$; and $P_4 = q$ copies of $\{[1 \succ X \succ 2], [2 \succ 1 \succ X]\}$.

As illustrated in Figure 2, 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 profile (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, alternative 2 becomes the Condorcet winner as illustrated in Figure 2. Note that all votes in P_1 prefer 2 to 1. Therefore, this constitutes a GNSP.

Suppose the GNSP-**Cd_α instance is a yes instance.** A group of agents, whose corresponding profile is 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 votes in P_1 prefer a to 1. Therefore, after P^* is removed, 1 beats all alternatives in X in head-to-head competition and gets a Copeland score of at least q. But, 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, $|P^*| \le \frac{q}{3}$. Suppose this is not true. Then at least $\frac{q}{3} + 1$ votes are removed, all of which prefer 2 to 1. So, 2 cannot beat 1 in the head-to-head competition in $P - P^*$. Also, $P^* \subseteq P_1 \cup P_4$ 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 $q + \alpha$. 2 gets a Copeland score of at most $q + \alpha$. Therefore, 2 cannot be the winner, which is a contradiction.

Third, $P^* \subseteq P_1$, and the 3-element subsets corresponding to votes in P^* are non-overlapping. Alternative 2 cannot lose to any alternative $a \in X$ in head-to-head competition, otherwise, 2's Copeland score is not strictly larger than 1's (which is at least q). 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 head-to-head competition as $\text{WMG}_{P-P^*}(2 \to a) \le (|P^*| - 2) - 2 - (\frac{q}{3} - 3) \le -1$, and cannot be the winner. So, 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 $|P^*| \le \frac{q}{3} - 1$. We show that 2's Copeland score is lower than 1' score (at least q), which is a contradiction.

• Case 1: $|P^*| = \frac{q}{3} - 1$. Therefore, 2 is tied with a such that $a \in S_j$ and $R_{S_j} \in P^*$ for some $j \le q$. 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 $4 + \alpha(q - 3)$. With $\alpha < \frac{q-4}{q-3}$, 2's score is lower than 1's score.

• Case 2: $|P^*| = \frac{q}{3} - 2$. In this case, 2 is defeated by all $a \in S_j$ for $R_{S_j} \in P^*$, 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: $|P^*| \le \frac{q}{3} - 3$. 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. □

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

Proof sketch. The proof is similar to the proof of Theorem 1. The problem is trivially in NP. Then, the NP-hardness is proved by the following reduction to RXC3 (the full proof is in Appendix B.2): Given a RXC3 instance $(X, \mathcal{S})$, we construct the following GNSP-MM with $q + 4$ alternatives $\{1, 2, 3, 4, 5, \ldots, q + 4\}$, where for every $4 \le i \le q + 4$, alternative i corresponds to x_{i-4}. The profile consists of four parts, whose WMG is illustrated in Figure 3.

Specifically, P_1 consists of q votes that correspond to the sets in $\mathcal{S}$: for every $j \le q$, there is a vote R'_S defined as $R'_S = S_j \succ 1 \succ 2 \succ (X \setminus S_j) \succ 4 \succ 3$. It is not hard to verify that $\text{MM}(P) = 2$, whose min-score is 0 (via alternatives 1 and 3).

When the RXC3 instance is a yes instance with solution $\mathcal{S}^*$, let $P^* \subset P_1$ denote the votes that correspond to $\mathcal{S}^*$. Then the WMG for $(P - P^*)$ is illustrated in Figure 3. It's not hard to find that 1, 2, and 3 share the max min-score of $-\frac{q}{3}$. Due to the lexicographic tie-breaking, $\text{MM}(P - P^*) = 1$, which implies a GNSP yes instance. When GNSP-MM is a yes instance with abstention group P^*, following similar reasoning as in the proof of Theorem 1, we can find that this happens only when $\text{MM}(P - P^*) = 1$ and P^* corresponds to a 3-cover of the RXC3 instance. □

The following theorem, stated without proof (the proof is similar to that of Theorem 1 and can be found in Appendix

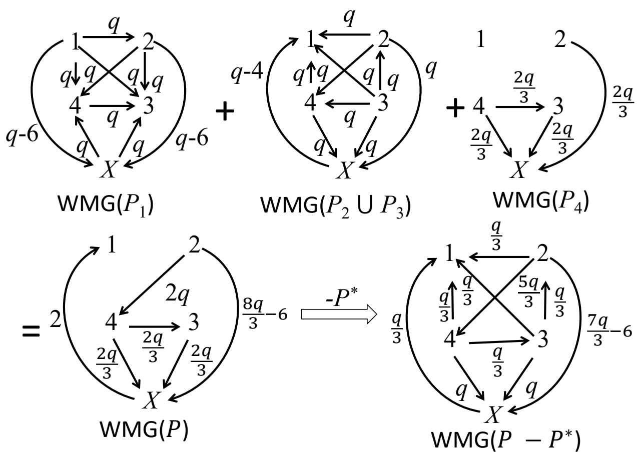

Figure 3: WMG of P for Maximin.

B.3), shows that verifying group no-show paradox in Condorcetified positional scoring rules is NP-complete under any tie-breaking mechanism.

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

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, MPSR+LEX *or* MPSR+FA.

Proof sketch. The hardness proof uses a reduction from RXC3 and is similar to 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. The full proof is in Appendix B.4.

Alternatives: there are in total $3q+3$ 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 $b_i \succ \bar{b}_i$ and $d_1 \succ d_2 \succ \dots \succ d_q$ in tie-breaking.

Profile: The profile P (shown in Table 1) consists of the following votes, where the top preferences are specified and the remaining alternatives ("*others*") are ranked arbitrarily. Both i and j in the table are in $\{1, 2, \dots, q\}$.

	Number of votes	Votes
P_1	$12q$	$c \succ w \succ$ *others*
P_2	$12q - 1$	$w \succ c \succ$ *others*
P_3	$10q + \frac{2q}{3}$	$d_0 \succ w \succ c \succ$ *others*
P_4	$12q - 2$ for each j	$d_j \succ w \succ c \succ$ *others*
P_5	$6q + 4i - 2$ for each i	$b_i \succ \bar{b}_i \succ w \succ c \succ$ *others*
	2 for each i	$b_i \succ d_0 \succ w \succ c \succ$ *others*
P_6	$6q + 4i - 6$ for each i	$\bar{b}_i \succ b_i \succ w \succ c \succ$ *others*
	2 for each $j \in S_i$	$\bar{b}_i \succ d_j \succ w \succ c \succ$ *others*
P_7	1 for each i	$\bar{b}_i \succ b_i \succ c \succ w \succ$ *others*

Table 1: Profile P of STV.

Note that STV$(P) = w$. 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 $q+1$, 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 (with solution $\mathcal{S}^*$). Let $I = \{i \leq q : S_i \in \mathcal{S}^*\}$, then agents with corresponding votes in P_7 whose top choices are b_i such that $i \in I$ have an incentive to (jointly) abstain from voting. After they abstain from the 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. At the beginning of round $q+1$, the plurality scores of the remaining alternatives are as in the following table. Therefore, as shown in Table 2, w is eliminated in round $q+1$, whose votes transfer to c. Finally, c will be the winner.

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

Table 2: STV score at $q+1$ round in $P - P^*$.

Suppose the GNSP-STV instance is a yes instance. We prove that the RXC3 instance is also a yes instance in two steps. First, the new winner must be c. c and w are adjacent in all votes. Thus, when w is eliminated, all of its votes are transferred to c, and c is ranked at the top in at least $24q - 1$ votes (in $P_1 \cup P_2$). This makes c the new winner. Second, the absent votes in P_7 constitute a solution to the RXC3 instance. Note that only agents whose votes are in P_1 and P_7 have incentives to abstain. Let I denote the indices i's of agents who abstain their votes from P_7 whose top-ranked preferences are $\bar{b}_i$. If $\mathcal{S}^* = \{S_i : i \in I\}$ is not a RXC3 solution, d_j will be eliminated in round $q+1$ and transfer all its votes to w for some j not in any S_i. Then w will be the winner, which is a contradiction. So $\mathcal{S}^*$ must be a solution. □

4 Algorithms for Verifying GNSP

We first present a general search-based algorithm that works for any rule as a baseline. Then, we present integer linear programming (ILP)-based algorithms for different voting rules.

Search-based algorithm. A brute-force breadth-first-search can enumerate all possible group abstentions in a breadth-first manner. In the worst case (when GNSP does not occur in a profile), the algorithm checks all possible combinations of group abstentions. From Proposition 1, the number of all possible group abstention combinations is bounded by $(\frac{n}{m!})^{m!}$, which determines the search run-time. Although the run-time is polynomial-time for a fixed m, the degree of the polynomial may be very high. Thus, this algorithms becomes too expensive for large n. We discuss possible improvements on this algorithm in Appendix C.

ILP formulations. To circumvent the computational inefficiency faced in the search-based approach for large n, we consider ILP-based algorithms for determining the GNSP-r problem. ILP is an optimization paradigm that finds integer solutions for optimizing a specific objective function under linear or discrete constraints. Some of our ILP formulations for different voting rules have conditional constraints, or constraints containing products of integer and binary variables. We refer to Bradley *et al.*; Conforti *et al.* [1977; 2014] for exposition to such techniques. We formulate ILPs for the group no-show paradox problem for four voting rules – Copeland, Maximin, Black's rule, and STV. In our formulations, we have both integer and binary variables and our objective is the size of the smallest group of agents that can abstain from voting to change the outcome for any profile P. Below, we describe the ILPs for Copeland and present the ILP formulations for other voting rules in Appendix C.

Copeland. First, we define the variables for the ILP formulations. Assume, that for profile P and voting rule r, $r(P) = a$. To verify if GNSP occurs for any alternative $b \neq a$, we need to check all agents j with preference $b \succ_{V_j} a$. For any linear order $R_i \in \mathcal{L}(\mathcal{A})$, assume that $|\{j \mid V_j = R_i\}| = n_i$. That is, n_i is the number of agents with preference R_i. Let $R_{b\succ a} = \{R \mid b \succ_R a\}$, the set of rankings where b is preferred over a. Now, for any linear order $R_i \in R_{b\succ a}$, agents may strategically abstain from voting. We denote x_i as the actual number of agents who vote R_i, with $x_i \leq n_i$. Thus, $\mathbf{x} = \{x_i\}_{R_i \in R_{b\succ a}}$ are the variables for the ILPs. We call P_x the alternate profile that is created by x_i agents voting for each linear order, $R_i \in R_{b\succ a}$. For a profile P, the group no-show paradox occurs for voting rule r, when there is a solution, $\mathbf{x}$ such that $r(P_x) = b$ for some $b \neq a$. Minimizing the objective function $\sum_{R_i \in R_{b\succ a}} (n_i - x_i)$ gives the smallest group size with an incentive to abstain.

Any alternative $b \neq a$, b will be the Copeland winner after abstention if b's Copeland score becomes the highest. For any pair of alternatives c, d, $P_x[c \succ d]$ in the updated profile P_x can be defined as follows-

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

Here $\mathbf{1}_{c\succ d}(R_i)$ is an indicator function that states whether $c \succ d$ in linear order R_i. There can be three scenarios: (1) $P_x[c \succ d] > P_x[d \succ c]$, (2) $P_x[c \succ d] < P_x[d \succ c]$, (3) $P_x[c \succ d] = P_x[d \succ c]$. Let, q_{cd}, r_{cd} be binary variables that will be true for scenario (1) and (3) respectively. It follows immediately that $q_{cd} + q_{dc} + r_{cd} = 1$.

The Copeland score for an alternative c is:

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

This equation imposes a constraint for the Copeland score for each alternative. The following constraints on the auxiliary variables are also needed for all c and d: $q_{cd} = 1 \implies P_x[c \succ d] > P_x[d \succ c]$; $r_{cd} = 1 \implies P_x[c \succ d] = P_x[d \succ c]$; and $q_{cd} + r_{cd} + q_{dc} = 1$.

Finally, if b is the new Copeland winner, we get the following constraint on the scores

$$\mathrm{CS}_{P_x}(b) \geq \mathrm{CS}_{P_x}(c) \;\; \forall c \neq b$$

Note that these winner constraints could be sightly different due to the tie-breaking mechanism used. We enumerate all possible winners, and check the corresponding ILP ($m - 1$ alternatives in total) for a solution. We tried an alternative formulation by using additional auxiliary variables for each alternative being the winner and use more conditional constraints, which performed worse in practice.

Tie-breaking Both the search-based and ILP algorithms are presented for anonymous voting rules, and in particular the experiments were done with lexicographic tie-breaking. However, they are readily adaptable to non-anonymous voting rules, for example, ones that use fixed agent tie-breaking. For the search algorithm, it is trivial to keep track of the fixed agent's preference. For the ILP formulations, different tie-breaking methods will result in slightly altered constraints.

Run-time consideration. ILP solving is an NP-complete problem, and for the ILP formulation we would have a worst-case run-time that is an exponential of $m!$ (See Appendix C for a brief discussion). But, powerful software packages like Gurobi can solve many such problems using heuristic methods efficiently. The number of primary variables (x_i) for Copeland (and the other voting rules as well) is upper bounded by the number of unique rankings in a profile, which in turn is bounded by $\min(n, m!)$. So, we expect the run-time to be dependent on $\min(n, m!)$. Additionally, we expect run-time to increase with m, as there are $O(m^2)$ auxiliary variables. Empirical results on the run-times of the algorithms are presented in Section 5 and Appendix D.

5 Experiments

In this section, we present results on the performance of the two algorithms and the likelihood of GNSP in both synthetic and real-world election data. All experiments were implemented in Python 3 and were run on a Windows laptop with 3.2 GHz AMD Ryzen 7 5800 CPU and 16 GB memory and the Gurobi solver was used for solving ILPs.

Synthetic datasets. We run a number of experiments on synthetic ranking data to test the run-time of our various algorithms and calculate the likelihood of GNSP for different voting rules under different conditions. To create the ranking data, for each sample profile, all agent rankings are i.i.d. samples from the same distribution. We run experiments on a number of popular models for ranking, such as Impartial Culture (IC), single-peaked preferences [Conitzer, 2009], Mallows [Mallows, 1957] and more. For each model, for different values of n between 10 and 1000 and $m \in [3, 10]$, we sample 1,000 preference profiles and calculate the occurrences of GNSP and the run-times. For space constraint, we mention only the IC and Mallows experiments here while describing the others in Appendix D. Mallows models are parameterized by ϕ, which mean the dispersion value. Smaller values of ϕ mean more correlation between the votes whereas higher values mean more random votes. $\phi = 1$ coincides with the aforementioned IC model. The Mallows model allows us to check likelihood for different levels of consensus among the voting agents. We also only present results for Copeland$_{0.5}$ but briefly mention important points from additional experiments, all of which are detailed in Appendix D.

Likelihood of GNSP. We compute the empirical likelihood of GNSP by counting the number of sample profiles where GNSP occurs. Figure 4 shows the likelihood of GNSP under Copeland$_{0.5}$. It can be seen that more consensus among the agents (low ϕ value) leads to a lower likelihood of GNSP. This behavior is also seen for other voting rules. In particular

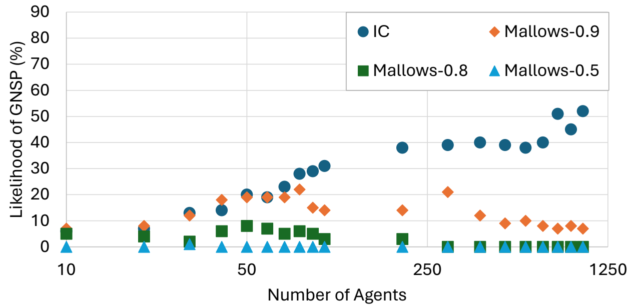

Figure 4: Likelihood of GNSP under $\text{Copeland}_{0.5}$ for $m = 4$.

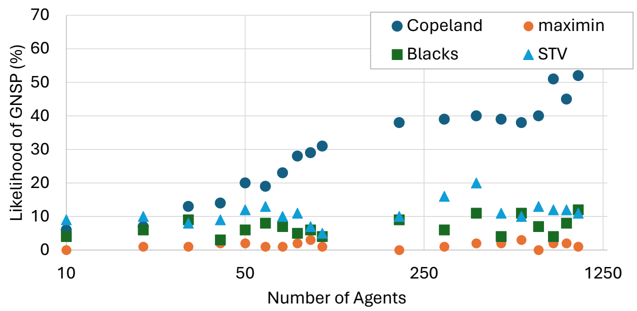

Figure 5: Likelihood of GNSP for Different Voting Rules under IC distribution for $m = 4$.

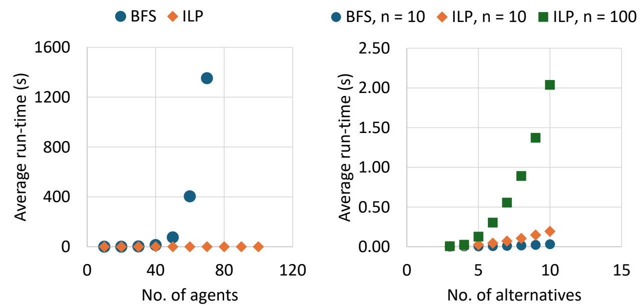

Figure 6: Run-time for the BFS and ILP algorithms. *Left:* Run-time vs No. of agents for $m = 4$. *Right:* Run-time vs No. of agents for $n = 10$ and $n = 100$, The BFS does not finish running for $n = 100$ and is not reported.

for the IC model, likelihood keeps increasing with n. However, this is not an effect that is seen for all voting rules or all models. As a comparison, Plassmann and Tideman [2014] computed the likelihood of no-show paradox for 3 alternatives under a spatial model and noticed that the likelihood decreases as the number of agents increases. So, we suspect the distribution of agent preferences plays an important part. In general, we have seen that ranking models where there is more randomness among the agents leads to more likelihood of GNSP. Another consistent trend is that GNSP likelihood increases with number of alternatives. One interesting result was that for single-peaked preferences, only STV shows very high likelihood of GNSP. Figure 5 shows the likelihood of GNSP under different voting rules for the IC distribution. We see that Copeland has the highest likelihood of GNSP, followed by STV, then Black's rule, and Maximin has the lowest likelihood. This observation of Maximin being more robust to GNSP is seen for all the ranking models that we have tested. More likelihood results for other voting rules and other models can be found in Appendix D.

Run-time of algorithms. Figure 6 (left) shows that the run-time for the ILP-based algorithm does not increase significantly with n for small m, while the BFS run-time increases exponentially in n, and becomes prohibitively high even for $n = 50$. Figure 6 (right) shows that the run-time of BFS does not increase significantly with the number of alternatives, m, when n is small. While the ILP algorithm's run-time increases with m for $n = 10$, it is not that significant. However, for $n = 100$, the ILP algorithm still manages to finish running reasonably fast while BFS fails to do so because of exponential blow-up of run-time. Additional experimental results (for other rules) are shown in Appendix D. These results illustrate that the BFS algorithm useful mostly when the number of agents is very small. On the other hand, the ILP algorithms are useful as long as the number of unique rankings is small and the number of alternatives is moderate, which includes a larger number of scenarios.

Real world data from PrefLib. We use all available election data from PrefLib [Mattei and Walsh, 2013] which have strict order-complete rankings for all agents. In total, there are 315 such profiles on PrefLib. We can verify the existence of GNSP in all of this profiles using either the BFS or ILP algorithm. Out of all 315 observed profiles, only one profile violates group participation for Copeland, Black's rule and STV, and we found no violations for Maximin. For Copeland, the occurrence was for an election with 30 agents and 11 candidates, where one voter abstaining causes a tie, with the tie-breaking result causing the no-show paradox. For both Black's rule and STV rule, the occurrence is for voting profiles with large number of alternatives (32 and 41 respectively) with very few agents (4 for both). So all three scenarios are kind of rare in the high number of candidates and unique preference rankings among the agents. We have included the data files for the profiles leading to paradoxes in the supplementary materials.

6 Discussion and Future Work

We prove that the group no-show paradox is computationally hard to verify for Copeland, Maximin, STV, and Condorcetified integer positional scoring rules, and provided search and ILP-based algorithms for computing them. Whether the results for Copeland, Maximin, and STV hold regardless of the tie-breaking rules remains an interesting open question. Studying the complexity and algorithms for other undesirable events, such as (group) manipulation is a natural direction for future work. In general, research in this direction helps establish a data-centric foundation for evaluating and designing voting rules.

Acknowledgments

We thank the anonymous reviewers for this and previous versions of this work for all their helpful comments. LX acknowledges NSF #1453542 and #2007476, an IBM-RPI AIRC fund, and a Google gift fund for support. The codes for our implementation of the verification algorithms is provided in the following repository: https://github.com/farhadmohsin/AxiomVerification.

Ethics Statement

The group no-show paradox can incentivize agents to abstain from voting. To make better group decisions, participation of agents is desirable, and thus the problem that we tackle in this paper has the broad goal getting better collective decisions. All experiments ran on real data uses anonymized preference data so there is no privacy concern for the work.

References

[Bartholdi and Orlin, 1991] John Bartholdi, III and James Orlin. Single transferable vote resists strategic voting. *Social Choice and Welfare*, 8(4):341–354, 1991.

[Bartholdi *et al.*, 1992] John J. Bartholdi, Craig A. Tovey, and Michael A. Trick. How hard is it to control an election? *Mathematical and Computer Modelling*, 16(8):27–40, 1992.

[Black, 1958] Duncan Black. *The Theory of Committees and Elections*. Cambridge University Press, 1958.

[Bradley *et al.*, 1977] Stephen P Bradley, Arnoldo C Hax, and Thomas L Magnanti. Applied mathematical programming. *(No Title)*, 1977.

[Brandl *et al.*, 2015] Florian Brandl, Felix Brandt, and Johannes Hofbauer. Incentives for participation and abstention in probabilistic social choice. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems*, page 1411–1419, 2015.

[Brandl *et al.*, 2017] Florian Brandl, Felix Brandt, and Johannes Hofbauer. Random assignment with optional participation. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pages 326–334, 2017.

[Brandl *et al.*, 2019] Florian Brandl, Felix Brandt, Christian Geist, and Johannes Hofbauer. Strategic abstention based on preference extensions: Positive results and computer-generated impossibilities. *Journal of Artificial Intelligence Research*, 66:1031–1056, 2019.

[Brandt *et al.*, 2017] Felix Brandt, Christian Geist, and Dominik Peters. Optimal bounds for the no-show paradox via sat solving. *Mathematical Social Sciences*, 90:18–27, 2017.

[Brandt *et al.*, 2021] Felix Brandt, Johannes Hofbauer, and Martin Strobel. Exploring the no-show paradox for condorcet extensions. In *Evaluating Voting Systems with Probability Models*, pages 251–273. Springer, 2021.

[Brandt *et al.*, 2022] Felix Brandt, Marie Matthäus, and Christian Saile. Minimal voting paradoxes. *Journal of Theoretical Politics*, 34(4):527–551, 2022.

[Conforti *et al.*, 2014] Michele Conforti, Gérard Cornuéjols, Giacomo Zambelli, Michele Conforti, Gérard Cornuéjols, and Giacomo Zambelli. *Integer programming*. Springer, 2014.

[Conitzer and Walsh, 2016] Vincent Conitzer and Toby Walsh. Barriers to Manipulation in Voting. In Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel Procaccia, editors, *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

[Conitzer, 2009] Vincent Conitzer. Eliciting single-peaked preferences using comparison queries. *Journal of Artificial Intelligence Research*, 35:161–191, 2009.

[Duddy, 2014] Conal Duddy. Condorcet's principle and the strong no-show paradoxes. *Theory and decision*, 77(2):275–285, 2014.

[Faliszewski and Rothe, 2016] Piotr Faliszewski and Jörg Rothe. Control and bribery in voting. In *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

[Felsenthal and Nurmi, 2016] Dan S Felsenthal and Hannu Nurmi. Two types of participation failure under nine voting methods in variable electorates. *Public Choice*, 168(1):115–135, 2016.

[Fishburn and Brams, 1983] Peter C. Fishburn and Steven J. Brams. Paradoxes of Preferential Voting. *Mathematics Magazine*, 56(4):207–214, 1983.

[Fishburn and Brams, 1984] Peter C Fishburn and Steven J Brams. Manipulability of voting by sincere truncation of preferences. *Public Choice*, 44(3):397–410, 1984.

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

[Gonzalez, 1985] Teofilo F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985.

[Jimeno *et al.*, 2009] José L Jimeno, Joaquín Pérez, and Estefanía García. An extension of the moulin no show paradox for voting correspondences. *Social Choice and Welfare*, 33(3):343–359, 2009.

[Kamwa *et al.*, 2021] Eric Kamwa, Vincent Merlin, and Faty Mbaye Top. Scoring Run-off Rules, Single-peaked Preferences and Paradoxes of Variable Electorate. working paper, February 2021.

[Lepelley and Merlin, 2001] Dominique Lepelley and Vincent Merlin. Scoring run-off paradoxes for variable electorates. *Economic Theory*, 17:53–80, 2001.

[Mallows, 1957] Colin L. Mallows. Non-null ranking model. *Biometrika*, 44(1/2):114–130, 1957.

[Mattei and Walsh, 2013] Nicholas Mattei and Toby Walsh. PrefLib: A Library of Preference Data. In *Proceedings of*

Third International Conference on Algorithmic Decision Theory, 2013.

[Moulin, 1988] Hervé Moulin. Condorcet's principle implies the no show paradox. *Journal of Economic Theory*, 45(1):53–64, 1988.

[Pérez *et al.*, 2012] Joaquín Pérez, José L Jimeno, and Estefanía García. No show paradox in condorcet k-voting procedures. *Group Decision and Negotiation*, 21(3):291–303, 2012.

[Pérez, 2001] Joaquín Pérez. The Strong No Show Paradoxes are a common flaw in Condorcet voting correspondences. *Social Choice and Welfare*, 18(3):601–616, 2001.

[Plassmann and Tideman, 2014] Florenz Plassmann and T. Nicolaus Tideman. How frequently do different voting rules encounter voting paradoxes in three-candidate elections? *Social Choice and Welfare*, 42(1):31–75, 2014.

[Ray, 1986] Depankar Ray. On the practical possibility of a 'no show paradox'under the single transferable vote. *Mathematical Social Sciences*, 11(2):183–189, 1986.

[Rothe and Schend, 2013] Jörg Rothe and Lena Schend. Challenges to Complexity Shields That Are Supposed to Protect Elections Against Manipulation and Control: A Survey. *Annals of Mathematics and Artificial Intelligence*, 68(1–3):161–193, 2013.

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

[Xia, 2020] Lirong Xia. The smoothed possibility of social choice. In *Advances in Neural Information Processing Systems*, pages 11044–11055, 2020.

[Xia, 2021] Lirong Xia. The semi-random satisfaction of voting axioms. In *Advances in Neural Information Processing Systems*, pages 6075–6086, 2021.