# Obvious Manipulability of Voting Rules

Haris Aziz[1] and Alexander Lam[1]

UNSW Sydney
{haris.aziz,alexander.lam1}@unsw.edu.au

**Abstract.** The Gibbard-Satterthwaite theorem states that no unanimous and non-dictatorial voting rule is strategyproof. We revisit voting rules and consider a weaker notion of strategyproofness called not obvious manipulability that was proposed by Troyan and Morrill (2020). We identify several classes of voting rules that satisfy this notion. We also show that several voting rules including $k$-approval fail to satisfy this property. We characterize conditions under which voting rules are obviously manipulable. One of our insights is that certain rules are obviously manipulable when the number of alternatives is relatively large compared to the number of voters. In contrast to the Gibbard-Satterthwaite theorem, many of the rules we examined are not obviously manipulable. This reflects the relatively easier satisfiability of the notion and the zero information assumption of not obvious manipulability, as opposed to the perfect information assumption of strategyproofness. We also present algorithmic results for computing obvious manipulations and report on experiments.

**Keywords:** Social choice · voting · manipulation · strategyproofness.

## 1 Introduction

Throughout history, voting has been used as a means of making public decisions based on the citizens' preferences. The ancient Greeks would give a show of hands to disclose their most preferred public official, and the winner of the election was chosen as the official with the most first preferences [4]; such a voting system is called the *plurality vote*. Many other voting systems have been developed over time, such as the Borda Count, developed by Jean-Charles de Borda in 1770. The Borda Count gives each candidate a score based on their position in the voters' preference orders. This system was opposed by Marquis de Condorcet, who instead preferred the Condorcet method, which elects the candidate that wins the majority of pairwise head-to-head elections against the other candidates [2]. However, voting systems are not just used in politics; voting theory is frequently used and studied in artificial intelligence to aggregate the preferences of multiple agents into a single decision.

The studies of electoral systems in social choice theory have been wrought with negative results. Arrow's impossibility theorem [1] showed that there exists no voting system with three reasonable requirements. In a similar vein, the Gibbard-Sattherthwaite theorem [10, 18] states that when there are at least

three alternatives, every unanimous voting rule is either dictatorial, meaning only one voter's preferences are taken into account, or prone to manipulative voting, meaning a voter can give an untruthful ballot to gain a more preferred outcome.

Such strategic behaviour is a commonly studied problem in mechanism design and social choice, as many mechanisms sacrifice efficiency or fairness to ensure strategyproofness. The original notion of strategyproofness fails to explain the variation we observe in voters' tendency to strategically vote in different electoral systems. This has motivated research toward alternative concepts of strategyproofness that may be able to capture such variations. One such notion is *not obvious manipulability*, recently theorized by Troyan and Morrill [23]. Whilst strategyproofness assumes agents have complete information over other agent preferences and the mechanism operation, not obvious manipulability assumes agents are 'cognitively limited' and lack such information. As such, they are only aware of the possible range of outcomes that can result from each mechanism interaction. Put simply, a mechanism satisfies *not obvious manipulability (NOM)* if no agent can improve its best case or worst case outcome under any manipulation. A mechanism is *obviously manipulable (OM)* if either an agent's best case or worst case outcome can be improved by some untruthful interaction.

The assumptions made for *not obvious manipulability* are suitable when applied to voting rules, as ballots are commonly hidden from the voters, restricting their ability to compute a desirable manipulation. In this paper, we explore which voting rules are obviously manipulable, and if so, what the conditions are for obvious manipulability.

*Contributions* Our main contribution is to apply the concept of obvious manipulations to the case of voting rules for the first time. We study which voting rules are obviously manipulable, and what conditions are required for obvious manipulability. Whilst many classes of voting rules including Condorcet extensions and strict positional scoring rules with weakly diminishing differences are not obviously manipulable, we show that certain voting rules, including $k$-approval, are obviously manipulable. We also characterize the conditions under which positional scoring rules are obviously manipulable in the best case. For the class of $k$-approval voting rules, we characterize the conditions under which the rules are obviously manipulable. Many of our results apply to large classes of voting rules including positional scoring rules or Condorcet extensions. Table 1 summarizes several of our results.

One of our insights is that certain rules are obviously manipulable when the number of alternatives is relatively large compared to the number of voters. We also look at the problem of checking whether a particular instance of a voting problem admits an obvious manipulation. For the class of positional scoring rules, we provide a general polynomial-time reduction to the well-studied *unweighted coalitional manipulation problem*. As a corollary, we show that the problem of checking the existence of an obvious manipulation is polynomial-time solvable for the $k$-approval rule. Finally, we report on experimental results on the fraction of instances that admit obvious manipulations for the $k$-approval rule.

| **NOM** | **OM** |
|---|---|
| **Does not admit a voter with veto power** | |
| **k-Approval** ($n > \frac{m-2}{m-k}$)<br>Plurality | **k-Approval** ($n \leq \frac{m-2}{m-k}$) |
| **Almost-unanimous**<br>Condorcet-extension<br>STV<br>Plurality with runoff | |
| Positional scoring rule ($n > \frac{s_1}{s_1-s_2} + 1$) | Positional scoring rule that<br>admits a voter with veto power (existence) |
| Positional scoring rule with weakly<br>diminishing differences<br>Borda rule | |

Table 1: List of rules and conditions for voting rules to be NOM or OM.

## 2 Related Work

Our paper belongs to the rich stream of work in social choice on the manipulability of voting rules. The reader is referred to the book by Taylor [21] that surveys this rich field. A comparison of the susceptibility of voting rules to manipulation has a long history in social choice. For example, one particular approach is to count the relative number of preference profiles under which voting rules are manipulable (see, e.g., [8]). Another approach is analyzing the maximum amount of expected utility an agent can gain by reporting untruthfully [3].

Our work revolves around the concept of obvious manipulations, which was proposed by Troyan and Morrill [23]. This concept was inspired by a paper on 'obviously strategyproof mechanisms' by Li [11]. The latter paper describes the cognitively-limited agent that is only aware of the range of possible outcomes ranging from each report. In the paper, Li then proposes the characterization of 'obvious strategyproofness', a strengthening of strategyproofness. A mechanism is defined as obviously strategyproof if each agent's worst case outcome under a truthful report is strictly better than their best case outcome under any untruthful report. Troyan and Morrill [23] studied obvious manipulations in the context of matching problems. In particular, they showed that whereas the Boston mechanism is obviously manipulable, many stable matching mechanisms (including those that are not strategyproof) are not obviously manipulable.

Other, weaker notions of strategyproofness specific to voting rules have been proposed in the literature. Slinko and White [19, 20] considered *safe strategic voting* to represent the coalitional manipulation of scoring rules. Assuming every member of the coalition reports the same ballot, a manipulation is a *safe strategic vote* if it guarantees an outcome which is weakly preferred over truth-telling. Another notion has also been proposed by Conitzer et al. [6], who state that a ballot *dominates* another ballot if it guarantees a weakly more preferred outcome. The authors define a voting rule as being *immune to dominating manipulations* if there are no ballots that dominate any voter's true preferences, and classify

the immunity of certain rules under varying levels of information known by the manipulator. In particular relevance to our paper, they find that certain voting rules such as Condorcet-consistent rules and the Borda count are immune to dominating manipulations under zero information. We remark that immunity to dominating manipulations under zero information is a weaker notion than not obvious manipulability, and thus our work investigates a stronger notion defining a voting rule's resistance to manipulation than some existing notions. For further discussion on the strategic aspects of voting with partial information, the reader is referred to Chapter 6 and 8 of the book by Meir [12], where similar concepts such as local dominance are discussed.

In many elections, voters often lack information of other voters' preferences. This has prompted a probabilistic perspective into the manipulability of voting rules, often assuming a uniform distribution over each preference ordering. In 1985, Nitzan showed that in point scoring rules, a manipulation is more likely to succeed as the number of outcomes increases, and the number of voters decreases [14]. A similar probabilistic perspective was used by Wilson and Reyhani [24]. Computer scientists have also extensively researched the computational complexity of calculating a manipulative ballot; as the number of voters and outcomes becomes large, it can be computationally infeasible to compute a manipulation if the problem is intractable (see, e.g. [5, 7]).

## 3 Preliminaries

We consider the standard social choice voting setting $(N, O, \succ)$ that involves a finite set $N = \{1, 2, \ldots, n\}$ of $n$ voters and a finite set $O = \{o_1, o_2, \ldots, o_m\}$ of $m$ outcomes. We also assume that $n \geq 3$ and $m \geq 3$. Each voter $i$ has a transitive, complete and reflexive preference ordering $\succ_i$ over the set of outcomes $O$. We denote the preference profile of each voter $i \in N$ as $\succ = (\succ_1, \ldots, \succ_n)$, and use $\mathcal{L}(O)^n$ to denote the set of all such profiles for a given $n$. For a given voter $i \in N$, we use $\succ_{-i} = (\succ_1, \ldots, \succ_{i-1}, \succ_{i+1}, \ldots, \succ_n)$ to denote the preference profile of the voters in $N \backslash \{i\}$. A voting rule $f \colon \mathcal{L}(O)^n \to O$ is a function that takes as input the preference profile and returns an outcome from $O$.

An outcome $o \in O$ is called a *possible outcome* under a voting rule $f$ if there exists some preference profile $\succ$ such that $f(\succ) = o$. Since we are considering voting rules that return a single outcome, we will impose tie-breaking over social choice correspondences (voting rules that return more than one outcome) to return a single outcome. Unless specified otherwise, we will assume a fixed tie-break ordering over the outcomes.

**Definition 1.** *A voting rule $f$ is* manipulable *if there exists some voter $i \in N$, two preference relations $\succ_i, \succ_i'$ of voter $i$, and a preference profile $\succ_{-i}$ of other voters such that $f(\succ_i', \succ_{-i}) \succ_i f(\succ_i, \succ_{-i})$. Such a manipulation is defined as a* profitable manipulation *for voter $i$. A voting rule is* strategyproof (SP) *if it is not manipulable.*

Under voting rule $f$, a given set of outcomes and a fixed number of voters, we denote by $B_{\succ_i}(\succ_i', f)$ the best possible outcome (under $i$'s preference $\succ_i$) when

she reports $\succ_i'$, over all possible preferences of the other voters. We also denote by $W_{\succ_i}(\succ_i', f)$ the worst possible outcome (under $i$'s preference $\succ_i$) when she reports $\succ_i'$, over all possible preferences of the other voters. We now present the central concept used in the paper, which has been adapted from the paper by Troyan and Morrill [23] to the field of voting.

**Definition 2.** *A voting rule $f$ is not obviously manipulable (NOM) if for every voter $i$ with truthful preference $\succ_i$ and every profitable manipulation $\succ_i'$, the following two conditions hold:*

$$W_{\succ_i}(\succ_i, f) \succeq_i W_{\succ_i}(\succ_i', f) \tag{1}$$
$$B_{\succ_i}(\succ_i, f) \succeq_i B_{\succ_i}(\succ_i', f). \tag{2}$$

If either condition does not hold, then we say the voting rule is *obviously manipulable*. Specifically, if (1) does not hold, then we say the voting rule is *worst case obviously manipulable*. Similarly, if (2) does not hold, then we say it is *best case obviously manipulable*.

## 4 Sufficient Conditions for not being Obviously Manipulable

In this section, we identify certain conditions that imply not obvious manipulability when satisfied by voting rules.

**Definition 3.** *For a given voting rule $f$ and a fixed number of voters $n$ and outcomes $m$, a voter $i$ has veto power if there exists a possible outcome $o \in O$ and report $\succ_i$ such that $f(\succ_i, \succ_{-i}) \neq o$ for all $\succ_{-i}$.*

Our first result is a sufficient condition for a voting rule being NOM.

**Lemma 1.** *If a voting rule is obviously manipulable, then it must admit a non-dictatorial voter with veto power.*

However, existence of a voter with veto power does not imply obvious manipulability. We will illustrate this later in the paper.

**Definition 4.** *A voting rule $f$ is* almost-unanimous *if it returns an outcome $o$ when $o$ is the most preferred outcome for at least $n-1$ voters. Almost-unanimity implies unanimity.*

**Theorem 1.** *For $n \geq 3$, no almost-unanimous voting rule is obviously manipulable.*

*Proof.* Note that an almost-unanimous voting rule is not dictatorial. By definition, a rule that is almost-unanimous cannot admit a voter with veto power. Hence it follows from Lemma 1 that for $n \geq 3$, no almost-unanimous voting rule is obviously manipulable. □

**Corollary 1.** *Any majoritarian (Condorcet extension rule) is NOM.*

Similarly, Theorem 1 applies to several voting rules including STV [22] and Plurality with runoff [13] that are almost-unanimous.

**Corollary 2.** *STV and Plurality with runoff are NOM.*

We have shown that many voting rules are not obviously manipulable, so we question whether there are any obviously manipulable voting rules. We next investigate positional scoring rules.

## 5 Positional Scoring Rules

In this section, we consider positional scoring rules, a major class of voting rules which assigns points to candidates based on voter preferences and chooses the candidate with the highest score. A formal definition of a positional scoring rule is given below.

**Definition 5.** *A positional scoring rule assigns a score to each outcome using the score vector $w = (s_1, s_2, \ldots, s_m)$, where $s_i \geq s_{i+1} \forall i \in \{1, 2, \ldots, m-1\}$ and $\exists i \in \{1, 2, \ldots, m-1\} : s_i > s_{i+1}$. Each voter gives $s_i$ points to their $i$th most preferred candidate, and the score of a candidate is the total number of points given by all voters. The candidate with the highest number of points is returned by the rule.*

Note that this positional scoring rule definition rules out unreasonable, pathological scoring vectors such as $(1, 2, 3)$. Several well-known rules fall in the class of positional scoring rules. For example if $s_i = m - i$ for all $i \in [m]$, the rule is the Borda voting rule. If $s_1 = 1$ and $s_i = 0$ for all $i > 1$, the rule is plurality. If $s_m = 0$ and $s_i = 1$ for all $i < m$, the rule is anti-plurality.

Next, we identify a sufficient condition for a positional scoring rule to be NOM.

**Theorem 2.** *A positional scoring rule is NOM if $n > \frac{s_1}{(s_1 - s_2)} + 1$.*

*Proof.* It is sufficient to show that for $n > \frac{s_1}{(s_1 - s_2)} + 1$, the rule is almost-unanimous. Any outcome $a$ that is the most preferred by $n - 1$ voters has a score of at least $(s_1)(n-1)$. We show that this score is greater than the score of any other candidate. The maximum score any other outcome $b$ can get is by being in the first position of one voter and second position of all other voters so its score is $(s_2)(n-1) + s_1$. The score of $a$ is greater than the maximum score of $b$ if and only if

$$\begin{aligned} & (s_1)(n-1) > (s_2)(n-1) + s_1 \\ \iff & (n-1)(s_1 - s_2) > s_1 \\ \iff & n > \frac{s_1}{(s_1 - s_2)} + 1. \end{aligned}$$

□

This result suggests that many positional scoring rules are NOM when there are sufficiently many voters, and that scenarios with few voters may be required for a positional scoring rule to be obviously manipulable.

### 5.1 k-Approval

The $k$-approval rule is a subclass of positional scoring rules that lets voters approve of their $k$ most preferred candidates, or voice their disapproval for their $m-k$ least preferred candidates. It is a scoring rule with weight vector $w = (1,\dots,1,0,\dots,0)$, where there are $k$ ones, $m-k$ zeroes and $0 < k < m$.

Note that the $k$-approval rule is the same as the plurality rule when $k = 1$, and it is the same as the anti-plurality rule when $m - k = 1$.

**Lemma 2.** *The $k$-approval rule (kApp) is obviously manipulable if $n \leq \frac{m-2}{m-k}$.*

*Proof.* Suppose there are $n$ voters, the number of outcomes $m$ is at least $n(m-k)+2$, voter $i$'s true preferences are

$$\succ_i\colon o_1 \succ_i o_2 \succ_i \cdots \succ_i o_{m-1} \succ_i o_m,$$

and the fixed tie-break ordering is

$$\succ_L\colon o_k \succ_L o_1 \succ_L o_2 \succ_L \cdots \succ_L o_{k+1} \succ_L o_{k+2} \succ_L \cdots \succ_L o_{m-1} \succ_L o_m.$$

Under a $k$-approval rule, any voter may disapprove of their $m-k$ least preferred outcomes. Since there are a total of $n(m-k)$ disapprovals and $m \geq n(m-k)+2$, by the pigeonhole principle, there are at least 2 outcomes with zero disapprovals. Therefore the selected outcome must be the tie-break winner of the outcomes with zero disapproval votes, as they are approved by every voter.

Under a truthful ballot $\succ_i$, voter $i$ disapproves of outcomes $\{o_{k+1},\dots,o_m\}$, so $W_{\succ_i}(\succ_i, kApp) \notin \{o_{k+1},\dots,o_m\}$. We therefore have $W_{\succ_i}(\succ_i, kApp) = o_k$ as at least two outcomes in $\{o_1,\dots,o_k\}$ must have zero disapproval votes, and $o_k$ has the highest tie-break priority.

If voter $i$ instead disapproves of the outcomes in $\{o_k\} \cup \{o_{k+1},\dots,o_m\} \setminus \{o_{i'}\}$, where $k+1 \leq i' \leq m$, then the worst case outcome satisfies $W_{\succ_i}(\succ'_i, kApp) \succ_i o_{k-1}$, as $o_{i'}$ always loses the tie-break with any outcome from $\{o_1,\dots,o_{k-1}\}$. We therefore have $W_{\succ_i}(\succ'_i, kApp) \succ_i W_{\succ_i}(\succ_i, kApp)$, concluding the proof. □

**Lemma 3.** *The $k$-approval rule (kApp) is NOM if $n > \frac{m-2}{m-k}$.*

*Proof.* Suppose that there are $n$ voters, $m \leq \frac{kn-1}{n-1}$ outcomes and without loss of generality that voter $i$'s true preferences are

$$\succ_i\colon o_1 \succ_i o_2 \succ_i \cdots \succ_i o_m.$$

We note that $m \leq \frac{kn-1}{n-1} \iff n(m-k) \geq m-1$, so there are at least $m-1$ disapproval votes as each of the $n$ voters disapproves of $m-k$ outcomes. We first show that under these conditions, the $k$-approval rule is not best case obviously

manipulable. Under $\succ_i$, voter $i$'s best case outcome of $B_{\succ_i}(\succ_i, kApp) = o_1$ is achievable by the voters voting such that $o_1$ has zero disapprovals and each of the other outcomes has at least one disapproval. Since $i$'s best case outcome is his first preference, it cannot be strictly improved by any manipulation.

We next show that in this scenario, the $k$-approval rule is not worst case obviously manipulable. By the pigeonhole principle, there must be at least one outcome with zero disapprovals. Under a truthful ballot, voter $i$ disapproves of outcomes $\{o_{k+1}, \ldots, o_m\}$, so his worst case outcome is $W_{\succ_i}(\succ_i, kApp) = o_k$, achieved by the other voters disapproving of outcomes $\{o_1, \ldots, o_{k-1}\}$. Now under any manipulation, at least one outcome from $\{o_{k+1}, \ldots, o_m\}$ must be approved by voter $i$. This results in $W_{\succ_i}(\succ'_i, kApp) \in \{o_{k+1}, \ldots, o_m\}$, as the other voters can vote such that every outcome except for voter $i$'s least preferred approved outcome has been disapproved at least once. We therefore have $W_{\succ_i}(\succ_i, kApp) \succeq_i W_{\succ_i}(\succ'_i, kApp)$, concluding our proof. □

*Remark 1.* We note that the obvious manipulability of $k$-approval when $m \geq n(m-k)+2$ and the not obvious manipulability of $k$-approval when $m = n(m-k)+1$ also holds in the case of weighted voters, as the argument relies on the number of outcomes exceeding the total number of disapprovals.

Based on the two lemmas proved above, we achieve a characterization of the conditions under which the $k$-approval rule is obviously manipulable.

**Theorem 3.** *The $k$-approval rule is obviously manipulable if and only if $n \leq \frac{m-2}{m-k}$.*

**Corollary 3.** *The plurality rule is NOM.*

Since plurality is generally considered to be one of easiest rules to manipulate, the corollary above underscores the strength of obvious manipulations. We give the following intuition for the result on $k$-approval. Suppose a small committee is applying the $k$-approval rule to select a prize winner out of many candidates, and that certain candidates will be approved by every voter. The manipulator may also have a general idea of these candidates conditional on their report. If a fixed tie-break method is used (such as selecting the oldest candidate), the manipulator may disapprove of the oldest candidate who would otherwise win, instead approving a younger candidate who would not be selected regardless.

### 5.2 Strict Positional Scoring Rules

In the previous section, we noted that the $k$-approval rule is obviously manipulable. This may lead to the question of whether the lack of strictly decreasing scoring weights contributes to the obvious manipulability of a positional scoring rule. Hence, we focus on strict positional scoring rules in the following section.

**Definition 6.** *A positional scoring rule with weight vector $w = (s_1, s_2, \ldots, s_m)$ is* strict *if $s_i > s_{i+1}$ for all $i \in \{1, 2, \ldots, m-1\}$.*

We first note a strict positional scoring rule can be obviously manipulable.

**Lemma 4.** *There exists a strict positional scoring rule that can admit a voter with veto power and is obviously manipulable.*

In the following lemma, we also find that a strict positional scoring rule is not necessarily obviously manipulable if it admits a voter with veto power.

**Lemma 5.** *There exists a class of strict positional scoring rules that can admit a voter with veto power but are NOM.*

**Definition 7.** *A strict positional scoring rule with* $w = (s_1, s_2, \ldots, s_m)$ *has* diminishing differences *if* $s_i - s_{i+1} > s_{i+1} - s_{i+2}$ *for all* $i \in \{1, 2, \ldots, m-2\}$. *We say it has* weakly diminishing differences *if* $s_i - s_{i+1} \geq s_{i+1} - s_{i+2}$ *for all* $i \in \{1, 2, \ldots, m-2\}$.

An example of such a rule is the Harmonic-Borda/Dowdall system used in Nauru, which has weight vector $w = (1, 1/2, \ldots, 1/m)$ [17]. It is more favourable towards candidates that are the top preference of many voters, and has been described as a scoring rule that "lies between plurality and the Borda count" [9].

Next, we prove that a strict positional scoring rule with weakly diminishing differences is NOM.

**Theorem 4.** *A strict positional scoring rule with weakly diminishing differences is NOM.*

**Corollary 4.** *The Borda and Harmonic-Borda/Dowdall rules are NOM.*

*Remark 2.* Lemma 5 exemplifies a class of strict positional scoring rules which do not satisfy weakly diminishing differences but are NOM.

### 5.3 Obvious Manipulability in the Best Case

Although our previous results focus on worst case obvious manipulability, it is possible for a positional scoring rule to be best case obviously manipulable.

**Lemma 6.** *Assuming* $m, n \geq 3$, *a positional scoring rule* $f$ *is best case obviously manipulable if and only if for some* $k > 1$, *the first* $k$ *elements of the scoring vector are the same and* $n \leq \frac{m-2}{m-k}$.

Next, we demonstrate a fundamental connection between best case obvious manipulations and worst case obvious manipulations.

**Theorem 5.** *Assuming* $m, n \geq 3$, *for any positional scoring rule, if a voter's preference relation* $\succ_i$ *admits a best case obvious manipulation, then it also admits a worst case obvious manipulation.*

*Proof.* Suppose for some positional scoring rule $f$ that a voter's preference relation $\succ_i$ admits a best case obvious manipulation. From Lemma 6, for some $k > 1$, the first $k$ elements of the scoring vector must be the same, and we have $n \leq \frac{m-2}{m-k}$. Consequently, any outcome selected under $f$ must be in the top $k$ outcomes of each voter's report. We say that a voter 'approves' his $k$ most preferred outcomes, and 'disapproves' of his $m-k$ least preferred outcomes. An outcome cannot be chosen by $f$ if it has a disapproval vote from at least one voter.

We now construct the set of feasible outcomes $O_f$ which can be selected under the voter's preference relation $\succ_i$ and some $\succ_{-i}$. Let $O_v$ be the $m-k$ disapproved outcomes by $i$ under $\succ_i$. Since any outcome with at least one disapproval vote cannot be chosen, no outcome in $O_v$ can be selected. Now consider the set $O \backslash O_v$. Suppose without loss of generality that $O \backslash O_v = \{o_1, \ldots, o_k\}$, with tie-break ordering $\succ_L$: $o_1 \succ_L \cdots \succ_L o_k$. Denote $c := (n-1)(m-k)$ as the number of disapproval votes that the other $n-1$ voters can distribute. For $j \in \{1, \ldots, c+1\}$, outcome $o_j$ can be selected if the other voters cast disapproval votes for outcomes $\{o_1, \ldots, o_{c+1}\} \backslash \{o_j\}$. Furthermore, outcomes $o_{c+2}, \ldots, o_k$ cannot be selected, regardless of how the other voters report. Therefore the set of feasible outcomes $O_f$ are the $c+1$ highest tie-breaking ranked outcomes of the set $O \backslash O_v$. Voter $i$'s best case outcome is its most preferred outcome in $O_f$, whilst its worst case outcome is its least preferred outcome in $O_f$. We denote $o_b := B_{\succ_i}(\succ_i, f)$ as $i$'s best case outcome, and $o_w := W_{\succ_i}(\succ_i, f)$ as $i$'s worst case outcome.

We now define the set of feasible outcomes $O'_f$ under any preference report by voter $i$. This is the $n(m-k)+1$ highest tie-break ranked outcomes of $O$. Now suppose $\succ_i$ admits a best case obvious manipulation. There must exist an outcome $o'_b \in O'_f \backslash O_f$ that $i$ prefers over $o_b$. Consider the set $O'_v = \{o_w\} \cup O'_f \backslash (O_f \cup o'_b)$. Since $o_w \notin O'_f \backslash O_f$ and $o'_b \notin O_f$, we have

$$\begin{aligned} |O'_v| &= |\{o_w\}| + |O'_f| - |O_f| - |\{o'_b\}| \\ &= 1 + n(m-k) + 1 - (n-1)(m-k) - 1 - 1 \\ &= m - k. \end{aligned}$$

We now deduce $i$'s worst case outcome $W_{\succ_i}(\succ'_i, f)$ under the manipulation $\succ'_i$ where voter $i$ disapproves of all outcomes from $O'_v$. Under $\succ'_i$, every outcome in $O'_f \backslash O_f$ except for $o'_b$ has a disapproval vote and therefore cannot be selected. The outcome $o'_b$ satisfies $o'_b \succ_i o_b$ and therefore cannot be the worst case outcome. Finally, $o_w$ has a disapproval vote, so by elimination, we have $W_{\succ_i}(\succ'_i, f) \in O_f \backslash \{o_w\}$. Since $o_w$ is voter $i$'s least preferred outcome in $O_f$, we have $W_{\succ_i}(\succ'_i, f) \succ_i o_w$, meaning that $\succ'_i$ is a worst case obvious manipulation. □

## 6 Computing Obvious Manipulations

In the previous parts of the paper, we focussed on understanding the conditions under which a voting rule is obviously manipulable. Next, we consider the problem of computing an obvious manipulation for a given problem instance.

We present algorithmic results for computing obvious manipulations under positional scoring rules.

| Obvious Manipulation (OM) |
|---|
| Input: Number of voters $n$, set of outcomes $O = \{o_1, o_2, \ldots, o_m\}$, preference relation $\succ_i$ of voter $i$, tie-break order $\succ_L$ and voting rule $f$. |
| Problem: Find a preference relation $\succ_i'$ such that $W_{\succ_i}(\succ_i', f) \succ_i W_{\succ_i}(\succ_i, f)$ or $B_{\succ_i}(\succ_i', f) \succ_i B_{\succ_i}(\succ_i, f)$. |

If we only consider the best case manipulation, we refer to the problem as Best-case Obvious Manipulation (BOM). If we only consider the worst case manipulation, we refer to the problem as Worst-case Obvious Manipulation (WOM).

We present algorithms for the obvious manipulation problems. The algorithms are based on reductions to the Constructive Coalitional Unweighted Manipulation (CCUM) that is well-studied in computational social choice (see e.g., [25, 26]). We now introduce the Constructive Coalitional Unweighted Manipulation (CCUM).

| Constructive Coalitional Unweighted Manipulation (CCUM) |
|---|
| Input: Voting rule $f$, set of outcomes $O$, distinguished candidate $o \in O$, set of voters $S$ that have already cast their votes and set of voters $T$ that have not cast their votes. |
| Problem: Is there a way to cast the votes in $T$ such that $o$ wins the election under $f$? |

We show that for any voting rule, there is a polynomial-time algorithm for computing a best case obvious manipulation if CCUM can be solved in polynomial time.

**Lemma 7.** *For any voting rule, there is a polynomial-time algorithm for* BOM *if* CCUM *can be solved in polynomial time.*

*Proof.* Denote $o_b := B_{\succ_i}(\succ_i, f)$. We can compute $o_b$ as follows. We fix the preference $\succ_i$ of voter $i$ and solve CCUM for each possible outcome while keeping all the other voters as manipulators. This can be checked in $|O|$ calls to an algorithm to solve CCUM. Next, we find $i$'s best possible outcome if she is allowed to report any other preference. This can be checked by solving CCUM for each possible outcome while keeping all the voters as manipulators. Let $o^*$ be the possible outcome that is most preferred with respect to $\succ_i$. The instance is best case obviously manipulable if and only if $o^* \succ_i o_b$. □

We then show that for any positional scoring rule, there is a polynomial-time algorithm for OM if CCUM can be solved in polynomial time.

**Lemma 8.** *For any positional scoring rule, there is a polynomial-time algorithm for* WOM *if* CCUM *can be solved in polynomial time.*

*Proof.* First we compute the worst case outcome $W_{\succ_i}(\succ_i, f)$ of $i$ when she reports the truth. This is easily computed by running an algorithm that solves CCUM with $i$'s report being fixed, and checking which outcomes are possible.

We check whether $i$ can improve her worst case outcome by misreporting. We denote $o_w := W_{\succ_i}(\succ_i, f)$ and $O_{bad} := \{o \in O : o_w \succ_i o\} \cup \{o_w\}$. We also denote $O_{good} := A \setminus O_{bad}$. We want to check whether $i$ can ensure that no outcome from $O_{bad}$ is selected irrespective of how the other voters vote. We define a misreport $\succ_i'$ as follows. In $\succ_i'$, the outcomes of $O_{good}$ are preferred over the outcomes of $O_{bad}$. In $O_{good}$ the outcomes are ordered so that higher (tie-break) priority outcomes come earlier. In $O_{bad}$ the outcomes are ordered so that higher priority outcomes come later. We solve CCUM with respect to $\succ_i'$ and check whether some outcome in $O_{bad}$ can be selected. If such an alternative cannot be selected, we return yes. Otherwise we return no. □

Combining the two lemmas above, we get the following.

**Theorem 6.** *For any positional scoring rule, there is a polynomial-time reduction from solving* OM *to solving* CCUM*.*

Conitzer and Walsh [5] discuss the computational complexity of CCUM for various different voting rules. In particular, CCUM can be solved in polynomial time for the $k$-approval problem. For example, Zuckerman et al. [26] present a greedy polynomial-time algorithm for computing CCUM. For the sake of completeness, we explicitly write this algorithm for the $k$-approval rule with a fixed tie-break ordering. The algorithm assigns approved outcomes to the manipulators as follows. First, it assigns the distinguished outcome as each manipulator's first preference. Each manipulator then approves the $k-1$ outcomes with the lowest scores. If there are more than $k-1$ tied outcomes, the ones with the lowest tie-break priority are selected.

**Corollary 5.** OM *can be solved in polynomial time for $k$-approval.*

### Experimental Results

Since the $k$-approval rule is obviously manipulable and obvious manipulations can be found in polynomial time, we further investigate these manipulations in an experiment. Below, we experimentally determine the effects of $k$, $m$ and $n$ on the proportion of obviously manipulable voter preferences under the $k$-approval rule. Assuming a fixed tie-break ordering, we generate 1 million randomly permuted voter preference orderings and determine what proportion of these orderings admit an OM for a given set of parameters. It suffices to simply consider individual preference orderings as the best- and worst-case outcomes (and therefore obvious manipulability) for an agent's preference relation are over all possible preferences of the other agents. Note that from Theorem 5, the set of WOM-admitting preference orderings is the same as the set of OM-admitting preference orderings.

**Effect of $n$:** Figure 1 depicts the results from our experiments determining the effect of the number of voters $n$ on the proportion of obviously manipulable preference orderings. The downwards trend is concurrent with the existing theory that the proportion of individually manipulable voting profiles approaches zero as the number of voters tends to infinity [16]. A significantly lower proportion of preference orderings admit a BOM than those that admit a WOM. These trends are consistent for other values of $m$ and $k$, though other figures are omitted due to space restrictions.

**Effect of $m$ and $m-k$:** In Figure 2, we show heat maps of the proportion of OM-admitting preferences for $m \in \{21, \ldots, 30\}$ and $m-k$ values for which the preference profile is obviously manipulable. It is more appropriate to consider the number of disapprovals $m-k$ than the number of approvals $k$, as the impact of $k$ is relative to its difference from the number of outcomes. For example, it is better to compare $m = 21, k = 20$ with $m = 30, k = 29$ than with $m = 30, k = 20$. For a fixed number of disapprovals, the proportion of OM-admitting preferences increases with the number of outcomes. This is likely because a lower proportion of the outcomes can be 'blocked' by the other voters under the worst case outcome. The proportion increases steadily then rapidly decreases as the number of disapprovals increases, suggesting that an intermediary number of disapprovals increases individual manipulative power in comparison to the manipulative coalition of the other voters.

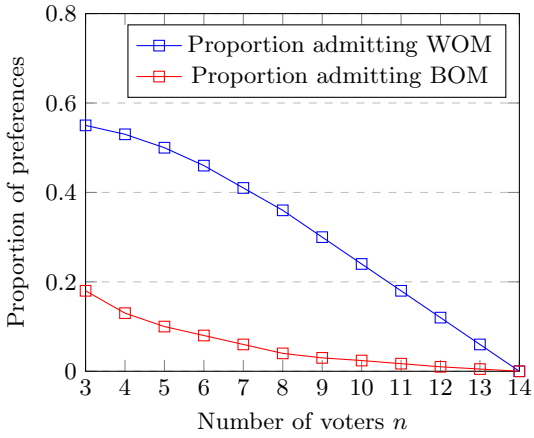


Fig. 1: Effect of $n$ on proportion of preferences that admit WOM and BOM ($k = 14, m = 15$)

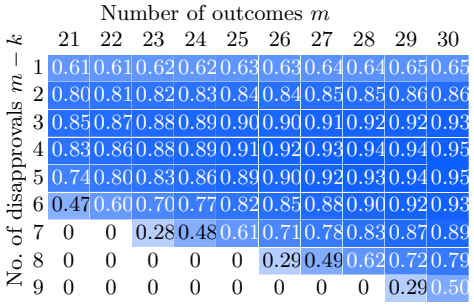


Fig. 2: Effect of $m$ and $m-k$ on proportion of OM-admitting preferences ($n = 3$)

## 7 Conclusion

In this paper, we initiated research on the obvious manipulability of voting rules. One of our key insights is that certain rules are obviously manipulable when the number of outcomes is relatively large as compared to the number of voters. The $k$-approval rule is an example of such a rule, and we have also shown that under the rule, an obvious manipulation can be computed in polynomial time. Despite all unanimous, non-dictatorial voting rules being manipulable for $n \geq 3$, most commonly used rules are NOM, suggesting that NOM is a significantly weaker notion than strategyproofness. We remark that in the positional scoring rules that we have classified as OM, the obvious manipulations are dependent on a fixed, deterministic tiebreak ordering which is standard in the voting literature. To gain further insights into which voting rules are more manipulable than others, a Bayesian approach could be used, in which voters have prior beliefs on the distribution of other votes. This approach lies between the perfect information of strategyproofness and the lack of information in NOM. As a new concept, NOM has currently been examined only for a handful of settings. It will be interesting to consider it when analyzing the strategic behaviour of agents in other settings such as fair division (see, e.g., [15]).

**Acknowledgements.** The authors thanks Anton Baychkov, Barton Lee and the anonymous reviewers of ADT 2021 for useful feedback.

## References

1 Arrow, K.: A difficulty in the concept of social welfare. Journal of Political Economy pp. 328–346 (1950)
2 Black, D.: Borda, Condorcet and Laplace. In: The Theory of Committees and Elections, chap. 18, pp. 156–162 (1986)
3 Carroll, G.: A quantitative approach to incentives : Application to voting rules (2011)
4 Chisholm, H.: Vote and voting. In: Encyclopaedia Britannica, p. 216 (1911)
5 Conitzer, V., Walsh, T.: Barriers to manipulation in voting. In: Brandt, F., Conitzer, V., Endriss, U., Lang, J., Procaccia, A.D. (eds.) Handbook of Computational Social Choice, chap. 6, Cambridge University Press (2016)

6 Conitzer, V., Walsh, T., Xia, L.: Dominating manipulations in voting with partial information. In: Proc. of the 25th AAAI Conference (2011)
7 Faliszewski, P., Procaccia, A.D.: AI's war on manipulation: Are we winning? AI Magazine pp. 53–64 (2010)
8 Favardin, P., Lepelley, D., Serais, J.: Borda rule, Copeland method and strategic manipulation. Review of Economic Design **7**(2), 213–228 (2002)
9 Fraenkel, J., Grofman, B.: The Borda count and its real-world alternatives: Comparing scoring rules in Nauru and Slovenia. Australian Journal of Political Science **49**(2), 186–205 (2014)
10 Gibbard, A.: Manipulation of voting schemes: A general result. Econometrica pp. 587–601 (1973)
11 Li, S.: Obviously strategy-proof mechanisms. American Economic Review pp. 3257–3287 (2017)
12 Meir, R.: Strategic Voting. Synthesis Lectures on Artificial Intelligence and Machine Learning (2018)
13 Niou, E.: Strategic voting under plurality and runoff rules. Journal of Theoretical Politics **13**(2), 209–227 (2001)
14 Nitzan, S.: The vulnerability of point-voting schemes to preference variation and strategic manipulation. Public Choice pp. 349–370 (1985)
15 Ortega, J.: Obvious manipulations in cake-cutting. CoRR **abs/1908.02988** (2019)
16 Peleg, B.: A note on manipulability of large voting schemes. Theory and Decision **11**(4), 401–412 (1979)
17 Reilly, B.: Social choice in the south seas: Electoral innovation and the Borda count in the Pacific Island countries. International Political Science Review **23**(4), 355–372 (2002)
18 Satterthwaite, M.A.: Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. J. Econ. Theory pp. 187–217 (1975)
19 Slinko, A., White, S.: Non-dictatorial social choice rules are safely manipulable. In: COMSOC'08, pp. 403–413 (2008)
20 Slinko, A., White, S.: Is it ever safe to vote strategically? Social Choice and Welfare pp. 403–427 (2014)
21 Taylor, A.D.: Social Choice and the Mathematics of Manipulation. Cambridge University Press (2005)
22 Tideman, N.: The single transferable vote. Journal of Economic Perspectives **9**(1), 27–38 (1995)
23 Troyan, P., Morrill, T.: Obvious manipulations. Journal of Economic Theory **185** (2020)
24 Wilson, M.C., Reyhani, R.: The probability of safe manipulation. In: COMSOC'10 (2010)
25 Xia, L., Zuckerman, M., Procaccia, A.D., Conitzer, V., Rosenschein, J.S.: Complexity of unweighted coalitional manipulation under some common voting rules. In: Proc. of the 21st IJCAI, pp. 348–353 (2009)
26 Zuckerman, M., Procaccia, A.D., Rosenschein, J.S.: Algorithms for the coalitional manipulation problem. Artificial Intelligence **173**(2), 392–412 (2009)