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 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. One of our insights is that certain rules are obviously manipulable when the number of alternatives is relatively large as compared to the number of voters. In contrast to the Gibbard-Satterthwaite theorem, many of the rules we examined were 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.

Keywords: Social choice · voting · manipulation · strategyproofness.

1 Introduction

Throughout history, voting has been used as a means of making public decisions based on the preferences of the citizens. 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 [Chisholm, 1911]; 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, which 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 [Black, 1986]. 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 [Arrow, 1950], proven in 1950, showed that there exists no voting system with three reasonable requirements. In a similar vein, Gibbard-Sattherthwaite theorem [Gibbard, 1973, Satterthwaite, 1975] states that when there are at least three alternatives, every 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 [2020]. Whilst strategy-proofness 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* if no agent can improve its best case or worst case outcome under any manipulation. A mechanism is *obviously manipulable* if either an agent's best case or worst case outcome can be improved by some untruthful interaction.

The assumptions made for *non 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 rule with weakly diminishing differences are not obviously manipulable, we show that certain voting rules including k-Approval are obviously manipulable. For the class of k-Approval voting rules, we characterize the conditions under which the rules are obviously manipulable. One of our insights is that certain rules are obviously manipulable when the number of alternatives is relatively large as compared to the number of voters.

Related Work Our paper belongs the rich stream of work in social choice on the manipulability of voting rules. The reader is referred to the book by Taylor [2005] 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 counti the relative number of preference profiles under which voting rules are manipulable (see, e.g., [Favardin et al., 2002]). Another approach is analysing maximum amount of expected utility an agent can gain by manipulating [Carroll, 2011].

Our work revolves around the concept of obvious manipulations, which was proposed by Troyan and Morrill [2020]. This concept was inspired by a paper on 'obviously strategy-proof mechanisms' by Li [2017]. 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 strategy-proofness', a strengthening of strategy-proofness. A mechanism is defined as obviously strategy-proof 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 [2020] studied obvious manipulation 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 strategy-proofness specific to voting rules have been proposed in the literature. Slinko and White [2008, 2014] considered *safe strategic voting* to represent 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 more preferred. For further discussion on strategic aspect of voting under uncertain, the reader is referred to Chapter 6 and 8 of the book by Meir [2018].

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 [Nitzan, 1985]. A similar probabilistic perspective was used by Wilson and Reyhani [2010].

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. Faliszewski and Procaccia [2010] and Conitzer and Walsh [2016]).

2 Preliminaries

We consider the standard social choice voting setting $(N, O, \succ)$ that involves a finite set $N = \{1, 2, \dots, n\}$ of n voters and a finite set $O = \{o_1, o_2, \dots, o_m\}$ of m outcomes. The preference profile $\succ = (\succ_1, \dots, \succ_n)$ specifies the transitive, complete and reflexive preference relation $\succ_i$ for each voter i over the alternative set O. We assume that $n \geq 3$ and $m \geq 3$, and that ties are broken based on lexicographic order. A voting rule f 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 alternative. Unless specified otherwise, we will assume lexicographic tie-breaking.

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 $B_{\succ_i}(\succ_i', f) := \sup_{\succ_{-i}} f(\succ_i', \succ_{-i})$ the best possible outcome (under i's preference $\succ_i$) when she reports $\succ_i'$ and other voters can report any preference. We also denote by $W_{\succ_i}(\succ_i', f) := \inf_{\succ_{-i}} f(\succ_i', \succ_{-i})$ the worst possible outcome (under i's preference $\succ_i$) when she reports $\succ_i'$ and other voters can report any preference. We now present the central concept used in the paper.

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 *obviously manipulable in the worst case*. Similarly, if (2) does not hold, then we say it is *obviously manipulable in the best case*.

3 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 and a fixed number of voters n and alternatives m, a voter i has veto power if, for any possible outcome $o \in O$, there exists a report $\succ_i$ that ensures o is not selected.*

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

Theorem 1. *If a voting rule is obviously manipulable, then it must admit a non-dictatorial vetoer.*

Proof. We show that if a voting rule f does not admit a non-dictatorial vetoer, then is not obviously manipulable. Consider a voting rule f that admits no vetoer i. The admission of a dictatorial voter implies that the rule is strategyproof, so it suffices to consider the case when f is not dictatorial.

First note that voter i's best outcome under truthful report $\succ_i$ is her most preferred alternative o in the set of possible outcomes under f. Such an outcome is achievable because o is a possible outcome and because i is not a vetoer. Therefore her best outcome under any untruthful report $\succ_i'$ cannot be strictly better than under a truthful report.

When i reports $\succ_i'$, her worst possible outcome with respect to her preference $\succ_i$ is her least preferred alternative from the set of possible outcomes. Such an outcome is achievable because f does not admit a vetoer. We therefore have $W_{\succ_i}(\succ_i, f) \succeq_i W_{\succ_i}(\succ_i', f)$ for all untruthful ballots $\succ_i'$. Therefore, f is NOM. □

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 alternative o when o is the most preferred alternative for all voters except one. Almost-unanimity implies unanimity. A rule that is almost-unanimous does not admit any vetoer. For $n \geq 3$, a majoritarian rule is almost unanimous.*

Theorem 2. *For $n \geq 3$, no almost-unanimous voting rule is obviously manipulable.*

Proof. Note that an almost-unanimous voting rule is not dictatorial. A rule that is almost-unanimous does not admit any vetoer. Hence it follows from Theorem 1 that for $n \geq 3$, no almost-unanimous voting rule is obviously manipulable. □

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

Proof. Any majoritarian rule is almost unanimous. Hence, the statement follows from Theorem 2. □

Similarly, Theorem 2 applies to several voting rules including STV and Plurality with runoff.

4 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 position scoring rule is given below.

Definition 5. *A position 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 ith 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.*

Several well-known rules fall in the class of position 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 3. *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 alternative 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 alternative 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

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

□

4.1 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 $w = (\underbrace{1,\dots,1}_{k \text{ ones}}, \underbrace{0,\dots,0}_{m-k \text{ zeros}})$, where $0 < k < m$.

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 1. *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

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

and the lexicographic ordering of the preferences is

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

Under a k-Approval rule, any voter may disapprove of their $m-k$ least preferred outcomes. Since there are a total of $n(m-k)$ vetoes 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 lexicographic winner of the outcomes that haven't been disapproved.

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 $W_{\succ_i}(\succ_i, kApp) = o_k$, achieved by the other voters vetoing each outcome of $\{o_1,\dots,o_{k-1}\}$ at most once, leaving o_k as the lexicographic tiebreak winner.

If voter i instead chooses to disapprove outcomes $\{o_k\} \cup \{o_{k+1},\dots,o_m\} \backslash \{o_{i'}\}$, where $k+1 \leq i' \leq m$, then the worst outcome satisfies $W_{\succ_i}(\succ_i', kApp) \succ_i o_{k-1}$, as $o_{i'}$ always loses the lexicographic tiebreak 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 2. *The k-Approval rule (kApp) is NOM if $n > \frac{m-2}{m-k}$.*

Proof. **Case 1 ($m = n(m-k)+1$):**
Suppose that there are n voters, $m = \frac{kn-1}{n-1}$ outcomes and without loss of generality that voter i's true preferences are

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

Recall that the k-Approval rule allows each voter to disapprove of $m-k$ outcomes. By the pigeonhole principle, there must be at least one outcome with zero disapprovals, so the chosen outcome must be the lexicographic winner of the outcomes with zero disapprovals.

Under $\succ_i$, voter i's best case outcome of $B_{\succ_i}(\succ_i, kApp) = o_1$ is trivially achievable by having the other $n-1$ voters disapprove of outcomes $\{o_2,\dots,o_k\}$. Since his best case outcome is his first preference, it cannot be strictly improved by any manipulation, so it suffices to prove that the worst case outcome cannot be strictly improved.

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 exactly once. We therefore have $W_{\succ_i}(\succ_i, kApp) \succeq_i W_{\succ_i}(\succ'_i, kApp)$, concluding our proof for the case where $m = n(m-k)+1$.

Case 2 ($m < n(m-k)+1$):
Again, voter i's best case outcome of $B_{\succ_i}(\succ_i, kApp) = o_1$ is trivially achievable by the voters voting such that o_1 has zero disapprovals and each of the other outcomes has at least one disapproval. It suffices to prove that the worst case outcome cannot be strictly improved.
Since each outcome of $\{o_{k+1}, \ldots, o_m\}$ has one disapproval from voter i, the worst case scenario must have the same chosen outcome as the worst case scenario where each outcome has at least one disapproval. We can therefore constrain our scenarios to ones meeting that condition. Now any possible manipulation by i must approve of at least one outcome from $\{o_{k+1}, \ldots, o_m\}$, and disapprove of at least one outcome from $\{o_1, \ldots, o_k\}$. However, since each outcome originally has one disapproval, any change in disapproval numbers by i can be reversed by the other voters disapproving of voter i's approved outcomes, and approving voter i's disapproved outcomes. 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 vetoes.

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

Theorem 4. *The k-Approval rule is obviously manipulable if and only if $n \leq \frac{m-2}{m-k}$.*

Proof. The statement follows from Lemma 1 and Lemma 2. □

Corollary 2. *The plurality rule is NOM.*

Proof. Note that for plurality, $k = 1$. Hence, $m \leq n(m-k)+1$ holds. □

Since plurality is generally considered one of rules that are easiest to manipulate, the corollary above underscores the strength of obvious manipulations. Alternatively, it is suggested that NOM is considerably weaker than strategyproofness.

4.2 Strict Positional Scoring Rules

In the previous section, we note that k-approval 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.

Definition 6. *A position 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 3. *There exists a strict positional scoring rule that can admit a voter with veto power and is obviously manipulable.*

Proof. Consider the scoring rule $w = (m+2, m+1, \ldots, 4, 0)$. Suppose we have $m = 4$, $w = (6, 5, 4, 0)$ and the following preferences:

$$\succ_i\colon o_1 \succ_i o_2 \succ_i o_3 \succ_i o_4,$$

$$\succ_j\colon \cdot \succ_j \cdot \succ_j \cdot \succ_j \cdot,$$

$$\succ_k\colon \cdot \succ_k \cdot \succ_k \cdot \succ_k \cdot.$$

We first show that this scoring rule can admit a voter with veto power. Here, voter i attempts to veto outcome o_4 by voting it last. We show that it is impossible for the other voters to vote such that outcome o_4 is chosen. Clearly, j and k must vote outcome o_4 as first preference. Now outcome o_1 must have a strictly lower score than o_4, so we must set it as the second preference of one voter and the last preference of the remaining voter.

$$\succ_j\colon o_4 \succ_j o_1 \succ_j \cdot \succ_j \cdot$$

$$\succ_k\colon o_4 \succ_k \cdot \succ_k \cdot \succ_k o_1$$

If either outcome o_2 or o_3 are set as j's third preference and k's second preference, then they will have a strictly higher score than outcome o_4. If they are set as the third preference of both j and k, then either o_2 would have a strictly higher score than o_4, or o_3 would have the same score as o_4, but be chosen by the lexicographic tiebreak. Therefore there does not exist a voting profile $\succ_{-i}$ that chooses outcome o_4, so it has been effectively vetoed by voter i.

We now show that this voting rule is obviously manipulable. Now instead suppose that the lexicographic order is

$$o_1 \succ_L o_2 \succ_L o_4 \succ_L o_3.$$

By setting the preferences as

$$\succ_i\colon o_1 \succ_i o_2 \succ_i o_3 \succ_i o_4,$$

$$\succ_j\colon o_4 \succ_j o_1 \succ_j o_3 \succ_j o_2,$$

$$\succ_k: o_4 \succ_k o_2 \succ_k o_3 \succ_k o_1,$$

$W_{\succ_i}(\succ_i, f) = o_4$ is achievable as it wins the lexicographic tiebreak with outcome o_3. Now suppose that voter i instead reports the manipulation

$$\succ_i': o_1 \succ_i o_3 \succ_i o_2 \succ_i o_4.$$

By a similar argument as above, it can be shown that it is impossible for the other voters to vote such that outcome o_4 is chosen. Therefore, $W_{\succ_i}(\succ_i', f) \succ_i W_{\succ_i}(\succ_i, f)$, concluding our proof. □

In the following lemma, we also find that a positional voting rule is not necessarily obviously manipulable if it is strict and admits a vetoer.

Lemma 4. *There exists a class of strict positional scoring rules that can admit a voter with veto power but are NOM.*

Proof. Consider the scoring rule $w = (\omega + m\epsilon, \omega + (m-1)\epsilon, \ldots, \omega + 2\epsilon, 0)$, where $\omega > 0$ is sufficiently large and $\epsilon > 0$ is sufficiently small. Suppose without loss of generality that voter i's preferences are:

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

Since the voting rule is strict, the best case outcome is trivial.

Case 1 ($m > n$):

We will show that this scoring rule admits a vetoer when the number of outcomes is greater than the number of voters. The highest score that outcome o_m can receive is $(n-1)(\omega + m\epsilon)$. By the pigeonhole principle, there must exist at least one outcome which is not voted last preference by any voter. Each of these outcomes always has a score greater than $n\omega$, which is greater than $(n-1)(\omega + m\epsilon)$. It can therefore be seen that o_m has been vetoed by voter i. Now suppose that the other $n-1$ voters report the following ballots:

$$o_{m-1} \succ \cdots \succ O \succ o_1$$

$$o_{m-1} \succ \cdots \succ O \succ o_2$$

$$\cdots$$

$$o_{m-1} \succ \cdots \succ O \succ o_{n-1},$$

where

$$O = \begin{cases} o_m & m = n+1 \\ o_{m-2} & m = n+2 \\ o_{m-2} \succ \cdots \succ o_n & m > n+2 \end{cases}$$

If $m = n+1$, then alternative o_{m-1} trivially has the highest score and is chosen. We now consider $m > n+1$. Clearly, the two alternatives with the highest scores are o_{m-1}

and o_{m-2}. We now show that the score of o_{m-1} is always greater than the score of o_{m-2}.

$$\begin{aligned}
&\text{Score}(o_{m-1}) > \text{Score}(o_{m-2}) \\
\iff &(\omega + 2\epsilon) + (n-1)(\omega + m\epsilon) > \omega + 3\epsilon + (n-1)(\omega + (m-n-1)\epsilon) \\
\iff &n\omega + \epsilon(2 + mn - m) > n\omega + \epsilon(4 + mn - n^2 - m) \\
\iff &n^2 > 4
\end{aligned}$$

which always holds as we assume $n \geq 3$. Therefore, voter i's worst case outcome under a truthful report is $W_{\succ_i}(\succ_i, w) = o_{m-1}$. Now any manipulation where o_{m-1} increases in preference gives it a strictly higher score, so it remains chosen. Any manipulation where o_{m-1} remains the same preference ranking can be reversed by the other voters making the same outcome changes in their ballots. Finally, any manipulation where o_{m-1} moves to a lower preference ranking results in o_m being unvetoed, making it the worst case outcome. Thus, this scoring rule is NOM when $m > n$.

Case 2 ($m \leq n$):

Now suppose the other $n-1$ voters vote o_m as their first preference, and that they vote such that every outcome is voted last preference by at least one voter. Excluding o_m, the outcome with the highest possible score is o_1, when it is voted second preference by $n-2$ of the other voters and last preference by 1 of the other voters. We now show that under this scenario, o_m has a higher score than o_1.

$$\begin{aligned}
&\text{Score}(o_m) > \text{Score}(o_1) \\
\iff &(n-1)(\omega + m\epsilon) > (\omega + m\epsilon) + (n-2)(\omega + (m-1)\epsilon) \\
\iff &\epsilon(mn - n) > \epsilon(mn - m - n + 2) \\
\iff &n > 2
\end{aligned}$$

which always holds as we assume $n \geq 3$. o_m is therefore voter i's worst case outcome. Using the same arguments as in Case 1, we deduce that o_m is also the worst case outcome under any manipulation by voter i. This scoring rule is therefore NOM under $m \leq n$. □

Definition 7. *A strict positional scoring rule with weight vector $w = (s_1, s_2, \ldots, s_m)$ 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 $i \in \{1, 2, \ldots, m-2\}$.*

A common example of such a rule is the Dowdall system, which has weight vector $w = (1, 1/2, \ldots, 1/m)$. It is more favourable towards candidates with many first preferences, and reduces the impact of voters randomly voting their late preferences due to the requirement of complete preferences.

Lemma 5. *Assuming $m, n \geq 3$ and lexicographic tiebreaks, a position scoring rule f is obviously manipulable in the best case if and only if for some $k > 1$, the first k elements of the scoring vector are the same and $k - 1 > (n-1)(m-k)$.*

Proof. Suppose voter i has preferences

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

Consider the instance where every voter reports the same preferences as i. If the first and second elements of the scoring vector are different, then $B_{\succ_i}(\succ_i, f) = o_1$ is achieved, which cannot be manipulated further. Therefore, obvious manipulability in the best case requires the first k elements of the scoring vector to be the same, where $k > 1$.

Now suppose the first k elements of the scoring vector are equal, causing o_1 to be tied with $k-1$ other outcomes. If any of those outcomes are of higher lexicographic order than o_1, then o_1 will not be selected in our instance. We now attempt to modify our instance so that o_1 remains the selected outcome.

We can modify our instance by having the other $n-1$ voters *replace* the other $k-1$ tied outcomes with one of the $m-k$ non-tied outcomes in its preference list. Each outcome only needs one replacement to become strictly worse than o_1. If $(n-1)(m-k) \geq k-1$, then enough replacements can be made, and we achieve $B_{\succ_i}(\succ_i, f) = o_1$. Therefore, obvious manipulability requires $k-1 > (n-1)(m-k)$.

If $k-1 > (n-1)(m-k)$, then there will remain at least one tied outcome. Suppose that the lexicographic order is

$$\succ_L\colon o_k \succ_L o_{k-1} \succ_L \cdots \succ_L o_2 \succ_L o_1 \succ_L o_m,$$

meaning none of the tied outcomes can be ignored in regards to replacement, as they all beat o_1 lexicographically. Voter i can manipulate its best case outcome by replacing the lexicographic tiebreak winner of the remaining tied outcomes with o_m. □

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

Theorem 5. *A strict positional scoring rule with weakly diminishing differences is NOM.*

Proof. Suppose we have weight vector $w = (s_1, s_2, \ldots, s_m)$, where $s_i - s_{i+1} \geq s_{i+1} - s_{i+2}$ for all $i \in \{1, 2, \ldots, m-2\}$. The rule is strict, so it is NOM in the best case because i's most preferred alternative gets selected if it is reported to be in the first position by all the voters.

Next, we show that that rule is NOM in the worst case. We do so by showing that whenever the voter i misreports, there exists a profile of the other voters under which i's least preferred alternative is selected.

Consider the scenario where voter i's truthful ballot is

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

and every other voter reports the reverse preference order

$$\succ_{-i}\colon o_m \succ_{-i} o_{m-1} \succ_{-i} \cdots \succ_{-i} o_2 \succ_{-i} o_1.$$

Alternative o_m has a score of $s_m + (n-1)s_1$, and for $i \in \{1, 2, \ldots, m-1\}$, o_i has a score of $s_i + (n-1)s_{m+1-i}$. We show that $s_m + (n-1)s_1 > s_i + (n-1)s_{m+1-i}$ for $i \in \{1, 2, \ldots, m-1\}$.

$$\begin{aligned} & s_m + (n-1)s_1 > s_i + (n-1)s_{m+1-i} \\ \iff & (n-1)(s_1 - s_{m+1-i}) > s_i - s_m \end{aligned}$$

Since $n \geq 3$, the inequality holds for $i = 1$. It suffices to show that $s_1 - s_{m+1-i} \geq s_i - s_m$ for all $i \in \{2, 3, \ldots, m-1\}$. Now due to weakly diminishing differences, the following inequalities hold:

$$\begin{aligned} s_1 - s_2 &\geq s_i - s_{i+1} \\ s_2 - s_3 &\geq s_{i+1} - s_{i+2} \\ &\ldots \\ s_{m-1-i} - s_{m-i} &\geq s_{m-2} - s_{m-1} \\ s_{m-i} - s_{m+1-i} &\geq s_{m-1} - s_m. \end{aligned}$$

If we sum up these inequalities, we have $s_1 - s_{m+1-i} \geq s_i - s_m$, so therefore $s_m + (n-1)s_1 > s_i + (n-1)s_{m+1-i}$ for $i \in \{1, 2, \ldots, m-1\}$, meaning alternative o_m has a strictly higher score than every other alternative. Any manipulation by i where o_m is not her last preference results in the outcome having a strictly higher score and remaining chosen under the same $\succ_{-i}$. If i reports the manipulation $\succ_i'$ such that o_m is still her last preference, it can be shown by the same argument that o_m remains her worst case outcome, achieved by the other voters reporting the reverse preference order. Therefore the rule is NOM in the worst case. □

Corollary 3. *The Borda and Dowdall rules are NOM.*

Remark 2. Lemma 4 exemplifies a class of strict positional scoring rules which do not satisfy weakly diminishing differences but are NOM.

5 Conclusion

In this paper, we initiated research on obvious manipulability of voting rules. 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 key insights is that certain rules are obviously manipulable when the number of alternatives is relatively large as compared to the number of voters. Despite all voting rules being manipulable for $n \geq 3$, most commonly used rules are NOM, suggesting that NOM is a significantly weaker notion than strategyproofness. This is expected as OM captures the absolute lack of information available to voters, a common occurrence in real world elections.

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.

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

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

As a new concept, NOM has currently been examined only for a handful of settings. It will interesting to consider it when analyzing the strategic behaviour of agents in other settings such as fair division.

Acknowledgments

The authors thanks Barton Lee for useful feedback.

Bibliography

K. Arrow. A difficulty in the concept of social welfare. *Journal of Political Economy*, pages 328–346, 1950.

D. Black. Borda, condorcet and laplace. In *The Theory of Committees and Elections*, chapter 18, pages 156–162. 1986.

Gabriel D. Carroll. A quantitative approach to incentives : Application to voting rules. 2011.

H. Chisholm. Vote and voting. In *Encyclopaedia Britannica*, page 216. 1911.

V. Conitzer and T. Walsh. Barriers to manipulation in voting. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 6. Cambridge University Press, 2016.

P. Faliszewski and A. D. Procaccia. Ai's war on manipulation: Are we winning? *AI Magazine*, pages 53–64, 2010.

P. Favardin, D. Lepelley, and J. Serais. Borda rule, copeland method and strategic manipulation. *Review of Economic Design*, 7(2):213–228, 2002.

A. Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, pages 587–601, 1973.

S. Li. Obviously strategy-proof mechanisms. *American Economic Review*, pages 3257–3287, November 2017.

R. Meir. *Strategic Voting*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2018.

S. Nitzan. The vulnerability of point-voting schemes to preference variation and strategic manipulation. *Public Choice*, pages 349–370, 1985.

M. A. Satterthwaite. Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *J. Econ. Theory*, pages 187–217, 1975.

A. Slinko and S. White. Non-dictatorial social choice rules are safely manipulable. In *COMSOC'08*, pages 403–413, 2008.

A. Slinko and S. White. Is it ever safe to vote strategically? *Social Choice and Welfare*, pages 403–427, 2014.

A. D. Taylor. *Social Choice and the Mathematics of Manipulation*. Cambridge University Press, 2005.

P. Troyan and T. Morrill. Obvious manipulations. *Journal of Economic Theory*, 185, 2020.

M. C. Wilson and R. Reyhani. The probability of safe manipulation. In *COMSOC'10*, 2010.