Information, interaction and manipulation in voting

Doctoral thesis

INFORMATION, INTERACTION AND MANIPULATION IN VOTING

Yuliya A. Veselova

2023

INFORMATION, INTERACTION AND MANIPULATION IN VOTING

Dissertation

To obtain the degree of Doctor at the Maastricht University,
on the authority of the Rector Magnificus,
Prof. dr. Pamela Habibović
in accordance with the decision of the Board of Deans,
to be defended in public
on Monday 18th of September 2023, at 13.00 hours

by

Yuliya A. Veselova

Supervisor
Prof. dr. H.J.M. Peters

Co-supervisor
Dr. A.J.A. Storcken

Assessment Committee
Prof. dr. A.J. Vermeulen (chair)
Prof. dr. U. Endriss (University of Amsterdam)
Dr. A. Mackenzie
Dr. Z. Terzopoulou (Jean-Monnet University, Saint-Etienne)

To my parents

Contents

1
Introduction

Collective decision making is a part of everyday life of the modern society, the earliest mention of voting dates back to antiquity. Since then, people not only invented a variety of voting procedures, or social choice rules, but also studied them formally with the help of mathematical and computer modelling. One of the problems with aggregating individual preferences into a collective choice is that people can misrepresent their preferences in order to achieve a more preferable voting result. This phenomenon is called manipulation and considered as negative, since due to manipulation the voting result becomes biased.

K. Arrow (Arrow, 1951) was the founder of an axiomatic approach to studying voting procedures proving that some set of reasonable properties of social choice rules is incompatible. Using this approach, A. Gibbard (Gibbard, 1973) and M. Satterthwaite (Satterthwaite, 1975) independently proved incompatibility of strategy-proofness and unanimity with non-dictatorship in case there are at least three alternatives. This fundamental result gave rise to a number of studies on manipulability of rules. Among them, there are works that reveal the degree of manipulability of rules depending on the number of voters and alternatives taking part. The degree of manipulability is defined as a proportion of voters' preference profiles with at least one voter having an incentive to manipulate.

The following thesis continues this line of research. Our aim was to enrich the basic model of manipulation (where only one voter manipulates, knowing all other voters' preferences and not thinking about their strategic actions). The model for manipulation under incomplete information was presented in (Reijngoud and Endriss, 2012) and studied in (Veselova, 2020). Suppose, that all individual preferences are collected before voting for an opinion poll. After that, information about preferences becomes available to voters, but in an aggregated form. For example, information about a ranking of alternatives according to a given rule. Having this information, a voter decides which preference order to declare, sincere or strategic. In Chapter 2 we present the extension of this model allowing coalitions of voters to manipulate. Thus, keeping the assumption of incomplete information we suggest a voter thinking of her allies and their actions. Then we answer the question how manipulability of rules changes when we switch from individual to coalitional manipulation and how different types of information affect incentives to manipulation in a society.

There is no doubt that groups of voters have more possibilities to influence the voting result than separate individuals. However, there exists a problem with coordinating their actions. Suppose, under complete information a group of individuals with identical preferences has an incentive to manipulate and they could achieve a better voting result. For example, all voters know what is the number of their allies and that by acting together they could make their best possible alternative the winner. Thus, they have an incentive to manipulate within a group. But what if some of them do not actually manipulate? If a voting result could become even worse for manipulating voters, than it was initially, then such manipulation is considered as unsafe. Thus, the next question under consideration is the following: which rules and conditions allow for an unsafe manipulation? If a voting rule is always safely manipulable, it means that it is easier to manipulate the result, since voters do not need to carefully coordinate their actions. This question is extensively answered in Chapter 3 of this thesis.

Having considered the problem of interaction of voters with identical preferences in manipulation, we need to mention though that there could be more than one group having an incentive to manipulate. Regardless of whether their preferences are similar or opposite to each other, their simultaneous manipulation could lead to a fail for some (or even all) of them. The model of manipulation under incomplete information is extended for allowing not only co-minded people to manipulate, but all voters having such an incentive. If manipulating voters take into account all other voters' manipulation, not only of their type, this leads to another level of uncertainty. Adding an assumption of incomplete information, we increase uncertainty even more. Thus, it is of interest to know how the combination of these two kinds of uncertainty influences voters' willingness to manipulate. Chapter 4 is devoted to the study of this question. In some cases rules become immune to manipulation and we find such cases.

Overview

This dissertation is organized as follows.

Chapter 2 considers the problem of coalitional manipulation in collective decision making and a probabilistic approach for solving it. We conduct computational experiments for calculating the degree of coalitional manipulability for various types of public information and compare it with individual maniulation. In a theoretical study we prove that under incomplete information individual and coalitional manipulability can be equal and consider asymptotic behavior of manipulability for plurality and Borda rule.

We address the issue of the safety of group manipulation in Chapter 3. For several voting rules we study conditions on the numbers of voters and alternatives which allow for an unsafe manipulation or which make manipulation always safe.

Chapter 4 is devoted to individual manipulability of social choice rules under incomplete information and for different assumptions about voters' behavior which constitute a behavioral model. With the help of computations we reveal how the type of information and behavioral model influence the relative manipulability of 12 social choice rules. It is formally proved that under certain conditions manipulability equals zero for scoring rules.

2

Manipulation by coalitions in voting with incomplete information

Adapted from: Y. A. Veselova *Manipulation by coalitions in voting with incomplete information*. In: Data Analysis and Optimization, Springer, 2023.

2.1 Introduction

We consider the problem of manipulation in collective decision making. It is well-known that voters can misrepresent their preferences in order to achieve a more preferable result. Of course, it is better when all voters want to declare their sincere preferences, otherwise, a collective decision would be biased and, consequently, would not reflect the preference of a society. Unfortunately, all social choice rules which have at least three possible outcomes are either manipulable or dictatorial. This result is called Gibbard-Satterthwaite theorem (Gibbard, 1973; Satterthwaite, 1975; Gärdenfors, 1976).

However, for a social choice rule to be vulnerable to manipulation it is enough to have only one situation and at least one voter having an incentive to misrepresent her preferences. Thus, different rules may be manipulable to a different extent. For this reason the most used approach to comparing the degree of manipulability of rules is measuring the share of situations (preference profiles) that admit manipulation by voters. This approach was first used by Nitzan (1985) and Kelly (1988) and further applied to the analysis of a big variety of rules in different models (Aleskerov and Kurbanov, 1999; Lepelley and Valognes, 2003; Pritchard and Wilson, 2007; Aleskerov et al, 2009; Aleskerov et al, 2011; Aleskerov et al, 2012; Maus et al, 2007, Peters et al, 2012; Slinko, 2006, Aleskerov et al, 2017).

The common assumption in publications of this line of research is that voters know each others' sincere preference, i.e. public information is reliable and complete. This is a rather strong assumption, but helps to simplify the comparative analysis of manipulability of social choice rules. Intuitively, incomplete information would make manipulation more difficult and rare.

A more realistic assumption is that voters have some information from opinion polls held before voting. This information could be represented, for example, by preferences of a subset of voters, or a list of candidate scores, or the winner of the election. A mathematical model for manipulation under poll information is presented by Reijngoud and Endriss (2012).

In the current research we apply this model to studying coalitional manipulability of social choice rules under different types of poll information. We consider the probability that in a randomly chosen preference profile there exists a coalition which has an incentive to manipulate under a given type of poll information. The formalization of coalitional manipulation in voting includes some assumptions: 1) voters form a coalition if they have the same preference; 2) all members of a coalition manipulate in the same way; 3) a coalition has an incentive to manipulate, if there exists an insincere strategy such that the coalition cannot become worse off and there is a chance of becoming better off with this strategy.

In our study the analysis of manipulation probability has three directions:

1) We analyze the power of a coalition: how coalitional manipulability differs from individual. Could coalitional manipulability be less than individual?

2) We compare manipulability of different social choice rules (we consider six popular rules which have polynomial complexity of calculating a winner: plurality rule, Borda rule, veto rule, runoff procedure, STV rule, and Copeland rule).

3) We study the role of information available to voters. How do different types of poll information affect coalitional manipulability?

We answer these questions via both theoretical investigation and computational experiments. We prove that for scoring rules (plurality, Borda, and veto rule in our analysis) the probability of coalitional manipulation is equal to the probability of individual manipulation if the only information available to voters is the information about the election winner after tie-breaking. Computational experiments are conducted in MatLab for all six rules and five poll information types for 3 alternatives and the number of voters from 3 to 15. It is shown that the probability of coalitional manipulation is almost always higher than individual manipulation and in many cases is very close to 1. The exceptions are the Borda rule and veto rule: for these the probability of coalitional manipulation could be less than the probability of individual manipulation for some public information types. This observation shows that manipulating with the same strategy is not optimal for coalition members in these cases. The veto rule even becomes almost immune to manipulation under incomplete information if there are more than 10 voters.

2.2 The Model

2.2.1 Definitions and Notations

There is a finite set of *voters* $N = \{1,...,n\}$ and a finite set of m *alternatives* X, $m \geq 3$. Each voter i has a strict *preference* P_i, which is a linear order, i.e. irreflexive, weakly complete and transitive binary relation on X. If voter i prefers an alternative a to an alternative b, we write aP_ib. The set of all linear orders is denoted by $L(X)$. A preference can also be represented in a form $P_i = (a,b,...c)$, which is equivalent to $aP_ibP_i...P_ic$. An *upper contour set* of an alternative a in a preference order P_i is $P_ia = \{b \in X : bP_ia\}$. Similarly, a *lower contour set* of a in P_i is $aP_i = \{b \in X : aP_ib\}$.

An ordered set of individual preferences, $\mathbf{P} = (P_1,...,P_n) \in L(X)^N$, is called a *preference profile*. A contraction of a preference profile onto the set $A \subseteq X$ is $\mathbf{P}/A = (P_1/A,...,P_n/A)$, where $P_i/A = P_i \cap (A \times A)$. A *coalition* is a subset of voters, $K \subseteq N$, $K \neq \emptyset$. A preference profile of coalition members is denoted by $\mathbf{P}_K$, and $\mathbf{P}_{-K}$ is preference profile of all other voters, $N \setminus K$. $\mathbf{P} = (\mathbf{P}_K, \mathbf{P}_{-K})$.

A vector of positions for alternative a is $v(a,\mathbf{P}) = (v_1(a,\mathbf{P}),...,v_m(a,\mathbf{P}))$, where $v_j(a,\mathbf{P})$ denotes the number of voters having a on the j-th position in a preference order, i.e. $j = 1 + |P_ia|$.

An $m \times m$ matrix of a *weighted majority graph* for a profile $\mathbf{P}$ is denoted by $WMG(\mathbf{P})$ and consists of elements

$$WMG(\mathbf{P})_{kl} = |\{i \in N : a_kP_ia_l\}|. \tag{2.1}$$

By μ we denote *majority relation*: $a_k\mu a_l$ if $WMG(\mathbf{P})_{kl} > WMG(\mathbf{P})_{lk}$.

A matrix of a majority graph is $MG(\mathbf{P})$, where

$$MG(\mathbf{P})_{kl} = \begin{cases} 1, & \text{if } a_k\mu a_l, \\ -1, & \text{if } a_l\mu a_k, \\ 0, & \text{otherwise.} \end{cases} \tag{2.2}$$

A *social choice correspondence* (SCC) is a mapping $C: L(X)^N \to 2^X \setminus \{\emptyset\}$. A *social choice rule* or simply *rule* is a mapping $F: L(X) \to X$. A rule can be obtained from SCC by using a tie-breaking rule $T : 2^X \setminus \{\emptyset\} \to X$. We consider an alphabetic tie-breaking rule: assume some linear order on X to be predefined, $aP_T bP_T c\ldots$, and when alternatives are tied, we choose the one which dominates all others by P_T (has a higher priority). Thus, a rule F is derived from SCC C, if $T(C(\mathbf{P})) = F(\mathbf{P})$. A *social ordering* is denoted by a weak order R (irreflexive, transitive, and negatively transitive binary relation), an element of the set of all weak orders on X, $W(X)$.

2.2.2 Poll Information Functions

It is assumed that an opinion poll is held before voting and it reveals voters' sincere preferences, $\mathbf{P}$. However, for some reasons not all information becomes available to voters. Instead of $\mathbf{P}$, voters get to know just $\pi(\mathbf{P})$, function π is called a poll information function (PIF). We consider the following types of PIF.

1. Profile: $\pi(\mathbf{P}) = \mathbf{P}$.
2. Score: $\pi(\mathbf{P}) = \mathbf{S}(\mathbf{P}) = (S(a_1, \mathbf{P}), \ldots, S(a_m, \mathbf{P}))$ assigns to each alternative its score (to be explained further) according to a given SWF F. For multi-stage procedures it is defined as a vector of vectors of scores for each stage.
3. Rank: $\pi(\mathbf{P}) = R$, returns a social ordering.
4. Winner: $\pi(\mathbf{P}) = C(\mathbf{P})$.
5. Unique winner (1Winner): $\pi(\mathbf{P}) = F(\mathbf{P})$

2.2.3 Individual manipulation

Thus, a voter i has information $\pi(\mathbf{P})$ about a preference profile $\mathbf{P}$ and knows her own preference order. A set of preference profiles of $N \setminus \{i\}$ consistent with her knowledge is called *information set* and defined as follows

$$W_i^{\pi(\mathbf{P})} = \{\mathbf{P'}_{-i} \in L(X)^{N \setminus \{i\}} : \pi(P_i, \mathbf{P'}_{-i}) = \pi(\mathbf{P})\}. \tag{2.3}$$

Given two PIFs π and π', if $\forall \mathbf{P} \in L(X)^N$ $\forall i \in N$ $W_i^{\pi(\mathbf{P})} \subseteq W_i^{\pi'(\mathbf{P})}$, then π is *at least as informative as* π'. Of course, the most informative is Profile-PIF.

Then, when is a voter willing to manipulate, i.e. misrepresent her preference in order to achieve a more preferable result? It is assumed that if there is at least one possible situation in which manipulation makes her better off and nothing changes in all other possible situations, then a voter has an incentive to manipulate under PIF π (Reijngoud and Endriss, 2012).

Definition 2.1. *Given a rule F and a preference profile $\mathbf{P}$, we say, that voter i has an incentive to π-manipulate under F, if there exists $\tilde{P}_i \in L(X)$ s.t.*
i) there is no $\mathbf{P'}_{-i} \in W_i^{\pi(\mathbf{P})}$, s.t. $F(\mathbf{P})\, P_i\, F(\tilde{P}_i, \mathbf{P'}_{-i})$;
ii) there exists $\mathbf{P'}_{-i} \in W_i^{\pi(\mathbf{P})}$, s.t. $F(\tilde{P}_i, \mathbf{P'}_{-i})\, P_i\, F(\mathbf{P})$.

Thus, if a voter has an incentive to π-manipulate in $\mathbf{P}$, is does not mean that her manipulation will be successful in this very profile.

Example 2.1. *Consider a case with 3 alternatives, 3 voters and among them voter 1 with preference aP_1bP_1c. PIF is 1Winner and $\pi(\boldsymbol{P}) = c$. Tie breaking rule is such that $T(\{a,b,c\}) = c$. A rule is plurality, which means that we choose alternatives that are ranked first by maximum number of voters.*

$$W_i^{\pi(\boldsymbol{P})} = \left\{ \begin{array}{cc} P_2 & P_3 \\ \hline c & c \\ a & a \\ b & b \end{array}, \begin{array}{cc} P_2 & P_3 \\ \hline c & c \\ b & b \\ a & a \end{array}, \begin{array}{cc} P_2 & P_3 \\ \hline c & c \\ a & b \\ b & a \end{array}, \begin{array}{cc} P_2 & P_3 \\ \hline c & c \\ b & a \\ a & b \end{array}, \right\} \bigcup \tag{2.4}$$

$$\bigcup \left\{ \begin{array}{cc} P_2 & P_3 \\ \hline b & c \\ c & a \\ a & b \end{array}, \begin{array}{cc} P_2 & P_3 \\ \hline c & b \\ a & c \\ b & a \end{array}, \begin{array}{cc} P_2 & P_3 \\ \hline b & c \\ c & b \\ a & a \end{array}, \begin{array}{cc} P_2 & P_3 \\ \hline c & b \\ b & c \\ a & a \end{array}, \begin{array}{cc} P_2 & P_3 \\ \hline b & c \\ a & a \\ c & b \end{array}, \begin{array}{cc} P_2 & P_3 \\ \hline c & b \\ a & a \\ b & c \end{array}, \begin{array}{cc} P_2 & P_3 \\ \hline b & c \\ a & b \\ c & a \end{array}, \begin{array}{cc} P_2 & P_3 \\ \hline c & b \\ b & a \\ a & c \end{array} \right\} \tag{2.5}$$

For all $\boldsymbol{P'}_{-1} \in W_1^{\pi(\boldsymbol{P})}$ when added to P_1 the winner after tie-breaking is the same, c. There are two groups of profiles in $W_1^{\pi(\boldsymbol{P})}$: 1) profiles that lead to the unique winner c and 2) profiles that lead to a tie $C(P_1, \boldsymbol{P'}_{-1}) = \{a,b,c\}$. If voter 1 changes her preference to $b\tilde{P}_1a\tilde{P}_1c$, then with profiles of the first group *voter 1 cannot change anything, but with profiles of the second group the result becomes b, which is more preferable for voter 1 than c. Thus, in every preference profile* $(P_1, \boldsymbol{P}'_{-1})$, *such that* $\boldsymbol{P}'_{-1} \in W_1^{\pi(\boldsymbol{P})}$ *voter 1 has an incentive to 1Winner-manipulate.*

Definition 2.2. *A rule* F *is called susceptible to individual* π*-manipulation if there exists a profile* $\mathbf{P} \in L(X)^N$ *and a voter* $i \in N$ *who has an incentive to* π*-manipulate in* $\mathbf{P}$ *under* F*. If a rule* F *is not susceptible to individual* π*-manipulation, it is immune to individual* π*-manipulation.*

Let $I_{ind}(m,n,\pi,F)$ be the probability that in a preference profile, randomly chosen from $L(X)^N$ there exists at least one voter who has an incentive to π-manipulate under F.

2.2.4 Coalitional manipulation

We assume that voters form a manipulating coalition if they have identical preferences. A coalition of voter i is denoted by K and it consists of all voters having the same preference as voter i. However, π is the only information available to voters, each voter does not know exactly who is in her coalition. In each preference profile $\mathbf{P}'$ of voter i's information set there is a set K of her coalition members (allies).

Then, a voter is willing to manipulate within a coalition when there is a strategy $\tilde{P}$ (insincere preference), such that the voting result is not less preferable in all profiles and is more preferable in at least one profile of her information set assuming that all members of her coalition vote $\tilde{P}$ (denoted by $\tilde{\mathbf{P}}'_K$) in each possible preference profile $\mathbf{P}'$. More formally,

Definition 2.3. *Given a rule* F *and a preference profile* $\mathbf{P}$, *we say, that voter* i *has an incentive to* π*-manipulate within a coalition*[1]*, if there exists* $\tilde{P} \in L(X)$

[1]The definition of coalitional manipulation differs from a standard one due to simplification we made: voters have identical preferences and manipulate in the same way. In a general framework, voters may have different preferences and manipulate differently.

s.t.
i) there is no $\mathbf{P}' = (P_i, P'_{-i})$, $P'_{-i} \in W_i^{\pi(\mathbf{P})}$ *s.t.* $F(\mathbf{P})\, P_i\, F(\tilde{\mathbf{P}}'_K, \mathbf{P}'_{-K})$*;*
ii) there exists $\mathbf{P}' = (P_i, P'_{-i})$, $P'_{-i} \in W_i^{\pi(\mathbf{P})}$ *s.t.* $F(\tilde{\mathbf{P}}'_K, \mathbf{P}'_{-K})\, P_i\, F(\mathbf{P})$*, where* $K = \{j \in N : P'_j = P_i\}$ *and* $\tilde{\mathbf{P}}'_K$ *is a preference profile of a coalition* K*, s.t. for all* $k \in K$ $P_k = \tilde{P}$.

If voter i has an incentive to π-manipulate within a coalition, then we similarly say that the whole coalition has an incentive to π-manipulate.

Definition 2.4. *A rule* F *is called susceptible to coalitional* π*-manipulation if there exists a profile* $\mathbf{P} \in L(X)^N$ *and a voter* $i \in N$ *who has an incentive to* π*-manipulate within a coalition in* $\mathbf{P}$*. If a rule* F *is not susceptible to coalitional* π*-manipulation, it is immune to coalitional* π*-manipulation.*

Denote by $I_{coal}(m,n,\pi,F)$ the probability that in a preference profile, randomly chosen from $L(X)^N$ there exists at least one voter who has an incentive to π-manipulate within a coalition under F.

If in a preference profile there exists a voter having an incentive to π-manipulate (individually or within a coalition), then it is also called *π-manipulable* (or manipulable under π). Otherwise, a preference profile is called *non-manipulable* under π.

2.2.5 Social choice correspondences

Here we give the definition of social choice correspondences that we focus on in this chapter. For each of them we need to specify how a ranking R and scores are computed to use them in *Rank*-PIF and *Score*-PIF. For each rule, $c \in C(\mathbf{P})$ iff there is no a such that aRc, i.e. $C(\mathbf{P})$ consists of all undominated alternatives in R.

- *Scoring rules.* A scoring rule is defined by a scoring vector $s = (s_1, ..., s_m)$, where s_j denotes the score assigned to an alternative for the j-th position in individual preferences. The total score of each alternative $a_j \in X$ is calculated as $S(a_j,\mathbf{P}) = \sum_{h=1}^{m} s_h \cdot v_h(a_j,\mathbf{P})$. Then, R is defined as follows: $\forall a_k, a_l \in X$ $a_k R a_l \Leftrightarrow S(a_k,\mathbf{P}) > S(a_l,\mathbf{P})$.

- *Plurality*: $s_{Pl} = (1,0,...,0)$.
- *Veto* (Antiplurality): $s_V = (1,...,1,0)$.
- *Borda*: $s_B = (m-1, m-2, ..., 1, 0)$.

- *Run-off procedure*. It has two stages:
1) The plurality score is calculated for each alternative. A first-stage vector of scores

$$S^1(\mathbf{P}) = (S^1(a_1,\mathbf{P}),...,S^1(a_m,\mathbf{P})),$$

where $S^1(a_j,\mathbf{P}) = \langle s_{Pl}, v(a_j,\mathbf{P})\rangle$. If $\exists\, a_k \in X$ s.t. $S^1(a_k) > n/2$, then social ordering is $a_k R a_j$ $\forall a_j \in X \setminus \{a_k\}$ and procedure terminates. Otherwise, procedure moves on to the stage two.
2) Two alternatives with maximal number of scores are chosen:
$a_k = argmax_{a_j \in X}(S^1(a_j,\mathbf{P}))$, $a_l = argmax_{a_j \in X\setminus\{a_k\}}(S^1(a_j,\mathbf{P}))$.
If there are ties, they are broken according to the alphabetical tie-breaking rule T. Then a second-stage vector of scores is calculated: $S^2(\mathbf{P}) = (S^2(a_k,\mathbf{P}), S^2(a_l,\mathbf{P}))$, where

$$S^2(a_k,\mathbf{P}) = \langle s_{Pl}, v(a_k, \mathbf{P}/\{a_k,a_l\})\rangle,$$

$$S^2(a_l,\mathbf{P}) = \langle s_{Pl}, v(a_l, \mathbf{P}/\{a_k,a_l\})\rangle.$$

In a social ordering an alternative with a higher score is considered better, $a_k R a_l$ if $S^2(a_k,\mathbf{P}) > S^2(a_l,\mathbf{P})$ and $a_l R a_k$ if $S^2(a_l,\mathbf{P}) > S^2(a_k,\mathbf{P})$. Both of them are better than all other alternatives, $\forall a_j \in X \setminus \{a_k, a_l\}$ $a_l R a_j, a_k R a_j$. The output of Score-PIF is $S(\mathbf{P}) = (S^1(\mathbf{P}), S^2(\mathbf{P}))$.

- *Single Transferable vote (STV)*. This is a multi-stage procedure, which we define in an iterative form.
0) $t := 1$, $X^t := X$, $\mathbf{P}^t := \mathbf{P}$.
1) $\forall a_j \in X^t$ $S^t(a_j,\mathbf{P}) := \langle s_{Pl}, v(a_j,\mathbf{P}^t)\rangle$.
2) If $\exists a_j \in X^t$ s.t. $S^t(a_j,\mathbf{P}) > n/2$, then $\forall a_h \in X^t \setminus \{a_j\}$ $a_j R a_h$, the

procedure terminates. Else $A := argmin_{a \in X^t}(S^t(a, \mathbf{P}))$.
3) If $A = X^t$, then the procedure terminates. Otherwise, alternatives of A are eliminated, $t := t+1$, $X^t := X^{t-1} \setminus A$, $\mathbf{P}^t := \mathbf{P}/X^t$; for all $x \in X^t$ and $a \in A$ it holds xRa; go to step 1. The output of Score-PIF is a vector of vectors $S(\mathbf{P}) = (S^1(\mathbf{P}), ..., S^{t*}(\mathbf{P}))$, where $t*$ is the number of cycles done by procedure.

- *Copeland*. A majority graph is computed. Then scores of alternatives are computed as follows

$$S(a_k, \mathbf{P}) = \sum_{l=1}^{m} MG(\mathbf{P})_{kl}.$$

A social ordering R is defined as usual: $\forall a_k, a_l \in X$ $a_k R a_l \Leftrightarrow S(a_k, \mathbf{P}) > S(a_l, \mathbf{P})$.

2.3 Theoretical results

In this section we prove some statements about the probability of individual and coalitional manipulation under incomplete information for any number of voters and alternatives. Before proving theorems, let us introduce some notations and consider an auxiliary statement, Lemma 1. Let d denote the number of preference profiles for n voters and m alternatives, i.e. $d = (m!)^n$, and $d_F(a)$ is the number of profiles in $L(X)^N$ where alternative a wins according to a rule F. Further, let $z(a)$ be the number of preference profiles in $L(X)^N$ where no voter has alternative a on the last position in a preference order and let $z_F(a)$ denote the number of preference profiles where alternative a wins according to a rule F and does not take the last position in any preference order.

Lemma 2.1. *For any alternative a and any rule F, s.t. for every $x \in X$ $\lim_{n\to\infty} d_F(x)/d = 1/m$, $\lim_{n\to\infty} z_F(a)/d_F(a) = 0$.*

Proof. The total number of preference profiles for n voters and m alternatives is $d = (m!)^n$. The number of preference profiles where no voter has alternative

a on the last place in a preference order does not depend on a rule or alternative and equals $z(a) = (m! - (m-1)!)^n$. Thus, the share of preference profiles where no voter has alternative a on the last place in a preference order:

$$\frac{z(a)}{d} = \frac{(m! - (m-1)!)^n}{(m!)^n} = \left(\frac{m! - (m-1)!}{m!}\right)^n = \left(1 - \frac{1}{m}\right)^n. \tag{2.6}$$

$$\frac{z(a)}{d_F(a)} = \frac{z(a)}{d} \cdot \frac{d}{d_F(a)}. \tag{2.7}$$

Since the rule is such that for any $x \in X$ $\lim_{n\to\infty} d_F(x)/d = 1/m$, $\lim_{n\to\infty} d/d_F(x) = m$. Using this and equation (2.5), we have

$$\lim_{n\to\infty} \frac{z(a)}{d_F(a)} = 0. \tag{2.8}$$

As $z_F(a)/d_F(a) < z(a)/d_F(a)$, $z_F(a)/d_F(a)$ also tends to 0 as n goes to infinity. □

To introduce the following lemma, we need two new notations. Let $g(a,b)$ denote the number of preference profiles with no voters having a on the last position and any alternative but b on the first position. And let $g_F(a,b)$ denote the number of preference profiles with the same property and where a wins according to F.

Lemma 2.2. *For any two alternatives a and b and any rule F, s.t. for every $x \in X$ $\lim_{n\to\infty} d_F(x)/d = 1/m$, $\lim_{n\to\infty} g_F(a,b)/d_F(a) = 0$.*

Proof. Let us denote the set of preference orders with a on the last position by A, and the set of preference orders with b on the first position by B. Then, $|A| = (m-1)!$ and $|AB| = (m-2)!$. Thus, $|A \cap \bar{B}| = (m-1)! - (m-2)! = (m-2)!\,(m-2)$. Finally, $|\overline{A \cap \bar{B}}| = |\bar{A} \cup B| = m! - (m-2)!\,(m-2)$. The number of preference profiles of n voters with preferences only of $\bar{A} \cup B$, i.e. $g(a,b)$, is $(m! - (m-2)!\,(m-2))^n$.

$$\frac{g(a,b)}{d} = \frac{(m! - (m-2)!\,(m-2))^n}{(m!)^n} = \left(1 - \frac{m-2}{m(m-1)}\right)^n. \qquad (2.9)$$

$$\frac{g(a,b)}{d_F(a)} = \frac{g(a,b)}{d} \cdot \frac{d}{d_F(a)}. \qquad (2.10)$$

Since for any $x \in X$ $\lim_{n\to\infty} d_F(x)/d = 1/m$, $\lim_{n\to\infty} d/d_F(x) = m$. Thus,

$$\lim_{n\to\infty} \frac{g(a,b)}{d_F(a)} = 0. \qquad (2.11)$$

As $g_F(a,b)/d_F(a) < g(a,b)/d_F(a)$, $g_F(a,b)/d_F(a)$ also tends to 0 as n goes to infinity. □

Now let us introduce some simplifying notations. Let $S(a)$ be the initial scores of a, i.e. $S(a,\mathbf{P})$, and $\tilde{S}(a)$ be the scores of a after manipulation of an individual or a group (depending on the context), i.e. $S(a,(\tilde{P}_i,\mathbf{P}_{-i}))$ or $S(a,(\tilde{\mathbf{P}}_K,\mathbf{P}_{-K}))$. The first result concerns individual manipulation under *Winner*-PIF for plurality rule.

Theorem 2.1. *For any $m \geq 3$ $\lim_{n\to\infty} I_{ind}(m,n,Winner,Plurality) = 1 - 1/m$ with alphabetic tie-breaking.*

Proof. (1) Let $X = \{a_1,a_2,...,a_m\}$ and $a_1 P_T a_2 P_T ... P_T a_m$. The PIF is $\pi(\mathbf{P}) = C(\mathbf{P})$. The result $C(\mathbf{P})$ could consist of one alternative, i.e. $C(\mathbf{P}) = \{a_k\}$, $k \in \{1,2,...,m\}$, or there can be a draw.

(2) First, consider the case $C(\mathbf{P}) = \{a_1\}$. Then $S(a_1) \geq S(a_j) + 1$ $\forall j \neq 1$. If any voter manipulates in favor of some other alternative a_h, it could not win, since $\tilde{S}(a_1) \geq \tilde{S}(a_h)$ and $a_1 P_T a_h$. Thus, in case of a tie, $\tilde{S}(a_1) = \tilde{S}(a_h)$, a_1 wins. So, all profiles with the unique winner a_1 are not susceptible to individual manipulation under $\pi(\mathbf{P}) = C(\mathbf{P})$.

(3) If $C(\mathbf{P}) = \{a_k\}$, $k \in \{3,4,...,m\}$, then $S(a_k) \geq S(a_j) + 1 \; \forall j \neq k$. Suppose there is a voter i who thinks that a_k is the worst alternative. Let a_h be the best alternative for voter i among alternatives that tie-break against a_k. Now prove that this voter has an incentive to *Winner*-manipulate in favor of a_h. Take $P_i = (a_l,...,a_h,...,a_k)$, where $a_h \neq a_l$, and $\tilde{P}_i = (a_h,...,a_l,...,a_k)$. Then, $\tilde{S}(a_h) = S(a_h)+1$, and $\tilde{S}(a_l) = S(a_l)-1$. Thus, $\tilde{S}(a_k) \geq \tilde{S}(a_h)$ and $F(\tilde{P}_i, \mathbf{P}'_{-i}) = a_k$ or $F(\tilde{P}_i, \mathbf{P}'_{-i}) = a_h$. So, condition i) from Definition 2.1 is satisfied.

To prove that condition ii) is also satisfied, we construct a preference profile $\mathbf{P}'_{-i} \in W_i^{\pi(\mathbf{P})}$ for $n \geq 6$, such that $F(\tilde{P}_i, \mathbf{P}'_{-i}) \, P_i \, F(\mathbf{P})$, i.e. $F(\tilde{P}_i, \mathbf{P}'_{-i}) = a_h$. Let $S(a_k,(P_i,\mathbf{P}'_{-i})) = \lfloor n/2 \rfloor$, $S(a_h,(P_i,\mathbf{P}'_{-i})) = \lfloor n/2 \rfloor - 1$, and $S(a_l,(P_i,\mathbf{P}'_{-i})) = n - 2\lfloor n/2 \rfloor + 1$, where $\lfloor x \rfloor$ is an integer part of x. Thus, $F(P_i, \mathbf{P}'_{-i}) = a_k$ and $F(\tilde{P}_i, \mathbf{P}'_{-i}) = a_h$.

So, if $C(\mathbf{P}) = \{a_k\}$, $k \in \{3,4,...,m\}$, $n \geq 6$ and there is at least one voter that has a_k one the lowest position in a preference order, then a preference profile $\mathbf{P}$ is susceptible to individual manipulation under $\pi(\mathbf{P}) = C(\mathbf{P})$.

(4) If $C(\mathbf{P}) = \{a_2\}$, then $S(a_2) \geq S(a_j) + 1 \; \forall j \neq k$ and we prove that a voter with preferences $P_i = (a_l,...,a_h,...,a_2)$ has an incentive to *Winner*-manipulate in favor of a_h in the same way as in (3), but with the only difference that $a_l \neq a_1$. If a voter having a_k on the last position also has a_1 on the first position, she does not have an incentive to *Winner*-manipulate, since the only alternative that tie-breaks against a_2 is a_1.

(5) The proportion of profiles with a single-valued outcome for plurality rule tends to 1 as n goes to infinity.[2] Since the rule is neutral (it means, it treats all the alternatives equally), the chance of winning for each of them tends to $1/m$.

(a) As we derived earlier, when the winner is $F(\mathbf{P}) = a_k$, $k \in \{3,4,...,m\}$, then manipulation is possible in profiles with at least one voter having a_k on the last place. The number of such profiles is $d_F(a_3) - z_F(a_3) + ... + d_F(a_m) - z_F(a_m)$.

[2] We refer to Gehrlein and Fishburn (1981). It is shown that the probability of a tie between any pair of alternatives with plurality rule tends to 0 as the number of voters goes to infinity (by Central Limit Theorem).

(b) If $F(\mathbf{P}) = a_2$, then manipulation is possible in profiles with at least one voter having a_2 on the last place and any alternative but a_1 on the first place. The number of such profile is $d_F(a_2) - g_F(a,b)$.

(c) If $F(\mathbf{P}) = a_1$, then individual manipulation is impossible.

Summing up, the number of preference profiles manipulable under *Winner*-PIF is not less than $d_F(a_2) - g_F(a,b) + d_F(a_3) - z_F(a_3) + ... + d_F(a_m) - z_F(a_m)$ and not greater than $d_F(a_2) + ... + d_F(a_m)$. By Lemma 2.1, for all $a \in X$, $\lim_{n\to\infty}(d_F(a) - z_F(a))/d_F(a) = 1$ and by Lemma 2.2, $\lim_{n\to\infty}(d_F(a_2) - g_F(a_2,a_1))/d_F(a_2) = 1$. Therefore, $\lim_{n\to\infty}(d_F(a_2) - g_F(a_2,a_1) + d_F(a_3) - z_F(a_3) + ... + d_F(a_m) - z_F(a_m))/d = 1 - 1/m$ and $\lim_{n\to\infty} I_{ind}(m,n,Winner,Plurality) = 1 - 1/m$.

$\square$

Thus, with infinite n only $1/m$ of profiles will be non-manipulable under *Winner*-PIF. In (Veselova, 2020) there is an asymptotic result for $I_{ind}(m,n,\mathit{1Winner},Plurality)$ which tends to 1 with n going to infinity. If a voter manipulates within a coalition under *Winner*-PIF, then again asymptotic probability equals 1.

Theorem 2.2. *For any* $m \geq 3$ $\lim_{n\to\infty} I_{coal}(m,n,Winner,Plurality) = 1$ *with alphabetic tie-breaking.*

Proof. Let $X = \{a_1, a_2, ..., a_m\}$ and $a_1 P_T a_2 P_T ... P_T a_m$. The PIF is $\pi(\mathbf{P}) = C(\mathbf{P})$.

Let us prove that all preference profiles with a single winner and a voter having the winning alternative on the last position in a preference order are *Winner*-manipulable within a coalition. If $C(\mathbf{P}) = \{a_k\}$, $k \in \{1,2,...,m\}$, then $S(a_k) \geq S(a_j) + 1\ \forall j \neq k$.

Consider voter i with preference $P_i = (a_l, a_h, ..., a_k)$. Let us prove that this voter has an incentive to *Winner*-manipulate within a coalition. Take $\tilde{P} = (a_h, a_l, ..., a_k)$, where only two alternatives, a_l and a_h are switched comparing to P_i. Thus, $\tilde{S}(a_k) = S(a_k)$, $\tilde{S}(a_l) = S(a_l) - |K|$, and $\tilde{S}(a_h) = S(a_h) + |K|$.

If $S(a_k) - S(a_h) > |K|$, then $F(\tilde{\mathbf{P}}'_K, \mathbf{P}'_{-K}) = a_k$. If $S(a_k) - S(a_h) < |K|$, then $F(\tilde{\mathbf{P}}'_K, \mathbf{P}'_{-K}) = a_h$. If $S(a_k) - S(a_h) = |K|$, then either $F(\tilde{\mathbf{P}}'_K, \mathbf{P}'_{-K}) = a_h$ or $F(\mathbf{P}) = a_k$ depending on a tie-breaking order. Thus, condition i) from Definition 2.3 is satisfied.

Let us show that condition ii) is also satisfied. We need to prove that there exists $\mathbf{P}' = (P_i, P'_{-i})$, $P'_{-i} \in W_i^{\pi(\mathbf{P})}$ s.t. $F(\tilde{\mathbf{P}}'_K, \mathbf{P}'_{-K}) P_i F(\mathbf{P})$, where $K = \{j \in N : P'_j = P_i\}$ and $\tilde{\mathbf{P}}'_K$ is a preference profile of a coalition K, s.t. for all $k \in K$ $P_k = \tilde{P}$. Let us construct $P'_{-i} \in W_i^{\pi(\mathbf{P})}$ for $n \geq 10$, such that $S(a_k, (P_i, P'_{-i})) = \lfloor n/2 \rfloor - 1$, $S(a_l, (P_i, P'_{-i})) = \lfloor n/2 \rfloor - 2$, and $S(a_h, (P_i, P'_{-i})) = |K| = 2\lfloor n/2 \rfloor + 3$. The result after manipulation of the coalition K is $F(\tilde{\mathbf{P}}'_K, \mathbf{P}'_{-K}) = a_h$. Thus, for $n \geq 10$ such profile exists.

As a consequence of Lemma 1, the share of profiles with at least one voter having the winning alternative on the last place in a preference order tends to 1. Furthermore, the share of profiles that result in a tie tends to zero. Thus, $\lim_{n \to \infty} I_{coal}(m, n, Winner, Plurality) = 1$. □

Theorem 2.3. *For any $m \geq 3$ $\lim_{n \to \infty} I_{coal}(m, n, Winner, Borda) = 1$ with alphabetic tie-breaking.*

Proof. (1) Again, we prove that if the winner is unique, $C(\mathbf{P}) = \{a_k\}$, then a voter having a_k on the last place in a preference order has an incentive to *Winner*-manipulate within a coalition in Borda rule. Let us fix the tie-breaking order $a_1 P_T a_2 P_T ... P_T a_m$ and assume that voter i's preference is $P_i = (a_h, a_l, ..., a_k)$. If $C(\mathbf{P}) = \{a_k\}$, then for all $j \in \{1, 2, ..., m\}$, $j \neq k$, $S(a_k) \geq S(a_j) + 1$.

(2) Consider a preference order $\tilde{P} = (a_l, a_h, ..., a_k)$ (switch the best and the second-best alternatives in P_i). If a coalition K of voter i manipulates with $\tilde{P}$, then $\tilde{S}(a_h) = S(a_h) - |K|$, $\tilde{S}(a_l) = S(a_l) + |K|$, $\tilde{S}(a_j) = S(a_j)$ for all $a_j \in X \setminus \{a_h, a_l\}$. Preference profiles $\mathbf{P}' = (P_i, P'_{-i})$, where $P'_{-i} \in W_i^{\pi(\mathbf{P})}$, are divided into three groups. The first one consists of profiles $\mathbf{P}'$, s.t. $|K| > S(a_k) - S(a_l)$. For these profiles $F(\tilde{\mathbf{P}}'_K, \mathbf{P}'_{-K}) = a_l$. For the second group it holds $|K| < S(a_k) - S(a_l)$, so, $F(\tilde{\mathbf{P}}'_K, \mathbf{P}'_{-K}) = a_k$ in this case. In the third group there

are profiles where $|K|= S(a_k)-S(a_l)$, in this case $C(\tilde{\mathbf{P}}'_K, \mathbf{P}'_{-K}) = \{a_l, a_k\}$, so the result is either a_k or a_l depending on a tie-breaking. So, condition i) from Definition 2.3 is satisfied.

(3) Let us show that condition ii) is also satisfied by constructing a preference profile $\mathbf{P}' = (P_i, P'_{-i})$, $P'_{-i} \in W_i^{\pi(\mathbf{P})}$ from the first group. Consider the following preferences: $P_1 = (a_h, a_l, ..., a_k)$ (preference of voter i and her coalition members), $P_2 = (a_h, a_k, ..., a_l)$, $P_3 = (a_k, a_h, ..., a_l)$, $P_4 = (a_k, a_l, ..., a_h)$, $P_5 = (a_l, a_k, ..., a_h)$, all other alternatives of $X \setminus \{a_k, a_l, a_h\}$ can be distributed in any way in P_1, P_2, P_3, P_4, P_5.

(3.1) Then, for n being even we construct a preference profile $\mathbf{P}' = (2P_1, P_2, P_3, qP_4, qP_5)$, which means that a preference order P_1 is replicated twice in $\mathbf{P}'$, P_2 - once, etc., and q is a positive integer. In this profile, $S(a_k) = q(2m-3)+2m-3$, $S(a_l) = q(2m-3)+2m-4$, $S(a_h) = 4m-5$, for all other alternatives $a_j \in X \setminus \{a_k, a_l, a_h\}$ it holds $S(a_j) \leq (m-3)(4+2q)$. Thus, $S(a_k)-S(a_l) = 1$, $|K|= 2$, and $S(a_k) > S(a_h)$. An inequality $S(a_k) > S(a_j)$ for all $a_j \in X \setminus \{a_k, a_l, a_h\}$ holds if $q > 2m/3-3$ and $S(a_k) > S(a_h)$ holds for $q > (2m-2)/(2m-3)$.

(3.2) For n being odd we construct a profile $\mathbf{P}' = (2P_1, 2P_2, (q+1)P_4, qP_5)$, where q is a positive integer. For this profile, $S(a_k) = q(2m-3)+3m-5$, $S(a_l) = q(2m-3)+3m-6$, $S(a_h) = 4m-4$, for all other alternatives $a_j \in X \setminus \{a_k, a_l, a_h\}$ it holds $S(a_j) \leq (m-3)(5+2q)$. Thus, $S(a_k)-S(a_l) = 1$, $|K|= 2$, and $S(a_k) > S(a_h)$. An inequality $S(a_k) > S(a_j)$ for all $a_j \in X \setminus \{a_k, a_l, a_h\}$ holds if $q > (2m-10)/3$ and $S(a_k) > S(a_h)$ holds for $q > (m+1)/(2m-3)$.

Thus, the condition ii) is satisfied for all $n > n^*$, where $n^* = 4+2q^*$, and q^* is the maximal number of $2m/3-3$, $(2m-2)/(2m-3)$, $(2m-10)/3$, and $(m+1)/(2m-3)$. Therefore, for all $n > n^*$, if in a preference profile there is a voter with a_k on the last place, this voter has an incentive to manipulate within a coalition in $\mathbf{P}$.

(4) Borda rule also satisfies the requirement of Lemma 1, i.e. for any $x \in X$ we have $\lim_{n\to\infty} d_F(x)/d = 1/m$ by neutrality and zero ties probability (Marchant, 2001)). Thus, the share of profiles with at least one voter having the winning alternative on the last place tends to 1 as n goes to

infinity. Since all these profiles are *Winner*-manipulable within a coalition, $\lim_{n\to\infty} I_{coal}(m,n,Winner,Borda) = 1$. □

The next theorem shows that the probability of manipulation for scoring rules is the same when we consider individual or coalitional manipulation under *1Winner*-PIF. Let us introduce some more notations for the proof.

For a scoring vector $s = (s_1, s_2, ..., s_m)$, a *jump* is a non-zero difference between two adjacent scoring values. If s has r jumps, then this means that there are distinct $k_1, \dots, k_r \in \{1, \dots, m-1\}$ such that $s_{k_1} > s_{k_1+1}, \dots, s_{k_r} > s_{k_r+1}$, while all other differences are zero. We use the notation $\Delta_j = s_{k_j} - s_{k_j+1}$ for $j = 1, \dots, r$ to denote the non-zero differences between scoring values.

Theorem 2.4. *For any number of voters n and any number of alternatives m* $I_{ind}(m,n,1Winner,F) = I_{coal}(m,n,1Winner,F)$ *for scoring rules.*

Proof. Let $X = \{a_1, ..., a_m\}$. Consider a scoring rule with a a scoring vector $s = (s_1, s_2, ..., s_m)$, the first jump in s goes after k_1, and a voter voter i with a preference $a_1 P_i a_2 P_i ... P_i a_m$. Let us consider two cases: $F(\mathbf{P}) \in \{a_1, a_2, ..., a_{k_1+1}\}$ and $F(\mathbf{P}) \in \{a_{k_1+2}, ..., a_m\}$.

1) We prove that if $F(\mathbf{P}) \in \{a_1, a_2, ..., a_{k_1+1}\}$, then voter i has no incentive to *1Winner*-manipulate under F individually and within a coalition.

1.1) First, if $F(\mathbf{P}) = a_1$, then voter i has no incentive to *1Winner*-manipulate since there is no alternative better than a_1 and condition ii) of Definitions 2.1 and 2.3 cannot be satisfied.

1.2) Suppose that $F(\mathbf{P}) = b$, $b \in \{a_2, a_3, ..., a_{k_1+1}\}$ and i manipulates in favor of some a, s.t. aPb. If i puts alternative a higher (if a is not a_1), then nothing changes for a since $s_1 = ... = s_{k_1}$ and again condition ii) of Definition 2.1 is violated. Thus, i could only put b lower in $\tilde{P}_i$, but then some alternative $c \in \{a_{k_1+2}, ..., a_m\}$ goes higher in a preference order. If there are no jumps in s after $k_1 + 1$, then nothing changes for b and c and condition ii) is violated. If there are other jumps after $k_1 + 1$, then c gets plus A scores. Since the only information is $F(\mathbf{P}) = b$, i.e. $S(b, \mathbf{P}) \geq S(x, \mathbf{P})\ \forall x \in X$, there exists $\mathbf{P}'_{-i} \in W_i^{\pi(\mathbf{P})}$, s.t. $S(c,(\tilde{P}_i,\mathbf{P}'_{-i})) = S(c,\mathbf{P}') + A > S(x,(\tilde{P}_i,\mathbf{P}'_{-i}))$ $\forall x \in X \setminus \{c\}$. The same is true for manipulation within a coalition. If for some $\mathbf{P}' \in W_i^{\pi(\mathbf{P})}$ holds $S(c,(\tilde{P}_i,\mathbf{P}'_{-i})) = S(c,\mathbf{P}') + A > S(x,(\tilde{P}_i,\mathbf{P}'_{-i}))$ $\forall x \in X \setminus \{c\}$, then $S(c,(\tilde{\mathbf{P}}_K,\mathbf{P}'_{-K})) = S(c,\mathbf{P}') + |K|A > S(x,(\tilde{\mathbf{P}}_K,\mathbf{P}'_{-K}))$ $\forall x \in X \setminus \{c\}$. Therefore, condition i) of Definitions 2.1 and 2.3 is not satisfied and i does not have an incentive to *1Winner*-manipulate under a scoring rule when $F(\mathbf{P}) \in \{a_1,a_2,...,a_{k_1+1}\}$ either individually or within a coalition.

2) Now suppose that $F(\mathbf{P}) = c$ and $c \in \{a_{k_1+2},...,a_m\}$. Voter i cannot give alternatives from $\{a_1,a_2,...,a_{k_1}\}$ more scores, but can increase the score of a_{k_1+1} by Δ_1. So, let $\tilde{P}_i$ be obtained from P_i but a_{k_1+1} and $a \in \{a_1,a_2,...,a_{k_1}\}$ are switched. Thus, $S(a_{k_1+1},(\tilde{P}_i,\mathbf{P}'_{-i})) = S(a_{k_1+1},\mathbf{P}') + \Delta_1$ and $S(a,(\tilde{P}_i,\mathbf{P}'_{-i})) = S(a,\mathbf{P}') - \Delta_1$ and $S(x,(\tilde{P}_i,\mathbf{P}'_{-i})) = S(x,\mathbf{P}')$ for all $x \in X \setminus \{a,a_{k_1+1}\}$. So, either $S(a_{k_1+1},(\tilde{P}_i,\mathbf{P}'_{-i})) > S(x,(\tilde{P}_i,\mathbf{P}'_{-i}))$ for all $x \in X \setminus \{a_{k_1+1}\}$ and a_{k_1+1} wins or $S(c,(\tilde{P}_i,\mathbf{P}'_{-i})) > S(a_{k_1+1},(\tilde{P}_i,\mathbf{P}'_{-i}))$ and c wins. In case of a tie the result is again either a_{k_1+1} or c depending on a tie-breaking order. The same holds for coalitional manipulation, but $S(a_{k_1+1},(\tilde{P}_K,\mathbf{P}'_{-K})) = S(a_{k_1+1},\mathbf{P}') + |K|\Delta_1$ and $S(a,(\tilde{P}_K,\mathbf{P}'_{-K})) = S(a,\mathbf{P}') - |K|\Delta_1$. Thus, for all $\mathbf{P}' \in W_i^{\pi(\mathbf{P})}$ the result is not worse then $F(\mathbf{P})$ after manipulation of i or K and better for some $\mathbf{P}' \in W_i^{\pi(\mathbf{P})}$. Therefore, if $F(\mathbf{P}) \in \{a_{k_1+2},...,a_m\}$, voters with a preference $a_1 P_i a_2 P_i ... P_i a_m$ have an incentive to *1Winner*-manipulate both individually and within a coalition under F.

3) Thus, for any voter having an incentive to manipulate individually there is also an incentive to manipulate within a coalition under *1Winner*-PIF. At the same time, if a voter does not have an incentive to manipulate individually under *1Winner*-PIF, then there she has no incentive to manipulate within a coalition. It means that the set of individually manipulable profiles is the same as the set of profiles manipulable within a coalition. So, for scoring rules $I_{ind}(m,n,\mathit{1Winner},F) = I_{coal}(m,n,\mathit{1Winner},F)$. $\square$

As shown by Veselova (2020), the probability of manipulation for plurality rule under *1Winner*-PIF tends to 1. By Theorem 4, $I_{coal}(m,n,\mathit{1Winner},\mathit{Plurality})$ also tends to 1. On the other hand, it was proved that $I_{ind}(m,n,\mathit{1Winner},\mathit{Veto}) = 0$ by Reijngoud and Endriss (2012), and by Theorem 4 $I_{coal}(m,n,\mathit{1Winner},\mathit{Veto}) = 0$.

2.4 Computational experiments

This Section shows computed values of $I_{coal}(m,n,\pi,F)$ for all PIFs from Section 2.2 and all rules listed in Section 2.5. Moreover, we compare these values with $I_{ind}(m,n,\pi,F)$ computed in the work of Veselova (2020). We consider $m = 3$ and n from 3 to 15. All computations were done in MatLab (a code of the main program can be seen in Appendix A) . Results are represented in Figures 1.1-1.12.

We make the following observations.

- Except for veto rule and only one case with Borda rule, coalitional manipulability is not less than individual. Particularly, we observe a clear going-to-1 tendency not only for *1Winner*-PIF, but also for *Winner*-PIF (all rules except for veto) and *Rank*-PIF in some cases (plurality, Borda, runoff, STV).
- In all cases with non-zero individual manipulability of veto rule the values of coalitional manipulability are strictly lower than individual.
- Individual and coalitional manipulability under *1Winner*-PIF coincide not only for scoring rules, but for all rules under consideration. Moreover, for runoff and Copeland rule these values coincide under *Winner*-PIF.
- The observation 'less information - equal or higher manipulability' is still true in the coalition case for plurality rule, runoff, and STV. With little exceptions it holds for Copeland and with only one exception case for Borda rule. For veto rule the opposite is true: 'less information - equal or less manipulability'.
- A rule is called strongly computable from π-images if a voter knowing $\pi(\mathbf{P})$ can compute the result of the rule for any way of her misrepresenting preference. One of results of the work by Veselova (2020) is that

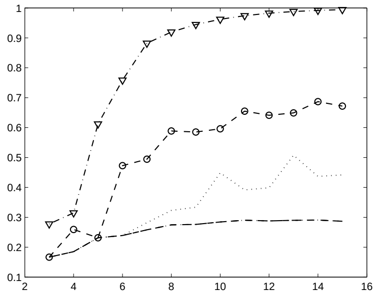


Figure 2.1: Plurality rule, individual

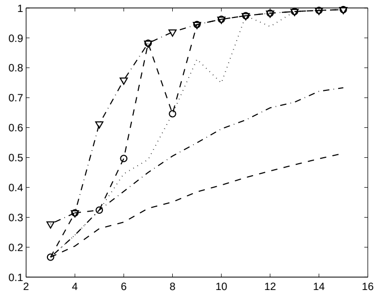


Figure 2.2: Plurality rule, coalitional

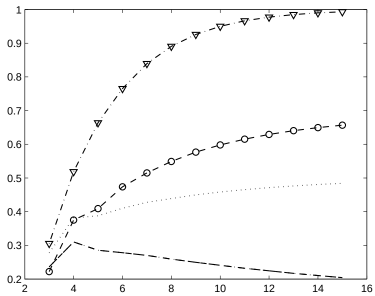


Figure 2.3: Borda rule, individual

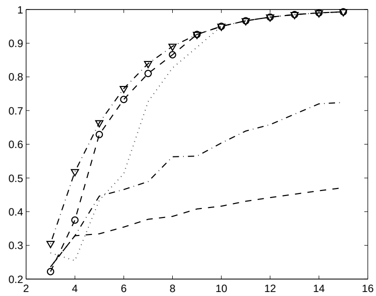
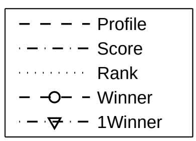


Figure 2.4: Borda rule, coalitional



individual manipulability under π does not change compared to a complete information case if the rule is strongly computable from π-images. The same does not hold for coalitional manipulation.

- With n growing coalitional manipulation of veto rule quickly becomes zero for any kind of incomplete information. It could be explained by the following argument. Manipulation in veto rule means switching the least preferred alternative and some other. The larger is the number of voters, the larger is the cardinality of the maximal possible coalition of voter i. The larger is the coalition, the bigger is the chance of making the least preferred alternative the winner by adding scores to it.

- Periodicity of manipulability index for Copeland rule is rather strong for *Winner*-PIF and *1Winner*-PIF, its amplitude is around 0.4-0.6. So, slight changes in the number of voters may lead to a considerable reduction in manipulation possibilities.

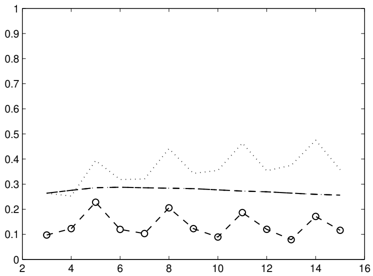


Figure 2.5: Veto rule, individual

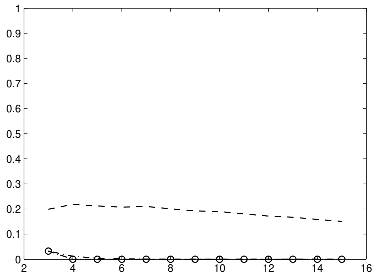


Figure 2.6: Veto rule, coalitional

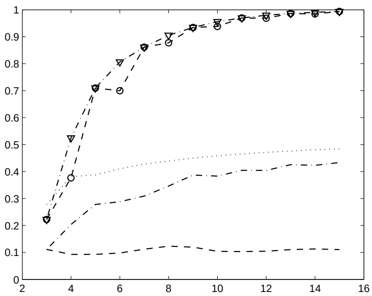


Figure 2.7: Runoff, individual

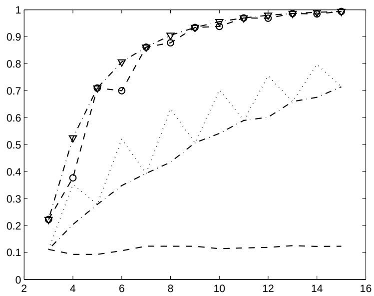
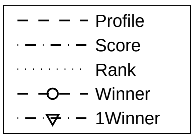


Figure 2.8: Runoff, coalitional



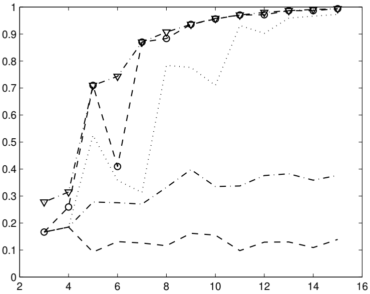


Figure 2.9: STV, individual

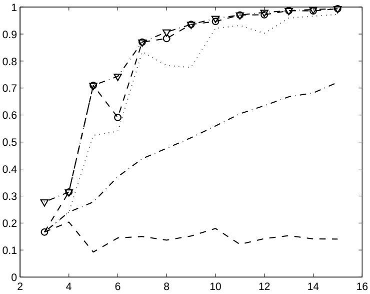


Figure 2.10: STV, coalitional

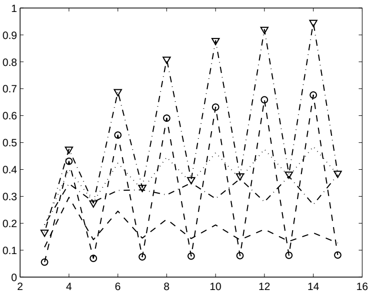


Figure 2.11: Copeland rule, individual

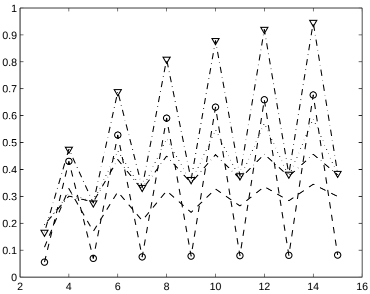
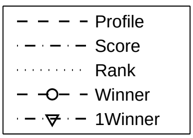


Figure 2.12: Copeland rule, coalitional



2.5 Conclusion

Studying individual manipulation is convenient for modeling and revealing incentives of separate voters. However, other voters might also take part in manipulation and this assumption can change voters' incentives. Every voter has a group of co-minded people and she can take them into account even if she does not know their exact number. For a single voter it is easier to predict actions of voters of her type. Having the same preference, they also have the same incentives. So, the aim of this work was to consider group actions of co-minded people in manipulation model under incomplete information and compare results with individual manipulation.

In the theoretical part, we considered asymptotic behavior of individual and coalitional manipulation probability for plurality rule and coalitional manipulation probability for Borda rule under *Winner*-PIF. Finally, we proved that individual and coalitional manipulation are equal for scoring rules under under *1Winner*-PIF. The computational part of the research illustrates theoretical findings for the case of 3 alternatives and, additionally, allows to observe the behavior of manipulation probabilities for other rules and PIFs.

This work is just the first attempt to combine informational aspect and manipulation by groups in one model. It sheds some light on the problem of their joint influence. Thus, we showed that incomplete information of the types that allow to compute the winner increases manipulability for plurality, Borda, runoff, STV, and Copeland rules. The effect of coalitional manipulation is the same. On the contrary, for veto rule manipulability decreases under incomplete information and considering also coalitional manipulation makes this effect stronger.

The question that we did not touch here is that some coalition members may decide not to manipulate and that is related to the question of the safety of coalitional manipulation, which the proceeding Chapter 3 is devoted to. Additionally, a more significant influence on incentives to manipulation may be expected from voters with a different preference, because they can also manipulate, counter-manipulate, etc. If we add such uncertainty about the actions of other manipulators and consider it together with incomplete information, results may be difficult to predict.

3
On the safety of group manipulation

Adapted from: H. Peters. and Y. Veselova *On the safety of group manipulation*. Social Choice and Welfare, 2023.

Obviously, groups of voters have even more opportunities to influence the result: they can unite in coalitions and coordinate their manipulation. In many applications, however, this coordination cannot and does not actually take place explicitly. Rather, a voter who aims to manipulate, may take into account that other voters with the same preference may also manipulate in the same way. In fact, this is what we assume in this paper. Given a (sincere) preference within a profile of preferences, we will use the word 'group' to indicate all voters who have this preference. We then say that a(ny) voter in this group has an incentive to manipulate if there is a(n insincere) preference such that, if all voters in this group report this preference, then the election result is better for them according to the true, sincere preference. In that case, a problem may arise if not all voters in the group participate in the manipulation, because if this happens the result may actually be worse than without manipulation. In other words, due to lack of or poor communication within a group of like-minded voters, manipulation may be harmful.

We will call manipulation 'safe' if this does not happen: even if not all voters in a group participate in the manipulation, the result is not worse than without manipulation. Otherwise, manipulation is 'unsafe'. We now provide an example of such an unsafe manipulation for the well-known Borda rule.

Example 3.1. Suppose that there are five alternatives, a,b,c,d,e. A preference profile with seven voters is given in the following table. The first line of the table shows the number of voters for each preference order occurring in the profile.

3	2	1	1
a	d	d	e
b	c	c	d
c	b	e	a
e	e	a	c
d	a	b	b

The Borda rule assigns 4 points to the top alternatives, 3 points to the second-ranked alternatives, etc., until 0 points to the last ranked alternatives, and these points are then added up to obtain the total scores. For the given preference profile, total scores are $S(a) = 15$, $S(b) = 13$, $S(c) = 16$, $S(d) = 15$, $S(e) = 11$. Thus, with sincere preferences, alternative c wins. For the group of three voters $K = \{1,2,3\}$ each one having preference (a,b,c,e,d) (i.e., a is preferred to b, b to c, etc.), there is no way to make a win, but they have an incentive to manipulate by reporting preference (b,a,e,c,d). If all voters in K report this preference, then the scores will be $S(a) = 12$, $S(b) = 16$, $S(c) = 13$, $S(d) = 15$, $S(e) = 14$, so that b is the winner of the election, and b is preferred over c by the members of K according to their sincere preference.

Now suppose that only one voter of K decides to manipulate. In this case, the scores are: $S(a) = 14$, $S(b) = 14$, $S(c) = 15$, $S(d) = 15$, $S(e) = 12$. Alternatives c and d have maximal scores. If we assume that d wins against c by tie-breaking, then the final outcome is d, but for the members of K outcome d is worse than c. Therefore, this group manipulation is unsafe. $\lhd$

If manipulation is unsafe, then this fact may prevent voters from voting strategically. However, the possibility of an unsafe manipulation depends on the rule, the number of voters and the number of alternatives. In this paper we consider a collection of well-known rules and investigate for which of these rules group manipulation can be unsafe, and which rules are only safely manipulable.

The concept of (un)safe manipulation has already been considered by Slinko and White (2014). However, their model differs from the one considered in this paper. In their approach, a voter i in a group K has an incentive to manipulate if there is some subset of K, including voter i, such that the election result improves for i if exactly the voters in this subset report a (the same) insincere preference. They call manipulation 'unsafe' if the result can get worse if some other subset, including i, deviates. The main result in Slinko and White (2014) is an extension of the Gibbard-Satterthwaite theorem: for each rule with at least three alternatives in its range, there is a preference profile and a voter who can safely individually manipulate – that is, this voter is not worse off if also some other voters with the same preference manipulate in the same way. We postpone a more elaborate comparison between our paper and Slinko and White (2014) until Section 3.6.1.

Following up on the model of Slinko and White (2014), several papers focus on different aspects of safe manipulation. Computational complexity of finding a safe strategic vote under k-approval and Bucklin rules was studied in Hazon and Elkind (2010). The same question for Borda rule and some classes of scoring rules was considered in Ianovski et al (2011). The asymptotic probability of a safe manipulation under the IAC assumption (all voting profiles are equally likely) for scoring rules is computed in Wilson and Reyhani Shokat Abad (2010). In an extension of the aforementioned model each manipulator thinks not only about his/her allies, but about all voters having an incentive to manipulate (they are called Gibbard-Satterthwaite-manipulators, or GS-manipulators). Then, a strategy is considered as 'safe' if for any manipulating subset of GS-manipulators, using this strategy is not worse than sincere voting. This kind of model was considered in Elkind et al (2015) and Grandi et al (2019). These references are just a few from the strand of literature on voting manipulation games. For a more detailed survey we refer the reader to Slinko (2019).

The rest of the paper is organized as follows. Section 3.1 presents the formal model and the rules that we consider: scoring rules, in particular Borda; run-off; Copeland; and single transferable vote. Section 3.2 considers scoring rules in general and Borda in particular, Section 3.3 considers the run-off rule, Section 3.4 the Copeland rule, and Section 3.5 single transferable vote. Section 3.6 concludes, in particular with a comparison between Slinko and White (2014) and our approach.

3.1 Definitions and notations

3.1.1 The framework

A society of $n \geq 3$ *voters*, $N = \{1, \dots, n\}$, decides which of m *alternatives* from the set X, $|X| = m \geq 3$, to choose.[1] Each voter has a *preference*, i.e., a linear order[2] on X. We denote the set of all preferences by $L(X)$. For $a, b \in X$

[1] The cases $n < 3$ or $m < 3$ are uninteresting for the purpose of this paper, as can easily be verified in the sequel.

[2] That is, an irreflexive, asymmetric, transitive and complete binary relation.

and $P \in L(X)$ we write aPb instead of $(a,b) \in P$. Also, we often write $P = (a_1,\ldots,a_m)$, meaning that $a_1P\ldots Pa_m$, where $X = \{a_1,\ldots,a_m\}$. A *preference profile* is a vector $\mathbf{P} = (P_1,\ldots,P_n) \in L(X)^N$ of individual preferences.

A *social choice correspondence* (SCC) is a map $C: L(X)^N \to 2^X \setminus \{\emptyset\}$ (where 2^X denotes the set of all subsets of X). A *social choice rule* or simply *rule* is a map $F: L(X)^N \to X$. Thus, a rule can be identified with a single-valued SCC. In this paper we will mainly consider social choice rules derived from social choice correspondences by tie-breaking according to a fixed linear order on X – we will be precise about this whenever this is needed.

For a preference profile $\mathbf{P}$, a preference $P \in L(X)$, and a subset $K \subseteq N$ such that $P_i = P \in L(X)$ for all $i \in K$, we also write $(P_K, \mathbf{P}_{-K})$ instead of $\mathbf{P}$. If, in particular, $K = \{k \in N : P_k = P\}$, then we call K the *group* of (any) voter $i \in K$ at $\mathbf{P}$. Thus, a group collects all voters with the same preference, for some preference in a preference profile.

The following definition captures the situation where a(ny) voter in a group prefers the alternative which results if all voters in that group vote insincerely using the same preference.

Definition 3.1. Voter $i \in N$ *has an incentive to manipulate* rule F at profile $\mathbf{P} \in L(X)^N$ if there is a $\tilde{P} \in L(X)$ such that $F(\tilde{P}_K, \mathbf{P}_{-K})\ P\ F(\mathbf{P})$, where K is the group of i at $\mathbf{P}$ (i.e., all voters who have common preference $P = P_i$ at $\mathbf{P}$).

Clearly, this definition implies that if voter i has an incentive to manipulate, then all members of i's group have an incentive to manipulate – using the same preference $\tilde{P}$. Therefore, we also say that group K *has an incentive to manipulate*.

We introduce some further terminology. A preference profile $\mathbf{P} \in L(X)^N$ is *manipulable under* rule F if there is a voter who has an incentive to manipulate at $\mathbf{P}$. A rule F is *manipulable* if there is a manipulable preference profile under F.

3.1.2 Safe and unsafe manipulations

Let F be a rule, and let $\mathbf{P} \in L(X)^N$ be a preference profile. Suppose that voter i belonging to group K has an incentive to manipulate F at $\mathbf{P}$ by preference $\tilde{P}$. We say that *manipulation with $\tilde{P}$ is unsafe for i at* $\mathbf{P}$ if there exists $M \subsetneq K$ such that $i \in M$ and $F(\mathbf{P})\, P_i\, F(\tilde{P}_M, \mathbf{P}_{-M})$. If such an M does not exist, then *manipulation with $\tilde{P}$ is safe for i at* $\mathbf{P}$. In words, a manipulation is safe if it never results in a worse alternative if not all members of the group join in the manipulation. Clearly, if $K = \{i\}$ then every manipulation is safe.

A preference profile $\mathbf{P} \in L(X)^N$ is *safely manipulable* (given F) if there is a voter for whom manipulation is safe with $\tilde{P}$ for some $\tilde{P} \in L(X)$. It is *unsafely manipulable* if there is a voter for whom manipulation with $\tilde{P}$ is unsafe for some $\tilde{P} \in L(X)$. A preference profile can be both safely and unsafely manipulable, even by the same voter.

The rule F is *safely manipulable* if there is a safely manipulable preference profile, and *unsafely manipulable* (UM) if there is an unsafely manipulable preference profile. Again, F can be both safely and unsafely manipulable. Rule F is *only safely manipulable* (OSM) if for every manipulable profile $\mathbf{P} \in L(X)^N$, $\mathbf{P}$ is not unsafely manipulable. Hence, F is OSM if it is not UM.

3.1.3 Social choice correspondences

In this subsection we introduce the social choice correspondences from which the rules to be studied in this paper, will be derived by tie-breaking.

Scoring correspondences

A *scoring vector* is a vector $s = (s_1, \ldots, s_m) \in \mathbb{R}^m$ such that $s_1 \geq \cdots \geq s_m \geq 0$ and $s_1 > s_m$. For a preference profile $\mathbf{P}$ and an alternative a, let $v_j(a, \mathbf{P})$ denote the number of voters having a at the j-th position (where voter i has a at the j-th position if $|\{b \in X : bP_ia\}| = j-1$). Then $S(a, \mathbf{P}) = \sum_{j=1}^m s_j v_j(a, \mathbf{P})$ is the total score of a at $\mathbf{P}$. The *scoring correspondence* F with scoring vector s assigns to each preference profile $\mathbf{P}$ the set $\{a \in X : S(a,\mathbf{P}) \geq S(a',\mathbf{P})$ for all $a' \in X\}$, i.e., the set of alternatives with maximal total score. Well-known examples are:

- *q-approval*: $s_1 = \cdots = s_q = 1$, $s_{q+1} = \cdots = s_m = 0$, where $q \in \{1,\ldots,m-1\}$; for $q = 1$ this is also called *plurality*, and for $q = m-1$ this is also called *veto* or *antiplurality*,
- *Borda*: $s = (m-1, m-2, \ldots, 1, 0)$.

Run-off

For a preference profile $\mathbf{P}$, two alternatives with maximal plurality scores (see Section 3.1.3) are chosen, if necessary using a tie-breaking rule. Among these two, say a and b, we choose the alternative(s) which win in a pairwise contest, that is, a is chosen if $|\{i \in N : aP_ib\}| \geq |\{i \in N : bP_ia\}|$ and b is chosen if $|\{i \in N : bP_ia\}| \geq |\{i \in N : aP_ib\}|$.

Copeland

For a preference profile $\mathbf{P}$, the *Copeland score* of an alternative a is the number

$$\left|\left\{b \in X : |\{i \in N : aP_ib\}| > \frac{n}{2}\right\}\right| - \left|\left\{b \in X : |\{i \in N : bP_ia\}| > \frac{n}{2}\right\}\right|.$$

Hence, the Copeland score of an alternative a is the number of alternatives beaten by a minus the number of alternatives that beat a, where x beats y if a strict majority of the voters prefers x over y. The Copeland correspondence chooses the alternatives with maximal Copeland score.

Single-transferable-vote, STV

For a preference profile $\mathbf{P}$, for each alternative a determine the number of voters who have a at top position, i.e., determine its plurality score $S(a,\mathbf{P})$ for scoring vector $(1,0,\ldots,0)$. If all nonzero plurality scores are equal, then STV assigns the set of all alternatives that occur at top, i.e., that have nonzero plurality score. If not all these nonzero plurality scores are equal then: if there is an alternative a with plurality score strictly higher than $n/2$, then STV assigns $\{a\}$; otherwise, leave out those alternatives that have minimal (possibly zero) plurality score. This results in a restricted preference profile with fewer alternatives. Now repeat this procedure until no more alternatives can be left out: STV assigns the remaining alternatives to $\mathbf{P}$. As an illustration, consider the following two profiles:

$$\begin{array}{ccccc} a & b & b & c & c \\ b & \cdot & \cdot & \cdot & \cdot \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{array} \quad \textit{and} \quad \begin{array}{cccccc} a & b & b & c & c & c \\ b & \cdot & \cdot & \cdot & \cdot & \cdot \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{array}$$

In the left profile, after eliminating alternatives with zero plurality score (if any), a is eliminated: this results in a profile where b has plurality score 3, so that STV assigns $\{b\}$. In the right profile, again after eliminating alternatives with zero plurality score (if any), also a is eliminated, so that STV assigns $\{b,c\}$.

3.2 (Un)safe manipulability of scoring rules

In this section we investigate the (un)safe manipulability of rules derived from scoring correspondences, so-called *scoring rules*. Our first main result concerns these rules in general, and our second result focuses on Borda (cf. Section 3.1.3).

As already mentioned, we need to apply tie-breaking in order to derive rules from social choice correspondences. We do this by fixing a linear order on the set of alternatives X and taking the maximal element according to this order from a set assigned by the correspondence. In what follows we will be more precise whenever this is needed.

For a scoring vector s, a *jump* is a non-zero difference between two adjacent scoring values. If s has r jumps, then this means that there are distinct $k_1,\ldots,k_r \in \{1,\ldots,m-1\}$ such that $s_{k_1} > s_{k_1+1},\ldots,s_{k_r} > s_{k_r+1}$, while all other differences are zero. We use the notation $\Delta_j = s_{k_j} - s_{k_j+1}$ for $j = 1,\ldots,r$ to denote the non-zero differences between scoring values.

Our first result concerns unsafe and only safe manipulability of scoring rules in general.

Theorem 3.1. *Let s be a scoring vector with r jumps, and let F be a scoring rule derived from the scoring correspondence associated with scoring vector s by tie-breaking. If $r = 1$, then F is only safely manipulable. If $r \geq 2$, then the results are as in the following table:*

<table>
<tr><td rowspan="3"></td><td colspan="3">2 jumps</td><td colspan="2" rowspan="2">3 or more jumps</td></tr>
<tr><td rowspan="2">$\Delta_1 > \Delta_2$</td><td colspan="2">$\Delta_1 \leq \Delta_2$</td></tr>
<tr><td>$k_1 = 2$,
$k_2 = 4$</td><td>otherwise</td><td>$\Delta_1 > \Delta_3$
or $\Delta_2 > \Delta_3$</td><td>otherwise</td></tr>
<tr><td>$m = 3$</td><td>$\forall n$: OSM</td><td colspan="2">$\forall n$: OSM</td><td colspan="2">(not applicable)</td></tr>
<tr><td>$m = 4$</td><td>$\exists n$: UM</td><td colspan="2">$\forall n$: OSM</td><td>$\exists n$: UM</td><td>$\forall n$: OSM</td></tr>
<tr><td>$m = 5$</td><td>$\exists n$: UM</td><td>$\forall n$: OSM</td><td>$\exists n$: UM</td><td colspan="2">$\exists n$: UM</td></tr>
<tr><td>$m \geq 6$</td><td>$\exists n$: UM</td><td colspan="2">$\exists n$: UM</td><td colspan="2">$\exists n$: UM</td></tr>
</table>

Table 1

Proof. (i) In this first part of the proof, we assume that there is an unsafe manipulation and derive conditions implied by this assumption. Let $\mathbf{P}$ be a preference profile and let $a,b,c \in X$ such that for voter i in group K (and, consequently, for all voters in K) we have aP_ibP_ic. Suppose that group K has an incentive to manipulate and manipulation is unsafe with $F(\mathbf{P}) = b$, $F(\tilde{P}_K,\mathbf{P}_{-K}) = a$, and $F(\tilde{P}_M,\mathbf{P}_{-M}) = c$ for some $\tilde{P} \in L(X)$ and $M \subsetneq K$. In words, b is the alternative chosen at $\mathbf{P}$, group K can achieve a by manipulating via $\tilde{P}$, but if only the voters in M deviate, the worse alternative c results.

For every alternative $x \in X$, let ε_x denote the change in score when a voter $i \in K$ changes from P_i to $\tilde{P}$, hence $\varepsilon_x |G| = S(x,(\tilde{P}_G,\mathbf{P}_{-G})) - S(x,\mathbf{P})$ for every $G \subseteq K$. Since $F(\mathbf{P}) = b$ we have

$$S(b,\mathbf{P}) \geq S(c,\mathbf{P}), \tag{3.1}$$

and since $F(\tilde{P}_M,\mathbf{P}_{-M}) = c$ we have $S(c,(\tilde{P}_M,\mathbf{P}_{-M})) \geq S(b,(\tilde{P}_M,\mathbf{P}_{-M}))$, hence

$$S(c,\mathbf{P}) + \varepsilon_c |M| \geq S(b,\mathbf{P}) + \varepsilon_b |M|. \tag{3.2}$$

If $\varepsilon_b \geq \varepsilon_c$, then by (3.1) and (3.2), $S(c,\mathbf{P}) = S(b,\mathbf{P})$ and $\varepsilon_b = \varepsilon_c$, which by tie-breaking implies $b = F(\tilde{P}_M,\mathbf{P}_{-M})$, a contradiction. Therefore, $\varepsilon_b < \varepsilon_c$. Similarly, $\varepsilon_c < \varepsilon_a$. Consequently, $\varepsilon_b < \varepsilon_c < \varepsilon_a$. The five possible (sign) cases are given in the following table:

Case	ε_b	ε_c	ε_a
1	$-$	$-$	0
2	$-$	$-,0,+$	$+$
3	0	$+$	$+$
4	$-$	$-$	$-$
5	$+$	$+$	$+$

Table 2

In the remainder of the proof, based on Table 2, we derive necessary conditions for an unsafe manipulation as in Part (i) to exist. In the last part, we show that when these conditions are not fulfilled, there can be an unsafe manipulation.

(ii) Suppose that $r = 1$, $\Delta_1 = s_k - s_{k+1}$. Then, for all $x \in X$, $\varepsilon_x = 0$ or $\varepsilon_x = -\Delta_1$ or $\varepsilon_x = \Delta_1$. This and $\varepsilon_b < \varepsilon_c < \varepsilon_a$ imply that in Table 2 the only possible case is Case 2. Hence, $\varepsilon_a = \Delta_1$ and $\varepsilon_b = -\Delta_1$ but this is not possible: indeed, if the score of a increases by Δ_1, then a moves from the bottom $m-k$ alternatives in P_i to the top k alternatives in $\tilde{P}$, but aP_ib, so, b is also among the bottom $m-k$ alternatives in P_i and therefore cannot decrease in score when going to $\tilde{P}$. Thus, in case of precisely one jump a scoring rule is only safely manipulable, and the first claim in the theorem is proved.

(iii) Suppose that $r = 2$, $\Delta_1 = s_{k_1} - s_{k_1+1}$, $\Delta_2 = s_{k_2} - s_{k_2+1}$. We go through all cases in Table 2 and consider all possible combinations of jumps for each ε_x, $x \in \{a,b,c\}$ in each case. Of course, throughout we use that initially b is chosen, then c, and at the end a, but we do not always spell out the details.

First, note that Cases 4 and 5 in Table 2 are not possible since these cases require at least three jumps to occur.

In Case 1, $\varepsilon_b < \varepsilon_c < 0 = \varepsilon_a$, there are two possibilities:

1.1 From P_i to $\tilde{P}$, b goes down one jump and c goes down one jump: this is only possible if $\Delta_1 > \Delta_2$, and then $\varepsilon_b = -\Delta_1$, $\varepsilon_c = -\Delta_2$, and $\varepsilon_a = 0$. This can be summarized as follows: $ab|c|\circ \to \circ, a|b|c$. [Here, $|$ denotes a jump, $ab|c|\circ$ contains the relevant information about P_i, and $\circ, a|b|c$ contains the relevant information about $\tilde{P}$. The small circles $\circ$ indicate other alternatives that are minimally available.]

1.2 b goes down two jumps and c goes down one jump. Then either $abc|\circ|\circ \to \circ, \circ, a|c|b$, hence $\varepsilon_b = -\Delta_1 - \Delta_2$, $\varepsilon_c = -\Delta_1$, and $\varepsilon_a = 0$; or $ab|c|\circ\circ \to \circ, a|\circ|b, c$, hence $\varepsilon_b = -\Delta_1 - \Delta_2$, $\varepsilon_c = -\Delta_2$, and $\varepsilon_a = 0$.

In Case 2, $\varepsilon_b < 0$ and $\varepsilon_a > 0$. Then $\varepsilon_b = -\Delta_2$ and $\varepsilon_a = \Delta_1$, For ε_c there are two possibilities:

2.1 $\varepsilon_c = \Delta_2$ and $\circ|ab|c \to a|\circ, c|b$. This is only possible if $\Delta_1 > \Delta_2$.

2.2 $\varepsilon_c = 0$, and $\circ|abc|\circ \to a|\circ\circ, c|b$ or $\circ|ab|\circ, c \to a|\circ\circ|b, c$.

Finally, in Case 3, $\varepsilon_b = 0$ and $\varepsilon_a, \varepsilon_c > 0$. There are again two possibilities:

3.1 $\varepsilon_a = \Delta_1$, $\varepsilon_c = \Delta_2$, and $\circ|ab|c \to a|b, c|\circ$ or $\circ|a|bc \to a|c|\circ, b$. This is only possible if $\Delta_1 > \Delta_2$.

3.2 $\varepsilon_a = \Delta_1 + \Delta_2$, $\varepsilon_c = \Delta_2$, and $\circ|\circ|abc \to a|c|\circ\circ, b$.

Based on these six possibilities, we can now examine the $r = 2$ cases in Table 1.

- If $m = 3$, then $P_i = a|b|c$, and therefore none of the Cases 1.1–3.2 applies. Hence, any manipulation in this case is safe.
- If $m = 4$ and $\Delta_1 > \Delta_2$, then Cases 1.1, 2.1, and 3.1 apply, and so there can be unsafe manipulations.
- If $m = 4$ and $\Delta_1 \leq \Delta_2$, then none of the Cases 1.1–3.2 applies. Hence, any manipulation in this case is safe.
- If $m = 5$ and $\Delta_1 \leq \Delta_2$, then from Cases 1.2, 2.2, and 3.2, it follows that unsafe manipulation may be possible for the following five jump combinations: $k_1 = 1, k_2 = 2$ (3.2); $k_1 = 1, k_2 = 3$ (2.2); $k_1 = 1, k_2 = 4$ (2.2); $k_1 = 2, k_2 = 3$ (1.2); $k_1 = 3, k_2 = 4$ (1.2). In the remaining case, $k_1 = 2, k_2 = 4$, no unsafe manipulation is possible.

- If $m = 5$ and $\Delta_1 > \Delta_2$, then all Cases 1.1–3.2 may apply and therefore all six jump combinations are possible, so that unsafe manipulation is possible for any of these combinations.

- If, finally, $m \geq 6$, then it is sufficient to consider the Cases 1.2, 2.2, and 3.2, to conclude that for any jump combination unsafe manipulation is possible.

(iv) We next consider the case $r = 3$. Then there must be at least four alternatives.

- If $m \geq 5$, then for each combination of three (or more) jumps it is possible to manipulate unsafely by using only two jumps as in Cases 1.2 and 2.2 Thus, unsafe manipulation may be possible for any jump combination.

- Now let $m = 4$. We consider the five cases in the Table 2.

  - Case 1 implies $P_i = a|b|c|\circ$ and $\tilde{P} = a|\circ|b|c$. Since $\varepsilon_b < \varepsilon_c < \varepsilon_a$, this implies $\Delta_2 > \Delta_3$. In this case, unsafe manipulation may be possible.

  - Case 2 implies $P_i = \circ|a|b|c$ and $\tilde{P} = a|\circ|c|b$ or $\tilde{P} = a|c|\circ|b$. Since $\varepsilon_b < \varepsilon_c < \varepsilon_a$, this implies $\Delta_1 > \Delta 3$ or $\Delta_1 > \Delta_2 + \Delta_3$. In turn, this implies that unsafe manipulation may be possible if $\Delta_1 > \Delta_3$.

  - Case 3 implies $P_i = \circ|a|b|c$ and $\tilde{P} = a|c|b|\circ$. Since $\varepsilon_b < \varepsilon_c < \varepsilon_a$, this implies $\Delta_1 > \Delta_2 + \Delta_3$. Under this condition, unsafe manipulation may be possible in this case.

  - Case 4 implies $P_i = a|b|c|\circ$ and $\tilde{P} = \circ|a|b|c$. Since $\varepsilon_b < \varepsilon_c < \varepsilon_a$, this implies $\Delta_2 > \Delta_3 > \Delta_1$. In this case therefore, an unsafe manipulation may exist, but by Case 1, $\Delta_2 > \Delta_3$ is already sufficient for this.

– Case 5 implies $P_i = \circ|a|b|c$ and $\tilde{P} = a|b|c\circ$. Since $\varepsilon_b < \varepsilon_c < \varepsilon_a$, this implies $\Delta_2 < \Delta_3 < \Delta_1$. In this case therefore, an unsafe manipulation may exist, but by Case 2, $\Delta_1 > \Delta_3$ is already sufficient for this.

Summarizing, an unsafe manipulation may exist if and only if $\Delta_1 > \Delta_3$ or $\Delta_2 > \Delta_3$.

(v) The OSM cases in Table 1 have now been proved. We complete the proof of the theorem by providing a procedure to construct a preference profile for any kind of unsafe manipulation.

Assume that we have a particular number of alternatives m, a given scoring vector s, and a group K of voters with preferences P s.t. $aPbPc$. Take any way of unsafe manipulation, $\tilde{P}$, corresponding to the given m and s from the previous part of the proof. Then, the chosen way of unsafe manipulation defines score differences for alternatives a, b, and c when one voter manipulates (switches from P to $\tilde{P}$). These score differences are: $\varepsilon_a = \sum_{j=1}^{r} \alpha_j \Delta_j$, $\varepsilon_b = \sum_{j=1}^{r} \beta_j \Delta_j$, and $\varepsilon_c = \sum_{j=1}^{r} \gamma_j \Delta_j$, where the $\alpha_j, \beta_j, \gamma_j$ are elements of $\{-1, 0, 1\}$. We need to prove that there exists a preference profile for some n such that members of K have an incentive to manipulate with $\tilde{P}$ and this manipulation is unsafe.

First, without loss of generality we assume tie-breaking according to $aP^t c$, $bP^t c$. Let the scores of alternatives be such that $S(b, \mathbf{P}) = S(c, \mathbf{P})$ and $S(c, (\tilde{P}_K, \mathbf{P}_{-K})) = S(a, (\tilde{P}_K, \mathbf{P}_{-K}))$. This, together with $\varepsilon_b < \varepsilon_c < \varepsilon_a$, implies that $F(\mathbf{P}) = b$, $F(\tilde{P}_K, \mathbf{P}_{-K}) = a$, and $F(\tilde{P}_M, \mathbf{P}_{-M}) = c$ for some $M \subseteq K$.

For the difference in scores for alternatives a and c before and after manipulation, we have $S(a, (\tilde{P}_K, \mathbf{P}_{-K})) - S(a, \mathbf{P}) = \varepsilon_a |K|$ and $S(c, (\tilde{P}_K, \mathbf{P}_{-K})) - S(c, \mathbf{P}) = \varepsilon_c |K|$. Since $S(c, (\tilde{P}_K, \mathbf{P}_{-K})) = S(a, (\tilde{P}_K, \mathbf{P}_{-K}))$ we have $S(a, \mathbf{P}) + \varepsilon_a |K| = S(c, \mathbf{P}) + \varepsilon_c |K|$ and, finally, $S(c, \mathbf{P}) - S(a, \mathbf{P}) = \varepsilon_a |K| - \varepsilon_c |K|$. So, $S(c, \mathbf{P}) - S(a, \mathbf{P}) = \sum_{j=1}^{r} \mu_j \Delta_j$ for some integers μ_j.

Summing up, in the required profile $\mathbf{P}$ it is needed that: (a) the score differences between a, b, and c are fixed, $S(b, \mathbf{P}) - S(c, \mathbf{P}) = 0$, $S(c, \mathbf{P}) - S(a, \mathbf{P}) =$

$\sum_{j=1}^{r} \mu_j \Delta_j$ for some integers μ_j; (b) other alternatives do not affect the result; (c) there are exactly $|K|$ voters with preferences P_i.

We now describe a procedure to generate a profile $\mathbf{P}$ with these properties. We first fix a preference profile for some set of voters K, where every member of K has the same preference P, say, $aPbPcPa_1Pa_2P...Pa_{m-3}$. Take any number of voters in K and include their preferences, P_K, in the profile $\mathbf{P}$ that we are constructing. Then we have condition (c) satisfied.

For the voters outside K we consider the following basic profile $B(a)$:

$$\begin{array}{cccc}
a_{m-3} & a_{m-4} & \cdots & a \\
a & a_{m-3} & \cdots & c \\
c & a & \cdots & b \\
b & c & \cdots & a_1 \\
a_1 & b & \cdots & a_2 \\
\vdots & \vdots & \vdots & \vdots \\
a_{m-4} & a_{m-5} & \cdots & a_{m-3}
\end{array}$$

Observe that in $B(a)$ the scores of all alternatives are equal, and that P does not occur. Suppose that we need to increase the score of alternative a by the amount Δ_l, which is the size of the l-th jump, following position k_l. Then we replace column (preference) k_l in $B(a)$ by $\tilde{R}$, where $\tilde{R}$ is obtained by switching positions k_l and k_l+1 in column k_l. This results in a profile $B'(a)$ where the scores of all alternatives except a and a_{m-3} are still equal (and equal to the scores in $B(a)$), the score of a has increased by Δ_l, and the score of a_{m-3} has decreased by Δ_l. Note that $\tilde{P} \neq P$ and, thus, P does not occur in $B'(a)$. So, we include $B'(a)$ in $\mathbf{P}$. If it is needed to increase the score of a by the size of another jump, we include $B''(a)$ constructed similarly, etc.

Similar constructions can be made for b and c, if we need to increase their scores, by starting from the most left columns $(a_{m-3}, b, a, c, a_1, \ldots, a_{m-4})$ and $(a_{m-3}, c, a, b, a_1, \ldots, a_{m-4})$ respectively. Doing this as many times as needed to satisfy conditions (a) and (b), in the end we obtain the required preference profile. Moreover, notice that we can choose any number of voters in K. So, if an unsafe manipulation exists for some m then it is always possible to find a profile with a group of only two voters having an incentive to manipulate unsafely. □

Observe that, although Theorem 3.1 identifies all scoring rules where an unsafe manipulation exists, it is silent about how many voters are needed to have such a manipulation. It is difficult to derive general results about this for (all) scoring rules and therefore, in the next theorem, we focus on the arguably most famous rules with at least two jumps, namely Borda rules (cf. Section 3.1.3). Note that, since all jumps at a Borda rule have equal size, the cases with less than 5 alternatives are covered by Theorem 3.1.

Theorem 3.2. *Let F be a Borda rule. If $m = 5$, then an unsafely manipulable profile exists if and only if $n \geq 4$. If $m \geq 6$, then an unsafely manipulable profile exists if and only if $n \geq 3$.*

Proof. (a) First let $m = 5$, $X = \{a,b,c,d,e\}$, and consider the following profiles $\mathbf{P}'$ and $\mathbf{P}''$ for $n = 4$ and $n = 5$, respectively:

$$\begin{array}{cccc} P'_1 & P'_2 & P'_3 & P'_4 \\ \hline a & a & e & e \\ b & b & c & d \\ c & c & d & c \\ d & d & b & a \\ e & e & a & b \end{array} \qquad \begin{array}{ccccc} P''_1 & P''_2 & P''_3 & P''_4 & P''_5 \\ \hline a & a & e & e & d \\ b & b & c & c & e \\ c & c & b & d & c \\ d & d & a & b & b \\ e & e & d & a & a \end{array} .$$

Suppose that tie-breaking is done according to the ordering $T = (e,c,a,b,d)$. Then $F(\mathbf{P}') = c$. If group $K = \{1,2\}$ changes their preferences to $\tilde{P} = (b,a,d,c,e)$, then $F(\tilde{P}_K, \mathbf{P}'_{-K}) = b$, which is preferred by the members of K to c. However, $F(\tilde{P}_{\{1\}}, \mathbf{P}'_{-\{1\}}) = e$, so that this manipulation is unsafe.

As to $\mathbf{P}''$, note that also $F(\mathbf{P}'') = c$ and $K = \{1,2\}$ can manipulate again by $\tilde{P} = (b,a,d,c,e)$. If only voter 1 manipulates, then again e results, so that also this manipulation is unsafe.

Thus, if $m = 5$ and $n = 4$ or $n = 5$, there exists an unsafe manipulation.

(b) We next show that no unsafe manipulation exists if $m = 5$ and $n = 3$. In this case, a possibly unsafely manipulating group can only consist of 2 members, say $K = \{1,2\}$. Suppose, indeed, that for some $a,b,c \in X$, aP_ibP_ic for all $i \in K$, and that there is a preference P_3 for voter 3 and a preference $\tilde{P}$ such that $F(\mathbf{P}) = b$, $F(\tilde{P}_{\{1,2\}}, P_3) = a$, and $F(\tilde{P}_{\{1\}}, \mathbf{P}_{\{2,3\}}) = c$. Note that, at $\mathbf{P}$, the Borda score of c must be strictly larger than the Borda score of a: if not, then the score of c should increase more than the score of a after manipulation by just one member of K, but then c would still win after manipulation by both members of K, a contradiction. Further, the score of a contributed by P_1 and P_2 is at least 4 more than the the score of c contributed by P_1 and P_2, since aP_ibP_ic for $i = 1,2$. In turn, these facts imply that the score of c contributed by P_3 is at least five more than the score of a contributed by P_3, which is impossible with five alternatives.

(c) Consider the case $m = 6$, $X = \{a,b,c,d,e,f\}$, and $n = 3$, and the profile $\mathbf{P}$ with $P_1 = P_2 = (a,b,c,d,e,f)$ and $P_3 = (c,b,f,d,e,a)$. Let $\tilde{P} = (a,e,d,c,b,f)$. Then $F(\mathbf{P}) = b$, $F(\tilde{P}_{\{1,2\}}, P_3) = a$, and (assuming that c beats a by tie-breaking) $F(\tilde{P}_{\{1\}}, \mathbf{P}_{\{2,3\}}) = c$, so that an unsafe manipulation exists in this case.

(d) Finally, the hitherto constructed profiles where an unsafe manipulation exists, can be extended with any number of alternatives, simply by adding those alternatives at the bottom of the preferences. Also, each of the manipulable profiles can be extended by any even number of agents 2ℓ: add ℓ times the pair of preferences $(a_1,\ldots,a_m)$ and $(a_m,\ldots,a_1)$, where $X = \{a_1,\ldots,a_m\}$, and note that this just adds equal scores for all alternatives. The proof of the theorem is now complete. $\square$

3.3 (Un)safe manipulability of run-off

For the definition of the run-off correspondence, see Section 3.1.3.

We start by observing that at a run-off rule it is impossible to manipulate in favor of the most preferable alternative.

Lemma 3.1. *Let F be a run-off rule, let $\mathbf{P}$ be a preference profile, let $i \in N$ and $a \in X$ such that aP_ix for all $x \in X \setminus \{a\}$ and $a \neq F(\mathbf{P})$, and let K be the group of i. Then there is no $\tilde{P} \in L(X)$ such that $a = F(\tilde{P}_K, \mathbf{P}_{-K})$.*

Proof. If a does not survive the first stage of the run-off procedure at $\mathbf{P}$, then it will also not survive the first stage at any $(\tilde{P}_K, \mathbf{P}_{-K})$. If a survives the first stage but not the second stage of the run-off procedure at $\mathbf{P}$, then for any $\tilde{P} \in L(X)$, either a does not survive the first stage at $(\tilde{P}_K, \mathbf{P}_{-K})$, or it does. In the latter case, since for every $x \in X$ we have $|\{j \in K : aP_jx\}| \geq |\{j \in K : a\tilde{P}_jx\}|$, it follows that a does not survive the second stage at $(\tilde{P}_K, \mathbf{P}_{-K})$. □

Our results for run-off rules are as follows.

Theorem 3.3. *Let F be a run-off rule.*

(a) *If $m = 3$, then F is only safely manipulable.*

(b) *If $m = 4$, then F is only safely manipulable if and only if $n \leq 5$.*

(c) *If $m \geq 5$, then F is only safely manipulable if and only if $n \leq 4$.*

Proof. We will prove the theorem for seven specific cases, depending on the numbers m and n of alternatives and voters, and then summarize how the theorem follows from these cases.

(1) Let $m = 3$, $X = \{a, b, c\}$, $\mathbf{P} \in L(X)^N$, and let K be a group with common preference aP_ibP_ic for every $i \in K$. If K can manipulate unsafely by $\tilde{P}$, then we must have $F(\mathbf{P}) = b$, $F(\tilde{P}_K, \mathbf{P}_{-K}) = a$, and $F(\tilde{P}_M, \mathbf{P}_{-M}) = c$ for some $M \subseteq K$. Such a manipulation, however, is excluded by Lemma 3.1. This proves Part (a) of the theorem.

(2) If $n = 3$ then for an unsafe manipulation a group of at least two members is required, but then their common top alternative is chosen by F. So F is only safely manipulable.

From now on, we assume that $m, n \geq 4$, $a, b, c, d \in X$, and the members of group K have a preference P with top alternative a and with bPc, cPd.

(3) Let $n = 4$. Assume, contrary to what we want to prove, that K has an unsafe manipulation at $\mathbf{P}$ via $\tilde{P}$. Then by Lemma 3.1 we may assume that $F(\mathbf{P}) = c$, $F(\tilde{P}_K, \mathbf{P}_{-K}) = b$, and $F(\tilde{P}_M, \mathbf{P}_{-M}) = d$ for some $M \subseteq K$. Since $|K| = 2$, the plurality score of a at $\mathbf{P}$ is 2, and the plurality score of c at $\mathbf{P}$ is 1 or 2. In the latter case, the plurality score of d at $\mathbf{P}$ is 0, but then $F(\tilde{P}_M, \mathbf{P}_{-M}) \neq d$ since $|M| = 1$, a is the top alternative of P, and b is the top alternative of $\tilde{P}$, and so d does not survive the first stage at $(\tilde{P}_M, \mathbf{P}_{-M})$, contradicting that $F(\tilde{P}_M, \mathbf{P}_{-M}) = d$. Therefore, we have that the plurality score of c at $\mathbf{P}$ is 1.

If the plurality score of d at $\mathbf{P}$ is 0, then as before, $F(\tilde{P}_M, \mathbf{P}_{-M}) \neq d$, a contradiction. Thus, the plurality score of d at $\mathbf{P}$ is 1. Since $F(\mathbf{P}) = c$, a and c survive the first stage at $\mathbf{P}$, which implies that the tie-breaking order P^t satisfies $cP^t d$ and $cP^t a$. Since the top alternative of $\tilde{P}$ is b, at $(\tilde{P}_M, \mathbf{P}_{-M})$ the alternatives a, b, c, d all have equal plurality score 1, and since $cP^t d$ and $cP^t a$, we have that b and c survive the first round, contradicting again that $F(\tilde{P}_M, \mathbf{P}_{-M}) = d$.

Hence, we have proved that for $m \geq 4$ and $n = 4$ there is no unsafe manipulation.

(4) Let $n = 5$ and $m = 4$. As in Part (3), assume that K has an unsafe manipulation at $\mathbf{P}$ via $\tilde{P}$, with $F(\mathbf{P}) = c$, $F(\tilde{P}_K, \mathbf{P}_{-K}) = b$, and $F(\tilde{P}_M, \mathbf{P}_{-M}) = d$ for some $M \subseteq K$. Clearly, the plurality score of a at $\mathbf{P}$ cannot be larger than 2, and therefore is equal to 2. In particular, $|K| = 2$, say $K = \{1, 2\}$. Since $F(\mathbf{P}) = c$ and $F(\tilde{P}_K, \mathbf{P}_{-K}) = b$, the plurality score of c at $\mathbf{P}$ is 1 or 2.

In the latter case, say that P_3 and P_4 have top alternative c. Since $F(\tilde{P}_{\{1\}}, \mathbf{P}_{-\{1\}}) = d$ and the top alternative of $\tilde{P}$ is b, we have that a, b, and d each have plurality score 1 at $(\tilde{P}_{\{1\}}, \mathbf{P}_{-\{1\}})$, and c and d survive the first stage. However, $cP_j d$ for $j = 2, 3, 4$, so that c finally wins, a contradiction.

Hence, the plurality score of c at $\mathbf{P}$ is 1. Since $F(\tilde{P}_{\{1\}}, \mathbf{P}_{-\{1\}}) = d$, the plurality score of d at $\mathbf{P}$ is at least 1, and since $F(\mathbf{P}) = c$, it is exactly 1. It follows that the plurality score of b at $\mathbf{P}$ is also 1. In turn, for the tie-breaking order P^t, this implies that $cP^t d$. But then, at $(\tilde{P}_{\{1\}}, \mathbf{P}_{-\{1\}})$, d does not survive the first round, contradicting that $F(\tilde{P}_{\{1\}}, \mathbf{P}_{-\{1\}}) = d$.

Hence, we have proved that for $m = 4$ and $n = 5$ there is no unsafe manipulation.

(5) Let $n = 5$ and $m = 5$, $X = \{a,b,c,d,e\}$. We show that there is an unsafe manipulation. Let $K = \{1,2\}$, and let $\mathbf{P}$ be a preference profile with $P_1 = P_2 = (a,b,c,d,e)$, and such that c, d, and e each have plurality score of 1 at $\mathbf{P}$, cP_ja for $j = 3,4,5$, and bP_5dP_5c. Let the tie-breaking order be $P^t = (c,d,e,a,b)$. Then $F(\mathbf{P}) = c$. If $\tilde{P}$ has top alternative b and $d\tilde{P}c$, then $F(\tilde{P}_{\{1,2\}}, \mathbf{P}_{-\{1,2\}}) = b$, and $F(\tilde{P}_{\{1\}}, \mathbf{P}_{-\{1\}}) = d$. So K has an unsafe manipulation.

(6) For $n \geq 6$ and $m = 4$ we construct unsafely manipulable profiles based on the following preferences: $P^1 = (a,b,c,d)$, $P^2 = (c,a,b,d)$, $P^3 = (d,b,c,a)$, $P^4 = (b,a,d,c)$. Let $\mathbf{P}_n$ denote a preference profile with n voters. For $j = 0,1,2,\ldots$ let $\mathbf{P}_{6+3j}$ such that it contains P^1, P^2, and P^3 each $2+j$ times: $\mathbf{P}_{6+3j} = ((2+j)P^1,(2+j)P^2,(2+j)P^3)$. Assume that the tie-breaking order is $P^t = (b,d,c,a)$. Then $F(\mathbf{P}_{6+3j}) = c$. If the voters with preference P^1 change to P^4, then b wins. If only one voter manipulates, then d wins.

Similarly, we consider preference profiles $\mathbf{P}_{7+3j} = ((2+j)P^1,(2+j)P^2,(3+j)P^3)$ for $j = 0,1,2,\ldots$; with the same tie-breaking rule, the same kind of unsafe manipulation exists. Finally, we consider profiles $\mathbf{P}_{8+3j} = ((2+j)P^1,(2+j)P^2,(4+j)P^3)$ for $j = 0,1,2,\ldots$, and tie-breaking order $P^t = (b,c,d,a)$: again the same kind of unsafe manipulation exists.

(7) Finally, we observe that for any unsafely manipulable profile we obtain an unsafely manipulable profile for more alternatives by simply adding those additional alternatives at the bottom of the preferences in the given profile.

The following table summarizes how the theorem follows from the seven parts of the proof.

	$n=3$	$n=4$	$n=5$	$n\geq 6$
$m=3$	OSM Part 1	OSM Part 1	OSM Part 1	OSM Part 1
$m=4$	OSM Part 2	OSM Part 3	OSM Part 4	UM Part 6
$m\geq 5$	OSM Part 2	OSM Part 3	UM Parts 5, 7	UM Parts 6, 7

□

3.4 (Un)safe manipulability of Copeland

Recall (Section 3.1.3) that the Copeland correspondence chooses the alternatives with maximal Copeland score, where the Copeland score of an alternative a at a preference profile $\mathbf{P}$ is the number $|\{b\in X : |\{i\in N : aP_ib\}|> \frac{n}{2}\}|-|\{b\in X : |\{i\in N : bP_ia\}|> \frac{n}{2}\}|$.

Theorem 3.4. *Let F be a Copeland rule. Then F is only safely manipulable if and only if $m=3$ or $n=3$.*

Proof. The proof proceeds in five parts.

(1) Suppose $n=3$. Since $|K|\geq 2$ is required for an unsafe manipulation, but in that case the top alternative of the voters in K is chosen, there exists no unsafe manipulation.

(2) Suppose $m=3$, $X=\{a,b,c\}$, let $\mathbf{P}$ be a preference profile, and let the voters in a group K have preferences $aPbPc$. For an unsafe manipulation by K we must have $F(\mathbf{P})=b$, $F(\tilde{P}_K,\mathbf{P}_{-K})=a$, and $F(\tilde{P}_M,\mathbf{P}_{-M})=c$ for some $M\subseteq K$ and $\tilde{P}\in L(X)$. We show that this is impossible. For alternative x denote by $S(x)$, $\tilde{S}(x)$, and $\bar{S}(x)$ the Copeland scores of x at $\mathbf{P}$, $(\tilde{P}_K,\mathbf{P}_{-K})$, and $(\tilde{P}_M,\mathbf{P}_{-M})$, respectively. Clearly we have $S(a)\geq\bar{S}(a)\geq\tilde{S}(a)$ and $S(c)\leq\bar{S}(c)\leq\tilde{S}(c)$. Since $F(\tilde{P}_K,\mathbf{P}_{-K})=a$ and $F(\tilde{P}_M,\mathbf{P}_{-M})=c$ we have $\bar{S}(c)\geq\bar{S}(a)$ and $\tilde{S}(a)\geq\tilde{S}(c)$. Hence $\tilde{S}(a)\geq\tilde{S}(c)\geq\bar{S}(c)\geq\bar{S}(a)\geq\tilde{S}(a)$, and therefore

$\tilde{S}(a) = \tilde{S}(c) = \bar{S}(c) = \bar{S}(a)$. This, however, is inconsistent with (any order of) tie-breaking. Thus, for $m = 3$ there exists no unsafe manipulation.

(3) We exhibit an unsafely manipulable preference profile for $m = 4$ and $n = 4$. Consider the profile $\mathbf{P}$ given by

P_1	P_2	P_3	P_4
c	c	b	b
a	a	a	a
b	b	c	c
d	d	d	d

The Copeland scores of a,b,c,d at $\mathbf{P}$ are, respectively, $1,1,1,-3$. With tie-breaking order $P^t = (a,b,c,d)$ we have $F(\mathbf{P}) = a$. If $K = \{3,4\}$ changes preferences to $\tilde{P} = (b,d,c,a)$, then the scores are $-1,1,1,-1$, so that $F(\tilde{P}_K, \mathbf{P}_{-K}) = b$. If $M = \{3\}$, then the scores are $0,1,2,-3$, so that $F(\tilde{P}_M, \mathbf{P}_{-M}) = c$. Hence, $\mathbf{P}$ is an unsafely manipulable preference profile.

(4) We exhibit an unsafely manipulable preference profile for $m = 4$ and $n = 5$. Consider the profile $\mathbf{P}$ given by

P_1	P_2	P_3	P_4	P_5
c	c	a	b	b
b	a	d	a	a
a	b	c	c	c
d	d	b	d	d

In this case, it is easy to verify that, with the same tie-breaking as in Part (3), $K = \{4,5\}$ can unsafely manipulate by $\tilde{P} = (b,d,c,a)$.

(5) Finally, if there are more than five agents then these agents can be added in pairs with opposite preferences to the unsafely manipulable profiles in Parts (3) and (4), to obtain such profiles with more than five agents. If there are more than four alternatives, then the additional alternatives can be added at the bottom of the unsafely manipulable profiles for four alternatives. This concludes the proof of the theorem. □

3.5 (Un)safe manipulability of Single-Transferable-Vote

For the definition of the Single-Transferable-Vote (STV) correspondence, see Section 3.1.3.

As for all the previous rules, group manipulation for STV with $m = 3$ is always safe:

Lemma 3.2. *Let F be the STV rule, and let $m = 3$. Then F is only safely manipulable.*

Proof. Let $X = \{a,b,c\}$, let $\mathbf{P}$ be a preference profile, and suppose there is an unsafe manipulation by group K, who have preference $P = (a,b,c)$, via $\tilde{P} \in L(X)$. Then we have $F(\mathbf{P}) = b$, $F(\tilde{P}_K, \mathbf{P}_{-K}) = a$, and $F(\tilde{P}_M, \mathbf{P}_{-M}) = c$ for some $M \subseteq K$. Then the top alternative of $\tilde{P}$ cannot be a, since this would not change the outcome, nor b, since this would either not change the outcome or lead to the elimination of a. Hence, the top alternative of $\tilde{P}$ is c. Denote by $S(x)$ the plurality score (i.e., number of top positions) of $x \in X$ at $\mathbf{P}$, and by $\tilde{S}(x)$ the plurality score of x at $(\tilde{P}_K, \mathbf{P}_{-K})$). Then $S(a) > \tilde{S}(a)$, $S(c) < \tilde{S}(c)$, and $S(b) = \tilde{S}(b)$.

We claim that $\tilde{S}(a) \geq \tilde{S}(b)$. Suppose not, i.e., $\tilde{S}(a) < \tilde{S}(b)$. Then, if $\tilde{S}(c) \geq \tilde{S}(a)$, alternative a will be eliminated at $(\tilde{P}_K, \mathbf{P}_{-K})$), contradicting $F(\tilde{P}_K, \mathbf{P}_{-K}) = a$. If $\tilde{S}(c) < \tilde{S}(a)$, then $S(c) < S(a)$ and $S(c) < S(b)$, and $\tilde{S}(c) < \tilde{S}(a) < \tilde{S}(b)$. This means that c is eliminated first, both at $\mathbf{P}$ and at $(\tilde{P}_K, \mathbf{P}_{-K})$); since, however, $F(\mathbf{P}) = F(P_K, \mathbf{P}_{-K}) = b$, this implies that also $F(\tilde{P}_K, \mathbf{P}_{-K}) = b$ as $P = (a,b,c)$ and $\tilde{P} = (c,\cdot,\cdot)$. This is a contradiction, and thus the claim is proved.

We next consider $\tilde{S}(c)$. If $\tilde{S}(a) = \tilde{S}(b)$, then $\tilde{S}(c) = \tilde{S}(a) = \tilde{S}(b)$, otherwise $a \neq F(\tilde{P}_K, \mathbf{P}_{-K})$. If $\tilde{S}(a) > \tilde{S}(b)$, then $\tilde{S}(c) \geq \tilde{S}(b)$, because otherwise again $a \neq F(\tilde{P}_K, \mathbf{P}_{-K})$ by a similar argument as in the second case of the preceding paragraph.

Finally, if $\tilde{S}(c) = \tilde{S}(a) = \tilde{S}(b)$, then for all $M \subseteq K$, c is eliminated in $(\tilde{P}_M, \mathbf{P}_{-M})$, a contradiction. If $\tilde{S}(c) > \tilde{S}(b)$, then there is some $M' \subseteq K$, such that $S(c)+|M'|-1 < S(b)$ and $S(c)+|M'| \geq S(b)$. For all G with $|G| < |M'|$, c is eliminated; and for all G with $|G| \geq |M'|$ the winner is a. This is again a contradiction, which concludes the proof of the lemma. □

Theorem 3.5. *Let F be the STV rule. Then F is only safely manipulable if and only if $m = 3$ or $n \leq 7$.*

Proof. The proof proceeds in several parts.

(1) By Lemma 3.2, if $m = 3$ then F is only safely manipulable.

(2) For $n \geq 8$ and $m = 4$, $X = \{a,b,c,d\}$, we construct unsafely manipulable profiles based on the following preferences: $P^1 = (a,b,c,d)$, $P^2 = (a,b,d,c)$, $P^3 = (b,a,c,d)$, $P^4 = (b,a,d,c)$, $P^5 = (b,c,d,a)$, $P^6 = (c,a,b,d)$, $P^7 = (d,b,c,a)$, and $P^8 = (d,c,a,b)$. Let $\mathbf{P}_n$ denote a preference profile of n voters.

(2.1) For $j = 0,1,2,\ldots$ consider a profile $\mathbf{P}_{8+4j} = (2P^1,(1+j)P^2,(1+j)P^5,(2+j)P^6,(2+j)P^8)$, meaning that preference P^1 occurs 2 times etc. The plurality scores (of the first round) are: $S_1(a) = 3+j$, $S_1(b) = 1+j$, $S_1(c) = 2+j$, $S_1(d) = 2+j$. Since b has minimal score, it is deleted. At the second round: $S_2(a) = 3+j$, $S_2(c) = 2j+3$, $S_2(d) = 2+j$, so that d is deleted. At the third round: $S_3(a) = 3+j$, $S_3(c) = 3j+5$, so that c wins. Suppose that the group of voters with preferences P^1 switch to P^3. Then $S_1(a) = 1+j$, $S_1(b) = 3+j$, $S_1(c) = 2+j$, $S_1(d) = 2+j$, so that a is deleted; $S_2(b) = 2j+4$, $S_2(c) = 2+j$, $S_2(d) = 2+j$, so that c and d are deleted, and thus b wins. If only one voter of the group manipulates, then the scores are: $S_1(a) = 2+j$, $S_1(b) = 2+j$, $S_1(c) = 2+j$, $S_1(d) = 2+j$, so, there is a complete tie and d wins provided that the tie-breaking order satisfies $dP^t a$, $dP^t b$, $dP^t c$.

(2.2) In a profile $\mathbf{P}_{9+4j} = (2P^1,(1+j)P^2,(1+j)P^5,(2+j)P^6,(3+j)P^8)$ with $j = 0,1,2,\ldots$ the same kind of manipulation by the same group leads to the same results, for any tie-breaking order.

(2.3) In a profile $\mathbf{P}_{14+4j} = (2P^1, (2+j)P^2, P^4, (1+j)P^5, (4+j)P^6, (4+j)P^8)$ with $j = 0, 1, 2, \ldots$ the result is c. Switching to P^3 for the group of voters having P^1 leads to b. When only one voter manipulates, d wins provided that $dP^t c$.

(2.4) For a profile $\mathbf{P}_{15+4j} = (2P^1, (2+j)P^2, P^4, (1+jP^5, (4+j)P^6, (5+j)P^8)$ with $j = 0, 1, 2, \ldots$ the result is c if $cP^t d$ and $cP^t a$. Switching to P^3 for the group of voters having P^1 leads to b and manipulation of only one member leads to d.

(2.5) Only two cases are left: $n = 10$ and $n = 11$. Let $\mathbf{P}_{10} = (2P^1, P^2, P^5, 3P^6, 3P^8)$ and $\mathbf{P}_{11} = (2P^1, P^2, P^4, 4P^6, 3P^7)$. In these profiles, the result is c, but if the voters having preferences P^1 switch to P^4, then the result changes to b, and in case of manipulation of one voter it changes to d.

Summing up, for $m = 4$ and $n \geq 8$ there exist unsafely manipulable profiles. This result also holds for $m > 4$: additional alternatives can be added at the bottom of all preferences and will not change the result.

(3) In this part of the proof we show that for $m \geq 4$ and $n \leq 7$ unsafely manipulable profiles do not exist. Let $\mathbf{P}$ be a preference profile. Take four alternatives $a, b, c, d \in X$, and let there be a group K with preference $aPbPcPd$ restricted to these alternatives, and with a their top alternative. Let $M \subseteq K$ and $\tilde{P} \in L(X)$. We use the notation $S(x)$ for the plurality score of x at $\mathbf{P}$, $\tilde{S}(x)$ for the plurality score of x at $(\tilde{P}_K, \mathbf{P}_{-K})$, and $\bar{S}(x)$ for the plurality score of x at $(\tilde{P}_M, \mathbf{P}_{-M})$. Assume that K has a manipulation via $\tilde{P}$.

(3.1) Since $|K| \geq 2$, $S(a) \geq 2$. Let $c = F(\mathbf{P})$. Then $S(c) \geq 2$. Hence $n > 3$. Consider the case $n = 4$, $S(a) = 2$ and $S(c) = 2$. Members of K cannot manipulate in favor of a (since voting for another alternative will lead to elimination of a), but they can make b winning by voting for b, which is better than c. This manipulation is safe, since $\bar{S}(a) = 1$, $\bar{S}(b) = 1$, and $\bar{S}(c) = 2$ and d cannot win in $(\tilde{P}_M, \mathbf{P}_{-M})$.

(3.2) Consider the case $n = 5$. The first-round scores are $S(a) = 2$, $S(c) = 2$, and there is some $x \in X$, $x \neq a$, $x \neq c$, s.t. $S(x) = 1$. Manipulation in favor of a is also impossible, so members of K vote for b. Again, d cannot be the winner at $(\tilde{P}_M, \mathbf{P}_{-M})$, even if $x = d$, since $\bar{S}(a) = \bar{S}(b) = \bar{S}(d) = 1$ and all these alternatives are eliminated in the first round.

(3.3) Consider the case $n = 6$. Again we have $S(a) = 2$, $S(c) = 2$, and members of K manipulating by voting for b. If $S(d) = 1$, then again d will be eliminated at $(\tilde{P}_M, \mathbf{P}_{-M})$. So, $S(d) = 2$. Let c be the STV winner at $\mathbf{P}$. Since $S(a) = S(c) = S(d) = 2$ it follows that $cP^t d$ (where P^t is the tie-breaking order). So, $\bar{S}(a) = 1$, $\bar{S}(b) = 1$, $\bar{S}(c) = 2$, and $\bar{S}(d) = 2$. Therefore in round 1 at $(\tilde{P}_M, \mathbf{P}_{-M})$ alternatives a and b are eliminated. But as $aPbPcPd$, the score of c in round 2 at $(\tilde{P}_M, \mathbf{P}_{-M})$ is at least 3. By $cP^t d$, it follows that d cannot be the STV winner at $(\tilde{P}_M, \mathbf{P}_{-M})$.

(3.4) Finally, consider the case $n = 7$. Since c is the STV winner at $\mathbf{P}$, $S(a) \neq 3$ and $S(d) \neq 3$. Otherwise c would be eliminated in the first round at $\mathbf{P}$. If $S(c) = 3$, then $\tilde{S}(a) = 0$, $\tilde{S}(b) = \tilde{S}(d) = 2$ and $\tilde{S}(c) = 3$. So, b is eliminated at $(\tilde{P}_K, \mathbf{P}_{-K})$ in the first round. Therefore, $S(a) = S(c) = S(d) = 2$ and $S(b) = 1$. Also, $|M| = 1$. Hence, $\bar{S}(a) = 1$ and $\bar{S}(b) = \bar{S}(c) = \bar{S}(d) = 2$. therefore, in round 1 at $(\tilde{P}_M, \mathbf{P}_{-M})$ alternative a is eliminated. Since all agents in K have preference $aPbPcPd$ it follows that in round two (after eliminating a) the score of b has increased by one to 3 whereas the scores of c and d are unchanged. Therefore, in round 2 at $(\tilde{P}_M, \mathbf{P}_{-M})$ alternatives c and d are eliminated. This contradicts that d is the STV winner in profile $(\tilde{P}_M, \mathbf{P}_{-M})$.

Thus, if $m \geq 4$ and $n \leq 7$, then there are no unsafe manipulations. This concludes the proof of the theorem. □

3.6 Concluding remarks

3.6.1 Relation with Slinko and White (2014)

In this subsection we compare our work with Slinko and White (2014) – henceforth SW.

For a rule F and a preference profile $\mathbf{P}$, according to SW a voter i with group K has an incentive to manipulate if there is a preference $\tilde{P} \in L(X)$ and a set $G \subseteq$

K with $i \in G$ such that $F(\tilde{P}_G, \mathbf{P}_{-G}) P_i F(\mathbf{P})$. Observe that SW do not require that all voters in K deviate to $\tilde{P}$. Clearly, if i has an incentive to manipulate in our sense (Definition 3.1), then i has an incentive to manipulate according to SW (simply take $G = K$), but the converse is not necessarily true.

Next, SW call such a manipulation by $\tilde{P}$ unsafe if there exists $M \subseteq K$ with $i \in M$ such that all members of M have an incentive to manipulate by $\tilde{P}$, but $F(\mathbf{P}) P_i F(\tilde{P}_M, \mathbf{P}_{-M})$; and safe if for all $U \subseteq K$ with $i \in U$, we have $F(\tilde{P}_U, \mathbf{P}_{-U}) P_i F(\mathbf{P})$ or $F(\tilde{P}_U, \mathbf{P}_{-U}) = F(\mathbf{P})$. Hence, if i has an incentive to manipulate by $\tilde{P}$ in our sense, so that, by the preceding paragraph, i also has an incentive to manipulate by $\tilde{P}$ according to SW, then if this manipulation is (un)safe in our sense (see Section 3.1.2), it is also (un)safe according to SW.

The definitions of (un)safely manipulable preference profiles and rules in SW are similar to ours (Section 3.1.2), so that we obtain the following corollary.

Corollary 3.1. *If a rule is safely (unsafely) manipulable for some m and n, then is is also safely (unsafely) manipulable according to SW.*

Thus, results about the (un)safety of manipulation in our sense are applicable to the model of SW. Unfortunately, we cannot directly adapt results in our paper about cases where we have only safe manipulations, to the model of SW in the same way. Indeed, in preference profiles where there is no manipulation in our sense there could still be voters having an incentive to manipulate according to SW, and this manipulation could be unsafe.

The main result in SW, their Theorem 2, says that for every onto and non-dictatorial rule F with range at least three there is a preference profile $\mathbf{P}$, a voter i, and a preference $\tilde{P}$, such that i has an incentive to manipulate and this manipulation is safe.

In the SW model, if voter i has an incentive to manipulate safely by $\tilde{P}$ in $\mathbf{P}$, this does not necessarily imply that the same voter has an incentive to manipulate in our model, since this safe manipulation according to SW still allows for $F(\tilde{\mathbf{P}}_K, \mathbf{P}_{-K}) = F(\mathbf{P})$. Thus, if a rule is safely manipulable according to SW, it does not follow directly from the definitions that the same holds in our model and for this reason Theorem 2 of SW does not carry over directly to our model. However, we can prove that if a rule F is manipulable in our model, then a safe manipulation also exists.

Theorem 3.6. *If a rule is manipulable, then it is also safely manipulable in our model.*

Proof. Let $\mathbf{P}$ be a manipulable profile, hence there are $\tilde{P} \in L(X)$ and $i \in N$ such that $F(\tilde{P}_K, \mathbf{P}_{-K}) P_i F(\mathbf{P})$. If this manipulation is safe, then we are done. If this manipulation is unsafe, then there is an $M \subset K$ with $i \in M$ such that $F(\mathbf{P}) P_i F(\tilde{P}_M, \mathbf{P}_{-M})$ and, consequently, $F(\tilde{P}_K, \mathbf{P}_{-K}) P_i F(\tilde{P}_M, \mathbf{P}_{-M})$. Consider the profile $\mathbf{P}' = (P_{K \setminus M}, \tilde{P}_M, \mathbf{P}_{-K})$. Now $K' = K \setminus M$ is a group and members of K' have an incentive to manipulate with $\tilde{P}$. Again, if this manipulation is safe, we are done. Otherwise, by the same reasoning there is $M' \subset K'$ that $F(\mathbf{P}') P_i F(\tilde{P}_{M'}, \mathbf{P}'_{-M'})$; and so on. This way we either find a safe manipulation or end up with a group of size one, and the single member of this group has a trivially safe manipulation. $\square$

3.6.2 Further remarks

We have considered the safety of group manipulation for several rules, and established conditions for the existence of safe and unsafe manipulations. Theorem 3.6 says that if a rule is manipulable (by a group), then it is safely manipulable. The situation is different for unsafe manipulation. For instance, scoring rules with one jump in a scoring vector turn out to be only safely manipulable, which means that they do not allow for unsafe manipulations at all. The other rules that we considered, are manipulable in an unsafe way. A more detailed analysis, however, shows that even for unsafely manipulable rules the existence of an unsafe manipulation depends on the number of voters and alternatives. For the rules under consideration in this paper, we have established exact bounds for these values. Moreover, even if we know that for the given values of m and n a social choice rule allows for an unsafe manipulation, this does not mean that any group manipulation is unsafe and, thus, risky. It only means that in some preference profile unsafe manipulation is possible. We do not have a general picture of how often unsafe manipulations occur.

4

Simultaneous manipulation under incomplete information

Adapted from: Veselova Y. and D. Karabekyan. *Simultaneous manipulation under incomplete information*. Working paper.

4.1 Introduction

In the majority of publications considering the probability of manipulation there are two important simplifying assumptions made: a) when deciding whether to manipulate or not voters do not take into account possible actions of other voters (so-called "naive" behavior) ; b) voters possess all information about preferences of each other. Although these assumptions are helpful when modeling and conducting experiments on manipulation, some important issues may be ignored by them.

The aim of this research is to combine these two aspects in one model and obtain exact results on manipulability of social choice rules. To deal with the first one we use a version of "safe" strategy used by Slinko and White (2014) and developed by Elkind et al (2015) and Grandi et al (2019). A voter having an incentive to manipulate individually with some strategy is called a Gibbard-Satterthwaite manipulator, or GS-manipulator. This strategy is considered as "safe" if for any possible action of other GS-manipulators this strategy does not lead to a worse result. Apart from the basic model with naive voters (Model 1) we consider two non-naive models of voters' beliefs. In Model 2 each GS-manipulator considers what the result of manipulation will be if all other GS-manipulators act strategically. If it occurs that voting sincerely is better than manipulation for a voter (provided that all other GS-manipulators do manipulate), then manipulation becomes risky and the voter looses an incentive to manipulate. In Model 3 voters believe that some potential manipulators may manipulate and others may not. This creates a higher level of uncertainty. And again if there is a risk of getting worse off by using a manipulation strategy instead of voting sincerely, a voter loses an incentive to manipulate.

Moreover, we add an assumption of incomplete information in the form that was considered in Chapter 2 and first presented by Reijngoud and Endriss (2012). It is assumed that all voters report their sincere preferences to an opinion poll held before voting. Then results of the opinion poll are made public. Since they are represented in an aggregated form (a result of a poll information function, PIF), voters do not know exactly each others' preferences. Thus, a voter has an incentive to manipulate under a given PIF if there is a strategy such that a voter has a chance of getting better off and has no chance of getting worse off with this strategy. We consider three types of PIF. The first one is the unique winner, the result of a rule after tie-breaking. The second one is the set of winners according to a social choice rule. And the third is the ranking of alternatives (a weak order) produced by a rule.

Thus, combining the uncertainty from incomplete information with the uncertainty about other manipulators' actions we get a serious obstacle for a potentially manipulating voter. However, if a voter has an incentive to manipulate despite all these difficulties, this adds to manipulability of a rule. Comparing manipulability of rules in this model is not the same as if voters possess all information and do not consider incentives of others (naive behavior). Moreover, in the literature there are not many papers studying the probability of manipulation with non-naive voters and with incomplete information separately (see Section 2). This work is the first one (to the best of our knowledge) considering both and their mutual influence. We conduct computational experiments calculating exact manipulability indexes for different combination of information types and voters' beliefs about others. Moreover, we prove that for any number of alternatives there is a specific number of voters such that for any greater number of voters manipulation is impossible for any scoring rule when voters have information about a unique winner of an election if voters take into account other manipulators' actions.

The chapter is organized as follows. Section 4.2 contains a literature review. In Section 4.3 we give formal definitions and notations including the description of rules. The next three sections describe and provide results for different behavioral models. Section 4.4 is devoted to Model 1, with naive voters' behavior. Sections 4.5 and 4.6 describe Models 2 and 3 with non-naive behavior of voters and their comparison with Model 1. Section 4.7 concludes.

4.2 Related literature

In the previous section we have considered a strand of literature on manipulability of rules when voters manipulate without thinking of others. Here we aim to consider more thoroughly the body of research devoted to interaction of voters, voting games, and informational aspects of voting for better understanding the place of our work in the literature.

If there is only one GS-manipulator among voters, then she is a pivotal voter and has an opportunity to influence a voting result on her own. If in a society there are several GS-manipulators, the voting result is difficult to predict due to the problem of their interaction. If a voter knows that other GS-manipulators may also decide to act strategically, can this affect her incentives to manipulation? This question was first considered in (Slinko and White, 2014) where each GS-manipulator considers the possibility that other voters with the same preferences (and, consequently, also being GS-manipulators) may strategise. These authors define a strategy to be "safe" if regardless of what subset of other co-minded agents manipulates there is no possibility for a voter to become worse off and for at least one subset she becomes better off. This direction was followed by Hazon and Elkind (2010) and Ianovski et al (2011) who studied computational complexity of finding a safe manipulative vote. The asymptotic probability of a safely manipulable profile for scoring rules was considered by Wilson and Reyhani Shokat Abad (2010).

The next step for a voter is to think not only about her allies, but also about other people who have an incentive to manipulate. So, an extension of this model considers all GS-manipulators as players in a voting game. Then a strategy chosen by a manipulating voter can be called "safe" if it is at least as good as sincere voting for any possible actions of other GS-manipulators. For simplicity it is usually assumed that each manipulator chooses between truth-telling and one strategy chosen according to some optimality principle. This kind of model was considered in (Elkind et al, 2015) and (Grandi et al, 2019). In these publications the existence of pure strategy Nash equilibria is studied for plurality and k-Approval rule with $k = 2, 3, 4$. Thus, this model of voters' behavior is the closest to Model 3 in our study. We also consider only GS-manipulators who choose between sincere voting and one manipulation strategy.

However, the set of players may not be restricted to the set of GS-manipulators. Voters which do not have an incentive to manipulate on their own may also be considered as players and pose a counter-threat to manipulators' actions. Pattanaik (1976b), Pattanaik (1976a) and Barberà (1980) study how coalitions of voters could counter-manipulate in response to individual manipulations and influence incentives of GS-manipulator, and the game between manipulator and counter-manipulator was considered by Grandi et al (2019).

And what if the set of players is the whole set of voters and the set of their strategies is not restricted? This framework is the most general and was considered many times, for example, by Moulin (1981), Myerson and Weber (1993) among the first. Both papers used Nash equilibrium as a solution concept, but faced the problem of a great multiplicity of equilibria. Since lots of these equilibria are weird, there appeared many papers suggesting different ways to eliminate them (see surveys by Meir, 2018 or Slinko, 2019). For example, one way is to assume that voters prefer to abstain or to vote sincerely when they are not pivotal (Desmedt and Elkind, 2010; Obraztsova et al, 2013). Another one is to refine the set of Nash equilibria (e.g. De Sinopoli, 2000; Sertel and Sanver, 2004; Desmedt and Elkind, 2010; Xia and Conitzer, 2010; Obraztsova et al, 2016). Moreover, it is possible to assume bounded rationality of voters, who may not think of other voters being strategic - and we come again to the aforementioned works of Slinko and White (2014), Elkind et al (2015) and Grandi et al (2019).

A topic which follows directly from the previous one is modeling voter levels of rationality. In the structural level-k models of Nagel (1995), Stahl and Wilson (1994) voters of level k of rationality believe that other voters are of level $k-1$. Thus, voters of level 0 do not strategize, voters of level 1 choose their best strategy in assumption that all other voters are of level 0, level 2 voters choose the best response believing that other voters are of level 1. A cognitive hierarchy (CH) model of Camerer et al (2004) has a difference that level-k voters believe that others can have any level from 0 to $k-1$. The CH-model was used in the work by Elkind et al (2020) which focuses on computational complexity of deciding whether a manipulation strategy weakly dominates a sincere vote for a level-2 voter. The CH-model is also applicable in our work. As in the work by Elkind et al (2020), we consider only the first three levels and assume that all voters not being GS-manipulators are of level 0.

A situation when voters do not know anything about actions of other voters, i.e. any voter can potentially submit any preference order, is equivalent to the zero-information case. Can a voter choose a strategy weakly dominating sincere voting in such a situation? Intuitively, no. Indeed, Moulin (1981) mentioned that for voters having no information about other voters' preferences the best strategy is to vote sincerely. For Condorcet-consistent rules and Borda rule it was formally proved by Conitzer et al (2011), and for non-manipulability of scoring rules authors give the bound which was strengthened in the work by Reijngoud and Endriss (2012). However, this is an extreme case and it seems more natural to assume that voters can predict actions of others to some extent (if they know their true preferences, like in the models mentioned above) or know something about preferences of a society (incomplete information).

Models of manipulation under incomplete information attract more attention in recent years. One of the first formal models for strategic voting with partial information was introduced by Conitzer et al (2011). The main focus of the paper was complexity of manipulation. A similar model was introduced by Reijngoud and Endriss (2012), but instead of partial orders, authors consider results of preelection opinion polls. This model was used for the analysis of manipulability of rules under various public information types by Veselova (2020). However, the main application sphere of opinion polls models is iterative voting, where they serve as a coordination device for voters (Myerson and Weber, 1993; Reijngoud and Endriss, 2012; Endriss et al, 2016; Meir et al, 2017). More complex models of information may include not only knowledge of other voters' preferences, but also knowledge about knowledge. For this purpose epistemic logic is used (Ditmarsch et al, 2012; Smaal, 2019).

Thus, the current work considers the model of individual manipulation by voters with bounded rationality under incomplete information. Each possible preference profile creates a game with GS-manipulators as players. As level 2 players in the work by Nagel (1995) and Stahl and Wilson (1994), in our behavioral Model 2 each manipulating agent thinks about others GS-manipulators as being level 1. And in Model 3 they admit that other GS-manipulators may be level 1 or 0 as in CH-models (Camerer et al, 2004; Elkind et al, 2020). In contrast to the mentioned works level 2 voters do not search for the best reply, but check whether their strategy for GS-manipulation under uncertainty still works when we add uncertainty about other voters' actions.

4.3 The Framework

4.3.1 The Model

Let $N = \{1,...,n\}$ be a set of *voters* which have preferences over a set of *alternatives* X, $|X| = m$. $P_i \subseteq X \times X$ is a *preference order* of agent i and it is assumed to be a linear order, i.e. irreflexive, weakly complete and transitive binary relation on X. The set of all linear orders on X is denoted by $L(X)$. A *preference profile* of all voters is denoted by $\mathbf{P} = (P_1,...,P_i,...,P_n)$ and a preference profile of all voters except i is $\mathbf{P}_{-i}$. A contraction of a preference profile onto the set $A \subseteq X$ is $\mathbf{P}/A = (P_1/A,...,P_n/A)$, where $P_i/A = P_i \cap (A \times A)$. The set of all preference profiles is $L(X)^N$ and includes $(m!)^n$ elements.

A mapping $C : L(X)^N \to 2^X \setminus \emptyset$ is called a *social choice correspondence* (SCC). If the result of a SCC contains more than one alternative, then a *tie-breaking rule* (TBR) is used, $T : 2^X \setminus \emptyset \to X$. We use alphabetic tie-breaking: let some linear order on X to be predefined, and when alternatives are tied, we choose the one which dominates all others by P_T, i.e $T(A) = \{a \in A | \forall x \in A, x \neq a\ (a,x) \in P_T\}$. The composition of functions C and T, i.e. $T \circ C$ is denoted by F and is called a *social choice rule* or simply *rule*.

By $v_j(a,\mathbf{P})$ we denote the number of voters having a on the j-th position in preferences (the most preferred alternative gets the 1st position). A vector of positions for an alternative a is $v(a,\mathbf{P}) = (v_1(a,\mathbf{P}),...,v_m(a,\mathbf{P}))$. For a contracted preference profile dimensions of $v(a,\mathbf{P})$ are the same, but the last $m - |A|$ elements of this vector are zeros.

By μ we denote *majority relation*: $a_k \mu a_l$ if $|\{i \in N : a_k P_i a_l\}| > |\{i \in N : a_l P_i a_k\}|$.

A matrix of a majority graph is $MG(\mathbf{P})$, where

$$MG(\mathbf{P})_{kl} = \begin{cases} 1, & \text{if } a_k \mu a_l, \\ -1, & \text{if } a_l \mu a_k, \\ 0, & \text{otherwise.} \end{cases} \tag{4.1}$$

Although a result of a SCC is a set of alternatives which are considered as the best ones, we could also define a ranking of alternatives, or a social ordering, based on this rule. For some rules this ranking is embedded in the procedure, e.g. rules with a scoring function. For other rules such ranking of alternatives could be defined explicitly (see Section 2.2). Thus, a *social ordering* is denoted by a weak order R (irreflexive, transitive, and negatively transitive binary relation), an element of the set of all weak orders on X, $W(X)$.

Similar to Reijngoud and Endriss (2012), Veselova (2020), Endriss et al (2016), we use the *poll information function* $\pi(\mathbf{P})$ (PIF) that shows what kind of information is known by a voter about $\mathbf{P}$. In this paper, we consider 4 types of PIFs.

1. *1Winner*. Information only about the unique winner after the TBR, $\pi_{1Winner}(\mathbf{P}) = F(\mathbf{P})$

2. *Winner*. Information only about the winner(s) before tie-breaking, $\pi_{Winner}(\mathbf{P}) = C(\mathbf{P})$

3. *Rank*. Information about the ranking of alternatives, $\pi_{Rank}(\mathbf{P}) = R$.

4. *Profile*. Information about a full profile is known. It is the classic case of complete information. $\pi_{Profile}(\mathbf{P}) = \mathbf{P}$.

Let $W_i^{\pi(\mathbf{P})}$ be the *information set* of voter i, the set of all possible preference profiles of other voters that are consistent with information $\pi(\mathbf{P})$ of voter i.

$$W_i^{\pi(\mathbf{P})} = \{\mathbf{P}'_{-i} \in L(X)^{N \setminus \{i\}} : \pi(P_i, \mathbf{P}'_{-i}) = \pi(\mathbf{P})\}. \tag{4.2}$$

We say that π is *at least as informative as* π' if for all $\mathbf{P} \in L(X)^N$ and for all $i \in N$ we have $W_i^{\pi(\mathbf{P})} \subseteq W_i^{\pi'(\mathbf{P})}$. In our list of PIFs, they go from the least informative (1Winners-PIF) to the most informative (Profile-PIF).

We define manipulation as follows.

Definition 4.1. *We say that a voter has an incentive to π-manipulate in a preference profile $\mathbf{P}$ under a rule F if there exists some preference $\tilde{P}_i$ such that i) either $F(\tilde{P}_i, \mathbf{P}'_{-i}) = F(\mathbf{P})$ or $F(\tilde{P}_i, \mathbf{P}'_{-i}) P_i F(\mathbf{P})$ for all $\mathbf{P}'_{-i}$ in $W_i^{\pi(\mathbf{P})}$ and ii) $F(\tilde{P}_i, \mathbf{P}'_{-i}) P_i F(\mathbf{P})$ for at least one $\mathbf{P}'_{-i}$ in $W_i^{\pi(\mathbf{P})}$.*

In other words, a voter will manipulate if for every possible profile of her information set she gets at least the same result and for at least one profile she gets a more preferable alternative, on the condition that all others vote sincerely.[1]

If at least one voter has an incentive to π-manipulate in $\mathbf{P}$ under F, then preference profile $\mathbf{P}$ is π-manipulable under F. A voter having an incentive to π-manipulate is called a π-*manipulator*. The set of all π-manipulators in profile $\mathbf{P}$ is denoted by $\Pi(\mathbf{P})$.

Definition 4.2. *A rule F is called susceptible to individual π-manipulation if there exists a profile $\mathbf{P} \in L(X)^N$ and a voter $i \in N$ who has an incentive to π-manipulate in $\mathbf{P}$ under F. If a rule F is not susceptible to individual π-manipulation, it is immune to individual π-manipulation.*

4.3.2 Social choice correspondences

Here we give formal descriptions of social choice correspondences, how the set of winners and the social ranking are determined. Scoring rules, run-off

[1] One can imagine an alternative definition of manipulation with only successful outcomes: a voter manipulates only if her manipulation leads to a success for all preference profiles of her information set. However, in this case manipulation becomes a very rare event and even impossible for some settings.

procedure, STV, and Copeland rule are defined the same way as in Chapter 2.

- *Maximin*. For each alternative $x \in X$ the number of scores is computed as follows $S(x,\mathbf{P}) = min_{a \in X}|\{i \in N : xP_ia\}|$. Then alternatives are ranked according to the number of scores: for all $a,b \in X$ i) $aRb \Leftrightarrow S(a,\mathbf{P}) > S(b,\mathbf{P})$. Alternatives with maximum score win, i.e. $x \in C(\mathbf{P}) \Leftrightarrow x \in argmax_{a \in X} S(a,\mathbf{P})$.

- *Baldwin's rule*. Multistage procedure.
  0) $t := 1$, $X^t := X$, $\mathbf{P}^t := \mathbf{P}$.
  1) For all $a \in X^t$ count Borda score $S^t(a,\mathbf{P}^t) := s_B \cdot v(a,\mathbf{P}^t)$.
  2) Find alternatives with minimum score $A := argmin_{a \in X^t}(S^t(a,\mathbf{P}))$.
  3) If $A = X^t$, then $C(\mathbf{P}) = X^t$ and the procedure terminates. Otherwise, alternatives of A are eliminated, $t := t+1$, $X^t := X^{t-1} \setminus A$, $\mathbf{P}^t := \mathbf{P}/X^t$; for all $x \in X^t$ and $a \in A$ it holds xRa; go to step 1.

- *Nanson's rule*. For each alternative
  0) $t := 1$, $X^t := X$, $\mathbf{P}^t := \mathbf{P}$.
  1) For all $a \in X^t$ $S^t(a,\mathbf{P}^t) := s_B \cdot v(a,\mathbf{P}^t)$.
  2) Compute the average score

$$\bar{r}^t = \sum_{a \in X^t} S^t(a,\mathbf{P}^t)/|X^t|. \tag{4.3}$$

  3) Find alternatives that have the score lower than $\bar{r}^t$: $A := \{a \in X^t | S^t(a,\mathbf{P}) < \bar{r}^t\}$.
  4) If A is empty, then $C(\mathbf{P}) = X^t$ and the procedure terminates. Otherwise, alternatives of A are eliminated, $t := t+1$, $X^t := X^{t-1} \setminus A$, $\mathbf{P}^t := \mathbf{P}/X^t$; for all $x \in X^t$ and $a \in A$ it holds xRa; go to step 1.

- *Black's procedure*. Procedure chooses a Condorcet winner $CW(\mathbf{P}) = [a | \neg\exists x \in X, x\mu a]$ if it exists, and then for all $x \in X$ $CW(\mathbf{P})Rx$. Otherwise, the Borda rule is applied.

- *Kemeny's rule*. Let the distance between linear orders be a function $d(P_i,P_j) = |(P_i \setminus P_j) \cup (P_j \setminus P_i)|$. then R is an ordering such that $R = argmin_{R' \in L(X)} \sum_{i \in N} d(R',P_i)$. The top alternative of R is the winner.

- *Threshold rule*. Alternatives are ordered on the basis of their position vectors: aRb if $v_m(a,\mathbf{P}) < v_m(b,\mathbf{P})$, or if there exist $k \leq m$ such that $v_i(a,\mathbf{P}) = v_i(a,\mathbf{P})$, $i = k-1,...,m$, and $v_k(a,\mathbf{P}) < v_k(b,\mathbf{P})$. In words, we compare the number of worst positions of alternatives. If they are equal, then we compare the number of second-worst positions, and so on. Undominated alternatives are winners: $x \in C(\mathbf{P}) \Leftrightarrow \neg\exists a \in X$, such that aRx.

4.4 Model 1: naive manipulation

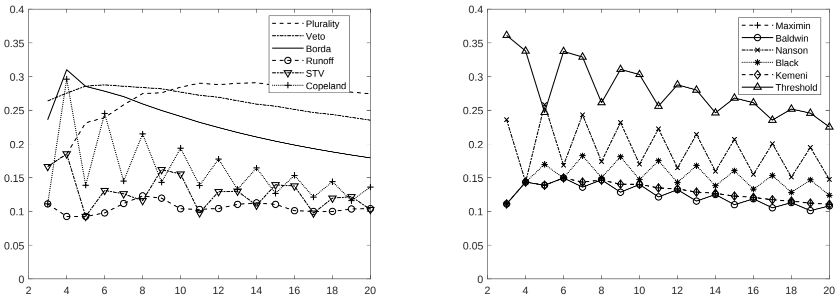
In the Definition 4.1 proposed by Reijngoud and Endriss (2012) a voter does not think about possible actions of others. We would like to consider other assumptions on voters' behavior and formalize this in the term "behavioral model". We consider three behavioral models: Model 1 suggests that voters behavior is naive, they do not think about actions of others (definition of manipulation is the same as the basic one); in Model 2 voters check whether their manipulation strategy still works when all other π-manipulators manipulate as well; in Model 3 voters think that there could be some other voters manipulating. In this section we discuss the basic model of individual manipulation under incomplete information, which assumes that a voter does not think about incentives of other voters. Let us define it formally.

Definition 4.3. *A voter has an incentive to π-manipulate in Model 1 ($M1$) in $\mathbf{P}$ under F if and only if she has an incentive to π-manipulate in $\mathbf{P}$ under F.*

To compare the degree of manipulability of social choice rules one needs some measure. Although in the case of complete information there exist different manipulability measures (Veselova, 2020), we will use the simplest one, which is the proportion of preference profiles where manipulation is possible.

$I^M(m,n,\pi,F)$ – the share of preference profiles where at least one voter has an incentive to π-manipulate in model M under a rule F.

We computed the values of $I^{M1}(m,n,Profile,F)$ in MATLAB for $m=3$, and n from 3 to 20 changing the rule and the type of PIF. A code of the main program calculating indices for this chapter can be seen in Appendix B. First we provide computational results for the simplest model, individual naive manipulation with complete information. Probability of individual naive manipulation for plurality, Borda, veto, runoff, STV, and Copeland rule for $n=3,\ldots,15$ and different PIFs has already appeared in Chapter 2 where we compared it with coalitional manipulation. Here we extend the number of voters and the set of rules and aim to compare these computations with further observations on Model 2 and 3. In order not to overload figures with graphs we illustrate them on two pictures for each figure (see Fig.4.1).

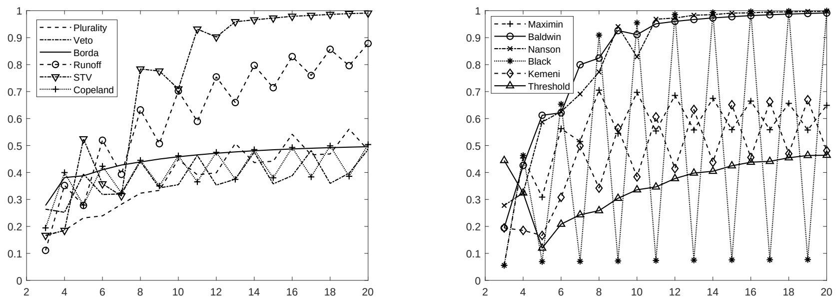


a) b)

Figure 4.1: The values of $I^{M1}(3,n,Profile,F)$

As can be seen, all the computed values of I^{M1} are not greater than 0.4. For most rules the trend is slowly decreasing and among the least manipulable rules are runoff procedure, STV and Baldwin rule. Now change information type to *Rank*-PIF.

Lets us consider I^{M1} for *Rank*-PIF, Fig.4.2. The less information is available to voters, the larger are voters' information sets. Profiles of the information set of voter *i*, $W_i^{\pi(\mathbf{P})}$, all give the same ranking when voter *i* does not manipulate. However, for all rules under consideration the result after manipulation is



a) b)

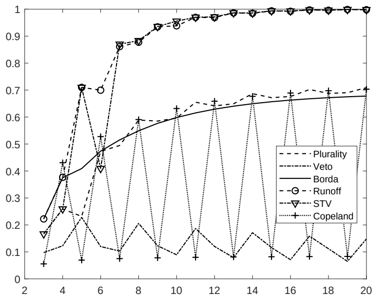
Figure 4.2: The values of $I^{M1}(3, n, Rank, F)$

not always the same for all profiles of voter's information set $W_i^{\pi(\mathbf{P})}$. In other words, information about ranks of alternatives does not allow a voter to compute the result of manipulation for any strategy she chooses (using the term from Reijngoud and Endriss, 2012, these rules are not strongly computable from Rank-images).

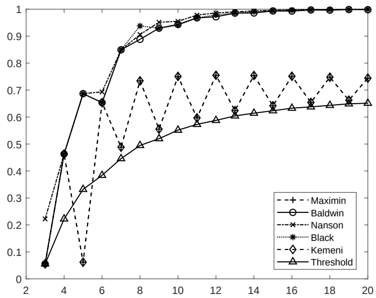
According to the definition, using a manipulation strategy must lead to a better result in some profiles of $W_i^{\pi(\mathbf{P})}$ and must not lead to a worse in others. The voter cannot distinguish between different profiles of her information set. So, if voter i has an incentive to manipulate in $\mathbf{P}$, then all profiles of $W_i^{\pi(\mathbf{P})}$ are manipulable, even those where voter i cannot really change anything. A reader can find a more detailed explanation of this effect in (Veselova, 2020).

For most cases the values of I^{M1} for *Rank*-PIF are greater than for *Profile*-PIF. And the general trend of $I^{M1}(3, n, Winner, F)$ is increasing for most rules. Moreover, for STV, Baldwin, and Nanson rules this index is very close to 1 when $n > 12$.

For *Winner*-PIF (Fig.4.3) even more rules show graphs going to 1 with growing n. These are the same as for *Rank*-PIF plus Black's procedure, maximin, and runoff. Except for Copeland and veto rule, all other rules demonstrate



a)

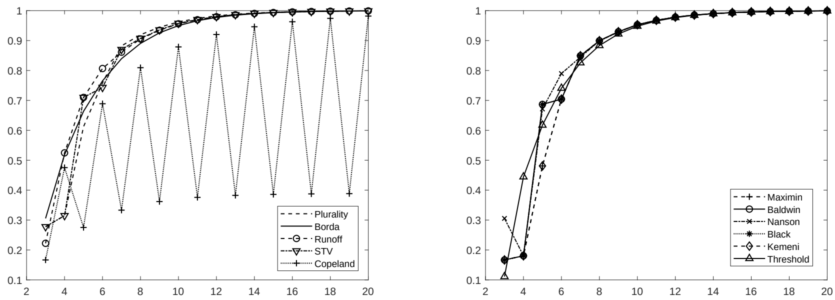


b)

Figure 4.3: The values of $I^{M1}(3, n, Winner, F)$ with naive model of manipulation

values of I^{M1} for *Winner*-PIF greater than for *Rank*-PIF for almost all n. The graph for Copeland rule has a higher amplitude and manipulability of veto rule decreased considerably for all n.

For the least informative PIF, *1Winner*-PIF (Fig.4.4), all rules except for Copeland and veto, merge near 1 for $n > 6$. Peaks of the graph for Copeland rule also approach 1, and veto rule disappears from the figure due to the zero-manipulability for *1Winner*-PIF (Reijngoud and Endriss, 2012).



a) b)

Figure 4.4: The values of $I^{M_1}(3, n, 1Winner, F)$

4.5 Model 2: manipulation with respect to all other π-manipulators

In this section we investigate what will change if a manipulating voter takes into account manipulations of others. In general, each π-manipulator has a set of strategies with that she has an incentive to π-manipulate. For simplicity, for each π-manipulator j in $\Pi(\mathbf{P}) \setminus \{i\}$ we fix one strategy $\hat{P}_j$ that she may use or not, which we will refer to as *π-manipulation strategy*.[2] In other words, voter i thinks that another voter j can π-manipulate only with a strategy $\hat{P}_j$. A π-manipulation strategy $\hat{P}_j$ is chosen by the principle of the best winning alternative and in case of equality we choose it alphabetically.

We denote by $\hat{\mathbf{P}}$ a preference profile obtained from $\mathbf{P}$ with the difference that all voters from $\Pi(\mathbf{P})$ use their π-manipulation strategy.

Definition 4.4. *We say that a voter i has an incentive to π-manipulate in Model 2 ($M2$) in a preference profile $\mathbf{P}$ under a rule F if there is a strategy $\tilde{P}_i$*

[2] Otherwise, we should have considered all combinations of strategies for π-manipulators. This is not only computationally hard, but also not very interesting. For the 3-alternatives case if there are several manipulation strategies, then most likely they are equivalent (always give the same result).

such that:
i) voter i has an incentive to π-manipulate with $\tilde{P}_i$ in $\boldsymbol{P}$ under F;
ii) it is either $F(\tilde{P}_i, \hat{\boldsymbol{P}}'_{-i}) = F(P_i, \hat{\boldsymbol{P}}'_{-i})$ or $F(\tilde{P}_i, \hat{\boldsymbol{P}}'_{-i}) P_i F(P_i, \hat{\boldsymbol{P}}'_{-i})$ for all $\boldsymbol{P}'_{-i}$ in $W_i^{\pi(\boldsymbol{P})}$.

In words, condition ii) requires that a strategy $\tilde{P}_i$ is still not worse than truth-telling provided that all other π-manipulators decide to manipulate. Thus, if some strategy $\tilde{P}_i$ for voter i fails to dominate truth-telling, we need to check another strategy, and so on. If none of them is dominant, then voter i does not have an incentive to π-manipulate in Model 2. So, although we fix manipulation strategies for other π-manipulators, we still need to consider all manipulation strategies for voter i. However, for the 3-alternatives case this occurs to be excessive, since different manipulation strategies do not differ in terms of the result.

	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Plurality	.06	.04	.03	.04	.03	.02	.03	.02	.02	.02	.02	.01	.01	.01	.01	.01	.01	.01
Veto	.06	.07	.08	.09	.09	.08	.08	.08	.07	.07	.07	.06	.06	.06	.05	.05	.05	.04
Borda	.00	.04	.03	.03	.02	.02	.01	.01	.01	.01	.01	.01	.01	.00	.00	.00	.00	.00
Run-off	.00	.00	.00	.00	.00	.01	.01	.00	.00	.00	.01	.00	.00	.00	.00	.00	.00	.00
STV	.06	.04	.00	.01	.01	.00	.03	.00	.00	.01	.00	.00	.01	.00	.00	.01	.00	.00
Copeland	.00	.04	.00	.03	.00	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
Maximin	.00	.00	.00	.00	.00	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
Baldwin	.00	.00	.00	.00	.00	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
Nanson	.00	.00	.07	.01	.06	.03	.04	.03	.03	.02	.02	.02	.01	.01	.01	.01	.01	.01
Black	.00	.00	.02	.00	.03	.01	.03	.01	.04	.01	.03	.01	.03	.01	.03	.01	.02	.01
Kemeny	.00	.00	.00	.00	.00	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
Threshold	.17	.04	.02	.12	.05	.05	.04	.05	.07	.04	.05	.06	.04	.04	.05	.04	.03	.04

Table 4.1: $I^{M1}(3, n, Profile, F) - I^{M2}(3, n, Profile, F)$

The first series of experiments refers to the complete information case. However, instead of values $I^{M2}(3, n, Profile, F)$ we show results for $I^{M1}(3, n, Profile, F) - I^{M2}(3, n, Profile, F)$, which is more illustrative. Particularly, it shows in which proportion of naively manipulable preference profiles (manipulable in M1) a threat of having something bad as the result in case of simultaneous manipulation destroys all incentives to manipulate. As can be seen from Table 4.1, this situation occurs most often under veto and threshold rule. There is almost no difference between Model 1 and Model 2 for runoff procedure, maximin, Baldwin and Kemeny rules under Profile-PIF. Except for veto and threshold rules, all other values do not exceed 0.07 and most of them are very close to 0. This suggests that having a threat to loose when all other manipulators do manipulate cannot be a serious obstacle to strategic voting, simply because it is quite rare.

Then we change information type to less and less informative and two opposite effects start to work together. On the one hand, the probability to meet a π-manipulable profile grows (see Tables.4.2-4.4) in general for many rules. On the other hand, the risk to result with something worse than initially due to simultaneous manipulation combined with uncertainty about preferences of others makes many profiles non-manipulable.

	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Plurality	.00	.07	.12	.12	.18	.23	.25	.36	.31	.33	.43	.37	.39	.47	.40	.42	.50	.42
Veto	.03	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Borda	.28	.19	.08	.08	.07	.06	.07	.08	.09	.10	.10	.11	.11	.12	.12	.12	.13	.13
Run-off	.11	.35	.28	.52	.29	.42	.36	.54	.27	.42	.24	.26	.10	.21	.00	.09	.00	.00
STV	.00	.07	.52	.24	.00	.72	.23	.36	.74	.37	.00	.58	.27	.00	.00	.21	.00	.00
Copeland	.19	.25	.21	.21	.25	.22	.27	.22	.29	.23	.29	.23	.30	.23	.30	.23	.30	.24
Maximin	.06	.21	.02	.16	.17	.15	.08	.09	.08	.00	.08	.00	.08	.00	.08	.00	.08	.00
Baldwin	.19	.19	.30	.32	.19	.40	.31	.27	.30	.32	.29	.33	.28	.34	.29	.32	.29	.32
Nanson	.28	.17	.23	.30	.00	.28	.29	.05	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Black	.06	.22	.02	.13	.03	.32	.04	.31	.03	.41	.04	.41	.04	.41	.04	.33	.04	.33
Kemeny	.19	.16	.15	.21	.37	.32	.42	.29	.45	.31	.48	.33	.49	.34	.50	.35	.51	.36
Threshold	.06	.17	.11	.19	.24	.23	.30	.34	.33	.37	.40	.39	.42	.44	.43	.45	.46	.45

Table 4.2: $I^{M2}(3,n,Rank,F)$

We list the most important observations and put them into groups.

Rank-PIF:

- Veto rule is non-manipulable for $n \geq 4$ and Nanson's is non-manipulable for $n \geq 11$;
- Borda and Maximin rules are the least manipulable;

	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Plurality	.00	.10	.12	.14	.26	.28	.25	.33	.33	.30	.37	.36	.34	.39	.37	.36	.40	.38
Veto	.03	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Borda	.22	.31	.15	.15	.16	.17	.19	.20	.22	.23	.24	.25	.26	.27	.27	.28	.28	.29
Run-off	.22	.38	.71	.70	.65	.59	.55	.59	.25	.50	.35	.20	.30	.26	.00	.24	.00	.00
STV	.00	.10	.71	.29	.00	.60	.28	.14	.00	.38	.00	.00	.00	.00	.00	.00	.00	.00
Copeland	.06	.37	.00	.29	.00	.32	.00	.35	.00	.36	.00	.38	.00	.38	.00	.39	.00	.39
Maximin	.06	.22	.02	.13	.44	.12	.21	.09	.24	.00	.27	.00	.28	.00	.30	.00	.31	.00
Baldwin	.06	.22	.64	.13	.17	.42	.00	.09	.00	.08	.00	.08	.00	.08	.00	.07	.00	.07
Nanson	.22	.22	.25	.25	.00	.21	.04	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Black	.06	.22	.64	.13	.00	.32	.00	.45	.00	.17	.00	.16	.00	.15	.00	.07	.00	.07
Kemeny	.06	.22	.02	.13	.44	.12	.21	.09	.24	.00	.27	.00	.28	.00	.30	.00	.31	.00
Threshold	.06	.22	.32	.14	.19	.18	.23	.25	.24	.27	.29	.28	.30	.31	.30	.32	.32	.31

Table 4.3: $I^{M2}(3,n,Winner,F)$

- Plurality, Kemeny, and threshold rules are the most manipulable.

Winner-PIF:

- STV rule is non-manipulable for $n \geq 13$;
- For Copeland, maximin, Baldwin, Black's and Kemeny's rules there is an alternation of zero and non-zero values of manipulability index;
- Plurality, Borda, and threshold rules are the most manipulable.

1Winner-PIF:

- For all rules except for Copeland rule there is a value n' such that all values of $I^{M2}(m,n,\pi,F)$ are zeros for all $n \geq n'$.

We also need to mention that the computed values for Kemeny and maximin rules coincide for all PIFs except *Rank*-PIF. The coincidence is explained by the same results these rules give for the 3-alternative case. And the difference for *Rank*-PIF is caused by the different ways the *Rank*-PIF is constructed for these rules.

	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Plurality	.00	.15	.00	.17	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Veto	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Borda	.31	.27	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Run-off	.22	.52	.71	.71	.65	.32	.55	.45	.25	.26	.35	.00	.30	.00	.00	.00	.00	.00
STV	.00	.15	.71	.29	.00	.62	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Copeland	.17	.21	.00	.15	.00	.20	.00	.23	.00	.26	.00	.28	.00	.29	.00	.30	.00	.31
Maximin	.17	.18	.21	.57	.26	.63	.00	.24	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Baldwin	.17	.18	.43	.57	.17	.50	.00	.24	.00	.26	.00	.28	.00	.00	.00	.00	.00	.00
Nanson	.31	.18	.00	.57	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Black	.17	.18	.43	.57	.00	.63	.00	.24	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Kemeny	.17	.18	.21	.57	.26	.63	.00	.24	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Threshold	.11	.44	.62	.13	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00

Table 4.4: $I^{M2}(3,n,1Winner,F)$

From the tables it can be seen that if we get zero manipulability in some case and increase the level of uncertainty, then zero manipulability is preserved. For example, we fix a PIF and go from Model 1 to Model 2.

Proposition 4.1. *For any PIF π, for any rule F, the number of voters n, and the number of alternatives m if $I^{M1}(m,n,\pi,F)=0$, then $I^{M2}(m,n,\pi,F)=0$.*

Proof. The proof follows directly from the definitions of manipulation in Model 1 and Model 2. To have an incentive to manipulate in Model 2, a voter needs to have an incentive to manipulate in Model 1. If $I^{M1}(m,n,\pi,F)=0$, then no voter has in incentive to manipulate in Model 1, and, consequently, also not in Model 2. So, $I^{M2}(m,n,\pi,F)=0$. □

The main observation for Model 2 and *1Winner*-PIF is that for all rules except Copeland manipulability indexes become 0 when the number of voters exceeds a certain value. It turns out that the same holds for any given number of alternatives for any scoring rule. First we prove Lemma 4.1 and then use it in Theorem 11 stating this.

For a scoring vector s, a *jump* is a non-zero difference between two adjacent scoring values. If s has r jumps, then this means that there are distinct

$k_1,\ldots,k_r \in \{1,\ldots,m-1\}$ such that $s_{k_1} - s_{k_1+1} > 0,\ldots,s_{k_r} - s_{k_r+1} > 0$, while all other differences are zero. Let $\Delta_j = s_{k_j} - s_{k_j+1}$ for $j = 1,\ldots,r$ denote the j-th jump.

Lemma 4.1. *For any scoring rule F, any number of alternatives m, any voter i, any jump Δ_j, and any two distinct alternatives a_h and a_l there is a number of voters n^*, such that for all $n > n^*$ there exists $\boldsymbol{P} \in L(X)^N$ such that*
1) $P_i = (a_1,a_2,...,a_m)$;
2) for all $a_g \in X \setminus \{a_h,a_l\}$ $S(a_h,\boldsymbol{P}) > S(a_g,\boldsymbol{P})$;
3) $S(a_h,\boldsymbol{P}) - S(a_l,\boldsymbol{P}) = \Delta_j$ ($S(a_h,\boldsymbol{P}) - S(a_l,\boldsymbol{P}) = 0$).

Proof. 1) For simplicity, let voter i be the first with a preference order $P^1 = (a_1,...,a_h,...,a_l,...,a_m)$ (which means $a_1P^1a_hP^1a_lP^1a_m$, dots mean there can be other alternatives). First, we consider the case when $h < l$. Then we denote $P^2 = (a_l,...,a_1,...,a_h,...)$, $P^3 = (a_h,...,a_l,...,a_1,...)$ (alternatives a_1, a_h, and a_l are on the same places, but cycled according to a permutation $(a_1\, a_l\, a_h)$), $P^4 = (a_1,...,a_l,...,a_h,...,a_m)$ (the same as P^1, but a_h and a_l switched), $P^5 = (a_h,a_l,...)$, $P^6 = (a_l,a_h,...)$, $P^7 = (...,a_h|a_l,...)$ (the line $|$ denotes the position of j-th jump, Δ_j).

2) Let us prove by construction that there exists a profile with $S(a_h,\mathbf{P}) - S(a_l,\mathbf{P}) = \Delta_j$. For an odd n: $\mathbf{P}' = (P^1,P^4,qP^5,qP^6,P^7)$. For an even n: $\mathbf{P}'' = (P^1,P^2,P^3,qP^5,qP^6,P^7)$. A profile with $S(a_h,\mathbf{P}) = S(a_l,\mathbf{P})$ is constructed the same way by leaning out P^7.

3) Now we prove that the condition $\forall a_g \in X \setminus \{a_h,a_l\}$ $S(a_h,\mathbf{P}) > S(a_g,\mathbf{P})$ could be satisfied for the constructed profiles. First, in preferences of type P^5 and P^6 let all other $m-2$ alternatives be cycled. The number of scores got by a_h and a_l in (qP^5,qP^6) is $qs_1 + qs_2$. Let $h = [2q/(m-2)]$, which is the number of whole cycles in (qP^5,qP^6). The number of scores got by any alternative from $X \setminus \{a_h,a_l\}$ is not greater than $h(s_3 + ... + s_m) + (2q - h(m-2))s_3 \leq hs_m + (2q-h)s_3$. Since $s_1 \geq s_2 \geq s_3 \geq ... \geq s_m$ and $s_1 > s_m$, $qs_1 + qs_2 > hs_m + (2q-h)s_3$ and the difference is not less then $min(h,q)(s_1 - s_m)$ (so, for this difference to be positive for all alternatives in $X \setminus \{a_h,a_l\}$, there should be at least one cycle of $m-2$ alternatives, i.e. $m-2 < 2q$). Thus, by taking

q big enough we can make scores of a_h in $\mathbf{P}$ be higher than scores of any other alternative in each type of $\mathbf{P}$ constructed in 2). If the condition $\forall a_g \in X \setminus \{a_h, a_l\}$ $S(a_h, \mathbf{P}) > S(a_g, \mathbf{P})$ is satisfied for some q, then for $q' = q + 1$ it is also satisfied.

$$n' = 3 + min\{q : \forall a_g \in X \setminus \{a_h, a_l\} S(a_h, \mathbf{P}') > S(a_j, \mathbf{P}')\},$$

$$n'' = 4 + min\{q : \forall a_g \in X \setminus \{a_h, a_l\} S(a_h, \mathbf{P}'') > S(a_g, \mathbf{P}'')\},$$

$$n^* = max(n', n'').$$

Therefore, for all $n > n^*$ there exists a preference profile with $P_i = (a_1, a_2, ..., a_m)$, $\forall a_g \in X \setminus \{a_h, a_l\}$ $S(a_h, \mathbf{P}) > S(a_g, \mathbf{P})$ and $S(a_h, \mathbf{P}) - S(a_l, \mathbf{P}) = \Delta_j$.

4) If $h > l$, i.e. a_h is less preferred than a_l by voter i, then we switch alternatives a_h and a_l in P^1, P^2, P^3, and P^4 with all other parts of the proof staying the same. □

Theorem 4.1. *For any scoring rule F and any number of alternatives m there is a finite number of voters n^*, such that for all $n > n^*$ it holds $I^{M2}(m, n, 1Winner, F) = 0$.*

Proof. Let $X = \{a_1, ..., a_m\}$. Consider a scoring rule with a scoring vector $s = (s_1, s_2, ..., s_m)$, the first jump in s goes after s_k, $s_k - s_{k+1} = \Delta_1$.

1) We prove that voter i with preferences $a_1 P_i a_2 P_i ... P_i a_m$ has no incentive to manipulate (in Model 1) under *1Winner*-PIF if $F(\mathbf{P}) \in \{a_1, a_2, ..., a_{k+1}\}$.

1.1) If $F(\mathbf{P}) = a_1$, then there is no need for voter i to misrepresent preferences, since it is the best alternative for i.

1.2) Suppose that $F(\mathbf{P}) = b$, $b \in \{a_2, a_3, ..., a_{k+1}\}$ and i manipulates in favor of some a, such that aPb. If i puts alternative a higher (if a is not a_1), then nothing changes for a since $s_1 = ... = s_k$. Thus, i could only put b lower in $\tilde{P}_i$, but then some alternative $c \in \{a_{k+2}, ..., a_m\}$ goes higher. If $b = a_{k+1}$ and there are no jumps in s after $k+1$, then putting b lower will not change the scores of

b and c and no manipulation is possible in this case. In other cases b gets $-A$ scores and c gets $+A$ scores, where $A = \alpha_1\Delta_1 + \alpha_2\Delta_2 + ...$ and $\alpha_j \in \{0,1\}$.

Thus, if there exists $\mathbf{P}' = (P_i, \mathbf{P}'_{-i})$, $\mathbf{P}'_{-i} \in W_i^{\pi(\mathbf{P})}$, such that $S(b,\mathbf{P}') - S(c,\mathbf{P}') = min(\Delta_1, \Delta_2)$, then c wins in $(\tilde{P}_i, \mathbf{P}'_{-i})$, since it gets $+A$ scores and b gets $-A$ scores. By Lemma 4.1 such preference profile exists for all $n > \hat{n}$. It means that there is a chance of getting c as a result which is worse than b for i. Therefore, i does not have an incentive to *1Winner*-manipulate when $F(\mathbf{P}) \in \{a_1, a_2, ..., a_{k+1}\}$.

2) Now prove that if $F(\mathbf{P}) \in \{a_{k+2}, ..., a_m\}$, then voter i with preferences $a_1 P_i a_2 P_i ... P_i a_m$ has an incentive to *1Winner*-manipulate. Suppose, $F(\mathbf{P}) = c$ and c is on $k+t$-th place in P_i, where $t \in \{2, ..., m\}$. Then manipulation in favor of some $b \in \{a_{k+1}, ..., a_{k+t-1}\}$ is possible: voter i switches alternatives $a \in \{a_1, ..., a_k\}$ and b. Since it does not matter which $a \in \{a_1, ..., a_k\}$ to choose for switching with b, we can assume that it is a_k. By the principle of the best winning alternative b must be a_{k+1}. After this manipulation b gets $+\Delta_1$ scores and a gets $-\Delta_1$ scores, while scores of other alternatives do not change. Thus, if there exists $\mathbf{P}' = (P_i, \mathbf{P}'_{-i})$, $\mathbf{P}'_{-i} \in W_i^{\pi(\mathbf{P})}$, such that $S(c,\mathbf{P}') - S(b,\mathbf{P}') = \Delta_1$, then b wins in $(\tilde{P}_i, \mathbf{P}'_{-i})$ provided that bP_Tc. If cP_Tb, then we need a profile $\mathbf{P}'' = (P_i, \mathbf{P}''_{-i})$, $\mathbf{P}''_{-i} \in W_i^{\pi(\mathbf{P})}$, such that $S(c,\mathbf{P}'') = S(b,\mathbf{P}'')$ for b to win after getting $+\Delta_1$ scores. By Lemma 4.1, such profiles exist for all $n > \check{n}$. Therefore, in some preference profiles of i's information set b wins, and there is no risk of getting a worse alternative as a result, so, i has an incentive to manipulate in Model 1 under *1Winner*-PIF when $F(\mathbf{P}) \in \{a_{k+2}, ..., a_m\}$ for all $n > \check{n}$.

3) Take a voter i with preferences $a_1 P_i a_2 P_i ... P_i a_m$ having an incentive to manipulate (in Model 1). If in $W_i^{\pi(\mathbf{P})}$ there is at least one preference profile such that $F(P_i, \hat{\mathbf{P}}'_{-i}) P_i F(\tilde{P}_i, \hat{\mathbf{P}}'_{-i})$, then voter i has no incentive to manipulate in Model 2. We prove the existence of such a profile $\mathbf{P}'$ by construction.

For simplicity, denote $c = a_{k+t}$, $b = a_{k+1}$, $a = a_k$.

We construct $\mathbf{P}'$ with the following preference orders: preferences of voter i $P^1 = (...a|b...c...)$, $P^2 = (...b|a...c...)$, $P^3 = (...|c...a...)$, $P^4 = (...c|...a...)$ (in P^1 and P^2 c on $k+t$-th place, in P^3 and P^4 a on $k+t$-th place), $P^5 = (ca...|...)$, $P^6 = (ac...|...)$, $P^7 = (c|a...)$. Thus, we have voter i with a preference order P^1, having an incentive to manipulate in Model 1. Voters with preference order P^2 are also GS-manipulators, but their manipulation is an opposite one (putting b on $k+1$-th place and a on k-th). Voters with preference orders P^3, P^4, P^5, and P^6 do not have an incentive to manipulate in Model 1 since the winning alternative is not lower than $k+1$ place.

The way of construction depends on a tie-breaking between c and a. Consider two cases: aP_Tc and cP_Ta. For all the following cases we assume that the condition $\forall x \in X \setminus \{a,c\}$ $S(c) > S(x)$ is satisfied. Later (in part 6 of the proof) we will show that there is a finite number of voters n^*, such that for all $n > n^*$ this is true.

4) If aP_Tc, then $S(c) > S(a)$.
4.1) n is even, $k \in \{1,\dots,m-2\}$, $\mathbf{P}' = (P^1, 2P^2, 2P^3, P^4, qP^5, qP^6, 2P^7)$. For this profile, $S(c,\mathbf{P}') - S(a,\mathbf{P}') = 2\Delta_1$ and $F(\mathbf{P}') = c$. If all voters from $\Pi(\mathbf{P}')$ manipulate, then $S(c,(\tilde{P}_i,\hat{\mathbf{P}}'_{-i})) - S(a,(\tilde{P}_i,\hat{\mathbf{P}}'_{-i})) = \Delta_1$. If i does not manipulate, and $\Pi(\mathbf{P}') \setminus \{i\}$ does, then $S(c,(P'_i,\hat{\mathbf{P}}'_{-i})) = S(a,(P'_i,\hat{\mathbf{P}}'_{-i}))$ and $F(P'_i,\hat{\mathbf{P}}'_{-i}) = a$.

4.2) n is odd, $k \in \{1,\dots,m-2\}$, $\mathbf{P}' = (P^1, P^2, P^3, P^4, qP^5, qP^6, P^7)$. For this profile, $S(c,\mathbf{P}') - S(a,\mathbf{P}') = \Delta_1$ and $F(\mathbf{P}') = c$. If all voters from $\Pi(\mathbf{P}')$ manipulate, then $S(c,(\tilde{P}_i,\hat{\mathbf{P}}'_{-i})) - S(a,(\tilde{P}_i,\hat{\mathbf{P}}'_{-i})) = \Delta_1$. If i does not manipulate, and $\Pi(\mathbf{P}') \setminus \{i\}$ does, then $S(c,(P'_i,\hat{\mathbf{P}}'_{-i})) = S(a,(P'_i,\hat{\mathbf{P}}'_{-i}))$ and $F(P'_i,\hat{\mathbf{P}}'_{-i}) = a$.
5) If cP_Ta, then $S(a) \geq S(c)$.
5.1) n is even, $k \in \{1,\dots,m-2\}$, $\mathbf{P}' = (P^1, P^2, P^3, P^4, qP^5, qP^6)$. For this profile, $S(a,\mathbf{P}') = S(c,\mathbf{P}')$ and $F(\mathbf{P}') = c$. If all voters from $\Pi(\mathbf{P}')$ manipulate, then $S(c,(\tilde{P}_i,\hat{\mathbf{P}}'_{-i})) = S(a,(\tilde{P}_i,\hat{\mathbf{P}}'_{-i}))$ and again $F(\tilde{P}_i,\hat{\mathbf{P}}'_{-i}) = c$. If i does not manipulate and $\Pi(\mathbf{P}') \setminus \{i\}$ (which is only one voter with preferences P^2 in this case) does, then $S(c,(P'_i,\hat{\mathbf{P}}'_{-i})) < S(a,(P'_i,\hat{\mathbf{P}}'_{-i})) = S(a,\mathbf{P}') + \Delta_1$ and

$F(P_i', \hat{\mathbf{P}}'_{-i}) = a$.
5.2) n is odd, $k = 1$, $\mathbf{P}' = (P^1, 2P^2, 2P^3, P^4, (q+1), qP^6)$. For this profile, $S(c, \mathbf{P}') - S(a, \mathbf{P}') = \Delta_1$ and $F(\mathbf{P}') = c$. If all voters from $\Pi(\mathbf{P}')$ manipulate, then $S(a, (\tilde{P}_i, \hat{\mathbf{P}}'_{-i})) = S(a, \mathbf{P}') + 2\Delta_1 - \Delta_1$, $S(c, (\tilde{P}_i, \hat{\mathbf{P}}'_{-i})) = S(a, (\tilde{P}_i, \hat{\mathbf{P}}'_{-i}))$ and again $F(\tilde{P}_i, \hat{\mathbf{P}}'_{-i}) = c$. If i does not manipulate, $S(c, (P_i', \hat{\mathbf{P}}'_{-i})) < S(a, (P_i', \hat{\mathbf{P}}'_{-i})) = S(a, \mathbf{P}') + 2\Delta_1$ and $F(P_i', \hat{\mathbf{P}}'_{-i}) = a$.
5.3) n is odd, $k \in \{2, \ldots, m-2\}$, $\mathbf{P}' = (P^1, P^2, P^3, P^4, (q+1)P^5, qP^6)$. All relations between scores of c and a are the same as in 5.1).

6) Now we need to show that it is possible to have the condition $\forall x \in X \setminus \{a, c\}\ S(c) > S(x)$ satisfied and find out when it is possible. We assume that in preferences of type P^5 and P^6 all other $m-2$ alternatives are cycled, i.e. $(a\, c\, x_1\, x_2 \ldots x_{m-1})$, $(a\, c\, x_{m-1}\, x_1\, x_2 \ldots x_{m-2})$, etc. Let $h = [2q/(m-2)]$, which is the number of whole cycles in (qP^5, qP^6).

The number of scores got by c in (qP^5, qP^6) is $qs_1 + qs_2$. The number of scores got by any alternative from $X \setminus \{a, c\}$ is not greater than $h(s_3 + \ldots + s_m) + (2q - h(m-2))s_3 \le hs_m + (2q - h)s_3$. Since $s_1 \ge s_2 \ge s_3 \ge \ldots \ge s_m$ and $s_1 > s_m$, $qs_1 + qs_2 > hs_m + (2q - h)s_3$ and the difference is not less then $min(h, q)(s_1 - s_m)$. Thus, by taking q big enough we can make scores of c in $\mathbf{P}$ be higher than scores of any other alternative.

Let such number of voters that the condition $\forall x \in X \setminus \{a, c\}\ S(c) > S(x)$ is satisfied for all cases 4.1), 4.2), 5.1), 5.2) be denoted by $\grave{n}$. Summing up, having a fixed number of alternatives, voter i with preferences $a_1 P_i a_2 P_i \ldots P_i a_m$ has no incentive to manipulate (in Model 1) under a scoring rule and *1Winner*-PIF if $F(\mathbf{P}) \in \{a_1, a_2, \ldots, a_{k+1}\}$ for all $n > \hat{n}$, but has an incentive to manipulate if $F(\mathbf{P}) \in \{a_{k+2}, \ldots, a_m\}$ for all $n > \breve{n}$. However, voter i does not have an incentive to manipulate under the same conditions in Model 2 for all $n > \grave{n}$. Thus, $I^{M2}(m, n, 1Winner, F) = 0$ for all $n > n^* = max(\hat{n}, \breve{n}, \grave{n})$. □

4.6 Model 3: manipulating subsets of GS-manipulators

In this model, a voter assumes that each of the other π-manipulators may manipulate or not. Thus, it is necessary to consider all possible subsets K of $\Pi(\mathbf{P}) \setminus \{i\}$. Then voter i compares the result with using strategy $\tilde{P}_i$ and voting sincerely provided that voters from K manipulate and others do not.

It is quite natural to assume that not all potential manipulators will actually manipulate. The problem is that for a voter it is computationally hard to think about actions of all possible subsets of π-manipulators. We show, however, that Model 3 does not differ very much from the previous one. Results are certainly strengthened, but not essentially different. Let us give a formal definition of manipulation in Model 3.

Definition 4.5. *We say that a voter i has an incentive to π-manipulate in Model 3 (M3) in a preference profile $\boldsymbol{P}$ under a rule F if there is a strategy $\tilde{P}_i$ such that:*
i) voter i has an incentive to π-manipulate with $\tilde{P}_i$ in $\boldsymbol{P}$ under F;
ii) it is either $F(\tilde{P}_i, (\hat{\boldsymbol{P}}'_K, \boldsymbol{P}'_{-K-i})) = F(P_i, (\hat{\boldsymbol{P}}'_K, \boldsymbol{P}'_{-K-i}))$ or $F(\tilde{P}_i, (\hat{\boldsymbol{P}}'_K, \boldsymbol{P}'_{-K-i})) P_i F(P_i, (\hat{\boldsymbol{P}}'_K, \boldsymbol{P}'_{-K-i}))$ for all $\boldsymbol{P}'_{-i}$ in W_i^{π} and for all $K \subseteq \Pi(\boldsymbol{P}'_{-i})$, $K \neq \emptyset$.

Analogously, condition ii) requires that a strategy $\tilde{P}_i$ is still not worse than truth-telling if some of π-manipulators also decide to manipulate.

Again, we first consider the difference of manipulability indexes in Model 2 and Model 3 for the complete information case. On average, this difference has higher values than the difference between Model 1 and Model 2. This shows that the number of profiles where a subset of manipulators can spoil the result by their strategic actions is larger than the number of profiles where this can be done by all manipulators.

High values are more likely for bigger numbers of voters, since the number of GS-manipulators is greater in this case. The highest values of difference cor-

	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Plurality	.00	.00	.02	.03	.02	.03	.04	.03	.04	.04	.03	.04	.04	.03	.03	.03	.03	.03
Veto	.00	.01	.02	.03	.03	.05	.06	.07	.08	.09	.09	.10	.11	.11	.12	.12	.13	.13
Borda	.00	.00	.01	.03	.03	.05	.05	.06	.06	.07	.07	.07	.07	.07	.08	.08	.08	.08
Run-off	.00	.00	.00	.00	.01	.03	.02	.01	.01	.02	.02	.03	.02	.02	.02	.02	.03	.04
STV	.00	.00	.00	.00	.01	.00	.04	.07	.00	.02	.05	.02	.04	.07	.01	.03	.05	.03
Copeland	.00	.00	.00	.03	.00	.04	.00	.04	.00	.04	.01	.03	.01	.03	.01	.03	.01	.03
Maximin	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.01	.01	.01	.01	.01	.01	.01
Baldwin	.00	.00	.00	.00	.00	.00	.00	.00	.01	.00	.01	.01	.02	.01	.02	.01	.02	.02
Nanson	.00	.00	.01	.01	.08	.02	.12	.04	.14	.06	.16	.08	.16	.09	.17	.09	.17	.10
Black	.00	.00	.01	.00	.02	.00	.03	.01	.04	.01	.04	.02	.05	.02	.05	.03	.06	.03
Kemeny	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.01	.01	.01	.01	.01	.01	.01
Threshold	.00	.02	.00	.06	.05	.03	.11	.07	.05	.11	.09	.08	.12	.11	.10	.13	.12	.12

Table 4.5: $I^{M2}(3,n,Profile,F) - I^{M3}(3,n,Profile,F)$

respond to veto, Borda, Nanson's and threshold rules. And the least difference show maximin, Kemeny (which are the same again), and Baldwin rules.

Rank	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Plurality	.00	.07	.12	.12	.18	.23	.25	.36	.31	.33	.43	.37	.39	.47	.40	.42	.50	.42
Veto	.03	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Borda	.28	.19	.08	.08	.07	.06	.07	.08	.09	.10	.10	.11	.11	.12	.12	.12	.13	.13
Run-off	.11	.35	.28	.52	.29	.42	.36	.54	.27	.42	.24	.26	.10	.21	.00	.09	.00	.00
STV	.00	.07	.52	.24	.00	.72	.23	.22	.74	.37	.00	.58	.27	.00	.00	.21	.00	.00
Copeland	.19	.25	.21	.21	.25	.22	.27	.22	.29	.23	.29	.23	.30	.23	.30	.23	.30	.24
Maximin	.06	.21	.02	.16	.17	.11	.07	.00	.07	.00	.07	.00	.07	.00	.07	.00	.06	.00
Baldwin	.19	.19	.30	.32	.19	.28	.31	.18	.15	.23	.13	.22	.14	.23	.14	.24	.15	.25
Nanson	.28	.17	.19	.30	.00	.18	.25	.05	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Black	.06	.22	.02	.13	.02	.09	.03	.09	.03	.09	.03	.01	.04	.01	.04	.01	.04	.01
Kemeny	.19	.16	.15	.21	.36	.31	.42	.28	.45	.31	.48	.33	.49	.34	.50	.35	.51	.36
Threshold	.06	.17	.11	.19	.24	.22	.30	.34	.32	.37	.40	.38	.42	.44	.42	.45	.46	.45

Table 4.6: $I^{M3}(3,n,Rank,F)$

Since Model 3 assumes even higher level of uncertainty for voters than Model 2, zero-manipulability obtained for Model 2 is inherited by Model 3. We formulate this inheritance in the following two propositions.

Proposition 4.2. *For any PIF π, for any rule F, number of voters n, and*

Winner	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Plurality	.00	.10	.12	.14	.26	.28	.25	.33	.33	.30	.37	.36	.34	.39	.37	.36	.40	.38
Veto	.03	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Borda	.22	.31	.15	.15	.16	.17	.19	.20	.22	.23	.24	.25	.26	.27	.27	.28	.28	.29
Run-off	.22	.38	.71	.70	.37	.59	.00	.37	.00	.08	.00	.00	.00	.00	.00	.00	.00	.00
STV	.00	.10	.71	.29	.00	.60	.00	.00	.00	.15	.00	.00	.00	.00	.00	.00	.00	.00
Copeland	.06	.37	.00	.29	.00	.32	.00	.35	.00	.36	.00	.38	.00	.38	.00	.39	.00	.39
Maximin	.06	.22	.02	.13	.43	.09	.20	.00	.23	.00	.25	.00	.27	.00	.28	.00	.29	.00
Baldwin	.06	.22	.64	.13	.00	.30	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Nanson	.22	.22	.21	.25	.00	.18	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Black	.06	.22	.64	.13	.00	.09	.00	.09	.00	.08	.00	.00	.00	.00	.00	.00	.00	.00
Kemeny	.06	.22	.02	.13	.43	.09	.20	.00	.23	.00	.25	.00	.27	.00	.28	.00	.29	.00
Threshold	.06	.22	.32	.14	.19	.18	.22	.25	.23	.27	.29	.27	.30	.31	.29	.32	.32	.31

Table 4.7: $I^{M3}(3,n,Winner,F)$

number of alternatives m if $I^{M1}(m,n,F,\pi)=0$*, then* $I^{M3}(m,n,F,\pi)=0$.

Proof. The proof is the same as for Proposition 1, but with respect to Model 3. □

Proposition 4.3. *For any PIF* π*, for any rule* F*, number of voters* n*, and number of alternatives* m *if* $I^{M2}(m,n,F,\pi)=0$*, then* $I^{M3}(m,n,F,\pi)=0$.

Proof. If $I^{M2}(m,n,F,\pi)=0$, then in any preference profile $\mathbf{P}$ no voter i has an incentive to manipulate in Model 2. There are two cases. In the first case voter i does not have an incentive to manipulate in Model 1. Then, by Definition 4, she does not have an incentive to manipulate in Model 3. In the second case, voter i has an incentive to manipulate in Model 1, but there is $\mathbf{P}'_{-i}$ in W_i^{π} such that $F(P_i,\hat{\mathbf{P}}'_{-i})P_iF(\tilde{P}_i,\hat{\mathbf{P}}'_{-i})$ for all $\tilde{P}_i$. Thus, condition ii) from Definition 4 is not satisfied, and voter i does not have an incentive to manipulate in Model 3. Consequently, $I^{M3}(m,n,F,\pi)=0$. □

Finally, if we fix the behavioral model and go from a more informative PIF π' to a less informative PIF π'', then 0-manipulability obtained for π' is preserved for π''.

	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Plurality	.00	.15	.00	.17	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Veto	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Borda	.31	.27	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Run-off	.22	.52	.71	.57	.37	.32	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
STV	.00	.15	.71	.29	.00	.22	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Copeland	.17	.21	.00	.15	.00	.20	.00	.23	.00	.26	.00	.28	.00	.29	.00	.30	.00	.31
Maximin	.17	.18	.21	.57	.26	.20	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Baldwin	.17	.18	.43	.57	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Nanson	.31	.18	.00	.57	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Black	.17	.18	.43	.57	.00	.20	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Kemeny	.17	.18	.21	.57	.26	.20	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Threshold	.11	.44	.62	.13	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00

Table 4.8: $I^{M3}(3,n,1Winner,F)$

Proposition 4.4. *Suppose, π' is at least as informative as π''. Then for any $M \in \{M1, M2, M3\}$, for any rule F, number of voters n, and number of alternatives m if $I^M(m,n,F,\pi') = 0$, then $I^M(m,n,F,\pi'') = 0$.*

Proof. Let us introduce a notation specially for this proof: every voter has her own information set W_i^π, but in accordance with a behavioral model M voters may change their preferences (as in Model 2 and 3) or not (Model 1), so let $W_{i,M}^\pi$ be an information set of voter i with respect to the behavioral model M. Thus, $W_{i,M2}^\pi = \{\hat{P}'_{-i} \in L(X)^{N\setminus\{i\}} | P'_{-i} \in W_i^\pi\}$, $W_{i,M3}^\pi = \{(\hat{\mathbf{P}}'_K, \mathbf{P}'_{-K-i}) \in L(X)^{N\setminus\{i\}} | P'_{-i} \in W_i^\pi, K \subseteq \Pi(\mathbf{P}'_{-i}), K \neq \emptyset\}$. Now let $\hat{W}_{i,M1}^\pi = W_i^\pi$, $\hat{W}_{i,M2}^\pi = W_i^\pi \cup W_{i,M2}^\pi$, and $\hat{W}_{i,M3}^\pi = W_i^\pi \cup W_{i,M3}^\pi$.

If $I^M(m,n,F,\pi') = 0$, then for any voter i in every profile $\mathbf{P}$ and any strategy $\tilde{P}_i \neq P_i$ it is either i) $F(P_i, \mathbf{P}'_{-i}) = F(\tilde{P}_i, \mathbf{P}'_{-i})$ for all $\mathbf{P}'_{-i}$ in $W_i^{\pi'(\mathbf{P})}$; or ii) there exists $\mathbf{P}'_{-i}$ in $\hat{W}_{i,M}^{\pi'(\mathbf{P})}$ such that $F(P_i, \mathbf{P}'_{-i}) P_i F(\tilde{P}_i, \mathbf{P}'_{-i})$.

Consider case i). Since π' is at least as informative as π'', then for all $\mathbf{P} \in L(X)^N$ and for all $i \in N$ we have $W_i^{\pi'(\mathbf{P})} \subseteq W_i^{\pi''(\mathbf{P})}$. Thus, if for any voter i in every profile $\mathbf{P}$ and any strategy $\tilde{P}_i \neq P_i$ there is no such $\mathbf{P}'_{-i}$ in $W_i^{\pi'(\mathbf{P})}$ that $F(P_i, \mathbf{P}'_{-i}) P_i F(\tilde{P}_i, \mathbf{P}'_{-i})$ or $F(\tilde{P}_i, \mathbf{P}'_{-i}) P_i F(P_i, \mathbf{P}'_{-i})$, then there will not be such profiles in $W_i^{\pi''(\mathbf{P})}$. Indeed, if there is a profile $\mathbf{P}'_{-i}$ in $W_i^{\pi''(\mathbf{P})}$ such that $F(P_i,\mathbf{P}'_{-i})P_iF(\tilde{P}_i,\mathbf{P}'_{-i})$ or $F(\tilde{P}_i,\mathbf{P}'_{-i})P_iF(P_i,\mathbf{P}'_{-i})$, and this profile does not belong to $W_i^{\pi'(\mathbf{P})}$, then $\mathbf{P}'_{-i}$ is in another information set of i (for another preference profile $\mathbf{P}$). But this contradicts to the statement that for any voter i in every profile $\mathbf{P}$ and any strategy $\tilde{P}_i \neq P_i$ it is $F(P_i,\mathbf{P}'_{-i}) = F(\tilde{P}_i,\mathbf{P}'_{-i})$ for all $\mathbf{P}'_{-i}$ in $W_i^{\pi'(\mathbf{P})}$.

Consider case ii). Since π' is at least as informative as π'', we have that for all $\mathbf{P} \in L(X)^N$ and for all $i \in N$, $W_i^{\pi'} \subseteq W_i^{\pi''}$, and, consequently, $\hat{W}_i^{\pi'} \subseteq \hat{W}_i^{\pi''}$. If there exists $\mathbf{P}'_{-i}$ in $\hat{W}_{i,M}^{\pi}$, such that $F(P_i,\mathbf{P}'_{-i})P_iF(\tilde{P}_i,\mathbf{P}'_{-i})$, then $\mathbf{P}'_{-i}$ belongs to $\hat{W}_i^{\pi''}$ as well. Thus, voter i does not have an incentive to manipulate under PIF π'' in model M. Since that holds for any voter i in any preference profile $\mathbf{P}$ and any strategy $\tilde{P}_i \neq P_i$, we get $I^M(m,n,F,\pi'') = 0$. $\square$

In Tables 4.6-4.8 zero values obtained for Model 2 are highlighted with a darker grey, and those for Model 3 with a lighter grey. First, let us look at Table 4.6, which shows the manipulability index for *Rank*-PIF and behavioral Model 3. It can be seen that only veto and Nanson's rule demonstrate clear zero-tending dynamics of $I^{M3}(m,n,F,\pi)$, and values for maximin rule and Black's procedure are quite low. Zero-manipulability occurs in some cases for runoff, STV, and maximin. However, most of zero values are inherited from Model 2. Plurality, Kemeny, and threshold rules are the most manipulable.

Now move on to Table 4.7, corresponding to *Winner*-PIF. For plurality, Borda, and threshold rules if voters know information about winners before tie-breaking, they still can manipulate, even when they think that others may also manipulate. We view this as an important observation, since it puts these rules in contrast with the remaining ones, which give non-manipulability or zero manipulability alternating with non-zero manipulability from certain number of voters.

A slight difference in the quality of public information, i.e. voters do not know which alternatives were tied (in case there is a tie) and know only a final winner – this difference makes almost all rules impossible to manipulate from a certain value of n. Periodicity is preserved for Copeland rule, and the amplitude seems to be growing. It means that Copeland rules provides an opportunity to manipulate even under uncertainty of preferences and intentions of others when the number of voters is even. However, other rules give us a hope that manipulation actually does not make much sense when you know little.

4.7 Conclusion

We have reviewed many works devoted to the problem of manipulability of social choice rules. They considered the problem from probabilistic, computational, informational points of view. These aspects were considered separately and results are often unfavorable: rules turn out to be manipulable (Gibbard-Satterthwaite theorem type), and, moreover, highly manipulable (computed probabilities) or efficiently manipulable. The current study is the first one combining informational and behavioral aspects of manipulation and applying a probabilistic approach to it. Indeed, what does it mean to manipulate when you do not know exactly what others prefer and what they will actually do? If we were aimed to answer the question whether these rules are susceptible to manipulation, for example, under *1Winner*-PIF and Model 3, the answer would be positive for all rules except for veto rule. But this answer would not show us the influence of uncertainty which can be clearly seen only with a studying probabilities. Due to excessive computations we are able to observe the fact that manipulation becomes impossible in many cases when we want to take these uncertainties into account.

One may argue that computing all possible situations consistent with public information and predicting other manipulators' actions is too hard for a voter. Of course, it is, and this adds to the immunity of rules to manipulation in reality. However, the goal of this research was not to show that manipulation is hard, but to demonstrate that it does not make much sense.

Another possible objection is that there are still many cases when manipulation is probable. The answer here is following: we considered only the probability of that a voter will have an incentive to manipulate, but did not compute the probability of a successful manipulation. As shown by Veselova

(2020), this probability turns out to be rather small. So, even if we have manipulability around 0.3, the proportion of successfully manipulable profiles could be much smaller than 0.3.

The deeper analysis of the manipulation models from this study may also include computation and comparison of the probability of success and failure for each voter. This could help to understand whether a voter is assumed to be too risk-averse when she does not manipulate in a view of one unfavorable situation. However, considering only strategies that do not spoil the result whatever other voters do seems quite natural for modeling voters behavior. It guarantees that a voter will not loose. The fact revealed by this study that under very natural conditions it is impossible to find such strategy in a vast number of cases is definitely good news.

Appendix A

1 Main program for Chapter 2

```
ICNK=zeros(1,13);
for n=3:15
m=3;

%Enumerating all preference types
p=transpose(perms(1:m));

% Generating AECs
A=1:(factorial(m)-1)+n;
C=nchoosek(A,factorial(m)-1);
K=nchoosek(factorial(m)-1+n,factorial(m)-1);
AEC=zeros(K,factorial(m));

AEC(:,1)=C(:,1)-1;
for i=2:(factorial(m)-1)
    AEC(:,i)=C(:,i)-C(:,i-1)-1;
end
AEC(:,factorial(m))=factorial(m)+n-C(:,factorial(m)-1)-1;

% Transforming into representative preference profiles
P=zeros(3,n,K);
for i=1:K
    h=1;
    for j=1:6
        for k=1:AEC(i,j)
            P(:,h,i)=p(:,j);
            h=h+1;
        end
    end
end

%Computing Positions-information function
v=zeros(3,3,K);
for i=1:K
   for ia=1:3
```

```
        for k=1:3
            for l=1:n
                if P(k,l,i)==ia
                v(ia,k,i)=v(ia,k,i)+1;
                end
            end
        end
    end
end

%Computing Weighted Majority graph
WMG=zeros(3,3,K);
MG=zeros(3,3,K);
for i=1:K
    [WMG(:,:,i),MG(:,:,i)]=majority(P(:,:,i));
end

% Computing Score-Rank-Winners
ScorePlur=zeros(1,3,K);
RankPlur=zeros(1,3,K);
WinnerPlur=zeros(1,3,K);

ScoreBorda=zeros(1,3,K);
RankBorda=zeros(1,3,K);
WinnerBorda=zeros(1,3,K);

ScoreVeto=zeros(1,3,K);
RankVeto=zeros(1,3,K);
WinnerVeto=zeros(1,3,K);

ScoreRunoff=zeros(2,3,K);
RankRunoff=zeros(1,3,K);
WinnerRunoff=zeros(1,3,K);

ScoreStv=zeros(3,3,K);
RankStv=zeros(1,3,K);
WinnerStv=zeros(1,3,K);

ScoreCopeland=zeros(1,3,K);
RankCopeland=zeros(1,3,K);
WinnerCopeland=zeros(1,3,K);

for i=1:K
```

```
    [WinnerPlur(:,:,i),RankPlur(:,:,i),ScorePlur(:,:,i)]=
    plurality(P(:,:,i),n);
    [WinnerBorda(:,:,i),RankBorda(:,:,i),ScoreBorda(:,:,i)]=
    borda(P(:,:,i),n);
    [WinnerVeto(:,:,i),RankVeto(:,:,i),ScoreVeto(:,:,i)]=
    veto(P(:,:,i),n);
    [WinnerRunoff(:,:,i),RankRunoff(:,:,i),ScoreRunoff(:,:,i)]=
    runoff(P(:,:,i));
    [WinnerStv(:,:,i),RankStv(:,:,i),ScoreStv(:,:,i)]=
    stv(P(:,:,i));
    [WinnerCopeland(:,:,i),RankCopeland(:,:,i),
    ScoreCopeland(:,:,i)]=copeland(P(:,:,i));
end

% Manipulation check
MP=zeros(K,1);
Check=zeros(K,1);
CheckVoter=zeros(K,n);
VotersCoalitions=zeros(K,n,n);
Flag=zeros(K,1);
f=0;
for i=1:K
    if Check(i,1)==0
    j=0;
    while MP(i,1)==0 && j<n
        j=j+1;
        if CheckVoter(i,j)~=1
        f=f+1;
        pe=[1,1,1,1,1,1];
        pb=[0,0,0,0,0,0];
        for ii=1:K
            if isequal(P(:,:,i),P(:,:,ii))
                Coalition=zeros(1,n);
                for jc=1:n
                    if P(:,j,i)==P(:,jc,ii)
                       CheckVoter(ii,jc)=1;
                       Coalition(1,jc)=1;
                    end
                end
                for jc=1:n
                    if P(:,j,i)==P(:,jc,ii)
                       VotersCoalitions(ii,jc,:)=Coalition(1,:);
                    end
```

```
                    end
                    c=sum(Coalition(1,:));
                    if c~=0
                        ManipP=P(:,:,ii);
                        for jm=1:6
                            if pe(1,jm)==1
                              for jj=1:n
                                  if Coalition(1,jj)==1
                                      ManipP(:,jj)=p(:,jm);
                                  end
                              end
                              if lexTie(borda(ManipP(:,:),n),P(:,j,i))>
                              lexTie(borda(P(:,:,ii),n),P(:,j,i))
                                  pe(1,jm)=0;
                                  pb(1,jm)=0;
                              end
                              if lexTie(borda(ManipP(:,:),n),P(:,j,i))
                              <lexTie(borda(P(:,:,ii),n),P(:,j,i))
                                  pb(1,jm)=1;
                              end
                            end
                        end
                        Flag(ii,1)=f;
                    end
                end
            end
            if sum(pb)~=0
                for ii=1:K
                    if Flag(ii,1)==f
                        MP(ii,1)=1;
                        Check(ii,1)=1;
                    end
                end
            end
            if j==n
                Check(i,1)=1;
            end
            end
        end
        end
end

ICM=zeros(K,1);
```

```
for i=1:K
    if MP(i,1)==1
        ICM(i,1)=factorial(n)/(factorial(AEC(i,1))*
        factorial(AEC(i,2))*factorial(AEC(i,3))*
        factorial(AEC(i,4))*factorial(AEC(i,5))*
        factorial(AEC(i,6)));
    end
end

ICNK(n-2)=sum(ICM)/(6)^n;
end
```

Appendix B

1 Main program for Chapter 4

```
global m n p rule PIF;
m=3;
p=transpose(perms(1:m));

IND1=zeros(12,28,3);
IND2=zeros(12,18,3);
IND3=zeros(12,18,3);

for pif_ind=1:1
    if pif_ind==1
        PIF = @ winner;
    end
    if pif_ind==2
        PIF = @ winners;
    end
    if pif_ind==3
        PIF = @ rank;
    end

for rule_ind=1:1
    if rule_ind==1
        rule = @ plurality;
    end
    if rule_ind==2
        rule = @ veto;
    end
    if rule_ind==3
        rule = @ borda;
    end
    if rule_ind==4
        rule = @ runoff;
    end
    if rule_ind==5
        rule = @ stv;
    end
```

```
    if rule_ind==6
        rule = @ copeland;
    end
    if rule_ind==7
        rule = @ maximin;
    end
    if rule_ind==8
        rule = @ bolduin;
    end
    if rule_ind==9
        rule = @ nanson;
    end
    if rule_ind==10
        rule = @ black;
    end
    if rule_ind==11
        rule = @ kemeni;
    end
    if rule_ind==12
        rule = @ threshold;
    end

for n=6:6

A=1:(factorial(m)-1)+n;
C=nchoosek(A,factorial(m)-1);
K=nchoosek(factorial(m)-1+n,factorial(m)-1);
AEC=zeros(K,factorial(m));
CardinalityAEC=zeros(K,1);

AEC(:,1)=C(:,1)-1;
for i=2:(factorial(m)-1)
    AEC(:,i)=C(:,i)-C(:,i-1)-1;
end
AEC(:,factorial(m))=factorial(m)+n-C(:,factorial(m)-1)-1;

for i=1:K
    Multip=1;
    for l=1:factorial(m)
        Multip=Multip*factorial(AEC(i,l));
    end
    CardinalityAEC(i)=factorial(n)/Multip;
end
```

```
P=zeros(m,n,K);
for i=1:K
    P(:,:,i)=tranform(AEC(i,:));
end

CheckVoter=zeros(K,n);
Groups_Eq_Voters=zeros(K,n);
NG=0;
man=zeros(K,n);
manStr=zeros(K,n);
ManG=zeros(K,1);
ManS=zeros(K,1);


for i=1:K
    for j=1:n
        if CheckVoter(i,j)==0
            Eq_voters=zeros(K,n);
            man_strategies=zeros(K,factorial(m));

            for ii=1:K
                eq_voters=is_equivalent(P(:,:,i),j,P(:,:,ii));
                Eq_voters(ii,:)=eq_voters;
                CheckVoter(ii,:)=CheckVoter(ii,:)+eq_voters;
                    [man_strategies(ii,:)]=manipulation(P(:,:,ii),
                    eq_voters);
                end
            end
            eff_man_inf_set=min(man_strategies)+
            max(man_strategies);
            max_eff=0;
            str_max_eff=0;
            for h=1:factorial(m)
                if eff_man_inf_set(h)>=1
                    if sum(man_strategies(:,h))>max_eff
                        max_eff=sum(man_strategies(:,h));
                        str_max_eff=h;
                    end
                end
            end
            if str_max_eff>0
                manStr=manStr+str_max_eff*Eq_voters;
```

```
                NG=NG+1;
                Groups_Eq_Voters=Groups_Eq_Voters+Eq_voters*NG;
                ManS=max(ManS,man_strategies(:,str_max_eff));
            end
        end
    end
end

for i=1:K
   for j=1:n
        if Groups_Eq_Voters(i,j)~=0
            man2=man(i,:);
            man2(1,j)=0;
            ManGroups=subsets(man2(1,:));
            groupN=sum(man(i,:));
            ResGV=zeros(2^(groupN-1)-1,1);
            ResG=zeros(2^(groupN-1)-1,1);
            NResGV=zeros(2^(groupN-1)-1,1);
            NResG=zeros(2^(groupN-1)-1,1);
            for l=1:2^(groupN-1)-1
                    manG=ManGroups(l,:);
                    manGV=manG;
                    manGV(1,j)=1;
                    manStrGV=manStr(i,:).*manGV;
                    manStrG=manStr(i,:).*manG;
                    ResG(l,1)=sim_man_result(P(:,:,i), manStrG);
                    ResGV(l,1)=sim_man_result(P(:,:,i), manStrGV);
                    for h=1:m
                        if P(h,j,i)==ResG(l,1)
                            NResG(l,1)=h;
                        end
                        if P(h,j,i)==ResGV(l,1)
                            NResGV(l,1)=h;    %
                        end
                    end
                    if min(NResGV<=NResG)==0
                        GEV=Groups_Eq_Voters(i,j);
                        for ii=1:K
                            for jj=1:n
                                if Groups_Eq_Voters(ii,jj)==GEV
                                    Groups_Eq_Voters(ii,jj)=0;
                                end
                            end
```

```
                        end
                    end

            end
        end
    end
end
ManSM=max(Groups_Eq_Voters,[],2);
for i=1:K
    if ManSM(i)>0
        ManSM(i)=1;
    end
end

Man=max(man,[],2);
for i=1:K
    if Man(i)>0
        Man(i)=1;
    end
end

IND1(rule_ind,n-2,pif_ind)=sum(Man.*CardinalityAEC)/
(factorial(m))^n;
IND2(rule_ind,n-2,pif_ind)=sum(ManS.*CardinalityAEC)/
(factorial(m))^n;
IND3(rule_ind,n-2,pif_ind)=sum(ManSM.*CardinalityAEC)/
(factorial(m))^n;
end
end
end
```

Bibliography

[1] Aleskerov, F., D. Karabekyan, A. Ivanov, and V. Yakuba (2017). Manipulability of majoritarian rules by coalitions with the same first-ranked alternative. *In: Procedia Computer Science* **122**, 993-1000.

[2] Aleskerov, F., D. Karabekyan, M. R. Sanver, and V. Yakuba (2011). On the Degree of Manipulability of Multi-Valued Social Choice Rules. *Homo Oeconomicus*, **28**.

[3] Aleskerov, F., D. Karabekyan, M. R. Sanver, and V. Yakuba (2012). On the manipulability of voting rules: the case of 4 and 5 alternatives. *Mathematical Social Sciences* **64**(1), 67-73.

[4] Aleskerov, F. and E. Kurbanov (1999). Degree of manipulability of social choice procedures. *In: Current trends in economics*. Springer, 13–27.

[5] Aleskerov, F., D. Karabekyan, R. Sanver, and V. Yakuba (2009). Evaluation of the Degree of Manipulability of Known Aggregation Procedures under Multiple Choice (in Russian). *In: Journal of the New Economic Association*, **1**(1), 37–61.

[6] Barbera, S. (1980). Stable voting schemes. *In: Journal of Economic Theory*, **23**(2), 267–274.

[7] Camerer, C. F., T. H. Ho, and J. K. Chong (2004). A cognitive hierarchy model of games. *In: The Quarterly Journal of Economics*, **119**(3), 861-898.

[8] Conitzer, V., T. Walsh, and L. Xia (2011). Dominating Manipulations in Voting with Partial Information. *In: Proceedings of the TwentyFifth AAAI Conference on Artificial Intelligence Dominating*. **11**, 638-643.

[9] De Sinopoli, F. (2000). Sophisticated voting and equilibrium refinements under plurality rule. *Social Choice and Welfare* **17**(4), 655-672.

[10] Desmedt, Y. and E. Elkind (2010). Equilibria of plurality voting with abstentions. *In: Proceedings of the 11th ACM conference on Electronic commerce*, 347-356.

[11] Van Ditmarsch, H., J. Lang, and A. Saffidine (2012). Strategic voting and the logic of knowledge. *In: 11th International Conferenceon Autonomous Agents and Multiagent Systems (AAMAS 12)*, 1247-1248.

[12] Elkind, E., U. Grandi, F. Rossi, and A. Slinko (2015). Gibbard-Satterthwaite Games. *In: Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, 533-539.

[13] Elkind, E., U. Grandi, F. Rossi, and A. Slinko (2020). Cognitive hierarchy and voting manipulation in k-approval voting. *Mathematical Social Sciences*, **108**, 193–205.

[14] Endriss, U., S. Obraztsova, M. Polukarov, and J. S. Rosenschein (2016). Strategic voting with incomplete information. *In: International Joint Conferences on Artificial Intelligence*, AAAI Press, 236-242.

[15] Farquharson, R. (1969). *Theory of voting*, Oxford: Blackwell.

[16] Gardenfors, P. (1976). Manipulation of social choice functions. *In: Journal of Economic Theory*, **13**(2), 217-228.

[17] Gehrlein, W. V. and P. C. Fishburn (1981). Constant scoring rules for choosing one among many alternatives. *Quality and Quantity* **15**(2), 203-210.

[18] Gibbard, A. (1973). Manipulation of voting schemes: a general result. *Econometrica*. **41**(4), 587-601.

[19] Grandi, U., D. Hughes, F. Rossi, and A. Slinko (2019). Gibbard-Satterthwaite games for k-approval voting rules. *Mathematical Social Sciences*, **99**, 24-35.

[20] Hazon, N., E. Elkind (2010). Complexity of safe strategic voting. *In: International Symposium on Algorithmic Game Theory*, Springer, 210-221.

[21] Ianovski, E., L. Yu, E. Elkind, and M. C. Wilson (2011). The complexity of safe manipulation under scoring rules. *In: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, **11**, 246-251.

[22] Kelly, J. S. (1988). 4. Minimal manipulability and local strategy-proofness. *Social Choice and Welfare*, **5**(1), 81-85.

[23] Lepelley, D. and F. Valognes (2003). Voting rules, manipulability and social homogeneity. *Public Choice*, **116**(1-2), 165-184.

[24] Marchant, T. (2001). The probability of ties with scoring methods: Some results. *Social Choice and Welfare*, **18**(4), 709-735.

[25] Maus, S., H. Peters, T. Storcken (2007). Anonymous voting and minimal manipulability. *Journal of Economic Theory*, **135**, 533-544.

[26] Meir, R. (2018). Strategic voting. *In: Synthesis Lectures on Artificial Intelligence and Machine Learning*, **13**(1), 1-167.

[27] Meir, R., M. Polukarov, J. S. Rosenschein, and N. R. Jennings (2017). Iterative voting and acyclic games. *Artificial Intelligence*, **252**, 100-122.

[28] Moulin, H. (1981). Prudence versus sophistication in voting strategy. *Journal of Economic Theory*, **24**(3), 398-412.

[29] Myerson, R. B. and R. J. Weber (1993). A theory of voting equilibria. *American Political Science Review*, 102-114.

[30] Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review*, *85*(5), 1313-1326.

[31] Nitzan, S. (1985). The vulnerability of point-voting schemes to preference variation and strategic manipulation. *Public Choice*, **47**(2), 349-370.

[32] Obraztsova, S., E. Markakis, and D. R. M. Thompson (2013). Plurality voting with truth-biased agents. *In: International Symposium on Algorithmic Game Theory*, Springer Berlin Heidelberg, 26-37.

[33] Obraztsova, S., Z. Rabinovich, M. Polukarov, N. R. Jennings (2016). Trembling hand equilibria of plurality voting. *In: International Joint Conferences on Artificial Intelligence*, AAAI Press, 440-446.

[34] Pattanaik, P. K. (1976a). Counter-threats and strategic manipulation under voting schemes. *The Review of Economic Studies*, **43**(1), 11-18.

[35] Pattanaik, P. K. (1976b). Threats, counter-threats, and strategic voting. *Econometrica: Journal of the Econometric Society*, 91–103.

[36] Peters, H., S. Roy, and T. Storcken (2012). On the manipulability of approval voting and related scoring rules. *Social Choice and Welfare*, **39**(2-3), 399-429.

[37] Peters, H. and Y. Veselova (2023). On the safety of group manipulation. *Social Choice and Welfare*.

[38] Pritchard, G. and M. C. Wilson (2007). Exact results on manipulability of positional voting rules. *Social Choice and Welfare*, **29**(3), 487-513.

[39] Reijngoud, A. and U. Endriss (2012). Voter response to iterated poll information. *In: Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*, **2**, 635-644.

[40] Satterthwaite, M. A. (1975). Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, **10**(2), 187-217.

[41] Sertel, M. R. and M. R. Sanver (2004). Strong equilibrium outcomes of voting games are the generalized Condorcet winners. *Social Choice and Welfare*, **22**(2), 331-347.

[42] Slinko, A. (2006). How the size of a coalition affects its chances to influence an election. *Social Choice and Welfare*, **26**(1), 143-153.

[43] Slinko, A. (2019). Beyond Gibbard and Satterthwaite: Voting Manipulation Games. *In: The Future of Economic Design*, Springer, 131-138.

[44] Slinko, A. and S. White (2014). Is it ever safe to vote strategically? *Social Choice and Welfare*, **43**(2), 403-427.

[45] Smaal, K. E. M. (2019). Strategic manipulation in voting under higher-order reasoning.*MA thesis*, Institute for Logic, Language and Computation, University of Amsterdam.

[46] Stahl, D. O. and P. W. Wilson (1994). Experimental evidence on players' models of other players. *Journal of economic behavior and organization*, **25**(3), 309-327.

[47] Veselova, Y. A. (2020). Does Incomplete Information Reduce Manipulability? *Group Decision and Negotiation*, **29**(3), 523-548.

[48] Veselova, Y. A. (2023). Manipulation by coalitions in voting with incomplete information. *In: Data Analysis and Otimization*, Springer.

[49] Veselova Y. and D. Karabekyan. (2023) Simultaneous manipulation under incomplete information. *Working paper*.

[50] Wilson, M. and R. Reyhani Shokat Abad (2010). The probability of safe manipulation. *In: Proceedings of the Third International Workshop on Computational Social Choice*, 423-430.

[51] Xia, L. and V. Conitzer (2010). Stackelberg voting games: Computational aspects and paradoxes. *In: Proceedings of the AAAI Conference on Artificial Intelligence*. **24**(1), 421-426.

Impact of the thesis

This thesis contributes to the area of manipulability of social choice rules. The problem of manipulation is well-known in social choice theory, since A.Gibbard and M.Satterthwaite many authors studied this question from theoretical, experimental and computational perspective. However, classical approaches to modelling strategic behavior of voters in voting are quite static. The goal of our research is to make models which are more flexible, i.e. can take into account various parameters, and, as a consequence, which are more realistic.

We argue that information available to voters and their view of other voters' behavior are the crucial aspects that affect individual manipulation incentives. For example, we show that for many rules voters that have an incentive to manipulate exist almost in every possible situation if they have positive expectations about their coalition members and possess information only about the winners from an opinion poll. On the other hand, some rules do not guarantee that a voting result will not get worse if some of your allies do not support group manipulation. This fact constitutes an obstacle for collective manipulation of these rules. Other rules, as revealed by our research, do not pose such a threat to the manipulator, so they can be considered as more easily manipulable.

Finally, in addition to the uncertainty about the true preference profile (due to the incompleteness of information), we consider the uncertainty about all other manipulators' actions. We study how restrictive the combination of these two types of uncertainty is for manipulation - a question which has not been considered before. It turns out that for certain conditions rules do not allow for manipulation anymore.

All these results can be helpful when choosing a social choice mechanism for a given collection of parameters, such as the type of public information, the number of voters and alternatives, opportunities for communication for voters. Or, alternatively, they can show how to restrict the parameters if a social choice rule is fixed. The area of application of social choice theory has been widening in recent years due to the spread of information technologies. The problem of preference aggregation arises not only in human voting, but also in decision making with autonomous software agents, whose computational ability is much stronger. For this reason, it becomes especially important to know the formal restrictions existing in aggregation methods for their best use.

Summary

This thesis is devoted to the problem of strategic behavior in voting and develops mathematical models of manipulation in several directions, each considered by a separate chapter. We reveal how the type of public information and voters' view of incentives of others influences manipulability of rules and studied the question of the safety of group manipulation.

The first question under consideration is coalitional manipulability under incomplete information. Coalition members are assumed to have identical preferences and all voters possess some information about a real preference profile from an opinion poll held before voting. We consider 5 different types of poll information functions. Manipulability is defined as the probability that in a randomly chosen preference profile there exists a coalition which has an incentive to manipulate under a given type of poll information. We calculate the degree of coalitional manipulability for 3 alternatives and the number of voters from 3 to 15 for different types of information and compare it with individual manipulability values. We study asymptotic behavior of manipulability for plurality and Borda rule under *1Winner* and *Winner*-PIFs theoretically and prove that for *1Winner*-PIF coalitional and individual manipulability of scoring rules coincide.

Suppose that manipulation is done by a group of voters who have the same preferences. If a voting result is more preferable for voters of this group provided that they all use the same strategy (report the same insincere preference), then each of them has an incentive to manipulate. However, due to the lack of information or communication some of group members may not manipulate. If there is a chance that they will become worse off in case only a subset of the whole group manipulates, then manipulation is unsafe. For several voting rules we find necessary and sufficient conditions on the numbers of voters and alternatives which allow for an unsafe manipulation or which make manipulation always safe.

Finally, we study individual manipulation under incomplete information and different assumptions about voters' beliefs about behavior of others. Since voters do not know preferences of others exactly, there is an uncertainty about the situation they are in. As earlier, incomplete information is modelled by poll information functions (PIFs). Voters that have an incentive to manipulate under PIF π are called π-manipulators. We consider three behavioral models. In the first model incentives to manipulation do not depend on other voters' incentives. In the second model each π-manipulator takes into account that all other π-manipulators strategise, and in the third model that only a subset of them do so. Therefore, the uncertainty about a situation is combined with uncertainty about other voters' actions. With the help of computations we reveal how the type of information and behavioral model influence the relative manipulability of 12 social choice rules. In theoretical results, we state and prove some propositions about the inheritance of zero manipulability from one model to another. And, finally, we prove that there is a certain number of voters starting from which manipulability of a scoring rule is zero with public information about a winner after tie-breaking.

Acknowledgments

First of all, I would like to thank my parents, Aleksandr Yakushev and Alla Yakusheva, who gave me life and all the best that they could give. I am infinitely grateful to my husband, Dmitry, always providing me with incredible support, love and possessing unbearable patience. I thank my daughters, Sophia and Mirra, who always encourage me to be better. I would like to express my deepest gratitude to my supervisor at NRU HSE Fuad Aleskerov who raised my interest in social choice theory and showed me what the science is. In Maastricht University I was extremely lucky to have the best supervisors, Hans Peters and Ton Storcken. They always gave their advice and support showing a real interest in my research and, what is more, an open heart.

I would also like to thank Elena Veselova for her help, Daniel Karabekyan for fruitful discussions, Ulle Endriss for comments at the beginning of my PhD, and Zoi Terzopoulou for important remarks at the very end of it. Finally, I am grateful to my dear sister, Elena Yakusheva, for her love and friendship, which are extremely important to me.

Yuliya A. Veselova
Maastricht
August 28, 2023

About the author

Yuliya Aleksandrovna Veselova was born in 1989 in Leningrad. From 2003 till 2007 she studied at the Gymnasium of Obninsk where her mathematics teachers were Valentina Ivanovna Goncharova and Valentina Ivanovna Lyovina. She became a student of National Research University Higher School of Economics in Moscow in 2007, and obtained the degree of Bachelor of Applied Mathematics and Information Science in June 2011, the degree of Master in June 2013, and the degree of Candidate of Science under the supervision of Prof. Fuad Aleskerov in 2018 in the same university. From October 2016 she has visited School of Business and Economics, Maastricht University, Maastricht, the Netherlands, as an external Ph.D. student, under the supervision of Prof. dr. Hans Peters and Dr. Ton Storcken.