



## King's Research Portal

*Document Version*  
Peer reviewed version

[Link to publication record in King's Research Portal](#)

*Citation for published version (APA):*

Endriss, U., Obraztsova, S., Polukarov, M., & Rosenschein, J. S. (2016). Strategic voting with incomplete information. *AAAI Press / International Joint Conferences on Artificial Intelligence*.  
<https://eprints.soton.ac.uk/391559/>

### **Citing this paper**

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### **General rights**

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### **Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Strategic Voting with Incomplete Information

Ulle Endriss

University of Amsterdam  
The Netherlands

Svetlana Obraztsova

Hebrew University  
Jerusalem, Israel

Maria Polukarov

University of Southampton  
United Kingdom

Jeffrey S. Rosenschein

Hebrew University  
Jerusalem, Israel

## Abstract

Classical results in social choice theory on the susceptibility of voting rules to strategic manipulation make the assumption that the manipulator has complete information regarding the preferences of the other voters. In reality, however, voters only have incomplete information, which limits their ability to manipulate. We explore how these limitations affect both the manipulability of voting rules and the dynamics of systems in which voters may repeatedly update their own vote in reaction to the moves made by others. We focus on the Plurality, Veto,  $k$ -approval, Borda, Copeland, and Maximin voting rules, and consider several types of information that are natural in the context of these rules, namely information on the current front-runner, on the scores obtained by each alternative, and on the majority graph induced by the individual preferences.

## 1 Introduction

Most existing models of collective decision making adopt simplifying assumptions regarding the information available to individual decision makers about the preferences and intentions of their peers. For instance, the standard model employed for the analysis of strategic behaviour in elections presupposes that each voter has access to the full preferences of the other voters and is able to optimise their own ballot accordingly. This assumption is central to the classical Gibbard-Satterthwaite Theorem [Gibbard, 1973; Satterthwaite, 1975], which shows that essentially all reasonable voting rules are subject to strategic manipulation. In practice, however, voters cannot be assumed to possess all of this information, and they cannot be assumed to hold with certainty the information to which they do have access. Exploring the consequences of such richer models of decision making is of particular interest in the domain of multiagent systems, where we can model the information available to agents in precise terms.

In this paper, to represent the incomplete information available to agents, we employ a model developed by Conitzer *et al.* [2011] and Reijngoud and Endriss [2012], in which each agent is endowed with an *information set* describing the range of profiles of preferences of the other agents he considers possible. We focus on information sets that arise naturally in the

context of voting, such as the range of all profiles that are consistent with a given candidate winning the election.

Our contributions are twofold. In the first part of the paper, we establish results on the manipulability of voting rules under incomplete information, complementing previous results by Reijngoud and Endriss [2012]. We focus on the Plurality, Veto,  $k$ -approval, Borda, Copeland, and Maximin rules (all defined in Section 2). While the Gibbard-Satterthwaite Theorem shows that all of these rules are susceptible to strategic manipulation when the manipulator has full information, this may—in principle—cease to be the case when we restrict their access to information. We show that, nevertheless, for several natural restrictions to a manipulator’s information, manipulation can still occur for certain rules. At the same time, we are able to identify some combinations of voting rules and information sets where manipulation can be ruled out, thereby providing positive results of practical interest.

In the second part of the paper, we connect strategic voting under incomplete information with the growing literature on iterative voting [Meir *et al.*, 2010; Lev and Rosenschein, 2012; Reijngoud and Endriss, 2012; Reyhani and Wilson, 2012; Grandi *et al.*, 2013; Obraztsova *et al.*, 2015]. We consider scenarios where voters are permitted to update their ballot any number of times, in response to the limited information they have regarding the other votes. The question then arises whether such a process will converge to a state where the election winner does not change anymore. As is well known, under complete information, convergence can only be guaranteed in very specific circumstances. Arbitrary improvement moves are not guaranteed to converge under any voting rule, best-response dynamics lead to equilibria only for Plurality and Veto, and other rules require more restrictions on allowed manipulations to guarantee convergence. In contrast to these largely negative results, we are able to show that, when the only information voters have access to is the identity of the candidate currently leading the polls, we obtain convergence results for all the voting rules considered here.

The remainder of this paper is structured as follows. In Section 2 we recall the definitions of the aforementioned voting rules and model of incomplete information. Section 3 is devoted to our results on both susceptibility and immunity to strategic manipulation, for different types of incomplete information, Section 4 contains our results on iterative voting under incomplete information, and Section 5 concludes.

## 2 Preliminaries

In this section, we recall relevant concepts from voting theory [Taylor, 2005] and define voting scenarios with incomplete information as introduced by Reijngoud and Endriss [2012].

### 2.1 Voting Rules

Let  $V = \{1, \dots, n\}$  denote a set of  $n$  voters (or agents) electing a winner from a set  $C = \{c_1, \dots, c_m\}$  of  $m$  candidates (or alternatives). Let  $\mathcal{L}(C)$  be the set of all strict linear orders on  $C$ . Each voter  $i$  submits a vote (or ballot)  $b_i \in \mathcal{L}(C)$ , which may or may not coincide with his real preference order  $\succ_i \in \mathcal{L}(C)$ . A profile  $\mathbf{b} = (b_1, \dots, b_n) \in \mathcal{L}(C)^n$  is a vector of votes, one for each agent. We denote by  $\mathbf{b}_{-i}$  the profile of all votes except that of agent  $i$ , and we write  $(b'_i, \mathbf{b}_{-i})$  for the profile we obtain when we replace  $b_i$  in  $\mathbf{b}$  by  $b'_i$ .

A voting rule  $\mathcal{F} : \mathcal{L}(C)^n \rightarrow 2^C \setminus \{\emptyset\}$  takes a profile and returns a nonempty subset of candidates, called the winners of the election. In this paper, we focus on *resolute* voting rules  $\mathcal{F} : \mathcal{L}(C)^n \rightarrow C$ , which always return a single winner. Specifically, we assume ties are broken *lexicographically* (i.e., in favour of the candidate with the lowest index). Here are some of the most common voting rules [Taylor, 2005]:

- *Positional scoring rules*. Each PSR is associated with a scoring vector  $(s_1, \dots, s_m)$  of integers with  $s_1 > s_m$  and  $s_1 \geq s_2 \geq \dots \geq s_m$ . If a voter ranks a candidate at the  $j$ th position, that candidate receives  $s_j$  points. The score of a candidate is the sum of her points over all ballots. This family of rules includes *Plurality* with the scoring vector  $(1, 0, \dots, 0)$ , *Veto* with  $(1, 1, \dots, 1, 0)$ , *Borda* with  $(m-1, m-2, \dots, 0)$  and *k-approval* with  $(1, \dots, 1, 0, \dots, 0)$ , i.e.,  $k$  1's, followed by 0's.
- *Maximin*. The score of candidate  $c$  is the minimum number, across all other  $c' \in C$ , of voters who prefer  $c$  to  $c'$ .
- *Copeland*. The score of candidate  $c$  is the number of direct majority contests she wins (i.e., the number of other  $c' \in C$  for which the majority of voters prefer  $c$  to  $c'$ ), minus the number of majority contests he loses.

The election winner is always the candidate with the highest score. A *Condorcet winner* is a candidate that wins every majority contest, and a *weak Condorcet winner* is one that does not lose any such contest, although she may tie some.

### 2.2 Information Functions

Even after a profile of ballots has been elicited by the election chair, this information typically will not be communicated to the electorate, at least not in its entirety, e.g., due to privacy concerns. Let  $\mathcal{I}$  denote the set of all possible pieces of information that might be communicated to the electorate in view of a given profile. Following Reijngoud and Endriss [2012], an *information function* (IF) is a function  $\pi : \mathcal{L}(C)^n \rightarrow \mathcal{I}$  that maps profiles to elements of  $\mathcal{I}$ . We focus on the following natural choices for  $\mathcal{I}$  and the corresponding IF  $\pi$ :

- *Winner (W)*. Given voting rule  $\mathcal{F}$ , the corresponding W-IF returns the winning candidate for the input profile according to  $\mathcal{F}$ :  $\pi(\mathbf{b}) = \mathcal{F}(\mathbf{b})$ .

- *Score (S)*. Given voting rule  $\mathcal{F}$ , the corresponding S-IF returns for each  $c \in C$  her score  $s_{\mathcal{F}}(c, \mathbf{b})$  for profile  $\mathbf{b}$  under  $\mathcal{F}$  (we omit  $\mathcal{F}$  when clear from the context).
- *Majority Graph*. A majority graph is a directed graph with the nodes being the candidates: there is an edge  $(c, c')$  from  $c$  to  $c'$  if a majority of voters prefer  $c$  to  $c'$ . The MG-IF returns the majority graph of the profile.

Given the signal  $\pi(\mathbf{b})$ , and assuming each voter  $i$  knows how the IF  $\pi$  is defined, his *information set*  $\mathcal{W}_i^{\pi(\mathbf{b})}$  is the set of (partial) profiles he must consider possible in view of the information he holds after receiving  $\pi(\mathbf{b})$ :

$$\mathcal{W}_i^{\pi(\mathbf{b})} = \{ \mathbf{b}'_{-i} \in \mathcal{L}(C)^{n-1} \mid \pi(b_i, \mathbf{b}'_{-i}) = \pi(\mathbf{b}) \}$$

Conitzer *et al.* [2011] work with a similar model, but explicitly list the profiles a voter considers possible. Here we follow Reijngoud and Endriss [2012] and restrict attention to information sets that can be represented compactly via  $\pi$ .

We say that  $\pi_1$  is *at least as informative* as  $\pi_2$ , denoted  $\pi_1 \subseteq \pi_2$ , if  $\mathcal{W}_i^{\pi_1(\mathbf{b})} \subseteq \mathcal{W}_i^{\pi_2(\mathbf{b})}$  for all  $\mathbf{b} \in \mathcal{L}(C)^n$  and  $i \in V$ .

## 3 $\pi$ -Manipulability

In this section, we analyse the susceptibility of different voting rules to strategic manipulation by voters with incomplete information, encoded in the form of information sets. Along with a number of *susceptibility results* regarding the manipulability of certain rules given certain information, we also obtain two new *immunity results*, demonstrating the role of information in preventing strategic behaviour.

Suppose that before the election we conduct an opinion poll to elicit truthful preferences  $\succ = (\succ_i)_{i \in V}$ , and then communicate the result to the voters using IF  $\pi$ . In the subsequent election, voter  $i$  then has an incentive to manipulate if there is a scenario in his information set for which some new ballot would result in a better election outcome for him, and there is no scenario where the same new ballot would result in a worse outcome than the truthful outcome. Thus, manipulation is “safe”.<sup>1</sup> Formally, given profile  $\mathbf{b}$ , we say that voter  $i$  can use  $\pi$ -manipulation  $b_i \xrightarrow{i} b_i^*$  for profile  $\mathbf{b}$ , if there exists a partial profile  $\mathbf{b}'_{-i} \in \mathcal{W}_i^{\pi(\mathbf{b})}$  such that  $\mathcal{F}(b_i^*, \mathbf{b}'_{-i}) \succ_i \mathcal{F}(\mathbf{b})$ , and for all  $\mathbf{b}'_{-i} \in \mathcal{W}_i^{\pi(\mathbf{b})}$  it is the case that  $\mathcal{F}(b_i^*, \mathbf{b}'_{-i}) \succ_i \mathcal{F}(\mathbf{b})$  or  $\mathcal{F}(b_i^*, \mathbf{b}'_{-i}) = \mathcal{F}(\mathbf{b})$ .<sup>2</sup> Now, a resolute voting rule  $\mathcal{F}$  is *susceptible to  $\pi$ -manipulation* (or simply,  *$\pi$ -manipulable*) if there is a voter who can  $\pi$ -manipulate the truthful profile  $\succ = (\succ_i)_{i \in V}$ . If a rule is not susceptible to  $\pi$ -manipulation, then it is *immune to  $\pi$ -manipulation*.

If  $\pi$  returns the *full profile*, then  $\pi$ -manipulability reduces to the standard notion of manipulability. In this case, the Gibbard-Satterthwaite Theorem [Gibbard, 1973; Satterthwaite, 1975] says that any resolute voting rule for three or more candidates that is surjective and nondictatorial is susceptible to manipulation. Reijngoud and Endriss [2012]

<sup>1</sup>Note that, just as in the classical setting [Taylor, 2005], we only address the case of one voter at a time considering to manipulate.

<sup>2</sup>Consult van Ditmarsch *et al.* [2013] for a discussion of different types of manipulation in view of the knowledge of the manipulator.

prove a simple generalisation of this theorem, stating that for any IF  $\pi$  any resolute voting rule for three or more candidates that is surjective, nondictatorial, and strongly computable from  $\pi$ -images, is susceptible to  $\pi$ -manipulation. Here, a voting rule is *strongly computable from  $\pi$ -images* if upon learning  $\pi(\mathbf{b})$  for any profile  $\mathbf{b}$ , a voter  $i$  can compute the winners for any way of voting himself (i.e., not just for  $b_i$ ). Thus, for example, all PSRs are strongly computable from S-images.

In the sequel, we explore the  $\pi$ -manipulability of Plurality, Veto,  $k$ -approval, Borda, Copeland, and Maximin, under IFs  $\pi = \text{W, S, MG}$ . We start with the case of W-IF, intuitively providing the least amount of information.<sup>3</sup> Note that S-IF is more informative than W-IF for all our voting rules. For Copeland, MG-IF is more informative than S-IF. For other rules, some pairs of IFs cannot be ranked in terms of their informativeness. For example, under Plurality we cannot compute the winner from the majority graph, nor *vice versa*. For a given profile, if we increase the information available to a voter, manipulability could either increase (as the voter might be able to rule out some profiles where he would do worse when manipulating) or decrease (as he might also be able to exclude some profiles where he would benefit from manipulating). However, as we will see next, manipulability *across all profiles* increases as information increases.<sup>4</sup> This useful result applies to pairs of IFs that are both at least as informative as the W-IF (so they allow the voter to infer the winner).

**Lemma 1.** *Let  $\pi_1$  and  $\pi_2$  be two information functions such that  $\pi_1 \subseteq \pi_2 \subseteq \text{W-IF}$ . Then any voting rule that is  $\pi_2$ -manipulable is also  $\pi_1$ -manipulable.*

*Proof.* Suppose  $\mathcal{F}$  is  $\pi_2$ -manipulable, i.e., there exist a profile  $\mathbf{b}$ , a voter  $i$ , and a ballot  $b_i^*$  such that  $\mathcal{F}(b_i^*, \mathbf{b}_{-i}^*) \succ_i \mathcal{F}(\mathbf{b})$  for some  $\mathbf{b}_{-i}^* \in \mathcal{W}_i^{\pi_2(\mathbf{b})}$ , while  $\mathcal{F}(\mathbf{b}) \succ_i \mathcal{F}(b_i^*, \mathbf{b}'_{-i})$  for no  $\mathbf{b}'_{-i} \in \mathcal{W}_i^{\pi_2(\mathbf{b})}$ . We will show that then  $\mathcal{F}$  can also be manipulated by  $i$  using the same ballot  $b_i^*$  under IF  $\pi_1$  in case the truthful profile is  $(b_i, \mathbf{b}_{-i}^*)$ . First, as  $\mathbf{b}_{-i}^* \in \mathcal{W}_i^{\pi_2(\mathbf{b})}$ , we have  $\mathcal{W}_i^{\pi_2(\mathbf{b})} = \mathcal{W}_i^{\pi_2(b_i, \mathbf{b}_{-i}^*)}$  and thus, as  $\pi_2 \subseteq \text{W-IF}$ , we know that  $\mathcal{F}(\mathbf{b}) = \mathcal{F}(b_i, \mathbf{b}_{-i}^*)$ . As the actual profile is always consistent with the information received, we have  $(b_i, \mathbf{b}_{-i}^*) \in \mathcal{W}_i^{\pi_1(b_i, \mathbf{b}_{-i}^*)}$ . Thus, the first condition is satisfied: there exists a  $\mathbf{b}_{-i}^* \in \mathcal{W}_i^{\pi_1(b_i, \mathbf{b}_{-i}^*)}$  such that  $\mathcal{F}(b_i^*, \mathbf{b}_{-i}^*) \succ_i \mathcal{F}(\mathbf{b}) = \mathcal{F}(b_i, \mathbf{b}_{-i}^*)$ . As for the second condition, as  $\pi_1 \subseteq \pi_2$  we have  $\mathcal{W}_i^{\pi_1(b_i, \mathbf{b}_{-i}^*)} \subseteq \mathcal{W}_i^{\pi_2(b_i, \mathbf{b}_{-i}^*)}$  and thus  $\mathcal{W}_i^{\pi_1(b_i, \mathbf{b}_{-i}^*)} \subseteq \mathcal{W}_i^{\pi_2(\mathbf{b})}$ , meaning there is no  $\mathbf{b}'_{-i} \in \mathcal{W}_i^{\pi_1(b_i, \mathbf{b}_{-i}^*)}$  for which  $\mathcal{F}(b_i, \mathbf{b}'_{-i}) = \mathcal{F}(\mathbf{b}) \succ_i \mathcal{F}(b_i^*, \mathbf{b}'_{-i})$ .  $\square$

In the remainder of this section we prove several results on the  $\pi$ -manipulability of specific voting rules for specific choices of  $\pi$ , thereby complementing the partial picture drawn by Reijngoud and Endriss [2012]. Most of our results are negative, but for the case of the MG-IF we are able to identify two sets of sufficient conditions for immunity to manipulation.

<sup>3</sup>As an aside, note that even when  $\pi$  produces no information at all, there exist examples for voting rules, albeit not natural rules, that are susceptible to  $\pi$ -manipulation [Reijngoud and Endriss, 2012].

<sup>4</sup>We are grateful to an anonymous reviewer for pointing this out.

### 3.1 Winner Information Function

Reijngoud and Endriss [2012] proved that Plurality and Borda are W-manipulable, while Veto is immune to W-manipulation. We complement these results by showing that also  $k$ -approval, Copeland, and Maximin are W-manipulable.

**Theorem 2.** *The  $k$ -approval (for  $k \leq m - 2$ ), Copeland, and Maximin voting rules are all susceptible to W-manipulation.*

*Proof.* First, let  $\mathcal{F}$  be  $k$ -approval and assume that there is a voter  $i$  such that the winner in the truthful profile,  $\mathcal{F}(\succ)$ , is ranked below the  $(k + 1)$ <sup>st</sup> position in his preference order.<sup>5</sup> We now show that such a voter can W-manipulate by swapping the  $k$ <sup>th</sup> and  $(k + 1)$ <sup>st</sup> highest choices in his ballot. Indeed, since the current winner is less preferable to  $i$  than any candidate in the first  $k + 1$  positions, this manipulation is safe for any profile in  $\mathcal{W}_i^{\text{W}(\succ)}$ , and there exists a profile in this set in which the  $(k + 1)$ <sup>st</sup> preferred candidate of voter  $i$  needs only one additional point to win.

We similarly prove W-manipulability of Copeland and Maximin. Let voter  $i$  be such that the winner in the truthful profile is ranked below the second position in his preference order. Let  $i$  swap his first and second best choices. This manipulation does not affect the Copeland/Maximin score of the truthful winner and does not harm  $i$  in any profile in  $\mathcal{W}_i^{\text{W}(\succ)}$ , but there exists a profile in this set where  $i$ 's second best choice wins the election after the manipulation.  $\square$

### 3.2 Score Information Function

Since all PSRs are strongly computable from images under S-IF, it follows from the generalisation of the Gibbard-Satterthwaite Theorem [Reijngoud and Endriss, 2012] that these rules are S-manipulable. By Theorem 2 and Lemma 1, Maximin and Copeland are S-manipulable as well.

**Corollary 3.** *The Maximin and Copeland voting rules are susceptible to S-manipulation.*

### 3.3 Majority Graphs

We now turn to the case where the information made available to voters consists of the majority graph.

**Theorem 4.** *The Copeland and Borda voting rules are susceptible to MG-manipulation.*

*Proof.* For Copeland, as the winner can be computed given the majority graph, the result follows immediately from Theorem 2 together with Lemma 1.

For Borda, consider the following majority graph. Assume candidates  $c_1, c_2, \dots, c_{m-1}$  have the same number of incoming and outgoing edges between them, and candidate  $c_m$  is a Condorcet loser—i.e., it loses all pairwise majority contests. It follows then that  $c_m$  cannot be the winner, as Borda never elects a Condorcet loser [Fishburn and Gehrlein, 1976].

Let voter  $i$  be such that  $c_m \succ_i c_{m-1} \succ_i c_2 \succ_i c_1$ . Since his top choice is not winning, swapping it with the second best choice is a safe manipulation for voter  $i$ . We now show that there is a profile in the information set of voter  $i$  where this manipulation is beneficial for him. Let the division of

<sup>5</sup>Note that for Veto, with  $k = m - 1$ , this would be impossible.

Block-2			Block-3			
$w \dots w$	$w' \dots w'$		$w' \hat{c}_{m-2} \dots w' \hat{c}_{m-2}$	$\hat{c}_1 \hat{c}_{m-3} \dots \hat{c}_1 \hat{c}_{m-3}$		
$\hat{c}_1 \dots \hat{c}_1$	$\hat{c}_{m-2} \dots \hat{c}_{m-2}$		$\hat{c}_1 \hat{c}_{m-3} \dots \hat{c}_1 \hat{c}_{m-3}$			
$\vdots$	$\vdots$		$\vdots$	$\vdots$		
$\hat{c}_{m-2} \dots \hat{c}_{m-2}$	$\hat{c}_1 \dots \hat{c}_1$		$\hat{c}_{m-3} \hat{c}_1 \dots \hat{c}_{m-3} \hat{c}_1$			
$w' \dots w'$	$w \dots w$		$\hat{c}_{m-2} w \dots \hat{c}_{m-2} w$			
			$w w' \dots w w'$			

Table 1: Plurality under MG-IF. Preferences for case  $s \geq s'$ .

votes on the edges in the graph induced by  $c_1, c_2, \dots, c_{m-1}$  be  $\lfloor \frac{n}{2} \rfloor + 1$  against  $\lfloor \frac{n}{2} \rfloor - 1$  (for odd  $n$ ), and let it be  $n - 1$  against 1 on the edges between  $c_1, c_2, \dots, c_{m-1}$  and  $c_m$ . This implies that all the candidates  $c_1, c_2, \dots, c_{m-1}$  have equal Borda scores, and  $c_1$  wins by the tie-breaking. Now, after the manipulation by voter  $i$ , the score of candidate  $c_{m-1}$  increases, and she wins the election.  $\square$

We now turn to positive results. For sufficiently large values of  $n$ , Plurality is known to be immune to MG-manipulation [Reijngoud and Endriss, 2012]. Here we extend this result to all  $k$ -approval rules with  $k \leq m - 2$  (i.e., excluding Veto).

**Theorem 5.** *For sufficiently large numbers of voters  $n$ ,  $k$ -approval with  $k \leq m - 2$  is immune to MG-manipulation.*

*Proof sketch.* We give an alternative proof for Plurality that can be extended to  $k$ -approval, except for the Veto case. The idea is to show that, for any given majority graph, for any voter  $i$  and associated manipulation, there is a profile in his information set in which that manipulation may be harmful to  $i$ . To this end, we construct a profile with these properties:

- the majority graph is identical to the given MG;
- $w$  is a top choice alternative of any fixed voter  $i$ , and it wins the election in the constructed profile;
- $w'$  is the least favourite alternative of voter  $i$ , and by construction it either has the same score as  $w$  but loses to  $w$  by the tie-breaking, or has one point less than  $w$  if it is prioritised over  $w$  by the tie-breaking order.

Our construction consists of three blocks. Block-1 includes ballots that induce the given majority graph. We know, due to the McGarvey Theorem [Moon, 1968], that, for any given majority graph, such a profile exists. Now, let  $s$  be the score of  $w$  induced by voters in Block-1,  $s'$  the score of  $w'$  induced by voters in Block-1, and, finally,  $s^{\max}$  the maximum score of any candidate induced by the profile of Block-1. Also, let  $\hat{C} = \{\hat{c}_1, \dots, \hat{c}_{m-2}\} = C \setminus \{w, w'\}$  be the set of candidates enumerated according to the tie-breaking order, and  $\hat{C}'$  the same set enumerated in reverse order. Then, Block-2 and Block-3 (also depicted in Tables 1 and 2), will consist of  $4 \cdot s^{\max} + 8 + 2|s - s'|$  voters arranged as follows.

Block-2 consists of two sub-blocks, each containing  $2 \cdot s^{\max} + 4$  voters, with profile  $w \succ \hat{C} \succ w'$  repeated in the first sub-block, and  $w' \succ \hat{C}' \succ w$  repeated in the second.

Block-3 contains  $2 \cdot (s - s')$  voters and comes in two variants, depending on the relation between  $s$  and  $s'$ . Specifically, if  $s \geq s'$  then Block-3 consists of  $(s - s')$  pairs of voters, where the first voter has preference  $w' \succ \hat{C} \succ w$  and the second voter has preference  $\hat{C}' \succ w \succ w'$ . If, on the other hand,  $s' > s$ , then Block-3 consists of  $(s' - s)$  pairs of voters,

Block-2			Block-3			
$w \dots w$	$w' \dots w'$		$w \hat{c}_{m-2} \dots w \hat{c}_{m-2}$	$\hat{c}_1 \hat{c}_{m-3} \dots \hat{c}_1 \hat{c}_{m-3}$		
$\hat{c}_1 \dots \hat{c}_1$	$\hat{c}_{m-2} \dots \hat{c}_{m-2}$		$\hat{c}_1 \hat{c}_{m-3} \dots \hat{c}_1 \hat{c}_{m-3}$			
$\vdots$	$\vdots$		$\vdots$	$\vdots$		
$\hat{c}_{m-2} \dots \hat{c}_{m-2}$	$\hat{c}_1 \dots \hat{c}_1$		$\hat{c}_{m-3} \hat{c}_1 \dots \hat{c}_{m-3} \hat{c}_1$			
$w' \dots w'$	$w \dots w$		$\hat{c}_{m-2} w' \dots \hat{c}_{m-2} w'$			
			$w' w \dots w' w$			

Table 2: Plurality under MG-IF. Preferences for case  $s' > s$ .

where the first voter has the preference  $w \succ \hat{C} \succ w'$  and the second has the preference order  $\hat{C}' \succ w' \succ w$ .

Now, any manipulation by voter  $i$  will take a point from alternative  $w$  and make his least favourite candidate  $w'$  win the election in the constructed profile. So at least in this profile the manipulation is harmful for  $i$ .

The above idea extends to the case of  $k$ -approval; however, it requires a more complicated construction, which we omit from this paper for the sake of brevity. We do need to point out, though, that this construction requires at least two zero-score positions in the scoring rule—that is,  $k < m - 1$ .  $\square$

The Maximin rule is known to be susceptible to MG-manipulation, at least for the case of an even number of voters [Reijngoud and Endriss, 2012]. We now show that this negative result can be circumvented for profiles without a weak Condorcet winner (when interpreting this result, note that every Condorcet winner is also a weak Condorcet winner).

**Theorem 6.** *For profiles without a weak Condorcet winner, the Maximin voting rule is immune to MG-manipulation.*

*Proof.* Take any majority graph without weak Condorcet winners, and assume there exists a voter  $i$  that has an improving MG-manipulation. Thus, for some profile in voter  $i$ 's information set it is profitable to change the order between some alternatives  $c_j$  and  $c_k$  (w.l.o.g., assume that in the truthful ballot of voter  $i$  alternative  $c_j$  is ranked above  $c_k$ ).

Let  $c_i$  be the top alternative of voter  $i$ . We now show that there exists a profile in  $i$ 's information set such that the above manipulation is harmful to him. Again, we use the McGarvey Theorem [Moon, 1968] to construct a preference profile (with  $n$  even) that corresponds to the given majority graph such that the votes in each pairwise majority contest are as follows:

- The division of votes on the edge between  $c_j$  and  $c_k$  is  $\frac{n}{2} + 2$  against  $\frac{n}{2} - 2$  if  $c_k$  loses to the top alternative  $c_i$  due to tie-breaking, and it is  $\frac{n}{2} + 3$  against  $\frac{n}{2} - 3$  if  $c_k$  is prioritised over  $c_i$  by the tie-breaking order.
- The edge between  $c_k$  and  $c_i$  has the ratio of votes  $\frac{n}{2} + 2$  against  $\frac{n}{2} - 2$  if it is directed from  $c_k$  to  $c_i$ , and otherwise it has the ratio  $\frac{n}{2} - 1$  against  $\frac{n}{2} + 1$ .
- All other directed edges incoming to  $c_i$  have the ratio of votes  $\frac{n}{2} + 2$  against  $\frac{n}{2} - 2$ .
- All other directed edges incoming to  $c_k$  have the ratio of votes  $\frac{n}{2} + 1$  against  $\frac{n}{2} - 1$ .
- All other directed edges in the graph have the ratio of votes  $\frac{n}{2} + 3$  against  $\frac{n}{2} - 3$  (in the corresponding order).

Assume that  $c_k$  loses to  $c_i$  due to tie-breaking. In this case, the Maximin score of both  $c_i$  and  $c_k$  is  $\frac{n}{2} - 2$ , and  $c_i$  is the winner due to tie-breaking. However, after the manipulation by  $i$ ,

the division of votes on the edge between  $c_j$  and  $c_k$  becomes  $\frac{n}{2}+1$  against  $\frac{n}{2}-1$ , thus increasing the score of  $c_k$  and making her the winner, which is not beneficial for  $i$ . Similarly, if  $c_k$  wins against  $c_i$  due to tie-breaking, then the score of  $c_i$  is  $\frac{n}{2}-2$ , the score of  $c_k$  is  $\frac{n}{2}-3$ , and  $c_i$  wins the election. However, after the manipulation by  $i$ , the division of votes on the edge between  $c_j$  and  $c_k$  becomes  $\frac{n}{2}+2$  against  $\frac{n}{2}-2$ , thus increasing the score of  $c_k$  and making her the winner after tie-breaking with  $c_i$ , which again is not beneficial for  $i$ .  $\square$

## 4 Iterative Voting

As the results in the previous section demonstrate, although limiting the information that is communicated to voters may have a positive effect on the manipulability of a rule, there are still only very few cases in which we can fully rule out manipulation. Nevertheless, we now demonstrate that strategic behaviour under incomplete information converges to a stable outcome more often than in the case of complete information. We focus on the case of W-IF, providing only winner information, which is the most realistic IF we have considered, in the sense of requiring the least amount of information.

A *path* is a sequence  $(\mathbf{b}^0 \rightarrow \mathbf{b}^1 \rightarrow \dots)$  of profiles such that, for every  $k \geq 1$ , there exists a unique voter, say  $i$ , for which  $\mathbf{b}^k = (b'_i, \mathbf{b}^{k-1})$  for some  $b'_i \neq b_i^{k-1}$  in  $\mathcal{L}(C)$ . It is a  $\pi$ -*improvement path* if, for all  $k \geq 1$ , the move made by the unique deviator at step  $k$  is a  $\pi$ -manipulation. The set of allowed paths may be further restricted by a suitable *improvement dynamics*. For example, some previous work has assumed that voters always choose the *best* manipulation available to them. Iterative voting is based on such myopic improvement dynamics: the voters start by announcing an initial vote, and then repeatedly change their ballots in turn, one at a time. As in most previous work, we make the natural assumption that the initial profile is the *truthful* one—that is,  $\mathbf{b}^0 = (\succ_1, \dots, \succ_n)$ . We do not impose any restrictions on the order in which the voters apply their updates.

Following Obraztsova *et al.* [2015], we define an *iterative voting system with incomplete information*  $(\mathcal{F}, \pi, D)$  as the process based on the improvement dynamics  $D$  under the voting rule  $\mathcal{F}$  with IF  $\pi$ . We say that such a system *converges to a stable outcome* (though not necessarily a stable profile) if every  $\pi$ -improvement path that contains only moves allowed by  $D$  returns a stable winner after a finite number of steps, i.e., it either terminates or moves back and forth between several profiles that all produce the same winner.

We now generalise a condition for convergence, termed *function monotonicity* (FM), introduced by Obraztsova *et al.* [2015]. FM is based on the idea of exploiting a potential argument [Monderer and Shapley, 1996]. It uses a real-valued function  $G : \mathcal{L}(C)^n \rightarrow \mathbb{R}$  over the set of profiles that is required to (weakly) increase along any allowed improvement path. Formally, given system  $(\mathcal{F}, \pi, D)$  and profile  $\mathbf{b}$ , let

$$G(\mathbf{b}) = s_{\mathcal{F}}(\mathcal{F}(\mathbf{b}), \mathbf{b}) + \frac{m - \text{index}(\mathcal{F}(\mathbf{b}))}{m + 1},$$

where, for any candidate  $c$ , the number  $\text{index}(c)$  indicates her position in the tie-breaking order. Thus,  $G(\mathbf{b})$  is the score of the winner for profile  $\mathbf{b}$  under  $\mathcal{F}$ , increased by a number

between 0 and 1 that reflects the “bonus” score each candidate receives due to her position in the tie-breaking order. Then, FM holds if and only if, for any allowed improvement path  $(\mathbf{b}^0 \rightarrow \mathbf{b}^1 \rightarrow \dots)$ , we have  $G(\mathbf{b}^k) \leq G(\mathbf{b}^{k+1})$  for all  $k \geq 0$ .

The following result generalises the corresponding result of Obraztsova *et al.* [2015] to the case of incomplete information. Its proof is routine and omitted for the sake of brevity.

**Lemma 7.** *Every iterative voting system  $(\mathcal{F}, \pi, D)$  that satisfies FM converges to a stable outcome.*

Thus, to prove convergence it suffices to show that a given system satisfies FM. We first apply this technique to Copeland and Maximin under the unrestricted improvement dynamics allowing the full range of W-manipulations.

**Theorem 8.** *Under the winner information function W-IF, the Copeland and Maximin voting rules converge to a stable outcome for any kind of improvement dynamics.*

*Proof.* We will show that both Copeland and Maximin satisfy FM. The claim then follows from Lemma 7. Let  $\mathcal{F}$  stand for either one of them (we can use the same proof for both).

Call  $s_{\mathcal{F}}(c, \mathbf{b}) + (m - \text{index}(c))/(m + 1)$  the  $G$ -score of candidate  $c$  in profile  $\mathbf{b}$ . Thus,  $G$  defines a resolute voting rule that is equivalent to  $\mathcal{F}$  with lexicographic tie-breaking, so we can reason about  $G$  rather than  $\mathcal{F}$ . We will show that the  $G$ -score of the winner in a given profile does not decrease as we move to the next profile. Thus, in case the identity of the winner does not change when we move from  $\mathbf{b}^k$  to  $\mathbf{b}^{k+1}$ , we have  $G(\mathbf{b}^k) \leq G(\mathbf{b}^{k+1})$ . Furthermore, in case the identity of the winner does change, we get  $G(\mathbf{b}^k) < G(\mathbf{b}^{k+1})$ , as the  $G$ -score of the old winner cannot decrease and the  $G$ -score of the new winner must be high enough to beat the old one.

What remains to be shown is that, indeed, for both Copeland and Maximin the  $G$ -score of the current winner does not decrease during a manipulation. First, observe that, under W-IF, for these rules no voter will ever change the ranking of his least-preferred alternative. The reason is that, knowing only the identity of the winner, any kind of upgrade of the worst candidate risks increasing that candidate’s score and beating the former winner. So every voter will always rank his worst candidate at the bottom of his ballot<sup>6</sup> and thus have no opportunity to demote her even further. But this means that no manipulator can afford to allow the  $G$ -score of the current winner to decrease, as his worst candidate could currently be runner-up, so any kind of demotion of the current winner could cause his least-preferred candidate to win.  $\square$

We stress that we can only guarantee convergence to a stable *outcome*, not to a stable *profile*, i.e., it is possible that the voters will update their ballots indefinitely, although eventually these updates will not affect the identity of the winner anymore. For example, in an election in which candidate  $c$  is winning by a large margin, a (memory-less) voter  $i$  with true preferences  $a \succ_i b \succ_i c$  who only knows that  $c$  is winning but is unaware of her margin of victory will keep attempting to manipulate by swapping  $a$  and  $b$  in her ballot, not realising that he has no chance of affecting the outcome.

<sup>6</sup>Note that at this point we make use of the assumption that the initial profile is truthful (at least w.r.t. least preferred candidates).

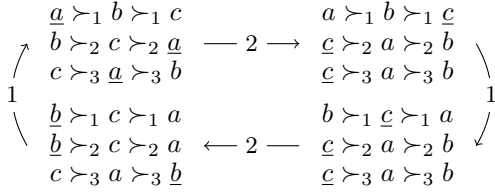


Figure 1: Cycle of safe W-manipulations from the truthful profile (top, left) for the PSR with scoring vector (3, 2, 0).

Theorem 8 cannot be extended to the PSRs, as the example in Figure 1 shows. Winners are underlined. In the initial truthful profile,  $a$  wins by tie-breaking. Then voter 2 (for whom  $a$  is worst) can safely manipulate and make  $c$  win. Then voter 1 manipulates, but unsuccessfully ( $c$  still wins). Now voter 2 manipulates again, making  $b$  win. Note that his manipulation is safe: there is no risk in demoting  $c$ , as  $a$  is demoted as well (losing more points than  $c$ ). Finally, voter 1 manipulates analogously, and we are back at the first profile.

Closer inspection of Figure 1 shows that the first two manipulation moves, while safe, arguably are not the most natural choices for our voters. For instance, in the initial profile voter 2 could have also manipulated by voting  $c \succ_2 b \succ_2 a$ , which would have achieved the same goal and which, intuitively, seems the more natural choice. This intuition can be formalised in a number of ways, e.g., by defining a suitable notion of “best” response. Here, instead, we opt for a much simpler solution. Under the *conservative improvement dynamics*, when choosing which of the available safe  $\pi$ -manipulations to play, a voter will always choose one that minimises the Kendall tau distance to his current ballot.

**Theorem 9.** *Under the winner information function W-IF and the conservative improvement dynamics, every positional scoring rule converges to a stable outcome.*

*Proof.* We use the same technique as for Theorem 8, i.e., we are done if we can show that no manipulator will ever remove his least preferred candidate from the bottom spot in his ballot. But this is clearly the case under W-IF and conservative improvements: if voter  $i$ 's least preferred candidate  $c$  is currently winning, no manipulation can make things worse, so  $i$  can choose one that keeps  $c$  at the bottom of his ballot. And if some other candidate is currently winning, it is not safe to put  $c$  in a position where she would get more than the minimum number of points, so  $i$  must keep  $c$  at a minimum-score spot and he can choose the bottom spot (the latter distinction only matters in case  $s_{m-1} = s_m$ , e.g., for Plurality).  $\square$

To put Theorems 8 and 9 into perspective, let us recall what other convergence results are known to date. We restrict attention to results that apply when agents initially vote truthfully and when ties are broken lexicographically (without these assumptions, results often are significantly weaker). Under complete information and best-response dynamics, Plurality converges [Meir *et al.*, 2010]. Lev and Rosensthein [2012] extended this result to the case of Veto, but also demonstrated that many other rules do not converge under these conditions [see also Reyhani and Wilson, 2012]. Several other authors introduced more restricted types of manipu-

lation moves, e.g., with voters attempting to manipulate only by swapping the first two candidates in their ballot [Grandi *et al.*, 2013] or by promoting their most preferred candidate amongst the current  $k$  front-runners [Reijngoud and Endriss, 2012], and were able to establish convergence results under those assumptions for a wider range of voting rules [see also Obraztsova *et al.*, 2015]. However, for the game-theoretical setting considered here, the only known positive results for the full information case concern Plurality and Veto.

The early work of Chopra *et al.* [2004] already hinted at the fact that the information each voter has access to will impact convergence, but these authors did not establish any concrete convergence results for specific voting rules. Reijngoud and Endriss [2012] used the same model of incomplete information as we do, but their convergence results concern heavily restricted forms of manipulation only. Meir *et al.* [2014] used a similar model of information, but focused only on Plurality. Thus, Theorems 8 and 9 significantly broaden the range of positive results regarding convergence in iterative voting.

## 5 Conclusion

We have studied the impact of incomplete information effects on the susceptibility of voting rules to strategic manipulation, and closed a number of gaps in existing research in this field. For one, although in many cases strategic behaviour by voters cannot be ruled out, even under heavy restrictions on the information they have access to, we have been able to obtain positive results on the immunity of certain voting rules when voters only have access to the majority graph, rather than to the full profile of individual preferences.

Our most important results concern the scenario of iterative voting, where we have been able to demonstrate that a process of repeated manipulations will converge for a number of voting rules when voters only have access to the information of which candidate is winning the election. This convincingly demonstrates the significance of the assumptions regarding information access we are willing to make, as similarly positive results so far have only been available for the very simplest of voting rules, namely Plurality and Veto.

Our results suggest a number of possible avenues for future work. First, both our study of manipulation under incomplete information and our study of convergence in iterative voting with incomplete information should be extended to a wider range of voting rules. Second, while we have considered particular forms of incomplete information that are natural abstractions of a full profile of preferences (and that might, e.g., get published in an opinion poll), other forms more closely associated with notions of uncertainty are also of interest (e.g., a voter may consider all those profiles possible that differ from some specific profile on at most  $k$  pairwise rankings). Such ideas have been explored by Meir *et al.* [2014] and could be applied to both the immunity and the convergence questions we have asked here. Third, in cases where convergence can be guaranteed, we should seek a better understanding of the properties of the election outcomes produced by iterative voting, in line with previous work by Reijngoud and Endriss [2012] and Grandi *et al.* [2013] for the case of restricted manipulation moves.

## Acknowledgments

This work has been partly supported by COST Action IC1205 on Computational Social Choice, by Israel Science Foundation grant number #1227/12, and by the Israeli Center of Research Excellence in Algorithms (I-CORE-ALGO).

## References

- [Chopra *et al.*, 2004] S. Chopra, E. Pacuit, and R. Parikh. Knowledge-theoretic properties of strategic voting. In *Proceedings of the 9th European Conference on Logics in Artificial Intelligence (JELIA-2004)*, 2004.
- [Conitzer *et al.*, 2011] V. Conitzer, T. Walsh, and L. Xia. Dominating manipulations in voting with partial information. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence (AAAI-2011)*, 2011.
- [van Ditmarsch *et al.*, 2013] H. van Ditmarsch, J. Lang, and A. Saffidine. Strategic voting and the logic of knowledge. In *Proceedings of the 14th Conference on Theoretical Aspects of Rationality and Knowledge (TARK-2013)*, 2013.
- [Fishburn and Gehrlein, 1976] P. C. Fishburn and W. V. Gehrlein. Borda's rule, positional voting, and Condorcet's simple majority principle. *Public Choice*, 28(1):79–88, 1976.
- [Gibbard, 1973] A. Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41(4):587–601, 1973.
- [Grandi *et al.*, 2013] U. Grandi, A. Loreggia, F. Rossi, K. B. Venable, and T. Walsh. Restricted manipulation in iterative voting: Condorcet efficiency and Borda score. In *Proceedings of the 3rd International Conference on Algorithmic Decision Theory (ADT-2013)*, 2013.
- [Lev and Rosenschein, 2012] O. Lev and J. S. Rosenschein. Convergence of iterative voting. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2012)*, 2012.
- [Meir *et al.*, 2010] R. Meir, M. Polukarov, J. S. Rosenschein, and N. R. Jennings. Convergence to equilibria in plurality voting. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI-2010)*, 2010.
- [Meir *et al.*, 2014] R. Meir, O. Lev, and J. S. Rosenschein. A local-dominance theory of voting equilibria. In *Proceedings of the 15th ACM Conference on Economics and Computation (EC-2014)*. ACM, 2014.
- [Monderer and Shapley, 1996] D. Monderer and L. Shapley. Potential games. *Games and Economic Behavior*, 14(1):124–143, 1996.
- [Moon, 1968] J. W. Moon. *Topics on Tournaments*. Holt, Rinehart and Winston, 1968.
- [Obraztsova *et al.*, 2015] S. Obraztsova, E. Markakis, M. Polukarov, Z. Rabinovich, and N. R. Jennings. On the convergence of iterative voting: How restrictive should restricted dynamics be? In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI-2015)*, 2015.
- [Reijngoud and Endriss, 2012] A. Reijngoud and U. Endriss. Voter response to iterated poll information. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2012)*, 2012.
- [Reyhani and Wilson, 2012] R. Reyhani and M. Wilson. Best reply dynamics for scoring rules. In *Proceedings of the 20th European Conference on Artificial Intelligence (ECAI-2012)*, 2012.
- [Satterthwaite, 1975] M. A. Satterthwaite. 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, 1975.
- [Taylor, 2005] A. D. Taylor. *Social Choice and the Mathematics of Manipulation*. Cambridge University Press, 2005.