# Strategic Manipulation with Incomplete Preferences: Possibilities and Impossibilities for Positional Scoring Rules

Justin Kruger
Université Paris-Dauphine
Paris, France
justin.g.kruger@gmail.com

Zoi Terzopoulou
Institute for Logic, Language and Computation
Amsterdam, The Netherlands
z.terzopoulou@uva.nl

## ABSTRACT

Many websites that recommend various services use crowdsourcing to collect reviews and rankings. These rankings, usually concerning a subset of all the offered alternatives, are then aggregated. Motivated by such scenarios, we axiomatise a family of positional scoring rules for profiles of possibly incomplete individual preferences. Many opportunities arise for the agents to manipulate the outcome in this setting. They may lie in order to obtain a better result by: $(i)$ switching the order of a ranked pair of alternatives, $(ii)$ omitting some of their truthful preferences, or $(iii)$ reporting more preferences than the ones they truthfully hold. After formalising these new concepts, we characterise all positional scoring rules that are immune to manipulation.

## KEYWORDS

Social Choice; Voting; Strategic Manipulation; Scoring Rules

**ACM Reference Format:**
Justin Kruger and Zoi Terzopoulou. 2020. Strategic Manipulation with Incomplete Preferences: Possibilities and Impossibilities for Positional Scoring Rules. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020,* IFAAMAS, 9 pages.

## 1 INTRODUCTION

The Internet seems to be a perfect environment for liars. Users can—and do—express falsehoods, using anonymity to avoid accountability. Another factor at play is the sheer amount of issues provided by a global network, far too many for any single person to have opinions about them all. So, a user can lie, not only in terms of the *quality*, but also in terms of the *quantity* of the information she holds. In fact, we can identify three types of lying: First, the user can assert something she believes to be false; second, she can refrain from asserting something that she believes to be true; third, she can assert something where in reality she has no opinion.

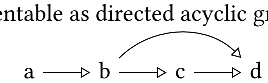
Although lying seems eminently possible for Internet users, there are many cases where it is nevertheless desired to solicit their true opinions, as when crowdsourcing reviews. In this paper we apply the three types of misrepresentation outlined above to preference aggregation over incomplete preferences. Specifically, we consider voting rules, which aim to select the best alternatives of a given set. Such rules, traditionally studied by economists, have recently received much attention from the community of computational social choice [6]. More generally, developments in technology have opened many new avenues for research concerning preference aggregation under the scope of multiagent systems and AI [8].

The study of incomplete preferences is one such avenue that has already been explored [2, 11, 12, 17, 19–22, 27]. In particular, Terzopoulou and Endriss [26] recently proposed a model that allows for incomplete preferences that are not necessarily transitive, covering cases like the one illustrated in the following example.

**Example 1.** You have been asked by a travel website to rank different hotels according to your preferences. You have only visited three hotels: *Sandy Cabins* (which is by the seaside), *Luxury Towers* (which is in the city), and *Snowy Chalets* (which is in the mountains). You know that you prefer *Sandy Cabins* to *Luxury Towers* when it is summer, and *Luxury Towers* to *Snowy Chalets* when it is winter. But you would only go to the seaside during summer, and you would only go to the mountains in winter, so comparing *Sandy Cabins* and *Snowy Chalets* does not make sense to you.

As well as providing an example of a non-transitive collection of pairwise preferences, Example 1 suggests that some preferences may be *intrinsically* incomplete.[1]

We abstract away from specific cases and refer to arbitrary alternatives as $a$, $b$, $c$,… Incomplete preferences over such alternatives are readily representable as directed acyclic graphs, for example:



An arrow $a \longrightarrow b$ means that $a$ is preferred to $b$. In text we will drop the tail of the arrow and write $a \triangleright b$.

Assuming that a group of agents report this kind of preferences, how should we determine the best, or winning, alternatives? We provide a contribution in this direction by generalising and characterising the class of positional scoring rules for incomplete preferences.[2] For now, suppose that we simply count how often an alternative is preferred to other alternatives, and select the alternatives with the highest count. For this method it may be profitable to lie by omission: Starting from the above preference, an agent could omit pairwise preferences where $b$ is preferred:

$$a \longrightarrow b \qquad c \longrightarrow d$$

Because this decreases the count of $b$, the new preference is a better ballot to submit in order to make $a$ win.

The above manipulation move would not be possible if the agents were required to submit complete preferences. Thus incomplete

[1] It may be argued that the agent in Example 1 has preferences over (hotel, season) pairs; nevertheless she must *express* these simply in terms of hotels, because this is what the website has specified as the ranking domain. Other agents may have different domains of preferences, like (hotel, weather) pairs. Although an interesting issue, we do not consider how preferences are translated across domains further than noting that it seems likely that expressed preferences can be non-transitive.

[2] Cf. the generalisation of scoring rules to multiwinner settings [13].

*Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 9–13, 2020, Auckland, New Zealand.* © 2020 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

preferences allow for novel types of manipulation: Besides *omitting* pairwise preferences, an agent may also invent preferences that she does not really hold, thereby *adding* pairwise preferences; she may also say she prefers one alternative to another when in fact the reverse is true, thereby *flipping* her preference on some pair of alternatives. For each of these, and their combinations, we provide necessary and sufficient conditions for positional scoring rules to make the given type of manipulation impossible.

**Related work.** The literature on manipulation in voting was pioneered by Gibbard [15] and Satterthwaite [24] who independently proved that, given some fairly weak conditions, no voting rule is immune to strategic manipulation. The conditions include requiring that the agents report complete preferences, as well as that the voting rule outputs a single winner (such voting rules are referred to as *resolute*). Later work has relaxed the single-winner condition, which leads to the question of how to extend preferences over alternatives to preferences over sets of alternatives. Early contributions concerning this issue were made by Gärdenfors [14] and Kelly [16], followed more recently by Duggan and Schwartz [10], Ching and Zhou [7], and Sato [23].[3]

Several papers have previously studied the aggregation of incomplete preferences as well [2, 11, 12, 17, 19–22, 26, 27]. A prevalent approach in computational social choice [2, 17, 27] is to consider the various *completions* of an incomplete preference profile and search for the alternatives that are *possible* or *necessary* winners. Such approaches are sensible if we assume that the preferences are completable [4], but this is not the case for preferences that are *intrinsically* incomplete, as described in Example 1.

Aggregation rules designed with genuinely incomplete preferences in mind were defined by Terzopoulou and Endriss [26]. These authors consider rules that assign weights to agents depending on the size of their truthful preferences, while we, in this paper, additionally explore positional scoring rules. Baumeister et al. [2] describe specific definitions of scoring rules as well, but restrict attention to truncated preferences and primarily provide complexity results, as opposed to our work which is of a more axiomatic and conceptual nature. Emerson [11] calls for more studies of voting with incomplete preferences and informally discusses a number of options for applying the Borda count on such settings.

Other papers have also tackled the problem of manipulation with incomplete preferences, obtaining similar results to ours. However, their starting concerns and definitions are somewhat different. Pini et al. [21] consider agents that can change their preference *in any way* and show that this makes it impossible to avoid *weak dictators* (i.e., agents that always have one of their most preferred alternatives among the winners). In contrast, our focus on specific types of manipulation of practical interest allows for some strong possibility results. The setting of Endriss et al. [12] differs from ours in that there agents can only report preferences of predetermined forms.

Implicit connections can also be drawn to work that considers complete preferences. For example, because Brandt [5] does not impose transitivity, taking the strict part of his preference relation leads to a similar framework to ours. His results concerning restricted versions of manipulation complement the results of our paper, as they apply to Condorcet-consistent rules—traditionally opposed to the scoring rules in which we are interested.

**Structure overview.** The basic voting model that we use is introduced in Section 2. Section 3 provides an axiomatisation of the class of positional scoring rules for incomplete preferences; it is complementary to but independent of the later sections. In Section 4 we formalise our notion of manipulation, which includes defining the three manipulation types of omission, addition, and flipping. We also characterise when positional scoring rules are immune to manipulation. Further results can be found in Sections 5 and 6, addressing combinations of manipulation types and single ones, respectively. We show that there is no positional scoring rule that prevents manipulation by omission and flipping or by omission and addition, but we obtain rules that prevent manipulation by both addition and flipping, and by omission alone. Section 7 concludes.

## 2 THE MODEL

In this section we present our basic framework.

Our scenario involves finite sets of *agents* $N = \{1, 2, \ldots, n\}$, with $n \geq 2$, and a finite set of *alternatives* $A = \{a, b, c, \ldots\}$, with $|A| \geq 3$. Every agent $i \in N$ holds *pairwise preferences* over the alternatives [26]. For example, $a \rhd_i b$ expresses that agent $i$ prefers $a$ to $b$, for $b \neq a$. The symbol $\rhd_i$ refers to agent $i$'s entire *preference*, which is a (possibly empty) set of strictly ordered pairs of alternatives:

$$\rhd_i = \{(a, b) \in A \times A : a \rhd_i b\}.$$

Given a preference $\rhd$, two alternatives $a$ and $b$ are *connected* if $a = b$, or if there is a path from $a$ to $b$ in the undirected version of the graph defined by $\rhd$. Similarly, $a$ and $b$ are said to be in an *undirected cycle* if that graph has a cycle that contains both, or if $a = b$.

We assume that preferences are *acyclic*: If $a_1 \rhd a_2 \rhd \cdots \rhd a_k$, then it cannot be the case that $a_k \rhd a_1$, for $2 \leq k \leq |A|$. We denote by $\mathcal{D}$ the set of all acyclic preferences over $A$. We stress that an acyclic preference $\rhd$ may not be transitive; there may exist alternatives $a, b, c \in A$ such that $a \rhd b$ and $b \rhd c$, but $a \not\rhd c$.

A *profile* $\blacktriangleright = (\rhd_1, \ldots, \rhd_n) \in \mathcal{D}^n$ collects the preferences of all agents in $N$. We write $(ab)$ as shorthand for the permutation on $A$ that swaps $a$ and $b$, and we write $\rhd_{(ab)}$ for the preference $\rhd$ with every occurrence of $a$ and $b$ switched. We also apply this notation to profiles: $\blacktriangleright_{(ab)}$ is the profile $\blacktriangleright$ with $a$ and $b$ switched. Given a subset of the agents $I \subseteq N$, a partial profile $\blacktriangleright_{-I}$ denotes the part of $\blacktriangleright$ where all agents besides those in $I$ report their preferences. By $(\rhd, \ldots, \rhd, \blacktriangleright_{-I})$ we denote the profile where agents in $N \backslash I$ report the same preferences as in $\blacktriangleright$, and where all agents in $I$ report the same preference $\rhd$. Two profiles $\blacktriangleright = (\rhd_1, \ldots, \rhd_n) \in \mathcal{D}^n$ and $\blacktriangleright' = (\rhd_{n+1}, \ldots, \rhd_{n+k}) \in \mathcal{D}^k$ can be combined to form

$$(\blacktriangleright, \blacktriangleright') = (\rhd_1, \ldots, \rhd_n, \rhd_{n+1}, \ldots, \rhd_{n+k}) \in \mathcal{D}^{n+k}.$$

A *voting rule* $F$ is a function that maps a profile $\blacktriangleright \in \mathcal{D}^n$, for any group $N$, to a nonempty subset of $A$, that is, the set of winners. Thus, $F(\blacktriangleright)$ may contain several, tied, winners.

We are specifically interested in voting rules that hinge on assigning points to alternatives. A *scoring function* is a function $s \colon (\mathcal{D} \times A) \to \mathbb{R}$ that assigns a score to every alternative in a

[3] We relax the single-winner condition because the positional scoring rules that we consider satisfy anonymity and neutrality, which immediately implies that the output must be irresolute. In the complete setting, positional scoring rules were axiomatised by Smith [25] and Young [29]. Myerson [18] generalised the axiomatisation to profiles of votes that could take any form over the set of alternatives.

given preference. We will write $s_{\triangleright}(a)$ as an abbreviation for $s(\triangleright, a)$. A scoring function is *non-trivial* if $s_{\triangleright}(x) \neq s_{\triangleright}(y)$ for some $x, y \in A$ and $\triangleright \in \mathcal{D}$. A *positional scoring function* ensures moreover the symmetrical treatment of all alternatives; one may thus think of scores assigned to positions in a graph. Formally, a scoring function $s$ is positional if and only if for all permutations $\sigma\colon A \to A$, preferences $\triangleright \in \mathcal{D}$, and alternatives $x \in A$, we have that $s_{\triangleright}(x) = s_{\triangleright_\sigma}(x_\sigma)$, where $x_\sigma = \sigma(x)$ and $\triangleright_\sigma = \sigma(\triangleright) = \{(a_\sigma, b_\sigma) : a \triangleright b\}$.

A scoring rule for incomplete preferences $F_s$, associated with a scoring function $s$, takes a profile of incomplete preferences and returns all alternatives with maximal scores, where the score of an alternative is the sum of the scores given by the scoring function across all agents. Formally, $F_s\colon \mathcal{D}^n \to 2^A \backslash \{\emptyset\}$ is defined by

$$F_s(\blacktriangleright) = \underset{x \in A}{\operatorname{argmax}} \sum_{i \in N} s_{\triangleright_i}(x).$$

Note that for distinct scoring functions $s$ and $s'$ it may be the case that $F_s = F_{s'}$. Then, we say that $s$ and $s'$ are *equivalent*. We withhold proof of the following in the interest of space.

**Proposition 1.** *Two scoring functions $s$ and $s'$ are equivalent if and only if there are real numbers $\{\alpha_\triangleright\}_{\triangleright \in \mathcal{D}}$ and $\beta > 0$ such that for all alternatives $x \in A$ and preferences $\triangleright$, $s_\triangleright(x) = \alpha_\triangleright + \beta \cdot s'_\triangleright(x)$.*

Here are some positional scoring functions that are new to the literature of incomplete preferences, of which domination scoring was informally described in the introduction.

**Domination scoring.** Define $\mathrm{ds}\colon (\mathcal{D} \times A) \to \mathbb{R}$ by

$$\mathrm{ds}_\triangleright(x) = |\{y \in A : x \triangleright y\}|.$$

**Cumulative scoring.** Define $\mathrm{cs}\colon (\mathcal{D} \times A) \to \mathbb{R}$ by[4]

$$\mathrm{cs}_\triangleright(x) = \begin{cases} 0 & \text{if } x \triangleright y \text{ for no } y \in A, \\ 1 + \sum_{\substack{y \in A \\ x \triangleright y}} \mathrm{cs}_\triangleright(y) & \text{otherwise.} \end{cases}$$

**Veto scoring.** Define $\mathrm{vs}\colon (\mathcal{D} \times A) \to \mathbb{R}$ by

$$\mathrm{vs}_\triangleright(x) = \begin{cases} 1 & \text{if there is some } y \in A \text{ such that } x \triangleright y, \\ 0 & \text{otherwise.} \end{cases}$$

**Stepwise scoring.** Define $\mathrm{ss}\colon (\mathcal{D} \times A) \to \mathbb{R}$ by

$$\mathrm{ss}_\triangleright(x) = \sum_{i=1}^{k} \bar{s}(C_i, C),$$

where $C = \{C_1, \dots, C_k\}$ is a partition of $A$ into sets of alternatives that are in the same undirected cycle, $x \in C$ for $C \in \mathcal{C}$, and $\bar{s} : (\mathcal{C} \times \mathcal{C}) \to \mathbb{R}$ is defined below (for $\triangleright_u^c$ denoting the undirected version of the relation $\triangleright^c \subseteq \mathcal{C} \times \mathcal{C}$ such that $C_i \triangleright^c C_j$ if and only if $x \triangleright y$ for some $x \in C_i$ and $y \in C_j$):

If $C_i = C_j$, then $\bar{s}(C_i, C_j) = 0$, and if $C_i \neq C_j$, then

$$\bar{s}(C_i, C_j) = \begin{cases} 1/2 & \text{if } C_i \triangleright_u^c C_1 \triangleright_u^c \dots C_m \triangleright^c C_j \text{ for some } m \geq 0 \\ -1/2 & \text{if } C_j \triangleright^c C_1 \triangleright_u^c \dots C_m \triangleright_u^c C_i \text{ for some } m \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Stepwise scoring assigns to two alternatives $x$ and $y$ the same score when they are in the same undirected cycle. Otherwise, if

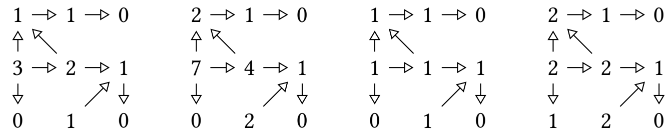



**Figure 1: Left to right; scores given by domination, cumulative, veto, and stepwise scoring for an example preference.**

$x$ is preferred to $y$, then $\mathrm{ss}_\triangleright(x) = \mathrm{ss}_\triangleright(y) + 1$. The cumulative-scoring function gives each alternative a score one greater than the summed score of all the alternatives to which it is preferred. For complete input profiles, the domination-scoring rule reduces to the Borda rule [3]; the cumulative-scoring rule reduces to the (complete) positional scoring rule with score vector $(2^{|A|-2}, \dots, 4, 2, 1, 0)$; the veto-scoring rule reduces to the veto rule; and the stepwise-scoring rule reduces to the trivial rule that returns the set of all alternatives for all input profiles. See Figure 1 for an illustration of the different scores these functions give for an example preference.

## 3 AXIOMATISATION OF SCORING RULES FOR INCOMPLETE PREFERENCES

Scoring rules, being intuitive and easy to understand, are widely used in practice, for example in political elections, sports competitions, and web applications. They also satisfy a number of appealing normative properties, traditionally known as *axioms*. The following are obvious translations of axioms from the complete [25, 29] to the incomplete framework.

**Anonymity.** For all sets of agents $N$, permutations $\pi\colon N \to N$, and profiles $\blacktriangleright = (\triangleright_1, \dots, \triangleright_n)$, it holds that $F(\pi(\blacktriangleright)) = F(\blacktriangleright)$, where $\pi(\blacktriangleright) = (\triangleright_{\pi(1)}, \dots, \triangleright_{\pi(n)})$.

**Neutrality.** For all permutations $\sigma\colon A \to A$ and profiles $\blacktriangleright = (\triangleright_1, \dots, \triangleright_n)$, it holds that $F(\sigma(\blacktriangleright)) = \sigma(F(\blacktriangleright))$, where $\sigma(\blacktriangleright) = (\sigma(\triangleright_1), \dots, \sigma(\triangleright_n))$ and $\sigma(F(\blacktriangleright)) = \{\sigma(a) : a \in F(\blacktriangleright)\}$.

**Reinforcement.** For all profiles $\blacktriangleright$ and $\blacktriangleright'$, if $F(\blacktriangleright) \cap F(\blacktriangleright') \neq \emptyset$, then $F(\blacktriangleright, \blacktriangleright') = F(\blacktriangleright) \cap F(\blacktriangleright')$.

**Continuity.** For all profiles $\blacktriangleright$ and $\blacktriangleright'$, there exists a positive integer $K$ such that, for every integer $k$ that is greater than $K$, it holds that $F(\underbrace{\blacktriangleright, \dots, \blacktriangleright}_{k}, \blacktriangleright') \subseteq F(\blacktriangleright)$.

Of these axioms, anonymity, reinforcement, and continuity hold for scoring rules [18]. This leaves neutrality, which is to be connected to positional scoring functions. However, non-positional and positional scoring functions may be equivalent.[5] Because of this, non-positional scoring functions may define neutral scoring rules. Let us define a positional scoring *rule* as one that can be represented by *some* positional scoring function. Lemma 1 helps bridge the gap between positional scoring rules and positional scoring functions, and leads into the axiomatisation (Theorem 1).

**Lemma 1.** *A scoring function $s$ is equivalent to some positional scoring function if for all alternatives $x, y \in A$ and preferences $\triangleright \in \mathcal{D}$,*

$$s_\triangleright(x) - s_\triangleright(y) = s_{\triangleright_{(xy)}}(y) - s_{\triangleright_{(xy)}}(x), \quad \textit{and} \tag{1}$$

$$s_\triangleright(z) - s_\triangleright(y) = s_{\triangleright_{(xy)}}(z) - s_{\triangleright_{(xy)}}(x) \quad \textit{for all } z \neq x, y. \tag{2}$$

[4] Note that the definition of the cumulative scoring is recursive.

[5] E.g., take $a, b \in A$: For $a \triangleright b$ let $s_\triangleright(a) = 1$, $s_\triangleright(b) = 0$; for $b \triangleright' a$, let $s_{\triangleright'}(a) = 2$, $s_{\triangleright'}(b) = 1$; set all other scores to 0.

*Proof.* By the fact that all permutations over $A$ can be expressed as successive permutations only involving two alternatives, (1) and (2) together imply that, for an arbitrary permutation function $\sigma$, preference $\rhd$, and alternatives $a$ and $b$,

$$s_{\rhd}(a) - s_{\rhd}(b) = s_{\rhd_\sigma}(a_\sigma) - s_{\rhd_\sigma}(b_\sigma). \tag{3}$$

Consider $|A|$ applications of (3) to an arbitrary $a \in A$:

$$\sum_{x\in A} s_{\rhd}(a) - s_{\rhd}(x) = \sum_{x\in A} s_{\rhd_\sigma}(a_\sigma) - s_{\rhd_\sigma}(x_\sigma).$$

This can be rewritten as follows:

$$|A| \cdot s_{\rhd}(a) - \sum_{x\in A} s_{\rhd}(x) = |A| \cdot s_{\rhd_\sigma}(a_\sigma) - \sum_{x\in A} s_{\rhd_\sigma}(x_\sigma).$$

So, given a scoring function $s'$ that satisfies (1) and (2), in order to show that $s'_{\rhd}(a) = s'_{\rhd_\sigma}(a_\sigma)$ it suffices to show that

$$\sum_{x\in A} s'_{\rhd}(x) = \sum_{x\in A} s'_{\rhd_\sigma}(x_\sigma). \tag{4}$$

Suppose that (1) and (2) hold for $s$. We want to find an equivalent positional scoring function $s'$. If $s$ is positional, we are done because it is equivalent to itself. If not, there are some $\rhd^*$, some permutation function $\sigma$, and some alternative $x$ such that $s_{\rhd^*}(x) \neq s_{\rhd^*_\sigma}(x_\sigma)$. Fix an alternative $a \in A$. Define $s'$ as follows: First set $s'_{\rhd^*}(x) = s_{\rhd^*}(x)$. Next, for each distinct permutation $\rhd^*_\rho$ of $\rhd^*$, let

$$s'_{\rhd^*_\rho}(y) = s_{\rhd^*_\rho}(y) + s_{\rhd^*}(a) - s_{\rhd^*_\rho}(a_\rho) \quad \text{for all } y \in A. \tag{5}$$

After going through the permutations of $\rhd^*$, there may still be some other $\rhd^\dagger$, $\omega$ and $y$ such that $s_{\rhd^\dagger}(y) \neq s_{\rhd^\dagger_\omega}(y_\omega)$, in which case the process needs to be repeated for $\rhd^\dagger$. As there are only finitely many preferences and the process will not consider the same preference twice, eventually there will be no more such cases. For all remaining preferences $\rhd$ set $s'_{\rhd} = s_{\rhd}$. Note that $s'$ is equivalent to $s$ and that (1) and (2) still hold.

We have above defined $s'$. Now take arbitrary $x \in A$, preference $\rhd$, and permutation function $\sigma$. We want to show that $s'_{\rhd}(x) = s'_{\rhd_\sigma}(x_\sigma)$. If the score $s$ was changed for no permutation of $\rhd$, then $s_{\rhd}(x) = s_{\rhd_\sigma}(x_\sigma)$ and indeed $s'_{\rhd}(x) = s'_{\rhd_\sigma}(x_\sigma)$ as required. Otherwise, there are some $\rhd^*$ for which $s_{\rhd^*} = s'_{\rhd^*}$, and permutation functions $\rho$, $\omega$ such that $\rhd^*_\rho = \rhd$ and $\rhd^*_\omega = \rhd_\sigma$. We can suppose that these are those permutation functions that were used in the construction of $s'$—if either is instead the identity permutation, the required equality trivially holds. Regardless, (4) holds:

$$\begin{aligned}
\sum_{x\in A} s'_{\rhd}(x) &= \sum_{x\in A} s'_{\rhd^*_\rho}(x_\rho) && \text{because alternatives appear exactly once}\\
&= \sum_{x\in A} s_{\rhd^*_\rho}(x_\rho) + s_{\rhd^*}(a) - s_{\rhd^*_\rho}(a_\rho) && \text{by (5)}\\
&= \sum_{x\in A} s_{\rhd^*}(x) && \text{by (3)}\\
&= \sum_{x\in A} s'_{\rhd^*}(x) && \text{by definition of } s'
\end{aligned}$$

The required equality concerning $\rhd^*$ and $\rhd^*_\omega = \rhd_\sigma$ can be obtained in an identical manner. □

**Theorem 1.** *A voting rule for possibly incomplete preferences in $\mathcal{D}$ is a positional scoring rule if and only if it satisfies anonymity, neutrality, reinforcement, and continuity.*

*Proof.* Every positional scoring rule obviously satisfies all the axioms of the statement, so the "only if" holds. For the "if"; a voting rule over incomplete preferences that satisfies anonymity, neutrality, reinforcement, and continuity has to be a scoring rule $F_s$ for some scoring function $s$ by Myerson [18]. If $s$ is trivial, then $F_s$ is positional. For non-trivial $s$, we show that (1) and (2) of Lemma 1 hold, and thus that the scoring rule $F_s$ is positional.

For (1), fix arbitrary alternatives $a, b \in A$ and preference $\rhd$. We will construct a profile where both $a$ and $b$ are winning, and such that if (1) did not hold they could not have the same score. For our construction we require a profile where $a$ and $b$ have the same summed score which is arbitrarily larger than the score of any other alternative. To do this, we first need that

$$\text{there exists a profile } \blacktriangleright \text{ such that } 2 \leq |F_s(\blacktriangleright)| < |A|. \tag{6}$$

Since $s$ is not trivial, there is some $\rhd'$ such that for nonempty $X \subsetneq A$, it holds that $s_{\rhd'}(x) = s_{\rhd'}(y) > s_{\rhd'}(z)$ for all $x, y \in X$, $z \in A \backslash X$. If $|X| > 1$ for some such set $X$, we have the required profile of (6). So suppose that $X = \{x\}$ for all relevant $X$. Pick $y$ such that $s_{\rhd'}(y) \geq s_{\rhd'}(z)$ for all $z \neq x$. It is true by neutrality that $y$ gets the unique highest score on $\rhd'_{(xy)}$ and that in the profile $(\rhd', \rhd'_{(xy)})$, $x$ and $y$ must be joint winners with the same summed score. So in this case we also have the required profile of (6).

Now, (6) provides a profile $\blacktriangleright$ where some distinct $x$ and $y$ are winning and some $z$ is not winning. Let $\Omega$ be the set of all permutations where $x$ and $y$ are fixed points. In the profile $(\omega(\blacktriangleright), \omega(\blacktriangleright_{(xy)}))_{\omega\in\Omega}$ the alternatives $x$ and $y$ are the unique winners. To see this, take some alternative $z \neq x, y$, and note that, by neutrality, for each profile $\omega(\blacktriangleright)$, alternative $x$ gets score at least as high as the score of $z$, and for some such profile $x$ gets strictly higher score than $z$; the same applies for $\omega(\blacktriangleright_{(xy)})$, so $z$ does not win; finally by neutrality $x$ cannot win without $y$ winning. Now permute $a$ with $x$ and $b$ with $y$. With enough copies of this profile we can create an arbitrarily large gap between the score of $a$ and $b$ and the scores of other alternatives. Note that permuting $a$ and $b$ does not change this profile.

Let $t$ be the maximal difference between the score of two alternatives in $\rhd$ and between the score of two alternatives in $\rhd_{(ab)}$. Formally,

$$t = \max\left(\max_{x,y\in A} s_{\rhd}(x) - s_{\rhd}(y),\ \max_{x,y\in A} s_{\rhd_{(ab)}}(x) - s_{\rhd_{(ab)}}(y)\right).$$

There is a profile $\blacktriangleright'$ where $a$ and $b$ get (the same) summed score at least $2t+1$ greater than all the other alternatives and $\blacktriangleright'_{(ab)} = \blacktriangleright'$. Let us add $\rhd$ and $\rhd_{(ab)}$ to $\blacktriangleright'$. By construction, no alternative other than $a$ and $b$ can have the maximal summed score. By neutrality, both $a$ and $b$ must be winning. But if (1) did not hold, only one of them would be winning. Thus part (1) of Lemma 1 must be satisfied.

The general procedure for part (2) of Lemma 1 is similar; though we must also start with arbitrary $c \in A \setminus \{a, b\}$. As before there is a profile $\blacktriangleright$ such that $F_s(\blacktriangleright) = \{a, b\}$ and $\blacktriangleright_{(ab)} = \blacktriangleright$. Define

$\blacktriangleright' = (\blacktriangleright_{(ac)}, \blacktriangleright_{(bc)})$, for which:

$$F_s(\blacktriangleright') = \{c\} \qquad \text{and} \qquad \blacktriangleright'_{(ab)} = \blacktriangleright'.$$

Note further that $a$ and $b$ receive the same score under $\blacktriangleright'$. Next define $\blacktriangleright'' = (\blacktriangleright, \blacktriangleright_{(ac)}, \blacktriangleright_{(bc)})$, for which:

$$F_s(\blacktriangleright'') = \{a, b, c\} \qquad \text{and} \qquad \blacktriangleright''_{(ab)} = \blacktriangleright''.$$

Suppose that (2) does not hold. Suppose in particular[6] that

$$s_{\rhd}(c) - s_{\rhd}(b) > s_{\rhd_{(ab)}}(c) - s_{\rhd_{(ab)}}(a).$$

Roughly speaking, $c$ performs "better" against $b$ under $\rhd$ than $c$ performs against $a$ under $\rhd_{(ab)}$. For an integer $k$, write $k.\rhd$ for $k$ copies of $\rhd$ and $k.\blacktriangleright$ for $k$ copies of $\blacktriangleright$. We can choose suitable integers $j$, $k$, $\ell$, and $m$ such that

- $F_s(j.\rhd, k.\blacktriangleright, l.\blacktriangleright', m.\blacktriangleright'') = c$, but also
- under $(j.\rhd_{(ab)}, k.\blacktriangleright, l.\blacktriangleright', m.\blacktriangleright'')$ the score of $a$ is greater than the score of $c$.

But $(j.\rhd, k.\blacktriangleright, l.\blacktriangleright', m.\blacktriangleright'')_{(ab)} = (j.\rhd_{(ab)}, k.\blacktriangleright, l.\blacktriangleright', m.\blacktriangleright'')$, which contradicts neutrality. □

# 4 MANIPULATION: DEFINITIONS AND CHARACTERISATIONS

In this section we describe the two main facets of manipulation: *when* it is profitable, and *how* it can be performed. We then provide necessary and sufficient conditions for positional scoring rules to be immune to manipulation.

## 4.1 When is it profitable to manipulate?

A preference $\rhd_i$ indicates which single alternatives are preferred to other single alternatives by agent $i$. But our voting rules output sets of alternatives. Thus, we need to determine when agent $i$ wants to manipulate from a set $X$ to a set $Y$, for $X, Y \subseteq A$. As a starting point, we assume that if $a \rhd_i b$, then agent $i$ will prefer $\{a\}$ to $\{b\}$.

**Definition 1.** Given a preference $\rhd$, the $m$-extension is a binary relation $\rhd^m$ over sets of alternatives such that

$$X \rhd^m Y \text{ if and only if } X = \{x\},\ Y = \{y\}, \text{ and } x \rhd y.$$

A *minimal* requirement of non-manipulability would be that no agent can change the outcome to a better one according to the $m$-extension of their preference.

There is a broad literature concerning more elaborate extensions of preferences from singletons to preferences over sets of objects (Barberà et al. [1] provide a survey).[7] Our intended interpretation of extensions is directly connected to the notion of manipulation. We define a *large* extension that does not represent the preferences of some particular agent, but rather the possible preferences that someone might conceivably have. The idea is that if *some* new alternative is preferred to *some* old alternative, an agent may desire to manipulate from the old to the new.

[6] The case of the inverse strict inequality is symmetric.

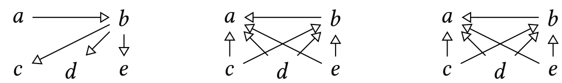
[7] Typically, extensions aim to describe the induced preferences over sets of objects, given various assumptions about the agent's psychology (e.g., being risk-adverse or risk-loving) and about how the alternative sets will actually be consumed (e.g., a random alternative must be used or all alternatives must be used).

**Definition 2.** Given a preference $\rhd$, the $\ell$-extension is a binary relation $\rhd^\ell$ over sets of alternatives such that

$$X \rhd^\ell Y \text{ if and only if there exist } x \in X \text{ and } y \in Y \text{ such that } x \rhd y \text{ and } \{x, y\} \nsubseteq X \cap Y.$$

Note that the binary relation given by the $\ell$-extension may not be acyclic; it may even contain symmetries. Preventing all manipulations of the $\ell$-extension would constitute a strong positive result.

**Example 2.** Suppose that there are three agents with sincere preferences as follows.




The summed domination scores are then:

| a | b | c | d | e |
|---|---|---|---|---|
| 1 | 5 | 4 | 4 | 4 |

The left agent can make $\{a, c, d, e\}$ winning. This is a possible manipulation according to the $\ell$-extension because $\{a, c, d, e\} \rhd^\ell \{b\}$. Note that $\{b\} \rhd^\ell \{a, c, d, e\}$ as well.

## 4.2 How can manipulation be performed?

We have described when we consider an outcome as more desirable for a manipulator; we still need to determine who the manipulators are, and how they can attempt to achieve this outcome.

Classic studies of manipulation consider a single agent acting alone. However, many applications of preference aggregation to the Internet require a broader definition of manipulative acts. A user can now easily create and handle several identities (all sharing the same truthful preference).[8] Our model allows for an—arbitrarily large—group of agents with identical preferences to manipulate.[9]

**Definition 3.** Consider a voting rule $F$, an extension $\star$, a subset of agents $I \subseteq N$ with the preference $\rhd$, and a profile

$$\blacktriangleright = (\rhd_1, \ldots, \rhd_n) = (\underbrace{\rhd, \ldots, \rhd}_{k}, \blacktriangleright_{-I}) \in \mathcal{D}^n.$$

Without loss of generality, suppose that $I = \{1, \ldots, k\}$. A $\star$-manipulation of $F$ by $I$ is possible if there exists some vector of preferences $(\rhd'_1, \ldots, \rhd'_k) \in \mathcal{D}^n$ such that $F(\rhd'_1, \ldots, \rhd'_k, \blacktriangleright_{-I}) \rhd^\star F(\blacktriangleright)$.

The rule $F$ is *manipulable* if a manipulation is possible by some group $I \subseteq N$, for some $N$, and *immune to manipulation* otherwise.

Definition 3 tells us who can manipulate. It does not tell us in what manner. We can sum up the three types of misrepresentation discussed in the introduction with reference to the following typical commitment of sworn testimony: that one tells the truth, the whole truth, and nothing but the truth. This can be decomposed into three parts, each prohibiting a different type of manipulation.

**Definition 4.** Consider a $\star$-manipulation of $F$ by $I$, as in Definition 3. Such a manipulation is by

**flipping** if $\rhd_i \setminus \rhd'_i = \{(x, y) : (y, x) \in \rhd'_i \setminus \rhd_i\}$ for all $i \in I$;
**omission** if $\rhd'_i \subseteq \rhd_i$ for all $i \in I$; or

[8] This is related to *false-name* manipulation [28] and so-called *Sybil attacks* [9].

[9] Besides its practical significance, allowing for group-manipulation gives stronger sufficient conditions for preventing manipulation. In the other direction, we only require at most two agents for the necessity proofs, and the second agent may be considered as a somewhat technical requirement in order to break ties.

**addition** if $\triangleright_i \subseteq \triangleright'_i$ for all $i \in I$.

We can also naturally define manipulation by *combinations* of the above. For instance, a manipulation is by a *combination of addition and omission* if the agents add some pairwise preferences and possibly omit some others. We say that two preferences $\triangleright$ and $\triangleright'$ *conform to* a certain (combination of) manipulation type(s) if $\triangleright$ and $\triangleright'$ satisfy the corresponding relations of Definition 4.

## 4.3 Necessary and sufficient conditions for immunity to manipulation

If scoring rules have certain forms, then they can be immune to manipulation even for the very large extension $\ell$ (Theorem 2). This follows from our sufficient conditions (which, as we will see, are not always easy to satisfy). In the other direction, we present necessary conditions for a rule to be immune to manipulation at least for the small extension $m$, thus establishing vital boundaries for non-manipulability (Theorem 3). Lemma 2 is critical for these results.

**Lemma 2.** *If a positional scoring function $s$ is not trivial, then for all positive numbers $t > 0$ and all alternatives $x$ and $y$, there exists a profile $\blacktriangleright$ such that for all $z \neq x, y$,*

$$\sum_{i \in N} s_{\triangleright_i}(x) = \sum_{i \in N} s_{\triangleright_i}(y) > t + \sum_{i \in N} s_{\triangleright_i}(z).$$

*Proof.* Take two arbitrary alternatives $x, y \in A$. For a non-trivial scoring function $s$ there is a preference $\triangleright$ for which we can order the alternatives using indices $j \in \{1, \ldots, |A|\}$ such that $s(a_j) \geq s(a_{j+1})$ and $s(a_1) > s(a_{|A|})$. Consider a permutation of this preference where $x$ is placed in the position of $a_1$, $y$ is placed in the position of $a_2$, some $z \neq x, y$ is placed in the position of $a_3$, etc. Now iteratively create new preferences which swap the positions of $x$ and $y$ and cycle through the positions of the other alternatives: After $2\,|A| - 4$ iterations this results in a collection of preferences for which $x$ is in the position of $a_1$ exactly $|A| - 2$ times and in the position of $a_2$ exactly $|A| - 2$ times; where $y$ is in those positions exactly as often as $x$; and where, for each $j = \{3, \ldots, |A|\}$, every other alternative is in the position $a_j$ exactly twice. Obviously $x$ and $y$ have the same summed score for these preferences. Similarly, all other alternatives have the same summed scores as each other. The inequalities over the scores of ordered alternatives imply that the summed score of $x$, and that of $y$, is some value $\delta > 0$ larger than the summed score of any other alternative. The required profile is created by taking $1 + \lceil t/\delta \rceil$ copies of the $2\,|A| - 4$ preferences. □

Given two preferences $\triangleright$ and $\triangleright'$ and alternatives $x, y \in A$, let us define the following inequality, intuitively stating that the difference between the scores of $x$ and $y$ in $\triangleright'$ is not strictly larger than in $\triangleright$:

$$s_\triangleright(x) - s_\triangleright(y) \geq s_{\triangleright'}(x) - s_{\triangleright'}(y). \quad (7)$$

**Theorem 2.** *The rule $F_s$, induced by the positional scoring function $s$, is immune to $\ell$-manipulation by a specific (combination of) type(s) if inequality (7) holds for all $\triangleright, \triangleright' \in \mathcal{D}$ and all $x, y \in A$ such that $\triangleright$ and $\triangleright'$ conform to the given type(s) and $x \triangleright y$.*

*Proof.* Aiming for a contradiction, suppose that the condition of the statement holds, but $F_s$ is $\ell$-manipulable by the given type(s). This means that there exist $X, Y \subseteq A$ such that

$$X = F(\blacktriangleright_{-I}, \triangleright'_1, \ldots, \triangleright'_k) \triangleright^\ell F(\blacktriangleright_{-I}, \triangleright, \ldots, \triangleright) = Y$$

for some subset $I = \{1, \ldots, k\} \subseteq N$ of a group $N$ and untruthful preferences $\triangleright'_j$ such that $\triangleright$ and $\triangleright'_j$ conform to the given type(s) for all $j \in \{1, \ldots, k\}$. Then, by Definition 2, there are $x \in X$ and $y \in Y$ such that $x \triangleright y$ and $\{x, y\} \nsubseteq X \cap Y$. We focus on the case where $x \notin Y$, since the case where $y \notin X$ is analogous. Because $y \in Y$, by the definition of the scoring rule we have that

$$\sum_{i \in I} s_\triangleright(x) + \sum_{i \in N \setminus I} s_{\triangleright_i}(x) < \sum_{i \in I} s_\triangleright(y) + \sum_{i \in N \setminus I} s_{\triangleright_i}(y).$$

Then, $x \in X$ implies that

$$\sum_{i \in I} s_{\triangleright'_i}(x) + \sum_{i \in N \setminus I} s_{\triangleright_i}(x) \geq \sum_{i \in I} s_{\triangleright'_i}(y) + \sum_{i \in N \setminus I} s_{\triangleright_i}(y).$$

It follows that

$$\sum_{i \in I} (s_{\triangleright_i}(x) - s_{\triangleright_i}(y)) < \sum_{i \in I} (s_{\triangleright'_i}(x) - s_{\triangleright'_i}(y))$$

which contradicts our hypothesis. □

We continue by providing our necessary conditions.

**Theorem 3.** *If the rule $F_s$, induced by the positional scoring function $s$, is immune to $m$-manipulation by a specific (combination of) type(s), then inequality (7) holds for all $\triangleright, \triangleright' \in \mathcal{D}$ and $x, y \in A$ such that $\triangleright$ and $\triangleright'$ conform to the given type(s) and $x \triangleright y$.*

*Proof.* We prove the contrapositive. Suppose there exist $\triangleright, \triangleright', x$, and $y$ as in the statement that satisfy $s_\triangleright(x) - s_\triangleright(y) < s_{\triangleright'}(x) - s_{\triangleright'}(y)$. This implies that the scoring function is not trivial.

Consider the following profile with four agents:

$$\blacktriangleright' = (\triangleright, \triangleright, \triangleright_{(xy)}, \triangleright'_{(xy)}).$$

From the inequality and because $s$ is positional, $y$ must have a higher score than $x$ in the profile $\blacktriangleright'$. However, if the first two agents change their preference to $\triangleright'$, then $x$ has a higher score than $y$. We can write this profile as $\blacktriangleright'' = (\triangleright', \triangleright', \triangleright_{(xy)}, \triangleright'_{(xy)})$.

Let
$$t = \max_{z \in A} \sum_{i \in N} s_{\triangleright_i}(z) - \sum_{i \in N} s_{\triangleright_i}(y),$$
and
$$t' = \max_{z \in A} \sum_{i \in N} s_{\triangleright'_i}(z) - \sum_{i \in N} s_{\triangleright'_i}(x).$$

We use Lemma 2 to create a profile $\blacktriangleright'''$ in which the scores of $x$ and $y$ are equal and greater than the score of any other alternative by at least the value $\max(t, t')$.

The profile that the manipulation happens *from* is $(\blacktriangleright', \blacktriangleright''')$, for which $y$ is winning; the profile that the manipulation happens *to* is $(\blacktriangleright'', \blacktriangleright''')$, for which $x$ is winning. □

Theorems 2 and 3 imply a characterisation result.

**Corollary 1.** *The rule $F_s$, induced by the positional scoring function $s$, is immune to $m$-manipulation by a specific (combination of) type(s) if and only if inequality (7) holds for all $\triangleright, \triangleright' \in \mathcal{D}$ and all $x, y \in A$ such that $\triangleright$ and $\triangleright'$ conform to the given type and $x \triangleright y$.*

The above characterisation result facilitates the detection of whether a particular positional scoring rule is immune to manipulation, focusing on pairs of preferences instead of full preference profiles.

## 5 MANIPULATION BY COMBINED TYPES

In this section we study whether it is actually possible for a non-trivial positional scoring rule to prevent combined types of manipulation. The answer we obtain is largely negative, but is also contingent to each particular manipulation type.

Our first remark concerns the combined type of addition and flipping, and is promising. The stepwise-scoring rule satisfies the condition of Theorem 2, instantiated for preferences $\rhd$ and $\rhd'$ that conform to flipping: Take two such preferences $\rhd$ and $\rhd'$ and two alternatives $a$ and $b$ such that $a \rhd b$. If $b \rhd' a$, then by definition of stepwise scoring, the difference between the scores of $a$ and $b$ cannot increase—in the best case it will remain the same, in the worst case it will decrease. If $a \rhd' b$, then the relevant difference will remain exactly the same. Analogous reasoning applies for addition, and straightforwardly for the combination of addition and flipping, leading to Proposition 2.

**Proposition 2.** *There exists a non-trivial scoring rule that is immune to $\ell$-manipulation by the combination of addition and flipping.*

However, an impossibility regarding immunity to manipulation by the combination of omission and flipping (Theorem 4 below) is implied by Lemmas 4 and 5. Lemma 3 is needed for the proof of the latter (and follows from Theorem 3 instantiated for omission, together with the fact that all positional scoring functions assign to all alternatives in the totally empty preference the same score).

**Lemma 3.** *If the rule $F_s$, induced by the positional scoring function $s$, is immune to $m$-manipulation by omission, then $s_\rhd(x) \geq s_\rhd(y)$ for all preferences $\rhd$ and alternatives $x, y$ with $x \rhd y$.*

**Lemma 4.** *If the rule $F_s$, induced by the non-trivial positional scoring function $s$, is immune to $m$-manipulation by omission, then for all complete preferences $\rhd$ there exist $x, y \in A$ such that $s_\rhd(x) \neq s_\rhd(y)$.*

*Proof.* Consider a positional scoring function $s$ and a rule $F_s$ immune to $m$-manipulation by omission. In every empty preference, all alternatives get the same score by $s$ (condition $(*)$). Suppose that all alternatives in every complete preference also get the same score by $F_s$ (condition $(**)$). We will show that all alternatives must have the same score in every preference $\rhd$, and thus $s$ must be trivial.

Consider an arbitrary preference $\rhd$ and two alternatives $x, y$ with $x \rhd y$. Theorem 3 instantiated for omission and applied on $(*)$ and $(**)$ implies that $s_\rhd(x) = s_\rhd(y)$. Thus, all connected alternatives will have the same score (condition$(***)$).

But all alternatives $x, y$ that are not connected must also be assigned with the same score: We can create two new preferences by connecting $x$ and $y$ by $(i)$ an arrow from $x$ to $y$ and $(ii)$ an arrow from $y$ to $x$. Theorem 3 instantiated for omission and applied on $(i)$, $(ii)$ and $(***)$ implies that the relevant scores must be the same, and the proof is concluded. □

**Lemma 5.** *If the rule $F_s$, induced by the positional scoring function $s$, is immune to $m$-manipulation by the combination of omission and flipping, then for all complete preferences $\rhd$ and alternatives $x, y$, it holds that $s_\rhd(x) = s_\rhd(y)$.*

*Proof.* Consider an arbitrary complete preference $\rhd$ such that $s_\rhd(x) \geq s_\rhd(y)$ whenever $x \rhd y$ (this must be the case by Lemma 3, if the rule $F_s$ is immune to manipulation by omission).

For some ordering of the alternatives $\{a_1 \ldots, a_m\} = A$, and without drawing the transitive arrows for simplicity, the preference $\rhd$ will be of the following form, where the scores assigned to the alternatives are mentioned below them (for $\gamma \in \mathbb{R}$):

$$\begin{array}{ccccccccc} a_1 & \longrightarrow & a_2 & \longrightarrow & a_3 & \longrightarrow & \cdots & \longrightarrow & a_m \\ \gamma & & \gamma - \delta_1 & & \gamma - \delta_1 - \delta_2 & & \cdots & & \gamma - \sum_{i=1}^{m-1} \delta_i \end{array}$$

We know that $\delta_i \geq 0$, and will show that $\delta_i = 0$, for all $1 \leq i \leq m-1$.

Consider flipping the preference between $a_1$ and $a_2$ as follows:

$$\begin{array}{ccccccccc} a_2 & \longrightarrow & a_1 & \longrightarrow & a_3 & \longrightarrow & \cdots & \longrightarrow & a_m \\ \gamma & & \gamma - \delta_1 & & \gamma - \delta_1 - \delta_2 & & \cdots & & \gamma - \sum_{i=1}^{m-1} \delta_i \end{array}$$

Because the difference in scores between $a_2$ and $a_3$ cannot increase (by Theorem 3 instantiated for flipping), we must have that $\delta_1 = 0$.

Now flip the preference between $a_1$ and $a_3$:

$$\begin{array}{ccccccccc} a_2 & \longrightarrow & a_3 & \longrightarrow & a_1 & \longrightarrow & \cdots & \longrightarrow & a_m \\ \gamma & & \gamma & & \gamma - \delta_2 & & \cdots & & \gamma - \sum_{i=2}^{m-1} \delta_i \end{array}$$

In order for the difference in scores between $a_2$ and $a_1$ to not increase (again by Theorem 3), it must hold that $\delta_2 = 0$.

We repeat this process, "moving" $a_1$ towards the bottom of the preference in steps, and obtaining $\delta_i = 0$ for all $1 \leq i \leq m$. □

**Theorem 4.** *Only the trivial positional scoring rule is immune to $m$-manipulation by the combination of omission and flipping.*

Next, we directly prove an impossibility result regarding immunity to manipulation by the combination of addition and omission.

**Theorem 5.** *Only the trivial positional scoring rule is immune to $m$-manipulation by the combination of addition and omission.*

*Proof.* Consider a rule $F_s$, induced by a positional scoring function $s$, that is immune to $m$-manipulation both by addition and by omission. We will prove that for any preference $\rhd$, $s_\rhd(x) = s_\rhd(y)$ for all $x, y \in A$. If $\rhd$ is empty, we are done. So suppose there is a pair $(a, b) \in \rhd$. For each such pair $(x, y)$ in $\rhd$, define the (singleton) preference $\rhd_{xy} = \{(x, y)\}$. Theorem 3 for omission and for addition implies that for all $(x, y)$ in $\rhd$,

$$s_\rhd(x) - s_\rhd(y) = s_{\rhd_{xy}}(x) - s_{\rhd_{xy}}(y). \tag{8}$$

Since the scoring rules we consider are positional, by Equation (8) it will follow that for all $(x, y), (x', y')$ in $\rhd$,

$$s_\rhd(x) - s_\rhd(y) = s_\rhd(x') - s_\rhd(y'). \tag{9}$$

Consider two alternatives $a$ and $b$ such that $a \rhd b$ and define the preference $\rhd' = \{(a, b), (b, c), (a, c)\}$ for some $c \in A \setminus \{a, b\}$. Again by Theorem 3 for addition and omission, $s_{\rhd_{ab}}(a) - s_{\rhd_{ab}}(b) = s_{\rhd'}(a) - s_{\rhd'}(b)$. By Equation (9) applied to the pairs $(a, b)$, $(b, c)$, and $(a, c)$, it follows that $s_{\rhd'}(a) = s_{\rhd'}(b) = s_{\rhd'}(c)$.

We thus know that $s_\rhd(x) = s_\rhd(y)$ whenever $x$ and $y$ are connected. Suppose now that $x$ and $y$ belong to two different connected components of $\rhd$. Then, the preferences $\rhd' = \rhd \cup \{(x, y)\}$ and $\rhd'' = \rhd \cup \{(y, x)\}$ will be acyclic and thus well-defined. By the same reasoning as above, we know that $s_{\rhd'}(x) = s_{\rhd'}(y)$ and $s_{\rhd''}(x) = s_{\rhd''}(y)$. Since $\rhd \subset \rhd'$ and $\rhd \subset \rhd''$, it follows by Theorem 3 for omission that $s_\rhd(x) - s_\rhd(y) \leq s_{\rhd'}(x) - s_{\rhd'}(y) = 0$ and that $s_\rhd(y) - s_\rhd(x) \leq s_{\rhd''}(y) - s_{\rhd''}(x) = 0$. So, $s_\rhd(x) = s_\rhd(y)$ and we have concluded the proof. □

Finally, note that Theorems 4 and 5 are strong in the sense that their proofs do not require that specific manipulations employ both omission and flipping (or addition and omission) at the same time.

## 6 MANIPULATION BY SINGLE TYPES

After having investigated manipulation by combined types and obtained two important impossibility results in Section 5, we now wonder whether specific types of manipulation (by addition, omission, or flipping) can at least be prevented by positional scoring rules. We bring very good news. For every single manipulation type, there exists some positional scoring rule immune to manipulation.

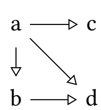
First, recall that the stepwise-scoring rule prevents manipulation by both addition and flipping (Proposition 2). So we immediately know that the following results hold.[10]

**Proposition 3.** *There exists a non-trivial positional scoring rule that is immune to $\ell$-manipulation by addition.*

**Proposition 4.** *There exists a non-trivial positional scoring rule that is immune to $\ell$-manipulation by flipping.*

Note though that the stepwise-scoring rule must be manipulable by omission (otherwise the impossibilities of Section 5 would fail). Here is an example illustrating this:

**Example 3.** In the preference below, omitting the arrow from $a$ to $d$ increases the difference between the stepwise scores of $a$ and $b$.




Next, Proposition 5 holds because the cumulative-scoring rule satisfies the condition of Theorem 2, instantiated for preferences $\rhd$ and $\rhd'$ that conform to omission: If $a$ is preferred to $b$ with respect to a preference $\rhd$, then by removing pairwise preferences from $\rhd$ the cumulative score of $a$ will be reduced at least as much as the cumulative score of $b$.

**Proposition 5.** *There exists a non-trivial positional scoring rule that is immune to $\ell$-manipulation by omission.*

But omission is the only manipulation type to which the cumulative-scoring rule is immune:

**Example 4.** In the left preference below, we can increase the difference between the cumulative scores of $a$ and $b$ by adding an arrow from $a$ to $d$; in the right preference, we can flip the arrow connecting $a$ and $b$ to increase the difference between $b$ and $c$.

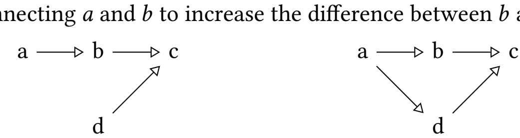



Figure 2 graphically depicts our observations. Note that examples for manipulation under the domination-scoring rule are easy to find, but the existence of rules that are immune to manipulation by flipping and not by addition remains a conjecture.

[10] We also find that the veto-scoring rule satisfies the condition of Theorem 2, instantiated for preferences that conform to addition. Indeed, if $a \rhd b$ for some alternatives $a, b$, then the veto score of $a$ must be 1, the veto score of $b$ must be 0, and the two scores will remain the same under every addition of pairwise preferences to $\rhd$. But veto scoring is manipulable by omission and by flipping.

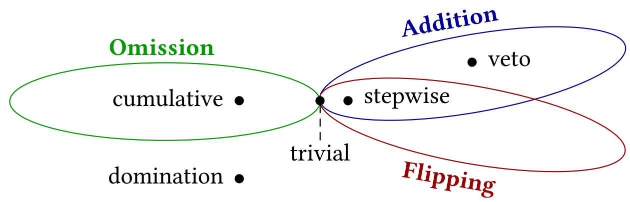



**Figure 2: The space of positional scoring rules, categorised with respect to their immunity to manipulation by the different types of omission, addition, and flipping.**

## 7 CONCLUSION

We have investigated the problem of strategic manipulation in voting for settings where the truthful preferences of the agents may be incomplete. Specifically, we have formalised the three ways in which the agents may attempt to misrepresent their preferences—by adding, omitting, or flipping pairwise comparisons.

In certain situations one may be more worried about one type of manipulation than another. For instance, in crowdsourcing settings, part of the mechanism designer's problem is not only to get the agents to tell the truth, but also to incentivise them to fully complete the task. Lying by omission may deprive the system of important data. On the other hand, in social situations, lying by omission intuitively seems more acceptable than outright misrepresentation. So, some manipulations may be considered tolerable, but this may also lead to an inverse conclusion for a mechanism designer: If concealing a fragment of their truth feels more acceptable to agents in a particular decision-making scenario than creating a brand new lie or completely twisting their preferences, it may be thought as more important to protect against the more likely type of lie.

We have axiomatised the class of positional scoring rules for incomplete preferences and have shown that immunity to manipulation by a single type is always achieved by some such rule. This is a very strong set of results, proven to hold not only for individual manipulators, but also for groups of agents that may manipulate simultaneously. Unfortunately, as far as combinations of types are concerned, manipulation can be prevented only for addition and flipping; all non-trivial positional scoring rules are susceptible to manipulation by omission together with addition or flipping.

It remains an open question whether our impossibilities can be circumvented outside the family of positional scoring rules, and how those rules that are immune to manipulation can be characterised axiomatically. Note also that in this paper we have been concerned with a slightly different notion of an impossibility than the one that is usually met in the literature and refers to the existence of dictators; we have only excluded trivial scoring rules that always output the whole set of alternatives. Clearly, there is room for further research pertaining to our model.

## ACKNOWLEDGMENTS

This research was conducted during a visit of Zoi Terzopoulou at Paris Dauphine university, partly funded by the COST action *European Network for Game Theory*. We also thank Ulle Endriss and the anonymous AAMAS reviewers for their very helpful comments.

# REFERENCES


[1] Salvador Barberà, Walter Bossert, and Prasanta K Pattanaik. 2004. Ranking Sets of Objects. In *Handbook of Utility Theory*. Springer, 893–977.

[2] Dorothea Baumeister, Piotr Faliszewski, Jérôme Lang, and Jörg Rothe. 2012. Campaigns for lazy voters: truncated ballots. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. 577–584.

[3] Jean-Charles de Borda. 1784. Memoire sur les Elections au Scrutin. *Histoire de l'Academie Royale des Sciences* (1784).

[4] Craig Boutilier and Jeffrey S. Rosenschein. 2016. Incomplete Information and Communication in Voting. In *Handbook of Computational Social Choice*. Cambridge University Press.

[5] Felix Brandt. 2011. Group-strategyproof irresolute social choice functions. In *Proceedings of the 22nd International Joint Conference On Artificial Intelligence (IJCAI)*. 79–84.

[6] Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia (Eds.). 2016. *Handbook of Computational Social Choice*. Cambridge University Press.

[7] Stephen Ching and Lin Zhou. 2002. Multi-valued strategy-proof social choice rules. *Social Choice and Welfare* 19, 3 (2002), 569–580.

[8] Carmel Domshlak, Eyke Hüllermeier, Souhila Kaci, and Henri Prade. 2011. Preferences in AI: An overview. *Artificial Intelligence* 175, 7–8 (2011), 1037–1052.

[9] John R Douceur. 2002. The Sybil attack. In *Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS)*. 251–260.

[10] John Duggan and Thomas Schwartz. 2000. Strategic manipulability without resoluteness or shared beliefs: Gibbard-Satterthwaite generalized. *Social Choice and Welfare* 17, 1 (2000), 85–93.

[11] Peter Emerson. 2013. The original Borda count and partial voting. *Social Choice and Welfare* 40, 2 (2013), 353–358.

[12] Ulle Endriss, Maria Silvia Pini, Francesca Rossi, and K. Brent Venable. 2009. Preference Aggregation over Restricted Ballot Languages: Sincerity and Strategy-Proofness. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI)*. 122–127.

[13] Piotr Faliszewski, Piotr Skowron, Arkadii Slinko, and Nimrod Talmon. 2016. Committee Scoring Rules: Axiomatic Classification and Hierarchy.. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*. 250–256.

[14] Peter Gärdenfors. 1976. Manipulation of social choice functions. *Journal of Economic Theory* 13, 2 (1976), 217–228.

[15] Allan Gibbard. 1973. Manipulation of voting schemes: A general result. *Econometrica* 41, 4 (1973), 587–601.

[16] Jerry S Kelly. 1977. Strategy-proofness and social choice functions without singlevaluedness. *Econometrica* 45, 2 (1977), 439–446.

[17] Kathrin Konczak and Jérôme Lang. 2005. Voting procedures with incomplete preferences. In *Proceedings of the IJCAI-05 Multidisciplinary Workshop on Advances in Preference Handling*.

[18] Roger B Myerson. 1995. Axiomatic derivation of scoring rules without the ordering assumption. *Social Choice and Welfare* 12, 1 (1995), 59–74.

[19] MS Pini, F Rossi, KB Venable, and T Walsh. 2007. Incompleteness and incomparability in preference aggregation. In *Proceedings of the 20th International Joint Conference On Artificial Intelligence (IJCAI)*. 1372–1377.

[20] Maria Silvia Pini, Francesca Rossi, Kristen Brent Venable, and Toby Walsh. 2006. Strategic voting when aggregating partially ordered preferences. In *Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems (AAAMAS)*. 685–687.

[21] Maria Silvia Pini, Francesca Rossi, Kristen Brent Venable, and Toby Walsh. 2008. Aggregating partially ordered preferences. *Journal of Logic and Computation* 19, 3 (2008), 475–502.

[22] Maria Silvia Pini, Francesca Rossi, Kristen Brent Venable, and Toby Walsh. 2011. Incompleteness and incomparability in preference aggregation: Complexity results. *Artificial Intelligence* 175, 7-8 (2011), 1272–1289.

[23] Shin Sato. 2008. On strategy-proof social choice correspondences. *Social Choice and Welfare* 31, 2 (2008), 331–343.

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

[25] John H Smith. 1973. Aggregation of preferences with variable electorate. *Econometrica* 41, 6 (1973), 1027–1041.

[26] Zoi Terzopoulou and Ulle Endriss. 2019. Aggregating Incomplete Pairwise Preferences by Weight. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*.

[27] Lirong Xia and Vincent Conitzer. 2011. Determining possible and necessary winners given partial orders. *Journal of Artificial Intelligence Research* 41 (2011), 25–67.

[28] Makoto Yokoo, Yuko Sakurai, and Shigeo Matsubara. 2001. Robust Combinatorial Auction Protocol Against False-name Bids. *Artificial Intelligence* 130, 2 (2001), 167–181.

[29] H Peyton Young. 1975. Social Choice Scoring Functions. *Journal on Applied Mathematics* 28, 4 (1975), 824–838.