Sincerity and Manipulation under Approval Voting*

Ulle Endriss
Institute for Logic, Language and Computation
University of Amsterdam

(To appear in *Theory and Decision*)

Abstract

Under approval voting, each voter can nominate as many candidates as she wishes and the election winners are those candidates that are nominated most often. A voter is said to have voted sincerely if she prefers all those candidates she nominated to all other candidates. As there can be a set of winning candidates rather than just a single winner, a voter's incentives to vote sincerely will depend on what assumptions we are willing to make regarding the principles by which voters extend their preferences over individual candidates to preferences over sets of candidates. We formulate two such principles, *replacement* and *deletion*, and we show that, under approval voting, a voter who accepts those two principles and who knows how the other voters will vote will never have an incentive to vote insincerely. We then discuss the consequences of this result for a number of standard principles of preference extension in view of sincere voting under approval voting.

Keywords: Approval voting, Manipulation, Ranking sets of objects

1 Introduction

Approval voting (AV) is a voting procedure that allows each voter to *approve* of as many candidates as she desires and that declares the candidate(s) collecting the most approvals the winner(s) of the election (Weber, 1978; Brams and Fishburn, 1978, 2007; Laslier and Sanver, 2010). AV has been strongly advocated by some, and severely criticised by others. While there may be no clear-cut conclusion to this debate, it is certainly true that AV has found its place as one of the major election systems studied in economic theory and political science and it is also widely used in practice. Brams (2008), for instance, cites the use of AV to elect the secretary-general of the United Nations and its adoption by numerous professional societies, including the American Mathematical Society (AMS), the Institute for Operations Research and Management Sciences (INFORMS), and the Society for Social Choice and Welfare.

As discussed, for instance, by Sanver (2010), despite its prominent position amongst the major voting procedures, AV does in fact not directly fit into the standard model of voting used in social

*Parts of this work have been presented at the 11th Conference on Theoretical Aspects of Rationality and Knowledge (TARK) in Brussels in June 2007, the Social Choice Colloquium at the University of Tilburg in September 2007, the Dagstuhl Seminar on Computational Issues in Social Choice in October 2007, and the 7th International Meeting on Logic, Game Theory and Social Choice (LGS) in Bucharest in July 2011, as well as local seminars at the Universities of Amsterdam, Padova, and Paris-Dauphine. The insightful feedback received from the participants of these meetings is gratefully acknowledged. Thanks are also due to Steve Brams and an anonymous reviewer for comments on an earlier version of this paper, and to Remzi Sanver and Bill Zwicker for illuminating discussions on preference extensions.

choice theory. In this model, each voter is assumed to have a ranking of the candidates representing her preferences; and voters vote by reporting such a ranking (truthfully or otherwise) to the election chair. AV does not fit this model, because in AV a ballot is a *subset* of the set of candidates rather than a *ranking* of those candidates. It is easy to see that it is impossible to encode the information required to recover a ranking from an AV ballot: there are more ways to rank n candidates than there are to pick a subset from a set of n candidates (for $n > 3$). This means that the question of whether a voter will vote *truthfully*, central to much of voting theory, is not a meaningful question for AV: if we cannot fully express our preferences, we can certainly not be expected to do so truthfully. This problem is well known (see, e.g., Niemi, 1984). Therefore, instead of truthfulness we will be interested in a less demanding notion, namely sincerity. In AV, a ballot is called *sincere* if the voter in question prefers each of the approved candidates to each of the disapproved candidates (Brams and Fishburn, 2007). That is, any given preference order will give rise to multiple sincere ballots to choose from.

Suppose a voter has learned how everybody else will vote in a forthcoming AV election. The question we seek to answer in this paper is this: *Are there reasonable assumptions under which the set of best responses of such a voter will always include a sincere ballot?*

When there is more than one candidate with a maximal number of approvals, then AV produces a set of tied winners (from which we may have to select a single candidate using a suitable tie-breaking rule). Therefore, a voter considering to manipulate an election will have to reason over alternative sets of winning candidates. The assumptions we shall work with concern the principles by which a voter will extend her preferences over individual candidates to a preference order over (nonempty) sets of candidates when considering what ballot to submit. This is a well-studied problem in social choice theory, often referred to as *ranking sets of objects* (see, e.g., Barberà et al., 2004; Gärdenfors, 1976; Kelly, 1977; Kannai and Peleg, 1984; Nitzan and Pattanaik, 1984; Puppe, 1995; Can et al., 2009; Erdamar and Sanver, 2009; Geist and Endriss, 2011) and we shall be examining incentives to vote sincerely under AV for a range of different such principles proposed in the literature.

The remainder of this paper is organised as follows. In Section 2 we introduce our model, which involves defining the approval voting procedure and the notion of sincerity, and we state the question we shall address in this paper in more detail. A central aspect of the model concerns the assumptions made on how voter preferences over individual candidates extend to preferences over sets of candidates. Section 3 reviews a range of axioms proposed in the literature for modelling this problem. Section 4 then presents our results on the existence of sincere best responses under AV for different sets of assumptions on how preferences are extended to set preferences. Section 5 reviews related work, and in Section 6 we summarise our results and conclude.

2 The Model

In this section we define our model and state the main problem we shall address. (An important detail of the model, namely how voters extend their preferences over individual candidates to sets of candidates, will largely be relegated to Section 3.)

2.1 Preferences

Recall that a *preorder* is a binary relation that is reflexive and transitive; a *weak order* is a preorder that is complete; and a *total order* is a weak order that is antisymmetric (Roberts, 1979).

Throughout this paper, let $\mathcal{X}$ be a finite set of two or more *candidates* and let $n = |\mathcal{X}|$ denote the number of candidates. Let $\underline{\mathcal{X}} = 2^{\mathcal{X}} \setminus \{\emptyset\}$ denote the set of *nonempty sets of candidates*.

We consider a finite set of *voters*. Each voter has *preferences over individual candidates*, modelled as a *total order* $\dot{\geqslant}$ on $\mathcal{X}$. That is, we write $a \mathrel{\dot{\geqslant}} b$ to express that the voter in question likes candidate a at least as much as candidate b. We write $a \mathrel{\dot{>}} b$ (strict preference) if $a \mathrel{\dot{\geqslant}} b$ but not $b \mathrel{\dot{\geqslant}} a$.

Each voter furthermore has *preferences over election outcomes* (before tie-breaking), i.e., over sets of candidates that the tie-breaking rule may choose from. This is modelled as a *weak order* $\succcurlyeq$ on $\underline{\mathcal{X}}$. We write $A \succ B$ if $A \succcurlyeq B$ but not $B \succcurlyeq A$ (strict preference), and $A \sim B$ if both $A \succcurlyeq B$ and $B \succcurlyeq A$ (indifference). We shall sometimes refer to $\succcurlyeq$ simply as a voter's preference order over sets of alternatives, but it is important to note that $\succcurlyeq$ only plays a role in the special situation where a voter is choosing between different sets of candidates from which one winner should be selected by the tie-breaking rule. That is, $\succcurlyeq$ depends both on the voter's preference relation $\dot{\geqslant}$ and on her beliefs regarding the nature of the tie-breaking rule in use.[1] While a voter's $\succcurlyeq$ is in part dependent on her $\dot{\geqslant}$, we assume that *we do not know* how to derive $\succcurlyeq$ from $\dot{\geqslant}$ (besides also not knowing $\dot{\geqslant}$ itself). For example, even if we were to learn that $a \mathrel{\dot{>}} b \mathrel{\dot{>}} c$, then this usually does not provide us with any hints regarding the relative ranking of $\{a, c\}$ versus $\{b\}$ in terms of $\succcurlyeq$. This issue will be discussed in detail in Section 3, where we review several principles that one might reasonably adopt for extending a total order $\dot{\geqslant}$ on $\mathcal{X}$ to a weak order $\succcurlyeq$ on $\underline{\mathcal{X}}$.

Later, we will require some notation to conveniently refer to the best and the worst candidates from a given set. For any set $A \in \underline{\mathcal{X}}$, we write $\max(A)$ for the (unique) maximal element in A, i.e., for that $a \in A$ for which $a \mathrel{\dot{\geqslant}} b$ for all $b \in A$ (with $\dot{\geqslant}$ always clear from the context). Similarly, we write $\min(A)$ for the (unique) minimal element in A.

Beyond her preference order $\dot{\geqslant}$, a voter might also have a utility (or valuation) function $u : \mathcal{X} \to \mathbb{R}$ mapping individual candidates to numerical values. Such a utility function induces a preference order $\dot{\geqslant}$, namely: $a \mathrel{\dot{\geqslant}} b$ if and only if $u(a) \geqslant u(b)$. But u on its own does *not* induce a corresponding weak order $\succcurlyeq$ on sets. To obtain $\succcurlyeq$ we also need to make assumptions regarding the tie-breaking rule and the voter's attitude towards risk. For instance, if we assume that the tie-breaking rule will select any of the candidates in a given set with equal probability, then we can compute the *expected* utility corresponding to a set. A risk-neutral voter would then rank sets of candidates according to their expected utility. If we make the assumption that voters will vote so as to maximise their expected utility (in view of a known tie-breaking rule), then we speak of *expected-utility maximisers*. This is a common assumption made in the literature (see, e.g., Myerson and Weber, 1993; Can et al., 2009). We stress that for much of this paper we do *not* assume that voters are endowed with utility functions and that they are expected-utility maximisers (the only result based on this assumption is Theorem 9).

2.2 Approval Voting

The system of *approval voting* (AV) consists of a method for balloting voters and a method for tallying the ballots received to determine the winner(s) of the election (Brams and Fishburn, 2007; Merrill and Nagel, 1987). A *ballot* $B \subseteq \mathcal{X}$ is a subset of the set of candidates. Each voter has to submit one such ballot. A candidate is a *winner* if he is included in at least as many ballots as any other candidate (that is, there can be more than one winner). The *outcome* of an election is the set of winners.

Submitting a ballot approving of all or no candidates amounts to *abstaining*. We do allow voters to abstain. A *proper* ballot is one that is not an abstention ballot.

In examples, we will use a *simplified notation* for writing ballots and election outcomes, from the

[1] In particular, we cannot use $\succcurlyeq$ to determine a voter's "most preferred set of candidates", i.e., our earlier observation that the concept of *truthful ballot* is not well-defined in approval voting is not affected by the introduction of $\succcurlyeq$.

perspective of a particular voter with a particular preference order $\dot{\geqslant}$. (Recall that both ballots and outcomes are subsets of $\mathcal{X}$.) The notation is best explained by means of a concrete example. Suppose there are 5 candidates. To denote the ballot that approves of the two top candidates and the bottom candidate according to our voter's preferences, we write `[11001]`. To denote outcomes, we use curly brackets rather than square brackets. For instance, {`01000`} is the election outcome in which our voter's second favourite candidate is the sole winner.

2.3 Sincere Best Responses

Given a voter's preference order $\dot{\geqslant}$, a ballot $B \subseteq \mathcal{X}$ is called *sincere* if the voter really does prefer each approved candidate to each disapproved candidate; that is, if $a \mathrel{\dot{\geqslant}} b$ for all $a \in B$ and all $b \in \mathcal{X} \backslash B$. This is the standard notion of sincerity in AV as defined by Brams and Fishburn (2007, p. 29). Observe that according to this definition, abstention ballots are considered sincere.[2]

Suppose we have fixed a principle of preference extension. Suppose furthermore a particular voter knows how all other voters are going to vote and is considering which ballot to submit. We are interested in her incentives to vote sincerely in this kind of situation. She has an incentive to vote insincerely, if there exists an outcome she can achieve by means of an insincere ballot that (strictly) dominates (according to her $\succcurlyeq$) all outcomes she can achieve by means of voting sincerely. *Vice versa*, if for every ballot (including every insincere ballot) we can find a (possibly different) sincere ballot that yields at least as preferred an outcome as the former, then our voter has no incentive to vote insincerely. Our goal in this paper is to identify assumptions under which this is the case. That is, we want to identify principles of preference extension under which a voter always has a *sincere best response* to any set of ballots reported by the other voters.

There are several statements in the literature that suggest that voters will always have such a sincere best response. Niemi (1984), for instance, states that "[u]nder AV, voters are never urged to vote insincerely" and de Sinopoli et al. (2006) make this claim more precise by stating that "for every pure strategy of the other players, the set of best replies contains a sincere strategy". On the other hand, Brams and Fishburn (1978) argue that "no voting system [including AV] is sincere", which in particular excludes the guaranteed existence of a sincere best response. The source of this discrepancy is the fact that different authors make different (and often implicit) assumptions about set preferences. Here we will discuss several of these principles and show that the existence of a sincere best response can be guaranteed under some of the stronger principles, while it fails under some of the weaker ones.

3 Principles of Preference Extension

A voter's a total preference order $\dot{\geqslant}$, declared over the set of candidates $\mathcal{X}$, is not sufficient to fully describe her preferences over election outcomes, which are nonempty sets of candidates. We therefore need to choose a suitable set of assumptions that govern how a voter's preference order $\dot{\geqslant}$ over individual candidates are to be lifted to a preference relation $\succcurlyeq$ over sets of candidates. This is a problem that has been studied extensively in social choice theory. Barberà et al. (2004) give an excellent overview of the most important principles proposed in the literature.[3]

[2]In fact, as Lemma 3 will show, whether or not we allow for abstention ballots and whether or not we consider them sincere ballots will not affect any of our results.

[3]Barberà et al. (2004) distinguish three types of scenarios: (1) the elements of the set are interpreted as as possible outcomes after some randomised choice; (2) the elements of the set represent opportunities and the decision maker in question can choose one of them; or (3) the sets themselves are the final outcomes. The kind of principles that are

3.1 Kelly Principle: No Information

A fundamental property that any reasonable choice of set preferences will satisfy is the *extension axiom*, which expresses that when a voter compares two singleton sets, her set preferences will directly correspond to her preferences over individuals.[4]

(EXT) $\{a\} \succ \{b\}$ if $a \dot{>} b$

Formally, axioms such as this are properties of triples $(\mathcal{X}, \dot{\geqslant}, \succcurlyeq)$ of a set of candidates $\mathcal{X}$, a total order $\dot{\geqslant}$ over those candidates, and a weak order $\succcurlyeq$ over nonempty sets of candidates. We say that $(\mathcal{X}, \dot{\geqslant}, \succcurlyeq)$ *satisfies* an axiom such as **(EXT)** if the conditions specified in the axiom hold for that set of candidates and that choice of relations. To simplify presentation, we will also often say that a weak order $\succcurlyeq$ *satisfies* an axiom, taking $\mathcal{X}$ and $\dot{\geqslant}$ as fixed.

Besides **(EXT)**, there are two further assumptions that we can safely make without any knowledge of the tie-breaking rule or the voter's attitude towards risk:

(MAX) $\{\max(A)\} \succcurlyeq A$
(MIN) $A \succcurlyeq \{\min(A)\}$

The former expresses that the voter never prefers a set over (the singleton consisting just of) the best element in that set; the latter says that she weakly prefers a set over the worst element in that set. By accepting all of **(EXT)**, **(MAX)**, and **(MIN)** together, we can express that A should be weakly preferred to B if all elements of A are weakly preferred to all elements of B and that A should be strictly preferred to B if all elements of A are strictly preferred to all elements of B. In reference to the work of Kelly (1977), Can et al. (2009) call this the *Kelly Principle*—a choice of terminology that we shall adopt here as well. We shall often write **(KEL)** as a shorthand for the conjunction of **(EXT)**, **(MAX)**, and **(MIN)**. All other approaches for lifting preferences discussed in the sequel will be refinements of the Kelly Principle.

3.2 Gärdenfors Principle: Tie-Breaking by a Rational Chair

The *Gärdenfors Principle* states that you should prefer set A over set B if you can obtain B from A by means of a sequence of operations that involve either removing the most preferred element of the set or adding a new element that is less preferred than those already in the set (Gärdenfors, 1976). Our statement of the Gärdenfors axioms follows Barberà et al. (2004).

(GF1) $A \cup \{b\} \succ A$ if $b \dot{>} a$ for all $a \in A$
(GF2) $A \succ A \cup \{b\}$ if $a \dot{>} b$ for all $a \in A$

We shall write **(GAR)** for the conjunction of **(GF1)** and **(GF2)**. The Gärdenfors Principle is strictly stronger (more restrictive) than the Kelly Principle. We state this well-known fact as a lemma. In the statement of that lemma, the proof of which is straightforward, and throughout the remainder of this paper, we say that a set of axioms Γ *entails* another set of axioms Δ, if for any triple $(\mathcal{X}, \dot{\geqslant}, \succcurlyeq)$ of a set of candidates $\mathcal{X}$, a total order $\dot{\geqslant}$ on $\mathcal{X}$, and a weak order $\succcurlyeq$ on the corresponding $\underline{\mathcal{X}}$, it is the case that, if $(\mathcal{X}, \dot{\geqslant}, \succcurlyeq)$ satisfies all of the axioms in Γ, then it will also satisfy all of the axioms in Δ.

relevant for us mostly fall under type (1); although a case can be made that some of of the principles belonging to type (2) are also of interest (for instance, if a voter is overly *optimistic* and will act *as if* she were able to choose her favourite winner from amongst a set of front-runners).

[4] Here and in the sequel, variables $a, b, \ldots$ range over $\mathcal{X}$, while variables $A, B, \ldots$ range over $\underline{\mathcal{X}}$. All axioms are understood to be universally closed, i.e., they are stated for *all* choices of a, b, A, B.

Lemma 1. (GAR) *entails* **(KEL)**, *but not vice versa.*

An attractive interpretation of the Gärdenfors Principle is the following. Suppose the election chair will be charged with breaking ties and our voter believes that he will do so *rationally* in the sense that there exists a total order over $\mathcal{X}$ (representing the chair's preferences) and the chair will select from any set A the unique element that is maximal with respect to that total order. The voter has certain beliefs as to which total order the chair will use, but we, the mechanism designers, do not know what these beliefs are. If we furthermore assume that the voter is an expected-utility maximiser, then this induces a particular preorder $\succcurlyeq$. Erdamar and Sanver (2009, Theorem 3.4) show that this $\succcurlyeq$ is the same $\succcurlyeq$ as the one generated by the Gärdenfors Principle. This, arguably, is a fairly realistic scenario, which further underlines the importance of the Gärdenfors Principle.

3.3 Kannai-Peleg Principle: Independence and Max-Min Preferences

Another important principle is given by the *independence axiom* of Kannai and Peleg (1984):

(IND) $\quad A \cup \{c\} \succcurlyeq B \cup \{c\} \quad$ if $A \succ B$ and $c \notin A \cup B$

This axiom expresses the intuitively appealing idea that adding the same additional candidate to each of two sets should not reverse a voter's preferences over those two sets. Kannai and Peleg prove a lemma that shows that if **(IND)** holds on top of the Gärdenfors Principle, then any set A is equally preferred as the set $\{\max(A), \min(A)\}$ consisting only of the best and the worst element in A. We model this latter property by the following axiom.

(MMX) $\quad A \sim \{\max(A), \min(A)\}$

If **(MMX)** is satisfied, then any judgements regarding set preferences can be made by only comparing the maxima and minima of sets.

The main theorem of Kannai and Peleg (1984) shows that for $|\mathcal{X}| \geqslant 6$ there can be no weak order that is consistent with both **(GAR)** and **(IND)**. This suggests that, despite its intuitive appeal, the independence axiom is too restrictive to model set preferences. **(MMX)** is weaker and Kannai and Peleg's impossibility does not persist if we use that axiom instead. We shall refer to the conjunction of **(GAR)** and **(MMX)** as the *Kannai-Peleg Principle*. Several preference extensions proposed in the literature are refinements of (what we call) the Kannai-Peleg Principle (Barberà et al., 2004).

3.4 Nitzan-Pattanaik Principle: Median-Based Preferences

Nitzan and Pattanaik (1984) suggest to rank alternative sets on the basis of their medians. The median of an odd-numbered set of candidates A is the candidate $a \in A$ such that the number of elements in A that are preferred to a is equal to the number of elements in A that are considered inferior to a. For even-numbered sets, there are two median elements. Formally, we define the median of a set $A \in \underline{\mathcal{X}}$ as the subset $\text{med}(A) = \{a \in A \mid \#\{b \in A \mid b \dot{\geqslant} a\} - \#\{b \in A\} \mid a \dot{\geqslant} b\} \in \{0, 1, -1\}\}$, which always has either one or two elements. We use this definition to formulate the following axiom:

(MED) $\quad A \sim \text{med}(A)$

We call the conjunction of **(GAR)** and **(MED)** the *Nitzan-Pattanaik Principle*. That is, under this principle, sets are compared by comparing their medians using the Gärdenfors Principle.

Like the Kannai-Peleg Principle, the Nitzan-Pattanaik Principle is appealing from a cognitive point of view. Indeed, it is very natural to assume that a voter's preferences over sets will depend on her preferences over a small number of distinguished elements within those sets, rather than the full sets. In the case of the Kannai-Peleg Principle, these distinguished elements are the best and the worst element in a set, while for the Nitzan-Pattanaik it is the middle-most element (or possibly the two middle-most elements).

3.5 Uniform Tie-Breaking with Expected-Utility Maximisers

Arguably the most natural choice for a tie-breaking mechanism is the *uniform tie-breaking rule*, where each of the front-runners is selected as the election winner with equal probability. Assuming that the uniform tie-breaking rule is used together with the assumption that voters are expected-utility maximisers yields a very attractive framework in which to analyse voters' preferences over sets of candidates. While these assumptions have been axiomatised in the literature (see, e.g., Fishburn, 1972; Can et al., 2009), for our purposes it is more convenient to give a direct definition.

Take a set of candidates $\mathcal{X}$ and a voter whose preferences are represented by a total order $\dot{\geqslant}$ on $\mathcal{X}$ and a weak order $\succcurlyeq$ on $\underline{\mathcal{X}}$. Then we say that this voter satisfies the assumption of being an expected-utility maximiser under uniform tie-breaking, if there exists a utility function $u : \mathcal{X} \to \mathbb{R}$ such that (*i*) u represents $\dot{\geqslant}$, i.e., for all $a, b \in \mathcal{X}$ we have $u(a) \geqslant u(b)$ if and only if $a \mathrel{\dot{\geqslant}} b$; and (*ii*) for all $A, B \in \underline{\mathcal{X}}$ we have $A \succcurlyeq B$ whenever the following condition holds:

$$\frac{1}{|A|} \cdot \sum_{a \in A} u(a) \quad \geqslant \quad \frac{1}{|B|} \cdot \sum_{b \in B} u(b)$$

3.6 Optimism and Pessimism

Adopting the terminology of Taylor (2005), we say that a voter is *optimistic* if she prefers set A over set B whenever she prefers the best element in A over the best element in B. We use the following axiom to express optimism:

(OPT) $\quad A \sim \{\max(A)\}$

We say that a voter is an optimist if she conforms to **(OPT)** on top of the Kelly Principle. A voter who believes that the person charged with breaking ties shares her own preferences is one example for someone who will have optimistic set preferences. Similarly, if the uniform tie-breaking rule is used and a voter is an expected-utility maximiser with a very skewed utility function, giving *much* higher utility to a than to b whenever $a \mathrel{\dot{>}} b$, then that voter will have set preferences that satisfy **(OPT)**.

A voter may also be *pessimistic* and judge sets of candidates in terms of the worst candidates in those sets. A pessimist is a voter that conforms to **(KEL)** and the following axiom:

(PES) $\quad A \sim \{\min(A)\}$

4 Existence of Sincere Best Responses

In this section we prove a number of theorems that show that, under certain assumptions on the principles for preference extension that voters conform to, we are able to guarantee that each voter will have a sincere best response to any combination of ballots chosen by the other voters. That is, if a would-be manipulator obtains information on how the others will vote, she will never have an incentive to exploit this knowledge by casting an insincere ballot.

4.1 An Example for Insincere Manipulation

Let us first give an example in which a voter will not have a sincere best response and where therefore manipulation by means of an insincere ballot *is* possible. Suppose in an election with four candidates, one voter knows how everybody else is going to vote and is considering to manipulate. Counting the ballots of the other voters, suppose our would-be manipulator's first choice has received 9 approvals, her second choice 10 approvals, her third choice 9 approvals, and her least favourite candidate 10 approvals (note that the number of voters is irrelevant). Using her own vote, she can now force a number of different election outcomes (before tie-breaking):

- Get outcome {`1101`} by voting with ballot `[1000]`.
- Get outcome {`0100`} by voting with ballot `[1100]`, `[1110]`, `[0100]`, or `[0110]`.
- Get outcome {`1111`} by voting with ballot `[1010]`.
- Get outcome {`0101`} by voting with ballot `[0101]`, `[0111]`, or `[1101]`.
- Get outcome {`0001`} by voting with ballot `[0001]`, `[0011]`, `[1001]`, or `[1011]`.
- Get outcome {`0111`} by voting with ballot `[0010]`.

All other outcomes are unattainable to her. Now, suppose our would-be manipulator conforms to the Gärdenfors Principle. Then some of the attainable outcomes will clearly not be attractive to her. For instance, {`0111`} is dominated by {`0100`}. In other words, she has certainly no incentive to vote by means of the (insincere) ballot `[0010]`. Outcomes {`0101`} and {`0001`} are also dominated.

However, the Gärdenfors Principle is not sufficiently strong to tell us which of {`1101`}, {`0100`}, and {`1111`} is the most preferred for our would-be manipulator. If it is {`1101`}, then we are in safe waters, because the ballot to get that outcome is sincere. If it is {`0100`}, then there is no problem either, because she can achieve this outcome by voting sincerely, e.g., using `[1100]`. But if {`1111`} happens to be the strictly most preferred outcome of the three, then our manipulator does have an incentive to vote insincerely. Hence, assuming only the Gärdenfors Principle, insincere manipulation *is* possible under AV. That is, we cannot guarantee the existence of a sincere best response.

4.2 Pivotal, Subpivotal and Insignificant Candidates

Suppose our would-be manipulator knows how all other voters are going to vote. Given those votes, we count the points obtained by each candidate. One or more of them will have received the highest number of approvals; we call them the *pivotal* candidates. Some will have received exactly one point less than the pivotal candidates; we call them the *subpivotal* candidates. We call the remaining candidates *insignificant*. Insignificant candidates have no chance of being elected, whatever ballot our manipulator chooses to submit.

We want to show that, under certain assumptions on her set preferences, for any (possibly insincere) ballot available to our would-be manipulator there exists a sincere ballot that yields at least as good an outcome. She has the following options:

(1) Get a nonempty (and not necessarily proper) subset of the pivotal candidates elected, by approving that subset together with any number of subpivotal or insignificant candidates.

(2) Get all of the pivotal candidates and a (possibly empty and not necessarily proper) subset of the subpivotal candidates elected, by approving that subset together with any number of insignificant candidates.

We now show that any ballot of the first type will be weakly dominated by a sincere ballot, even under the very weakest of assumptions (namely, the Kelly Principle). Hence, for our results later on we will only have to deal with (possibly insincere) ballots of the second type.

Lemma 2. *Suppose a voter whose preferences satisfy* **(KEL)** *learns about the ballots of the other voters. Then for any ballot that involves approving of a pivotal candidate there exists a sincere ballot that is at least as good a response.*

Proof. If our voter votes by approving of at least one pivotal candidate, then the outcome will be a subset of the set of pivotal candidates (scenario 1 above). Under the Kelly Principle, the outcome where her most preferred pivotal candidate is the sole winner weakly dominates all other such outcomes. So we only need to show that getting only the most preferred pivotal candidate elected is possible by means of a sincere ballot. But this is clearly the case: approving of that candidate and any candidates preferred to him (all of which will be subpivotal or insignificant) will achieve this. □

Observe that in case there is only one pivotal candidate and that candidate happens to be the least preferred candidate of our voter, then the strategy suggested in the proof above is to nominate *all* candidates, i.e., to abstain. While we do allow voters to abstain and while we do label abstention ballots as being sincere, one could object to this and argue that it is more interesting to prove the existence of sincere best responses that are not abstentions. Our next result shows that this is not a critical issue: if we accept (at least) the Kelly Principle, then our would-be manipulator can always do at least as well by voting by means of some *proper* sincere ballot as she can do by abstaining.

Lemma 3. *Suppose a voter whose preferences satisfy* **(KEL)** *learns about the ballots of the other voters. Then there exists a sincere ballot that does not amount to abstaining that is at least as good a response as abstaining.*

Proof. If our voter abstains, then the set of pivotal candidates will become the winners. If she instead votes by means of the sincere ballot that approves her most preferred pivotal candidate and all other candidates preferred to that candidate (if any), then only that most preferred pivotal candidate will win, which, by **(MAX)**, weakly dominates the abstention outcome. This sincere ballot is *proper* unless there is only a single pivotal candidate and that candidate happens to be the least preferred candidate. In that case, our voter can instead vote by approving only of her most preferred candidate. If that candidate is insignificant, then only the pivotal candidate will get elected (same outcome as under abstention); if that candidate is subpivotal, then it will get elected together with the pivotal candidate, which, by **(MIN)**, is at least as good as only getting the pivotal candidate. □

4.3 Optimism and Pessimism

The most restrictive assumptions on set preferences that we have discussed in Section 3 are optimism and pessimism. As we shall see next, if a voter is either optimistic or pessimistic, then she will always have a sincere best response. In fact, these two results will turn out to be corollaries of our main theorem (Theorem 6). However, as it is possible to give direct proofs that are much simpler than the proof of the general result, we include these direct proofs here so as to better illustrate our approach.

Theorem 4. *Under AV, upon learning the ballots of the other voters, a voter whose preferences satisfy* **(KEL)** *and* **(OPT)** *will always have a best response that is sincere.*

Proof. Lemma 2 applies. Thus, what remains to be shown is that for any ballot X that involves approving only of subpivotal and insignificant candidates (meaning that all pivotal and the approved subpivotal candidates win) there exists a sincere ballot Y that is at least as good a response. We distinguish two cases:

- First, suppose the most preferred winner for ballot X is a pivotal candidate (call that candidate c). Then the sincere strategy of only approving of c and all candidates preferred to c (if any) will result in the outcome $\{c\}$, which by **(MAX)** weakly dominates the outcome for X.
- Second, if the most preferred winner for ballot X is a subpivotal candidate, then the sincere strategy of approving of the most preferred subpivotal candidate c and all insignificant candidates preferred to c (if any) will result in c and all pivotal candidates winning, which under **(OPT)** weakly dominates the outcome for X.

Hence, under no circumstances will our voter have an incentive to not submit a sincere ballot. □

Theorem 5. *Under AV, upon learning the ballots of the other voters, a voter whose preferences satisfy* **(KEL)** *and* **(PES)** *will always have a best response that is sincere.*

Proof. For a pessimistic voter we can even give a *single* sincere ballot that will weakly dominate all strategies for a given situation. This strategy is to approve of the most preferred pivotal candidate c and all candidates preferred to c (if any). The outcome will be $\{c\}$, which for a pessimist dominates all other outcomes including at least one pivotal candidate. As only outcomes including pivotal candidates are feasible, this completes the proof. □

4.4 Main Theorem and Consequences

We now prepare the ground to prove our main theorem. We shall require two further axioms. The first of these is what we call the (weak) *replacement axiom.* The idea can be traced back to Sen (1991), who suggests an axiom whereby A should be weakly preferred to B, if the two sets have the same cardinality and there exists a surjective mapping $f : A \to B$ such that $a \dot{\geqslant} f(a)$ for all $a \in A$. Puppe (1995) points out that this is equivalent to a simpler axiom, postulating $(A \setminus \{a\}) \cup \{b\} \succcurlyeq A$ if $b \dot{>} a$ for $a \in A$ and $b \notin A$.[5] We shall use a weaker variant of Puppe's replacement axiom:[6]

(WRP) $\quad (A \setminus \{a\}) \cup \{b\} \succcurlyeq A$ or $A \cup \{b\} \succcurlyeq A$ or $A \setminus \{a\} \succcurlyeq A \quad$ if $b \dot{>} a$

The axiom expresses that when b is strictly preferred to a, then, in any given context, it is either beneficial to exchange a for b, to just obtain b, or to just give away a.[7]

Our second axiom, the *deletion axiom*, specifies a sufficient condition for being able to delete an element from a set of candidates without diminishing its attractiveness.

(DEL) $\quad A \setminus \{a\} \succcurlyeq A \quad$ if $A \succ \{a\}$

This is an intuitively appealing axiom, that, to the best of our knowledge, has not been considered in the literature before.[8] It expresses that the attractiveness of a set A will not be diminished if we delete

[5]The main axiom system considered in our own earlier work on sincerity in approval elections with small numbers of candidates is equivalent to the Gärdenfors Principle together with this replacement axiom (Endriss, 2007).

[6]The reasons why the original replacement axiom is not sufficient for our purposes will become clear in the proof of Theorem 8. Note that the weaker an axiom, the less contestable it is and the stronger any theorem that establishes the existence of sincere best responses for agents subject to that axiom.

[7]Note that the third disjunct on the lefthand side of **(WRP)** should be read as "$A \neq \{a\}$ and $A \setminus \{a\} \succcurlyeq A$", to avoid stating a preference of the empty set over A ($\succcurlyeq$ is only defined for pairs of nonempty sets).

[8]Note that, equivalently though less succinctly, we could have phrased the precondition for **(DEL)** as "if $A \succ \{a\}$ and $A\setminus\{a\} \neq \emptyset$". The reason is that the intended interpretation of $\succcurlyeq$ is as a relation on nonempty sets; so any statements about how the empty set relates to some other set are irrelevant.

one of its less preferable elements, where an element is judged to be "less preferable" if obtaining it on its own is strictly less preferable to obtaining A.[9]

Finally, we require a means of comparing the *degree of insincerity* of two ballots. Given any two ballots $B, B' \subseteq \mathcal{X}$, we write $B \gg B'$ and say that ballot B is strictly more insincere than ballot B' if and only if either B is insincere while B' is sincere, or both B and B' are insincere and one of the following two conditions is satisfied:

(*i*) $\max(\mathcal{X} \setminus B) \dot{>} \max(\mathcal{X} \setminus B')$ or
(*ii*) $\max(\mathcal{X} \setminus B) = \max(\mathcal{X} \setminus B')$ and $\#\{b \in B \mid \max(\mathcal{X} \setminus B) \dot{>} b\} > \#\{b \in B' \mid \max(\mathcal{X} \setminus B') \dot{>} b\}$.

Condition (*i*) says that a ballot becomes more sincere as the top disapproved candidate (that must dominate at least one approved candidate, and therefore may be considered an indicator of insincerity) becomes a less preferred candidate. Condition (*ii*) says that a ballot becomes more sincere as the number of approved candidates dominated by the top disapproved candidate reduces (while the top disapproved candidate himself remains the same). For example, we have `[1010]` $\gg$ `[1101]` by condition (*i*), and `[011]` $\gg$ `[010]` by condition (*ii*). Two ballots that cannot be compared via $\gg$ are considered equally insincere. Observe that the relation $\gg$ induces a ranking on ballots. In other words, $\gg$ is the strict part of a complete and transitive order on $2^{\mathcal{X}}$. The sincere ballots are exactly those that are minimal with respect to $\gg$. We stress that our reasons for introducing $\gg$ are purely technical: to prove our main theorem we shall require a total order on ballots that allows us to move from any (insincere) ballot to a sincere ballot in a finite number of steps. Having said this, the notion of degree of insincerity introduced here may also be of independent interest. For instance, we may want to distinguish different instances of manipulation that involve a severe form of insincerity from those that constitute only mildly insincere ballots. Related to this, Dowding and van Hees (2008), whose work we briefly review in Section 5.4, distinguish instances of election manipulation that are acceptable from those that are not.

We now state the theorem. It establishes the fact that any voter with preferences that satisfy **(WRP)** and **(DEL)** on top of the Kelly Principle will never have an incentive to not vote by means of a sincere ballot upon learning the ballots of the other voters. In other words, we can guarantee the existence of a sincere best response for any such voter. As we shall see, the theorem subsumes several similar results for intuitively appealing assumptions of set preferences, as discussed in Section 3.

Theorem 6. *Under AV, upon learning the ballots of the other voters, a voter whose preferences satisfy* **(KEL)**, **(WRP)** *and* **(DEL)** *will always have a best response that is sincere.*

Proof. Lemma 2 applies, so we only need to show that any strategy that involves approving of subpivotal and insignificant candidates only is weakly dominated by some sincere strategy. Recall that in case our voter votes by means of a ballot not approving of any pivotal candidates, the set of winners will be the union of the set of pivotal candidates and the set of approved subpivotal candidates. Let $B \subseteq \mathcal{X}$ be any insincere ballot not approving of a pivotal candidate and let $W \subseteq \mathcal{X}$ be the outcome forced by voting by means of B. We shall describe a procedure for turning B into a new ballot B' such that all of the following conditions are satisfied:

(a) the new outcome W' corresponding to B' weakly dominates W, i.e., $W' \succcurlyeq W$;
(b) the new ballot B' is less insincere than B, i.e., $B \gg B'$; and
(c) either B' is sincere or B' (like B) is an insincere ballot not approving of any pivotal candidates.

[9] The dual axiom, postulating $A \cup \{a\} \succcurlyeq A$ whenever $\{a\} \succ A$, may also be of independent interest, but is not required for our present purposes.

Condition (c) ensures that the same procedure will be applicable to B' as well. If we can define such a procedure, then the proof is complete, because it provides us with a method for turning any insincere ballot not approving of any pivotal candidates into a sincere ballot weakly dominating the former by means of a finite number of applications of that procedure.

We now describe the transformation procedure. Given B, define c as the most preferred candidate that is not approved by B, i.e., $c = \max(\mathcal{X} \setminus B)$. Note that, because B is insincere by assumption, c is well-defined and there exists a candidate that c is preferred to and that is approved. Then, depending on the type of c, apply one of the following transformations:

- First, suppose c is *insignificant*. Then let $B' = B \cup \{c\}$; that is, we approve of c in the new ballot and otherwise keep everything as it is. This satisfies all three conditions: (a) the outcome is not affected; (b) $B \gg B'$ by condition (i); and (c) B' does not approve of any pivotal candidates.
- Next, suppose c is *subpivotal*. We distinguish two cases:
  - If there is an approved subpivotal candidate c' that is less preferred than c, then proceed as follows. Note that $c' \in W$. By approving of c we can add c to W; and by disapproving c' we can delete c' from W. By **(WRP)**, we must have $(W \setminus \{c'\}) \cup \{c\} \succcurlyeq W$, $W \cup \{c\} \succcurlyeq W$, or $W \setminus \{c'\} \succcurlyeq W$.
    - ▹ If $(W \setminus \{c'\}) \cup \{c\} \succcurlyeq W$, then let $B' = (B \setminus \{c'\}) \cup \{c\}$. The new outcome will be $W' = (W \setminus \{c'\}) \cup \{c\}$. All three conditions will be satisfied: (a) $W' \succcurlyeq W$; (b) $B \gg B'$ by condition (i); and (c) B' does not approve of any pivotal candidates.
    - ▹ Otherwise, if $W \cup \{c\} \succcurlyeq W$, let $B' = B \cup \{c\}$. The new outcome will be $W' = W \cup \{c\}$. All three conditions will be satisfied: (a) $W' \succcurlyeq W$; (b) $B \gg B'$ by condition (i); and (c) B' does not approve of any pivotal candidates.
    - ▹ Otherwise, we must have $W \setminus \{c'\} \succcurlyeq W$. Then, let $B' = B \setminus \{c'\}$. The new outcome will be $W' = W \setminus \{c'\}$. All three conditions will be satisfied: (a) $W' \succcurlyeq W$; (b) $B \gg B'$ by condition (ii); and (c) B' does not approve of any pivotal candidates.
  - Otherwise, there is no such subpivotal candidate, i.e., all approved candidates that are less preferred than c are insignificant. In this case, in the new ballot simply disapprove of all of these insignificant candidates, i.e., $B' = \{b \in B \mid b \dot{>} c\}$. That is, c will cease to dominate any approved candidates. Again, all three conditions are satisfied: (a) the outcome is not affected; (b) $B \gg B'$ by condition (i); and (c) B' does not approve of any pivotal candidates.
- Finally, suppose c is pivotal. We again distinguish two cases:
  - If $\{c\} \succcurlyeq W$, then change the ballot completely and define $B' = \{a \in \mathcal{X} \mid a \dot{\geqslant} a^*\}$, where a^* is the most preferred pivotal candidate in $\mathcal{X}$. The new outcome will be $W' = \{a^*\}$. This satisfies all three conditions, because $W' \succcurlyeq W$ by **(KEL)** and B' is sincere.
  - Otherwise, as $\succcurlyeq$ is complete, it must be the case that $W \succ \{c\}$. Choose an approved candidate $c' \in B$ that is less preferred than c. By our assumptions, such a candidate c' does exist and cannot be pivotal. Now define the new ballot as $B' = B \setminus \{c'\}$. This satisfies two of our three conditions: (b) $B \gg B'$ by condition (ii); and (c) B' does not approve of any pivotal candidates. To check condition (a) we require a final case distinction:
    - ▹ If c' is insignificant, then the outcome is not affected and condition (a) is satisfied.

▷ If c' is subpivotal, then the new outcome will be $W' = W \setminus \{c'\}$. We obtain $W \succ \{c'\}$ from $W \succ \{c\}$ and $c \dot{>} c'$, and thus $W \setminus \{c'\} \succ W$ from **(DEL)**. Hence, condition (a) is satisfied also in this case.

This completes the proof of Theorem 6. □

Theorem 6 has a number of interesting consequences. For any assumptions on voter preferences that entail the set of axioms mentioned in the theorem, we immediately obtain the same kind of result for those assumptions. As a first corollary of this type, we show that we get a positive result for voters conforming to the Kannai-Peleg Principle, under which a voter's preferences over sets of candidates depend only on her preferences over the maxima and minima of those sets.

Theorem 7. *Under AV, upon learning the ballots of the other voters, a voter whose preferences satisfy* **(GAR)** *and* **(MMX)** *will always have a best response that is sincere.*

Proof. We need to show that axioms **(GAR)** and **(MMX)** together entail **(KEL)**, **(WRP)** and **(DEL)**. (The claim then follows from Theorem 6.) First, by Lemma 1, **(KEL)** is entailed by **(GAR)**.

Second, it is easy to check that **(GAR)** and **(MMX)** entail **(WRP)**. Indeed, replacing a by b (or just adding b or just removing a) in a set A where neither a or b are extremal will not affect the desirability of the set under **(MMX)**, while it will improve it if a or b are extremal.

Finally, **(DEL)** is also entailed by **(GAR)** and **(MMX)**: Assume **(GAR)** and **(MMX)**, and take any set $A \subseteq \mathcal{X}$ and any $a \in A$ such that $A \succ \{a\}$. Then a cannot be the maximal element in A, as that would contradict **(MAX)**. If a is the minimal element in A, then $A \setminus \{a\} \succ A$ by **(GF2)**. If a is neither the maximum nor the minimum of A, then $A \setminus \{a\} \sim A$ by **(MMX)**. Hence, in all cases **(DEL)** will be satisfied. □

Next, we prove a result for voters that conform to the Nitzan-Pattanaik Principle, according to which a voter's preferences over sets only depend on her preferences over the medians of those sets.

Theorem 8. *Under AV, upon learning the ballots of the other voters, a voter whose preferences satisfy* **(GAR)** *and* **(MED)** *will always have a best response that is sincere.*

Proof. We need to show that **(GAR)** and **(MED)** entail **(KEL)**, **(WRP)** and **(DEL)**. By Lemma 1, **(KEL)** is entailed by **(GAR)**.

Next, assume **(GAR)** and **(MED)**, and derive **(WRP)**: Take any set A and any two candidates a and b such that $b \dot{>} a$. If $a \notin A$ or $b \in A$, then **(WRP)** holds vacuously. So let $a \in A$ and $b \notin A$. If $|A|$ is odd and thus $\mathrm{med}(A)$ and $\mathrm{med}((A \setminus \{a\}) \cup \{b\})$ are both singletons, we obviously have $\mathrm{med}((A \setminus \{a\}) \cup \{b\}) \succcurlyeq \mathrm{med}(A)$ and thus $(A \setminus \{a\}) \cup \{b\} \succcurlyeq A$. If $|A|$ is even, then the situation is a little more subtle. In this case, let $\{m, m'\} = \mathrm{med}(A)$ with $m \dot{>} m'$. We make an exhaustive case distinction by considering all possible positions of a with respect to m and m':

- If $a \dot{>} m$, then $\mathrm{med}(A) = \mathrm{med}((A \setminus \{a\}) \cup \{b\})$. Thus, $(A \setminus \{a\}) \cup \{b\} \succcurlyeq A$ and **(WRP)** is satisfied.
- If $a = m$, then $\mathrm{med}(A \cup \{b\}) = \{m\}$. Thus, $A \cup \{b\} \succcurlyeq A$ and **(WRP)** is satisfied.[10]

[10] Observe that this is a case where the replacement axiom of Puppe (1995), corresponding to the first disjunct of **(WRP)**, would not be derivable using **(GAR)** and **(MED)** alone, because $\mathrm{med}((A \setminus \{a\}) \cup \{b\}) = \{m'', m'\}$ for some $m'' \dot{>} m$, and $\{m'', m'\}$ and $\{m, m'\}$ are not ranked by the Gärdenfors Principle. This is why we chose to introduce the weaker axiom **(WRP)**.

- If $m' \dot{\geqslant} a$, then $\text{med}(A \setminus \{a\}) = \{m\}$. Thus, $A \setminus \{a\} \succcurlyeq A$ and **(WRP)** is satisfied.

Note that a cannot be strictly between m and m', because in that case a itself would have to belong to the median set. Hence, we have indeed covered all possible cases.

Finally, we show that **(DEL)** is satisfied as well: Take any set A and any candidate $a \in A$ such that $A \succ \{a\}$. By **(MED)**, $\text{med}(A) \succ \{a\}$. First consider the case where $|A|$ is odd. Then $\text{med}(A) = \{m\}$ for some m with $m \dot{>} a$, and thus $\text{med}(A \setminus \{a\}) = \{m', m\}$ for some m' with $m' \dot{>} m$. $A \setminus \{a\} \succcurlyeq A$ then follows from **(GF1)**, as required. Now consider the case where $|A|$ is even. Then $\text{med}(A) = \{m, m'\}$ for some m and m' with $m \dot{>} m'$. It cannot be the case that $m \dot{>} a \dot{>} m'$, because that would contradict $\{m, m'\}$ being the median set. It can also not be the case that $a \dot{\geqslant} m$, because that would contradict $\text{med}(A) \succ \{a\}$. Hence, we must have $m' \dot{\geqslant} a$. This means that $\text{med}(A \setminus \{a\}) = \{m\}$. $A \setminus \{a\} \succcurlyeq A$ then follows from **(GF2)**, and we are done. □

As a final result, we show that if AV is used in combination with the uniform tie-breaking rule, voters whose preferences are based on utility functions and who are expected-utility maximisers will never have an incentive to vote insincerely upon learning the ballots of the other voters.[11]

Theorem 9. *Under AV with uniform tie-breaking, upon learning the ballots of the other voters, a voter who is an expected-utility maximiser will always have a best response that is sincere.*

Proof. To be able to apply Theorem 6, we need to show that the set preferences of a voter who is an expected-utility maximiser and who expects ties to be broken uniformly will satisfy the axioms **(KEL)**, **(WRP)**, and **(DEL)**. This certainly is the case for **(KEL)**. **(WRP)** is satisfied, because replacing a candidate with a preferred candidate will increase average utility. Finally, **(DEL)** is also clearly satisfied: $A \setminus \{a\} \succcurlyeq A$ can only be falsified if a yields above-average utility amongst the candidates in A, in which case the precondition $A \succ \{a\}$ would not hold. □

We emphasise that in Theorem 9 there is no assumption about the actual utility function underlying the preference order of the would-be manipulator. We only assume that some such function exists.

Importantly, none of the sets of assumptions made in Theorems 7–9 entail any of the other sets of assumptions made in these theorems. Hence, while each of them follows from Theorem 6, they are all independent of one another and will apply for different types of societies.

Finally, observe that **(KEL)** together with either **(OPT)** or **(PES)** entails all of **(KEL)**, **(WRP)** and **(DEL)**; thus Theorems 4 and 5 are indeed corollaries of Theorem 6, as announced at the beginning of Section 4.3. More immediately, Theorems 4 and 5 also follow from Theorem 7, because optimism and pessimism each entail both **(GAR)** and **(MMX)**.

5 Related Work

In this section, we review some closely related work on manipulation and sincerity under AV. In particular, this includes a number of contributions that are similar to our approach, but that are based on models that exhibit various (sometimes subtle) differences.

[11] A direct proof of Theorem 9 may be found in our earlier work (Endriss, 2007, Theorem 4).

5.1 Dichotomous Preferences and Small Numbers of Candidates

If voters have *dichotomous* preferences, then the impossibility of voting truthfully remarked upon in the introductory section disappears, and we can define a voting procedure for approval ballots to be strategy-proof if a voter never has an incentive to not vote by nominating all those candidates that are undominated according to her true preferences. Brams and Fishburn (1978, Theorem 3) show that AV is strategy-proof if voters are assumed to have dichotomous preferences. Of course, this immediately implies that in this case voters will always have a sincere best response.

Brams and Fishburn (1978), as part of the same theorem, also show that any voter with *trichotomous* preferences will always have a sincere best response, under assumptions on the corresponding set preferences that are equivalent to the Gärdenfors Principle. In particular, this means that for elections with at most three candidates, any voter conforming to the Gärdenfors Principle will always have a sincere best response. For elections with four candidates, as we have seen (in Section 4.1), we lose this guarantee, although it is possibly to regain it by making very weak additional assumptions on top of the Gärdenfors Principle (Endriss, 2007, Theorem 3).

5.2 Best Responses in the Presence of Uncertainty

Central to both our approach and that of Brams and Fishburn (1978) is that we analyse scenarios where a would-be manipulator faces the decision on whether or not to vote insincerely once she knows how all the other voters intend to vote. There is no uncertainty. If, instead, all that is known about the voting behaviour of the others are *mixed strategies* each randomising over several different ballots, then our results will not apply any longer. For instance, suppose there are four candidates, we use AV with uniform tie-breaking, and all voters are expected-utility maximisers (i.e., as in Theorem 9). Now suppose there is just one other voter besides the would-be manipulator, and that other voter is using the following mixed strategy: with 50% probability submit ballot `[1100]` and with 50% probability submit ballot `[0011]`. Then it would be in the best interest of the manipulator to vote by means of the (insincere) ballot `[1010]`, to ensure that for either one of the two possible cases her favourite amongst the two pivotal candidates will get elected. A model that allows us to analyse this kind of scenario has been studied, amongst others, by de Sinopoli et al. (2006).

As pointed out by Nuñez (2010), we can observe a similar effect in *Poisson voting games*, where the number of voters is unknown to the would-be manipulator. For example, in an election with four candidates, if the manipulator is unsure whether (1) she is the only voter or (2) there are two other voters with ballots leading to a tie between her third most preferred and least favourite candidate, then her best response is to approve of her first and third most preferred candidate. The structure of this example is very similar to above example with mixed strategies. In both cases the incentive to vote insincerely arises from the fact that the manipulator has to combine her best responses for two distinct situations within a single ballot.

Building on the work by Myerson and Weber (1993), Laslier (2009) analyses a model in which voters are expected-utility maximisers who base their voting strategies on their beliefs regarding the probability of a tie between exactly two front-runners (neglecting the remote but non-zero probability of a tie between three and more candidates). The beliefs about these probabilities could, for instance, be based on polling data. Laslier (2009) is able to show that in this model voters will vote sincerely, i.e., uncertainty does not necessarily lead to insincerity.

5.3 Optimistic and Cautious Perspectives on Mechanism Design

In recent work (Endriss et al., 2009), using a slightly different model than the one adopted here, we show that voters always have a best sincere response even under the Gärdenfors Principle (while in the present paper we were only able to obtain such general positive results under much stronger assumptions on $\succcurlyeq$). The difference between the two models concerns the interpretation of $\succcurlyeq$. In the present model, each voter is assumed to have a complete (weak) preference order $\succcurlyeq$ over candidate sets and the assumptions such as the Gärdenfors Principle (which in fact only induce a preorder rather than a weak order) encode those aspects of that actual order that we, the mechanism designers, know about. In the other model, $\succcurlyeq$ is assumed to be the voter's *actual* (incomplete) preference order over candidate sets. In that model, the assumption is that voters are not always able to compare all sets (or possibly even all individual candidates), and when they cannot compare two sets, one of which is associated with a sincere ballot and one with an insincere ballot, they will by default opt for the one associated with the sincere ballot. Hence, there are fewer potential reasons for manipulation, which explains why the positive results obtained can be more general under those assumptions.

The "optimistic" model of (Endriss et al., 2009) seems appropriate when we accept that voters may have limited cognitive abilities and thus incomplete preferences and when it is reasonable to assume that in case of incomparability they will by default opt for a sincere ballot. The "cautious" model adopted in the present paper, on the other hand, seems appropriate when we take the perspective of a mechanism designer who believes (or at least considers it possible) that voters have complete preferences over sets of candidates, but is only willing to make limited assumptions regarding the principles that voters use to lift their preferences to preferences over sets.

5.4 Sincere Manipulation

Some of our discussion of sincerity and degrees of insincerity is reminiscent of the work of Dowding and van Hees (2008). These authors challenge the view that manipulability is necessarily a bad thing. The good thing about manipulation is that it encourages reflection on the views of other members of society. In the context of ranked voting procedures, they propose a notion of *sincere manipulation*, not to be confused with the notion of *sincere ballot* discussed here. They argue that voting for the best possible alternative amongst those that are *feasible*, in the sense that they may still win the election, should not be considered a deplorable insincere act. What alternatives *are* feasible depends on the strategies of the other voters; so this definition cannot be applied to individual ballots in isolation. Roughly speaking, Dowding and van Hees (2008) define *insincere manipulation* as voting by reporting a preference that untruthfully ranks some candidate x over some other candidate y, but that results in y being elected. Their main technical result states that the plurality rule is immune to this kind of insincere manipulation. Thus, both the formal framework employed (which does not apply to nonranked procedures such as AV) and the kind of question investigated by Dowding and van Hees (2008) differ from our work.

6 Concluding Remarks

To conclude, let us briefly summarise our results. We have been able to show that, under certain assumptions concerning the principles that govern how voters extend their preferences over individual candidates to sets of candidates, the system of approval voting can offer protection against insincere voting in the sense that a voter who has been able to obtain complete information regarding the voting intentions of the others will never have an incentive to manipulate the election by means of entering a ballot that is not a sincere reflection of her true preferences.

Specifically, this is so if voters are known to base their preferences over sets only on the maxima and minima of those sets (Theorem 7), if they base them only on the medians of those sets (Theorem 8), or if ties are broken uniformly and voters are known to be expected utility-maximisers (Theorem 9), i.e., if they base their preferences over sets only on the "averages" of those sets. In all these cases, a voter will never have an incentive to vote insincerely. The same kind of guarantee can be given if voters are known to be either optimistic (Theorem 4) or pessimistic (Theorem 5). The most general result of this kind is Theorem 6. By Lemma 3, all our results are independent of whether or not we allow voters to abstain and whether or not we classify abstention ballots as sincere or insincere expressions of preference. At the same time, as is well known (see, e.g., Brams and Fishburn, 2007) and as is demonstrated by our example in Section 4.1, the same guarantees cannot be given under weaker assumptions on preference extensions.

Finally, we believe that the *deletion axiom* formulated in Section 4.4, which to the best of our knowledge has not yet been considered in the literature on extending preferences on sets to their powersets, may be of independent interest for that line of work. The axiom expresses the intuitively appealing principle that a decision maker should not object to removing an element from a set if she strictly prefers the full set to that element in isolation.

References

S. Barberà, W. Bossert, and P. Pattanaik. Ranking sets of objects. In *Handbook of Utility Theory*, volume 2. Kluwer Academic Publishers, 2004.

S. J. Brams. *Mathematics and Democracy: Designing Better Voting and Fair-Division Procedures.* Princeton University Press, 2008.

S. J. Brams and P. C. Fishburn. Approval voting. *The American Political Science Review*, 72(3): 831–847, 1978.

S. J. Brams and P. C. Fishburn. *Approval Voting.* Springer, 2nd edition, 2007.

B. Can, B. Erdamar, and M. R. Sanver. Expected utility consistent extensions of preferences. *Theory and Decision*, 67(2):123–144, 2009.

F. de Sinopoli, B. Dutta, and J.-F. Laslier. Approval voting: Three examples. *International Journal of Game Theory*, 35(1):27–38, 2006.

K. Dowding and M. van Hees. In praise of manipulation. *British Journal of Political Science*, 38(1): 1–15, 2008.

U. Endriss. Vote manipulation in the presence of multiple sincere ballots. In D. Samet, editor, *Proceedings of the 11th Conference on Theoretical Aspects of Rationality and Knowledge (TARK-2007)*, pages 125–134. Presses Universitaires de Louvain, 2007.

U. Endriss, M. S. Pini, F. Rossi, and K. B. Venable. Preference aggregation over restricted ballot languages: Sincerity and strategy-proofness. In C. Boutilier, editor, *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI-2009)*, pages 122–127, 2009.

B. Erdamar and M. R. Sanver. Choosers as extension axioms. *Theory and Decision*, 67(4):375–384, 2009.

P. C. Fishburn. Even-chance lotteries in social choice theory. *Theory and Decision*, 3(1):18–40, 1972.

P. Gärdenfors. Manipulation of social choice functions. *Journal of Economic Theory*, 13(2):217–228, 1976.

C. Geist and U. Endriss. Automated search for impossibility theorems in social choice theory: Ranking sets of objects. *Journal of Artificial Intelligence Research*, 40:143–174, 2011.

Y. Kannai and B. Peleg. A note on the extension of an order on a set to the power set. *Journal of Economic Theory*, 32(1):172–175, 1984.

J. Kelly. Strategy-proofness and social choice functions without single-valuedness. *Econometrica*, 45 (2):439–446, 1977.

J.-F. Laslier. The leader rule: A model of strategic approval voting in a large electorate. *Journal of Theoretical Politics*, 21(1):113–136, 2009.

J.-F. Laslier and M. R. Sanver, editors. *Handbook of Approval Voting*. Studies in Choice and Welfare. Springer-Verlag, 2010.

S. Merrill, III and J. Nagel. The effect of approval balloting on strategic voting under alternative decision rules. *The American Political Science Review*, 81(2):509–524, 1987.

R. B. Myerson and R. J. Weber. A theory of voting equilibria. *The American Political Science Review*, 87(1):102–114, 1993.

R. G. Niemi. The problem of strategic behavior under approval voting. *The American Political Science Review*, 78(4):952–958, 1984.

S. I. Nitzan and P. K. Pattanaik. Median-based extensions of an ordering over a set to the power set: Am axiomatic characterization. *Journal of Economic Theory*, 34(2):252–261, 1984.

M. Nuñez. Sincere scoring rules. THEMA Working Paper 2010-02, Université de Cergy-Pontoise, 2010.

C. Puppe. Freedom of choice and rational decisions. *Social Choice and Welfare*, 12(2):137–153, 1995.

F. S. Roberts. *Measurement Theory*. Addison-Wesley, 1979.

M. R. Sanver. Approval as an intrinsic part of preference. In *Handbook of Approval Voting*, Studies in Choice and Welfare, chapter 20. Springer-Verlag, 2010.

A. K. Sen. Welfare, preference and freedom. *Journal of Econometrics*, 50(1–2):15–29, 1991.

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

R. J. Weber. Comparison of voting systems. Cowles Foundation Discussion Paper No. 498/A, Yale University, 1978.