Automated Analysis of Social Choice Problems: Approval Elections with Small Fields of Candidates

Ulle Endriss

Institute for Logic, Language and Computation
University of Amsterdam

Abstract

We analyse the incentives of a voter to vote insincerely in an election conducted under the system of approval voting. Central to our analysis are the assumptions we make on how voters deal with the uncertainty stemming from the fact that a tie-breaking rule may have to be invoked to determine the unique election winner. Because we only make very weak assumptions in this respect, it is impossible to obtain general positive results. Instead, we conduct a fine-grained analysis using an automated approach that reveals a clear picture of the precise conditions under which insincere voting can be ruled out. At the methodological level, this approach complements other recent work involving the application of techniques originating in computer science and artificial intelligence in the domain of social choice theory.

1 Introduction

Voting theory and, more generally speaking, social choice theory, originated in economics and political science, but are becoming increasingly important for artificial intelligence (AI). There are two distinct reasons for this trend. First, as the formal study of collective decision making, social choice theory has the potential to make important contributions to the design and analysis of multiagent systems. Second, methods of computer science and AI have turned out to be very helpful in analysing problems of social choice, leading to the interdisciplinary research area known as *computational social choice* (Chevaleyre et al., 2007; Brandt et al., 2012). Maybe the clearest example for this kind of research is the large body of work devoted to the use of complexity theory for the analysis of the manipulation problem in voting, which has recently been reviewed by Faliszewski and Procaccia (2010). Other examples include the design of fast algorithms for computing the winners under complex voting rules (see, e.g., Conitzer et al., 2006) and the study of social choice problems with very large numbers of alternatives induced by the multi-attribute structure of decision problems arising in practice (see, e.g. Lang, 2004).

A collection of techniques from AI that has great potential for social choice theory but that has only received little attention to date is *automated reasoning*. It might be used to verify existing theorems in social choice theory and to search for new ones; it might be used to automatically check whether a given social choice rule meets a set of requirements; and it might assist in the design of new rules. So far, most efforts along this line have concentrated on one of the classical results in the field, Arrow's Theorem, which establishes the impossibility of devising a method for preference aggregation that simultaneously meets a small number of seemingly innocent requirements (Arrow, 1963). There have been attempts to prove (or at least verify proofs of) Arrow's Theorem (or special cases of Arrow's Theorem) using higher-order theorem provers (Nipkow, 2009; Wiedijk, 2007), first-order theorem provers (Grandi and Endriss, 2012), and SAT-solvers (Tang and Lin, 2009). One of the very few examples of using automated reasoning in other areas of computational social choice is our recent work on using a SAT-solver to automatically search for impossibility theorems in the domain of ranking sets of objects (Geist and Endriss, 2011).

In this paper we focus on yet another problem domain and we explore a much simpler approach.

1.1 The Problem: Sincerity and Manipulation in Approval Voting

One of the central questions in voting theory is under what circumstances a voter will have an incentive to *manipulate* the election by misrepresenting her true preferences. For instance, under plurality voting (where

each voter can award 1 point to one and only one candidate), if a voter favours a candidate from a small party who has no realistic chance of winning, then she has an incentive to manipulate and vote for her most preferred candidate from a mainstream party instead. The classical Gibbard-Satterthwaite Theorem shows that this problem cannot be avoided in general for any voting rule under which voters report a (possibly truncated) ranking of candidates (Taylor, 2005).

Under the system of *approval voting* (AV), a voter can approve of as many candidates as she likes and the candidate with the most approvals wins (Brams and Fishburn, 1978). That is, under AV voters report *sets* of candidates rather than *rankings*, which means that the Gibbard-Satterthwaite Theorem does not apply directly. In fact, if we continue to assume that true preferences are rankings of candidates, then the notion of truthful voting ceases to be well-defined (if you must report a set rather than a ranking, you certainly cannot report your truthful ranking). There is however a weaker notion of "good" behaviour in AV: a ballot B (a set of approved candidates) is called *sincere* if our voter does not prefer any of the candidates outside of B to any of those inside of B. We shall be interested in the following question: Assuming that a voter has obtained complete information on how all other voters are going to vote, under what circumstances can we ensure that she will never have an incentive to vote by means of an insincere ballot?

As it turns out, the answer to this question crucially depends on the assumptions we are willing to make on how a voter deals with the uncertainty arising from the fact that her vote might result in a tied election, meaning that the eventual winner must be chosen from a set of front-runners using a suitable tie-breaking rule. That is, when contemplating her ballot, the voter actually has to choose between several possible sets of front-runners rather than between several unique winners. On the other hand, her preferences are initially only defined over individual candidates. Thus, we need to reason about how a voter will extend her preferences to sets of candidates. If we make relatively strong assumptions regarding this matter, then we can obtain strong positive results showing that a fully informed voter will never have any cause to vote insincerely (Endriss, 2012). But these results break down when we weaken these assumptions. Our goal in this paper is to understand the case of weak assumptions regarding preference extension in more depth.

1.2 The Approach: Automated Analysis for a Fixed Number of Candidates

A simple but crucial insight is that we can abstract away from the number of voters. All that matters for the analysis of a manipulation situation is which candidates have obtained the most approvals, which are trailing by only 1 point each, and which are lagging behind even further (and thus certainly will not win). This means that it is possible, at least in principle, to exhaustively enumerate and check all relevant situations that could possibly arise, for any given number n of candidates. In this paper, we show that for small values of n this approach is also feasible in practice and that it allows us to derive interesting new results clarifying the precise conditions under which insincere voting might occur under AV.

The remainder of this paper is organised as follows. Section 2 introduces three principles for preference extension and sketches an algorithm for deciding whether one set of candidates dominates another under such a principle. Section 3 explains our approach to the automated analysis of the sincerity problem in AV with a fixed number of candidates and presents the results obtained. Section 4 concludes.

2 Preference Extensions

Let $\mathcal{X}$ be a finite set of two or more *candidates* and let $n := |\mathcal{X}|$. Let $\underline{\mathcal{X}} = 2^{\mathcal{X}} \backslash \{\emptyset\}$ denote the set of *nonempty sets of candidates*. We will consider a *voter* with a *preference relation $\geqslant$ over individual candidates* that is a *total order* (i.e., a binary relation on $\mathcal{X}$ that is reflexive, transitive, complete and antisymmetric). That is, we write $a \geqslant b$ to express that our voter likes candidate a at least as much as b. The corresponding strict preference relation is denoted by $>$; i.e., $a > b$ if $a \geqslant b$ but not $b \geqslant a$.

Suppose our voter can choose between two alternative outcomes of an election, A and B, but for either one of them two or more candidates might be tied for winning; i.e., A and B are elements of $\underline{\mathcal{X}}$ and a unique winner would eventually have to be chosen from A or B using a tie-breaking rule (e.g., choosing randomly or asking an arbiter). Can we predict whether our voter will prefer A or B, given that we only know $\geqslant$?

In general, we cannot. But there are some reasonable assumptions that we can make. We will now introduce three principles for extending a preference order $\geqslant$ on $\mathcal{X}$ to a preference order $\succcurlyeq$ on $\underline{\mathcal{X}}$. Here, $\succcurlyeq$ is assumed to be a *weak order* (i.e., a binary relation that is reflexive, transitive and complete). That is, while we assume that our voter can strictly rank all individual candidates, she might be indifferent between two sets of candidates. We write $\succ$ for the strict part of $\succcurlyeq$. This form of preference extension has been studied

extensively under the name of *ranking sets of objects* (see, e.g., Barberà et al., 2004; Geist and Endriss, 2011). Each of the three principles we present will (usually) not be sufficient to allow us to induce the full weak order $\succcurlyeq$, but it will allow us to rank at least some pairs of sets, i.e., it will allow us to predict the choices of our voter in at least some cases.

2.1 The Kelly Principle

According to Kelly (1977), we should prefer a singleton consisting only of a to a singleton consisting only of b if we prefer a to b (extension axiom); and we should like a set no more than its best element and no less than its worst element. The *Kelly Principle* is thus defined by three axioms:

(EXT)	$\{a\} \succ \{b\}$	if $a > b$
(MAX)	$\{\max(A)\} \succcurlyeq A$	where $\max(A) := a^\star$ such that $a^\star > a$ for all $a \in A \setminus \{a^\star\}$
(MIN)	$A \succcurlyeq \{\min(A)\}$	where $\min(A) := a^\star$ such that $a > a^\star$ for all $a \in A \setminus \{a^\star\}$

We write **(KEL)** for the conjunction of **(EXT)**, **(MAX)**, and **(MIN)**. The Kelly Principle amounts to very weak assumptions: even if we do not know anything at all about the tie-breaking rule or our voter's attitude towards risk, we can say for sure that she will conform to it.

2.2 The Gärdenfors Principle

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):

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

We write **(GAR)** for the conjunction of **(GF1)** and **(GF2)**. Note that whenever $A \succcurlyeq B$ under the Kelly Principle, then $A \succcurlyeq B$ also under the Gärdenfors Principle, i.e., **(GAR)** entails **(KEL)**. Accepting **(GAR)** is known to be equivalent to assuming that our voter believes that any ties will be broken by an outside arbiter according to some unknown but fixed preference order (Erdamar and Sanver, 2009).

2.3 The Sen-Puppe Principle

Expanding on a proposal by Sen (1991), Puppe (1995) defined the following axiom:

(SIF)	$(A \setminus \{a\}) \cup \{b\} \succcurlyeq A$ if $b > a$, with $a \in A$ and $b \notin A$

We call this the *single-flip axiom*. It states that replacing an element in a set with another element that is strictly more preferred should result in a set that is at least as good as the original one. Let us call the conjunction of **(GAR)** and **(SIF)** the Sen-Puppe Principle. It is more restrictive and less standard than those of Kelly and Gärdenfors, but still weaker than other common assumptions in the literature (Endriss, 2012).

2.4 Algorithms for Checking Dominance

Given two sets $A, B \in \underline{\mathcal{X}}$ and a total order $\geqslant$ on $\mathcal{X}$, we now want to devise an algorithm to check whether a given principle of preference extension suffices to infer that A *weakly dominates* B ($A \succcurlyeq B$) and whether it suffices to infer that A *strictly dominates* B ($A \succ B$).

    We can represent A and B as sequences of 0's and 1's. For instance, if there are 5 candidates, then $\{11001\}$ is the set consisting of our voter's two most preferred candidates as well as her least preferred candidate. We first discuss the case of strict dominance. For each of the three principles introduced, a necessary precondition for inferring $A \succ B$ is that we can divide each of the two lists into three (possibly empty) sublists, such that the first two sublists are of equal length, the second two sublists are of equal length (and thus also the third two sublists are), and that the third sublist corresponding to A and the first sublist corresponding to B only consist of 0's. The situation is summarised by the following diagram:

```
        A1              A2              A3
      ----------  |  ------------  |  000000000 }
{ ----------      |  ------------  |  ---------}
{ 0000000000      |  ------------  |
        B1              B2              B3
```

Furthermore, at least one of the lists `A1` and `B3` need to contain a 1. These are necessary requirements under all three principles. In addition, we need to check whether a division into sublists can be found that also satisfies the following principle-specific conditions regarding the two *middle parts* of the list (`A2` and `B2`):

(1) We can infer $A \succ B$ according to the Kelly Principle if and only if both `A2` and `B2` are the empty list for some division into sublists.

(2) We can infer $A \succ B$ according to the Gärdenfors Principle if and only if `A2` and `B2` are equal for some division into sublists.

(3) We can infer $A \succ B$ according to the Sen-Puppe Principle if and only if the following procedure succeeds for some division into sublists: Move simultaneously through both `A2` and `B2`, from left to right. At each position, apply the appropriate rule:

  - If the remaining two lists to the right of the current position are equal, stop and succeed (that is, succeed immediately if `A2` and `B2` were equal to begin with).
  - If either both lists have a 1 or both have a 0 at the current position, move on to the next position.
  - If there is a 1 at the current position in `A2` and a 0 in `B2`, then flip the next available 1 in `B2` to a 0 (stop and fail if there is no such 1), and move on to the next position.
  - If there is a 0 at the current position in `A2` and a 1 in `B2`, then stop and fail.

The algorithm for checking weak dominance is very similar. There are only two differences. First, it is not necessary for one of `A1` and `B3` to contain a 1. Second, for the Kelly Principle we can also infer $A \succcurlyeq B$ if for some division into sublists both `A2` and `B2` are equal to a list consisting of just a single 1.

For the sake of brevity, we omit a detailed proof of correctness of these algorithms. For the Kelly Principle, correctness is immediate. For the Gärdenfors Principle, note how the condition on `A1`/`B1` corresponds to **(GF1)**, that on `A3`/`B3` to **(GF2)**, and that on `A2`/`B2` to the fact that $\succcurlyeq$ is reflexive. For the Sen-Puppe Principle, observe that the rules for the comparison of `A2` and `B2` closely model **(SIF)**.

3 Sincerity and Manipulation under Approval Voting

We are now ready to present our results on insincere manipulation under AV.

3.1 The Case of Three Candidates

Let us first analyse the case of three candidates. Suppose a would-be manipulator knows how all other voters are going to vote. Given those votes, one or more candidates will have received the highest number of approvals; we call them the *pivotal* candidates. Some will have received exactly 1 point less than the pivotal candidates; we call them the *subpivotal* candidates. We call the remaining candidates *insignificant*; they have no chance of being elected, however our manipulator is going to vote.

There are $3^3 - 2^3 = 19$ different *situations:* each candidate must belong to one of the three groups, and at least one of them must be pivotal. Each of the rows in Table 1 corresponds to one such situation. Row (12) with label `S.P.I`, for example, represents the situation where our manipulator's favourite candidate is **s**ubpivotal, her second choice is **p**ivotal, and her last choice is **i**nsignificant. The columns of Table 1 correspond to the valid ballots available to her.[1] The two sincere ballots are shown on the left; the four insincere ballots are shown on the right. The table cells correspond to the election outcomes for a given situation and a given final ballot. For example, if our manipulator chooses the sincere ballot `[110]` (i.e., if she approves of her two most preferred candidates) in situation `S.P.I`, then outcome $\{010\}$ will be realised (i.e., her second favourite candidate will be the sole winner).

The manipulator knows in which situation we are and needs to choose a ballot. She does so by choosing (one of) the best outcome(s) attainable to her in the situation at hand. Depending on our assumptions on set preferences, we can exclude some possible choices. Suppose all we know is that the manipulator satisfies the Kelly Principle. In Table 1, we have underlined all those outcomes that, according to the Kelly Principle, are not strictly dominated by any other outcome in the same row. This shows that in most situations our voter will have no incentive to vote insincerely. For example, in situation (12) the unique top outcome is

---

[1] We do not list the *abstention ballots* `[000]` (approving no candidates) and `[111]` (approving all candidates). By a general result, such abstention ballots cannot influence the type of result we seek here (Endriss, 2012, Lemma 3).

	Sincere		Insincere Ballots			
Situation	[100]	[110]	[001]	[010]	[011]	[101]
(1) P.P.P	{100}	{110}	{001}	{010}	{011}	{101}
(2) P.P.S	{100}	{110}	{111}	{010}	{010}	{100}
(3) P.P.I	{100}	{110}	{110}	{010}	{010}	{100}
(4) P.S.P	{100}	{100}	{001}	{111}	{001}	{101}
(5) P.S.S	{100}	{100}	{101}	{110}	{111}	{100}
(6) P.S.I	{100}	{100}	{100}	{110}	{110}	{100}
(7) P.I.P	{100}	{100}	{001}	{101}	{001}	{101}
(8) P.I.S	{100}	{100}	{101}	{100}	{101}	{100}
(9) P.I.I	{100}	{100}	{100}	{100}	{100}	{100}
(10) S.P.P	{111}	{010}	{001}	{010}	{011}	{001}
(11) S.P.S	{110}	{010}	{011}	{010}	{010}	{111}
(12) S.P.I	{110}	{010}	{010}	{010}	{010}	{110}
(13) S.S.P	{101}	{111}	{001}	{011}	{001}	{001}
(14) S.I.P	{101}	{101}	{001}	{001}	{001}	{001}
(15) I.P.P	{011}	{010}	{001}	{010}	{011}	{001}
(16) I.P.S	{010}	{010}	{011}	{010}	{010}	{011}
(17) I.P.I	{010}	{010}	{010}	{010}	{010}	{010}
(18) I.S.P	{001}	{011}	{001}	{011}	{001}	{001}
(19) I.I.P	{001}	{001}	{001}	{001}	{001}	{001}

Table 1: Outcomes for each possible situation and ballot, for elections with three candidates.

{110} and that outcome is attainable by voting sincerely using [110] (besides being also attainable via one of the insincere ballots). In fact, rather surprisingly, there are *only two critical situations* where this is not the case. These are situations (11) and (13). In situation (11) it is conceivable that our voter's most preferred outcome is {111} (and not {110}), in which case she would have an incentive to vote insincerely using the ballot [101]. In situation (13), it is conceivable that her most preferred outcome is {011} (and neither {101} nor {111}), in which case she might vote using [010].

The following two "*plain axioms*" are directly derived from rows (11) and (13). They represent the minimal additional assumptions we need to make, above and beyond the Kelly Principle, if we want to rule out any incentives for our voter to vote insincerely.

(AX1) $\{a,b\} \succcurlyeq \{a,b,c\}$ if $a > b$ and $b > c$

(AX2) $\{a,c\} \succcurlyeq \{b,c\}$ or $\{a,b,c\} \succcurlyeq \{b,c\}$ if $a > b$ and $b > c$

For example, **(AX2)** excludes the possibility that {011} (corresponding to $\{b,c\}$ in the statement of the axiom) is strictly preferred to both {101} and {111}. Hence, in situation (13), one of the two sincere outcomes that are undominated according to the Kelly Principle will be most preferred amongst all feasible outcomes. To summarise our observations, inspection of Table 1 allows us to establish the following result:

Theorem 1. *Under AV with three candidates, upon learning the ballots of the other voters, a voter whose preferences satisfy* **(KEL)***,* **(AX1)***, and* **(AX2)** *will always have a best response that is sincere.*

That is, under the stated (very weak) assumptions on set preferences, no voter will ever have an incentive to vote by means of an insincere ballot when there are (at most) three candidates up for election. We get the same positive result for any other axiom system that entails the axioms referred to in Theorem 1 (for the special case of three candidates). In particular, we obtain the following result:

Theorem 2. *Under AV with three candidates, upon learning the ballots of the other voters, a voter whose preferences satisfy* **(GAR)** *will always have a best response that is sincere.*

Proof. This is an immediate corollary of Theorem 1, because **(GAR)** entails **(KEL)**, **(AX1)** and **(AX2)**. □

Theorem 2 is a known result. It immediately follows from Theorem 3 of Brams and Fishburn (1978), whose assumptions on preference extension are equivalent to **(GAR)**. Given that **(GAR)** is *not* entailed by **(KEL)** together with **(AX1)** and **(AX2)**, Theorem 1 is technically slightly stronger than Theorem 2.

Table 2: Algorithm to compute additional axioms required to rule out insincere voting for situation S.

3.2 Automated Derivation of Sincerity Results for a Fixed Number of Candidates

When there are four candidates and voters conform to the Gärdenfors Principle, then insincere voting *can* occur. To see this, consider the following example. Suppose your preferences are $a > b > c > d$; candidates b and d have received 10 approvals each (they are pivotal); and a and c have received 9 approvals each (they are subpivotal). Then you can force outcome $\{b\}$ (by voting for a, b and c), outcome $\{a, b, c, d\}$ (by voting for a and c), outcome $\{a, b, d\}$ (by voting for a), and several other outcomes that are dominated by one of the first three under the Gärdenfors Principle (e.g., voting only for c gives $\{b, c, d\}$, which is dominated by $\{b\}$). The Gärdenfors Principle is not strong enough to tell us which of those three outcomes you prefer the most. If it is, say, $\{b\}$, then we are fine, because you can achieve it by voting sincerely. But it *might* be the case that $\{a, b, c, d\}$ is your most preferred feasible outcome, in which case you have an incentive to vote insincerely (for a and c, but not for b).

Hence, we will not be able to generalise Theorem 2 to the case of four candidates. Instead, in analogy to our analysis of elections with three candidates, we now want to analyse *how many* scenarios there are where the Gärdenfors Principle cannot rule out manipulation with an insincere ballot. The problem, however, is that when there are more than three candidates it is not feasible anymore to write out and reliably check the kind of data shown in Table 1 by hand. For four candidates, there are already $3^4 - 2^4 = 65$ rows and $2^4 - 2 = 14$ columns to such a table; for five candidates there are $3^5 - 2^5 = 211$ rows and $2^5 - 2 = 30$ columns. Instead, we propose to generate and check this kind of data automatically.

Suppose we have fixed the number of candidates n and the set of assumptions we want to make regarding the extension of preferences to set preferences. Given n, we can generate all possible *situations* we need to consider: these can be represented by all lists of length n of the letters P, S, and I that include at least one copy of P each. We can also generate all possible *proper ballots:* these are all lists of length n of the numbers 0 and 1 that include at least one 0 and one 1 each. Those ballots that are lists in which all occurrences of 0 occur to the right of all occurrences of 1 are *sincere;* all others are *insincere.* Given any situation and any ballot, it is easy to compute the resulting *outcome* (the set of winning candidates). As we have seen in Section 2.4, for any two outcomes A and A', we can check whether A is *known to weakly dominate* A' according to the assumptions we have made regarding $\succcurlyeq$; and we can also check whether A is *known to strictly dominate* A' under these assumptions.

Now, to systematically check whether a particular scenario admits a situation where a voter would have an incentive to vote insincerely and to generate all additional plain axioms that would be required to rule out any such scenario, we execute the algorithm described in Table 2, once for every possible situation S. This algorithm will return a (possibly empty) list Γ of plain axioms (from which we can remove any duplicates):

(1) If the list Γ is empty, then this proves that our assumptions are sufficiently strong to guarantee the absence of incentives to vote insincerely.

(2) If Γ includes an unsatisfiable axiom (that is, an axiom with $\Omega' = \emptyset$), then this proves that our assump-

tions do allow for situations where a voter will have an incentive to vote insincerely. Furthermore, in this case it is *impossible* to rectify this problem by adding additional axioms.

(3) Otherwise, voters will sometimes have incentives to vote insincerely, but this problem *can* be rectified. We then have a proof that the original assumptions together with the axioms in Γ will guarantee that no voter ever has an incentive to vote insincerely.

We have implemented this algorithm and the algorithm for deciding dominance under the Kelly, the Gärdenfors, and the Sen-Puppe Principles described in Section 2.4 in PROLOG.[2]

3.3 The Case of Four Candidates

In the case of four candidates with voters conforming to the Gärdenfors Principle, our program produces the following two plain axioms as output:

(AX3)	$\{b\} \succcurlyeq \{a, b, c, d\}$ or $\{a, b, d\} \succcurlyeq \{a, b, c, d\}$	if $a > b > c > d$
(AX4)	$\{a, d\} \succcurlyeq \{a, c, d\}$ or $\{a, b, d\} \succcurlyeq \{a, c, d\}$ or $\{a, b, c, d\} \succcurlyeq \{a, c, d\}$	if $a > b > c > d$

The situation corresponding to **(AX3)** is S.P.S.P. This is the case familiar from the example of Section 3.2. The situation corresponding to **(AX4)** is S.S.S.P. In this case, the insincere ballot [1010] will produce the outcome $\{1011\}$, which the Gärdenfors Principle alone is not strong enough to show to be at least as preferred as the three outcomes attainable by means of sincere ballots.

Provided our program is a *correct implementation of the algorithm* (on this point, see Section 4), we can infer the following result:

Theorem 3. *Under AV with four candidates, upon learning the ballots of the other voters, a voter whose preferences satisfy* **(GAR)***,* **(AX3)***, and* **(AX4)** *will always have a best response that is sincere.*

That is, a minor refinement of the Gärdenfors Principle will rule out any incentives to vote insincerely for a voter who knows how the others are going to vote, even for elections with four candidates. Under the Sen-Puppe Principle the refinement required is even smaller:

Theorem 4. *Under AV with four candidates, upon learning the ballots of the other voters, a voter whose preferences satisfy* **(GAR)***,* **(SIF)***, and* **(AX3)** *will always have a best response that is sincere.*

Proof. The second disjunct of **(AX4)** is an instance of **(SIF)**, i.e., the latter entails the former. The claim then follows from Theorem 3. □

(SIF), and certainly the much weaker **(AX4)**, as well as **(AX3)** are all reasonable assumptions that will be justified in many practical cases. The Gärdenfors Principle itself is certainly widely accepted. That means, in practice, we will usually be able to exclude insincere manipulations from voters who have obtained full information on the voting intentions of others for elections with (up to) four candidates.

3.4 Quantitative Results

As we have seen, when the number of candidates is small, so are both the number of situations in which we must consider it possible that a voter might vote insincerely and the number of additional plain axioms that we would have to accept to be able to rule out insincere voting. Naturally, as we increase the number of candidates, the number of such "exceptions" will go up as well. To give an impression of *how much* the number of exceptions grows, Table 3 provides an overview of the relevant figures for elections with 2–7 candidates for our three preference extension principles. For each principle and each election size, the corresponding table cell shows two figures. The one on top is the number of critical situations (specifying for each candidate whether he is pivotal, subpivotal, or insignificant) in which a voter may benefit from insincere manipulation, and the one at the bottom is the number of additional plain axioms required. In no instance did our algorithm return an unsatisfiable axiom. That is, in all cases it is possible to amend the assumptions we started out with so as to rule out insincere manipulation, albeit in some cases a very large number of additional plain axioms will be required.

For comparison, Table 3 also shows the overall number of situations for each election size (which is $3^n - 2^n$). As we can see, as n grows, not only does the absolute number of critical situations increase, but the same is true for the ratio of critical situations over situations in general.

---

[2]The full program required to generate all the results in this paper consist of around 80 lines of code and is available from the author or may be downloaded from http://www.illc.uva.nl/~ulle/approval-voting/.

Number of candidates:	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$
Kelly Principle:	0	2	14	64	244	846
(KEL)	0	2	19	114	553	2372
Gärdenfors Principle:	0	0	2	18	100	444
(GAR)	0	0	2	24	173	972
Sen-Puppe Principle:	0	0	1	10	61	294
(GAR) + **(SIF)**	0	0	1	10	63	321
Number of situations:	5	19	65	211	665	2059

Table 3: Number of critical situations (top figure) and number of additional plain axioms required to rule out insincere voting (bottom figure), for different axiom systems and numbers of candidates (n).

4 Conclusion

We have seen several results highlighting conditions under which a voter in an approval election who has obtained full information on the voting intentions of all other voters (the classical manipulation scenario) will never have an incentive to vote insincerely. These results for *weak* principles of preference extension and *small* numbers of candidates complement earlier results for *strong* principles of preference extension and *arbitrary* numbers of candidates (Endriss, 2012). Interestingly, our results show that the kind of manipulation considered can usually be ruled out for elections with up to four candidates, while in the classical setting this is only possible for elections with at most two candidates (Taylor, 2005).

Just as interesting as the results themselves is the method we have used to obtain them, namely by automatically exploring the space of all possible voting situations. All of our results have been derived using a very short logic program. The correctness of such a program, and thus of our results, can be verified in a similar manner as a classical manual proof.

References

K. J. Arrow. *Social Choice and Individual Values*. John Wiley and Sons, 2nd edition, 1963.

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

F. Brandt, V. Conitzer, and U. Endriss. Computational social choice. In G. Weiss, editor, *Multiagent Systems*. MIT Press, 2012. In press.

Y. Chevaleyre, U. Endriss, J. Lang, and N. Maudet. A short introduction to computational social choice. In *Proc. 33rd Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM-2007)*. Springer-Verlag, 2007.

V. Conitzer, A. J. Davenport, and J. Kalagnanam. Improved bounds for computing Kemeny rankings. In *Proc. 21st National Conference on Artificial Intelligence (AAAI-2006)*, 2006.

U. Endriss. Sincerity and manipulation under approval voting. *Theory and Decision*, 2012. In press.

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

P. Faliszewski and A. D. Procaccia. AI's war on manipulation: Are we winning? *AI Magazine*, 31(4):53–64, 2010.

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.

U. Grandi and U. Endriss. First-order logic formalisation of impossibility theorems in preference aggregation. *Journal of Philosophical Logic*, 2012. In press.

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

J. Lang. Logical preference representation and combinatorial vote. *Annals of Mathematics and Artificial Intelligence*, 42(1–3):37–71, 2004.

T. Nipkow. Social choice theory in HOL: Arrow and Gibbard-Satterthwaite. *Journal of Automated Reasoning*, 43(3): 289–304, 2009.

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

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

P. Tang and F. Lin. Computer-aided proofs of Arrow's and other impossibility theorems. *Artificial Intelligence*, 173(11): 1041–1053, 2009.

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

F. Wiedijk. Arrow's Impossibility Theorem. *Formalized Mathematics*, 15(4):171–174, 2007.