

# Single Transferable Vote Resists Strategic Voting

John J. Bartholdi, III

*School of Industrial and Systems Engineering  
Georgia Institute of Technology, Atlanta, GA 30332*

James B. Orlin

*Sloan School of Management  
Massachusetts Institute of Technology, Cambridge, MA 02139*

November 13, 1990; revised April 4, 2003

## **Abstract**

We give evidence that Single Transferable Vote (STV) is computationally resistant to manipulation: It is NP-complete to determine whether there exists a (possibly insincere) preference that will elect a favored candidate, even in an election for a single seat. Thus strategic voting under STV is qualitatively more difficult than under other commonly-used voting schemes. Furthermore, this resistance to manipulation is inherent to STV and does not depend on hopeful extraneous assumptions like the presumed difficulty of learning the preferences of the other voters.

We also prove that it is NP-complete to recognize when an STV election violates monotonicity. This suggests that non-monotonicity in STV elections might be perceived as less threatening since it is in effect “hidden” and hard to exploit for strategic advantage.

# 1 Strategic voting

For strategic voting the fundamental problem for any would-be manipulator is to decide what preference to claim. We will show that this modest task can be impractically difficult under the voting scheme known as Single Transferable Vote (STV). Furthermore this difficulty pertains even in the ideal situation in which the manipulator knows the preferences of all other voters and knows that they will vote their complete and sincere preferences. Thus STV is apparently unique among voting schemes in actual use today in that it is computationally resistant to manipulation. It might be that this resistance can help protect the integrity of social choice: If the work to construct a strategic preference is excessive, this might mean that strategic voting is not practical, even though theoretically possible.

Following the conventions of [14] we formalize the fundamental problem of strategic voting as the following “yes/no” question.

## **EFFECTIVE PREFERENCE**

**GIVEN:** A set  $C$  of candidates; a distinguished candidate  $c$ ; and the set  $V$  of sincere, transitive, strict, and complete preferences of the voters.

**QUESTION:** Is there a preference ordering on  $C$  that when tallied with the preferences of  $V$  will ensure the election of  $c$ ?

A polynomial-time algorithm for EFFECTIVE PREFERENCE is one that will always answer the question correctly within a number of steps bounded above by a polynomial in the size of a description of the election (which is  $O(|V||C|)$ ). Polynomial-time algorithms are considered fast because the work to answer the question does not increase too rapidly as a function of the size of the problem; similarly, problems for which there exist polynomial time algorithms are considered “easy”. In contrast, an algorithm which requires exponential time quickly becomes impractical as the size of the problem increases; accordingly, such problems are considered “hard”. This theoretical distinction is widely borne out in practical experience [14].

The celebrated theorems of Gibbard, Satterthwaite, and Gärdenfors show that any voting scheme that is minimally fair is in principle susceptible to strategic voting; that is, there exist instances in which some voter has incentive to misrepresent his true preferences [13, 15, 28]. For a susceptible voting scheme, if there exists an algorithm that is guaranteed to answer EFFECTIVE PREFERENCE within polynomial time, then, following [5], we say the voting scheme is *vulnerable*; otherwise the voting scheme is *resistant*.

This paper may be read as a companion to [5], which proved that most voting schemes in common use are vulnerable to strategic voting. Here we prove that STV, a voting scheme in widespread practical use, is qualitatively different from the others in that it requires distinctly more effort to vote strategically. Thus STV might encourage sincere voting, since it can be difficult to do otherwise.

In [4] it was observed that voting schemes due to Dodgson and to Kemeny had the undesirable property that it is NP-hard to tell whether any particular candidate has won the election. Such schemes are hard to manipulate for the uninteresting reason that they are hard to operate. What is wanted is a voting scheme that supports quick computation for authorized use, such as determining winners, but erects computational barriers to abuse. In [5] there was displayed a scheme that computes winners quickly but resists manipulation. However it is a rather contrived scheme whose only use, as far as we know, has been as a tie-breaking rule by the International Federation of Chess. This paper, taken with the results of [5] suggest that STV is the only voting scheme in actual use that computes winners quickly (in polynomial time) but is inherently resistant to strategic voting.

Note that we are not arguing for the adoption of STV—it has troubling faults documented elsewhere [6, 9, 12, 17, 21]—rather we are contributing to the argument begun by [4, 5] that computational properties ought to be among the criteria by which a prospective voting scheme is evaluated. In this regard STV differs in an interesting and possibly helpful way from other common voting schemes.

## 2 Single Transferable Vote

Under STV each voter submits a total order of the candidates. STV tallies votes by reallocating support from “weaker” candidates to “stronger” candidates and excess support from elected candidates to remaining contenders. In comparison with other voting schemes in practical use, it is rather complicated. However, despite its relative complexity, STV or its variants are used in elections for the parliaments of the Republic of Ireland, Tasmania, and Malta; for the senates of Northern Ireland, Australia, and South Africa; for all local authorities in the Republic of Ireland and some in Australia and Canada; and in the United Kingdom for many public and professional institutions, trade unions, and voluntary societies [23]. In the U. S. A. it is used by Cambridge, Massachusetts for the election of its city council and school committee and by New York City for the election of its district school boards [1]. John Stuart Mill praised its advantages as being “. . . such and so numerous . . . that, in my conviction, they place [STV] among the greatest improvements yet made in the theory and practice of government” [19]. Subsequent analysis has tempered this enthusiasm somewhat, since STV, like any voting system, must be imperfect. Most troubling of its weaknesses is that it is possible for a candidate to change from a winner to a loser as a result of gaining more support [9]. Nevertheless, STV has much to commend it, most notably its tendency to guarantee proportional representation. Largely for this property, STV is championed by such organizations as The Electoral Reform Society of Great Britain and Ireland [23].

We formalize STV as an algorithm in Figure 1, with supporting procedures in Figures 2, 3, and 4. All variants we know differ most significantly in how they reallocate excess support from an elected candidate to the remaining contenders. The “pure” form of STV uses the procedure given in Figure 4. For comparison, Figure 5 gives the procedure used by the city of Cambridge, Massachusetts. Our complexity results will hold for all such variants of STV, independently of how excess votes are reallocated.

For a picturesque rather than algorithmic description of STV, see [10]. For

an official guide to one version of STV in practice, see *Rules for Counting Ballots under Proportional Representation*, City of Cambridge Election, 1941, reproduced in [1]. A good summary assessment of the strengths and weaknesses of STV may be found in [7]. Finally, [16] contains compilable source code for a program to tabulate ballots in an STV election. The program is documented and seems suitable for practical use, as it handles administrative complications such as ties and incompletely specified ballots.

### 3 Strategic voting under STV

Others have remarked on the apparent difficulty of strategic voting under an STV election. Most immediately, Steven J. Brams, in private communication to the first author, conjectured that strategic voting is computationally difficult under STV. In addition, Chamberlin found that by several empirical measures strategic voting under STV seemed more difficult than under other voting schemes he tested (Plurality, Borda, and Coombs) [8]. He observed that STV, in contrast to the other schemes, “. . . usually has a much more complex and election-specific manipulation strategy”. This is consistent with our results since the NP-completeness of EFFECTIVE PREFERENCE under STV means that there is no simple pattern, structure, or rule to simplify the construction of an effective preference. Thus, for each strategic voter, the search for an effective preference must be essentially enumerative search over an exponentially large set—even when the strategic voters are trying to coördinate their votes to achieve a common goal. In contrast, for Plurality, Borda, and Coombs it is always the case that a voter can, within polynomial time, either construct a strategic preference or else conclude that none exists [5].

We will prove that EFFECTIVE PREFERENCE is hard, even for the special case of STV in which all candidates vie for a single seat. In this case the voting scheme works by successively eliminating a candidate with the fewest votes and reallocating his support. This type of voting scheme is also called *alternative voting* or *successive elimination*. The following is strong theoretic-

```

procedure STV;
begin;
  InitializeScores();
  quota :=  $\lceil \frac{|V|}{k+1} + 1 \rceil$ ;
  repeat
    if any candidate has score  $\geq$  quota then
      status of candidate := elected;
      ReallocateVotesOfWinner( candidate );
    else
      with candidate of lowest score do
        status := defeated;
        ReallocateVotesOfLoser( candidate );
      end;
    end;
  until total of elected candidates and viable candidates =  $k$ ;
  for each remaining viable candidate do
    his status := Elected;
  end;
end;

```

Figure 1: STV, written in pseudo-Pascal. This tallies votes to elect  $k$  candidates from a set  $C$  based on the preferences of a set  $V$  of voters. A viable candidate is one who has not yet been defeated or elected.

```

procedure InitializeScores();
begin;
  for each candidate do
    his score := 1.0;
    his status := viable;
  end;
  for each ballot do
    its weight := 1.0;
    add its weight to the score of the first viable candidate on ballot;
  end;
end;

```

Figure 2: Procedure to initialize scores for the STV algorithm

```

procedure ReallocateVotesOfLoser( candidate );
begin;
  for each ballot do
    if candidate is the first viable candidate on the ballot then
      Increase score of next viable candidate on ballot by weight of ballot;
    end;
  end;
end;

```

Figure 3: Procedure ReallocateVotesOfLoser

```

procedure ReallocateVotesOfWinner( candidate );
begin;
  excess votes for candidate := score of candidate – quota;
  for each ballot do
    if candidate is the first viable candidate on the ballot then
      Set weight of this ballot := (weight) (excess/score);
      Increase score of next viable candidate on ballot by weight of ballot;
    end;
  end;
end;

```

Figure 4: Procedure ReallocateVotesOfWinner for the “pure” form of STV

```

procedure ReallocateVotesOfWinner( candidate );
begin;
  excess votes for candidate := score of candidate – quota;
  for j := 1 to excess do
    Draw a ballot randomly from among those on which candidate is the
    first viable candidate;
    Increase score of next viable candidate on ballot by weight of ballot;
  end;
end;

```

Figure 5: Procedure ReallocateVotesOfWinner for the variant of STV that is used in Cambridge, Massachusetts

cal evidence that there is no dependably fast algorithm to solve EFFECTIVE PREFERENCE under Successive Elimination.

**Theorem 1.** *EFFECTIVE PREFERENCE is NP-complete under Successive Elimination.*

*Proof.* First observe that EFFECTIVE PREFERENCE under Successive Elimination is in NP, since the effectiveness of a preference can be verified in polynomial time by tallying the election and checking whether  $c$  wins.

Now we prove that EFFECTIVE PREFERENCE under Successive Elimination is “as hard as” the following problem that is known to be NP-complete [14].

**3-COVER**

GIVEN: A set  $S$  with  $|S| = n$ ; subsets  $S_1, \dots, S_m \subset S$ , with  $|S_i| = 3$  for  $i = 1, \dots, m$ .

QUESTION: Does there exist an index set  $I$  with  $|I| = n/3$  and  $\bigcup_{i \in I} S_i = S$ ?

Now let  $\{S_1, \dots, S_m\}$  denote an instance of 3-COVER. We will create an STV election for which there exists an effective preference if and only if there exists a 3-cover for  $S$ . Our contrived election will be to select a single winner from among  $5m + n + 3$  candidates. In this election the candidates fall naturally into the following five groups.

1. “possible winners”  $c$  and  $w$ ;
2. “first losers”  $a_1, \dots, a_m$  and  $\bar{a}_1, \dots, \bar{a}_m$ ;
3. “second line”  $b_1, \dots, b_m$  and  $\bar{b}_1, \dots, \bar{b}_m$ ;
4. “ $w$ -bloc”  $d_0, d_1, \dots, d_n$ ;
5. “garbage collectors”  $g_1, \dots, g_m$ ;

The election we construct will have the following properties.

**Property 1:** Of the first  $3m$  candidates to be eliminated,  $2m$  of them are  $a_1, \dots, a_m$  and  $\bar{a}_1, \dots, \bar{a}_m$ .

**Property 2:** Let  $I = \{i : \bar{a}_i \text{ is eliminated prior to } a_i\}$ . Then  $c$  will win the election if and only if  $I$  is a 3-cover.

**Property 3:** For any specified index set  $I \subseteq \{1, \dots, m\}$ , there exists a preference which, when tallied with all the other preferences, guarantees that  $\bar{a}_i$  is eliminated prior to  $a_i$  if and only if  $i \in I$ . Furthermore, such a preference can be constructed by allocating the first  $m$  positions on the preference list as follows: if  $i \in I$  then place  $a_i$  in the  $i$ th position; otherwise place  $\bar{a}_i$  in the  $i$ th position.

Properties 1–3 together imply that a strategic voter can guarantee victory for  $c$  in this election if and only if he can solve the 3-cover problem.

Now we describe the details of our contrived election. The voters are as follows. (Wherever we leave preferences unspecified, they can be arbitrary.)

- $12m$  voters with preference  $(c, \dots)$ ;
- $12m - 1$  voters with preference  $(w, c, \dots)$ ;
- $10m + (2n/3)$  voters with preference  $(d_0, w, c, \dots)$ ;
- For each  $i = 1, \dots, n$ ,  $12m - 2$  voters with preference  $(d_i, w, c, \dots)$ ;
- For each  $i = 1, \dots, m$ ,  $12m$  voters with preference  $(g_i, w, c, \dots)$ ;
- For each  $i = 1, \dots, m$ :  $6m + 4i - 5$  voters with preference  $(b_i, \bar{b}_i, w, c, \dots)$  and, for each of the three  $j$  such that  $j \in S_i$ , 2 voters with preference  $(b_i, d_j, w, c, \dots)$ ;
- For each  $i = 1, \dots, m$ :  $6m + 4i - 1$  voters with preference  $(\bar{b}_i, b_i, w, c, \dots)$  and 2 voters with preference  $(\bar{b}_i, d_0, w, c, \dots)$ ;
- For each  $i = 1, \dots, m$ :  $6m + 4i - 3$  voters with preference  $(a_i, g_i, w, c, \dots)$ ; 1 voter with preference  $(a_i, b_i, w, c, \dots)$ ; and 2 voters with preferences  $(a_i, \bar{a}_i, w, c, \dots)$ ;

- For each  $i = 1, \dots, m$ :  $6m + 4i - 3$  voters with preference  $(\bar{a}_i, g_i, w, c, \dots)$ ; 1 voter with preference  $(\bar{a}_i, \bar{b}_i, w, c, \dots)$ ; and 2 voters with preferences  $(\bar{a}_i, a_i, w, c, \dots)$ ;

Now we consider the election.

*Votes for c and w.* Initially  $c$  has  $12m$  votes and  $w$  has  $12m - 1$  votes. Except for the  $12m$  initial supporters of  $c$ , all voters rank  $c$  immediately below  $w$ . Thus  $c$  cannot get another vote until  $w$  is eliminated. Moreover, if  $w$  gets another 2 votes, then  $w$  cannot be eliminated before  $c$ . Thus, to be effective, a strategic preference must ensure that  $w$  is eliminated from the election before  $c$ .

*The w-bloc.* Candidate  $d_0$  initially receives  $10m + 2n/3$  votes and each of the other  $d_i$  receives  $12m - 2$  votes. For every voter whose first choice is  $d_i$ , the second choice is  $w$ , so that if any  $d_i$  is eliminated prior to  $w$ , then  $w$  will obtain more than 2 additional votes, and  $c$  cannot win. Thus the strategic voter must ensure that each candidate  $d_i$  achieves a score of at least  $12m$ . We will contrive the remainder of the election so that this can be accomplished only if certain of the “second line” candidates are eliminated early and their votes reallocated to candidates in the w-bloc.

*The second line.* Candidate  $b_i$  initially receives  $6m + 4i + 1$  votes. If  $b_i$  is eliminated, his votes are reallocated as follows: for each  $j \in S_i$ , 2 votes go to  $d_j$  and the remainder go to  $\bar{b}_i$ . Similarly, candidate  $\bar{b}_i$  also receives  $6m + 4i + 1$  votes initially. If  $\bar{b}_i$  is eliminated, his votes are reallocated as follows: 2 votes go to  $d_0$  and the remainder go to  $b_i$ . The third choice of all initial supporters of  $b_i$  or  $\bar{b}_i$  is  $w$ .

Note that if  $b_i$  is eliminated before  $\bar{b}_i$ , then the revised score of  $\bar{b}_i$  will exceed  $12m$ . Similarly, if  $\bar{b}_i$  is eliminated before  $b_i$ , then the revised score of  $b_i$  will exceed  $12m$ . This means that at most one of  $b_i, \bar{b}_i$  can be eliminated before  $c$  or  $w$ .

*The First Losers.* The initial score of each candidate  $a_i, \bar{a}_i$  is  $6m + 4i$ , which is one less than  $b_i$  and  $\bar{b}_i$ . If  $a_i$  is eliminated, his votes are reallocated

as follows: 1 vote goes to  $b_i$ ; 2 votes go to  $\bar{a}_i$ ; and the remainder goes to  $g_i$ . Similarly, if  $\bar{a}_i$  is eliminated, his votes are reallocated as follows: 1 vote goes to  $\bar{b}_i$ ; 2 votes go to  $a_i$ ; and the remainder goes to  $g_i$ .

*The Garbage Collectors.* Each candidate  $g_i$  initially has  $12m$  votes.

Now we show how the sequence in which candidates are eliminated in the election can encode a 3-cover.

**Lemma 1.** *Exactly one of  $b_i, \bar{b}_i$  will be among the first  $3m$  candidates to be eliminated. Furthermore, candidate  $c$  will win the election if and only if the set  $J = \{j : b_j \text{ is among the first } 3m \text{ candidates to be eliminated}\}$  is the index of a set cover for the 3-cover problem.*

*Proof.* The first claim is true because the first  $3m$  candidates to be eliminated must be among the  $a_i, \bar{a}_i, b_i, \bar{b}_i$ ; and if one of  $b_i, \bar{b}_i$  is eliminated, then the other will have score exceeding  $12m$ .

To establish the second claim, suppose that  $J$  is the index of a set cover. Then  $|J| = n/3$ . Consider the election after the first  $3m$  candidates have been eliminated. Suppose  $i \in S_j$  and  $j \in J$ . Then  $b_j$  has been eliminated and  $d_i$  has received 2 of  $b_j$ 's votes. Thus the revised score of  $d_i$  is at least  $12m$ , which exceeds that of  $w$ . Since  $J$  is the index of a set cover, each  $d_i, i = 1, \dots, m$ , has a revised score of at least  $12m$ . Also,  $\bar{b}_j$  has been eliminated for  $j \notin J$ . Thus  $m - n/3$  of the  $b_j$ 's have been eliminated and the revised score of  $d_0$  is at least  $(10m + 2n/3) + 2(m - n/3) = 12m$ , which exceeds that of candidate  $w$ .

After  $3m$  candidates have been eliminated there remain only  $c, w$ , all the  $d_i$ , all the  $g_i, b_j$  for  $j \notin J, \bar{b}_j$  for  $j \in J$ , all of which excepting  $w$  have at least  $12m$  votes. Candidate  $w$ , with  $12m - 1$  votes, will be eliminated next. Candidate  $c$  will inherit his votes and his score will rise to at least  $24m - 1$ .

Next the  $d_i$  candidates are eliminated since each has only  $12m$  votes. The votes of each  $d_i$  are reallocated to  $c$ , whose score rises to at least  $(24m - 1) + (n + 1)(12m) = (12n + 36)m - 1$ .

Now no  $g_i$  has more than then  $24m + 8m = 32m$  votes. Similarly, no remaining  $b_i$  or  $\bar{b}_i$  has more than  $32m$  votes. Since each  $g_i$ ,  $b_i$ , or  $\bar{b}_i$  that is eliminated will have his votes reallocated to  $c$ , it must be that  $c$  is the eventual winner of the election.

Conversely, suppose that  $J$  is not the index of a set cover; we will show that  $c$  cannot win the election. If  $|J| > n/3$ , then fewer than  $m - n/3$  of the  $\bar{b}_j$ 's have been eliminated, and the revised score of  $d_0$  cannot exceed  $12m - 2$ . Thus  $d_0$  will be eliminated prior to  $w$ , and  $w$  will receive his votes and achieve a higher score than  $c$ . In such case  $w$  cannot be eliminated before  $c$  and so  $c$  cannot win the election.

If  $|J| \leq n/3$  and  $i \in S$  is not covered by the corresponding set cover, then the score of  $d_i$  has not increased and  $d_i$  must be eliminated before  $w$ . Thus  $w$  will receive his votes and achieve a higher score than  $c$ . Again  $w$  cannot be eliminated before  $c$  and so  $c$  cannot win the election.  $\square$

Now we show how to construct a preference that will control which of the  $b_i$  are among the first  $3m$  candidates to be eliminated.

**Lemma 2.** *Let  $I \subseteq \{1, \dots, m\}$  and consider the strategic preference in which the  $i$ th candidate is  $a_i$  if  $i \in I$  and  $\bar{a}_i$  otherwise. Then the following is the order in which the first  $3m$  candidates are eliminated.*

- The  $3i - 2$  candidate to be eliminated is  $\bar{a}_i$  if  $i \in I$ , and  $a_i$  otherwise;
- The  $3i - 1$  candidate to be eliminated is  $b_i$  if  $i \in I$ , and  $\bar{b}_i$  otherwise;
- The  $3i$  candidate to be eliminated is  $a_i$  if  $i \in I$ , and  $\bar{a}_i$  otherwise;

*Proof.* (By induction) Assume that the first  $3i - 3$  candidates have been eliminated and that these candidates were, for each  $j < i$ , both  $a_j$  and  $\bar{a}_j$  and exactly one of each pair  $b_j, \bar{b}_j$ .

**Case 1**  $i \in I$ . The next viable candidate on the strategic preference is  $a_i$ , who therefore has  $6m + 4i + 1$  votes. Candidate  $\bar{a}_i$  has  $6m + 4i$  votes and is eliminated, after which the revised score of  $b_i$  is  $6m + 4i + 1$ ; the revised

score of  $\bar{b}_i$  is  $6m + 4i + 2$ ; and the revised score of  $a_i$  is  $6m + 4i + 3$ . Next, candidate  $b_i$  is eliminated, followed by  $a_i$ . Thus, by induction, the elimination of candidates is as claimed.

**Case 2**  $i \notin I$ . The next viable candidate on the strategic preference is  $\bar{a}_i$ . Candidate  $a_i$  must be eliminated next, after which the revised score of  $\bar{b}_i$  is  $6m + 4i + 1$ ; the revised score of  $b_i$  is  $6m + 4i + 2$ ; and the revised score of  $\bar{a}_i$  is  $6m + 4i + 3$ . Next, candidate  $\bar{b}_i$  is eliminated, followed by  $\bar{a}_i$ . By induction, the elimination of candidates is as stated in the theorem.

□

Now by Lemmas 1 and 2 candidate  $c$  will win the election if and only if  $I$  is a solution to the set cover problem. □

Now since the single elimination voting scheme is a special case of STV in all its variants, we have our main result.

**Corollary 1.** *EFFEKTIVE PREFERENCE under STV is NP-hard independently of how the excess votes of a winning candidate are reallocated among the remaining contenders.*

## 4 An alternative formalization

To a certain extent the formalization of a problem is a matter of taste. Different formalizations emphasize different aspects of the problem. For example EFFEKTIVE PREFERENCE emphasizes the problem of a manipulator with a specific goal: to make  $c$  the winner. The following alternative formalization emphasizes that the manipulator simply wants a strategy that is “better than” sincere behavior.

### PREFERRED OUTCOME

GIVEN: A set  $C$  of candidates; the set  $V$  of sincere, transitive, strict, and complete preferences of the voters; and the sincere preferences of the manipulator

QUESTION: Can the manipulator achieve a preferred outcome in the election by voting other than his true preferences?

The proof of Theorem 1 can easily be modified to show NP-completeness of PREFERRED OUTCOME for STV. We need only add to the contrived election that the true preferences of the manipulator are  $(c, w, \dots)$ , so that if he votes sincerely  $w$  will win. There is only one preferred outcome for the manipulator—that  $c$  win—and so our proof applies equally well to this problem.

PREFERRED OUTCOME may present our result more dramatically, but it does not generalize well. EFFECTIVE PREFERENCE, however, supports natural generalization to coalitions of strategic voters. For example, we can study the computational problem of coordinating votes by asking whether there is any set of preferences that could be claimed by the members of a coalition that would elect  $c$ . This is the topic of a subsequent paper.

## 5 Non-monotonicity

We say that a voting procedure is *monotonic* if a candidate can never change from a winner to a loser by gaining support (that is, by rising in the preferences of some of the voters, with all other candidates retaining their relative rankings in the preferences of the individual voters). Doron and Kronick have shown that STV is non-monotonic: there exist elections for which a candidate can change from an STV winner to a loser simply as a result of gaining support [9]. We show that it is NP-complete to recognize when non-monotonicity occurs in STV elections. Thus it can be impractical to exploit non-monotonicity for strategic purposes. To some extent this weakens any argument against STV that is based on non-monotonicity. Since non-monotonicity can be hard to recognize, voters are unlikely to be regaled with examples of its pernicious effects in real elections. Furthermore, even if voters know that non-monotonicity is possible, its distortions, since hidden, might be perceived as “random” and even, since unlikely to be exploited, benign. It might be seen by voters as similar in spirit

to the randomness occasionally designed into voting systems, as in [1]. Thus non-monotonicity might be more easily tolerated than flaws that are easy to find. When flaws are easy to find, then they are more likely to be exploited by manipulators and they are easier to display to an outraged electorate. Indeed, when flaws easy to find, manipulative behavior might be encouraged, while if flaws are hard to find, manipulative behavior might be uneconomic.

Doron and Kronick argue that any non-monotonic voting procedure is “perverse”. They vividly make their point by imagining an STV election in which it is observed that non-monotonicity affected the outcome.

Most voters would probably be alienated and outraged upon hearing the hypothetical (but theoretically possible) election night report: “Mr. O’Grady did not obtain a seat in today’s election, but if 5,000 of his supporters had voted for him in second place instead of first place, he would have won!”

However, because it is NP-complete to recognize an instance of non-monotonicity, voters would almost certainly *not* hear any such news on election night. In fact they would probably never hear it—not because the claim was false, but because it would simply take too long to discover.

We formalize the problem of recognizing non-monotonicity as follows.

**NON-MONOTONICITY**

GIVEN: A set  $C$  of candidates; the set  $V$  of sincere, transitive, strict, and complete preferences of the voters; and a distinguished candidate  $c \in C$  who would be a loser if the election was tallied now.

QUESTION: Is there a subset of voters in  $V$  who can change  $c$  from a loser to a winner by lowering  $c$  in their preferences?

**Theorem 2.** *NON-MONOTONICITY is NP-complete for Successive Elimination.*

*Proof.* Our proof is similar to that of Theorem 1 and so we omit details. Again

let  $\{S_1, \dots, S_m\}$  denote an instance of 3-COVER. Consider the following election (which is almost the same as in the proof to Theorem 1).

- $12m$  voters with preference  $(c, w, \dots)$ ;
- $12m - 1$  voters with preference  $(w, c, \dots)$ ;
- $12m$  voters with preference  $(w', w, c, \dots)$ ;
- $10m + (2n/3)$  voters with preference  $(d_0, w, c, \dots)$ ;
- For each  $i = 1, \dots, n$ ,  $12m - 2$  voters with preference  $(d_i, w, c, \dots)$ ;
- For each  $i = 1, \dots, m$ ,  $12m$  voters with preference  $(g_i, w, c, \dots)$ ;
- For each  $i = 1, \dots, m$ :  $6m$  voters with preference  $(b_i, \bar{b}_i, w, c, \dots)$  and, for each of the three  $j$  such that  $j \in S_i$ , 2 voters with preference  $(b_i, d_j, w, c, \dots)$ ;
- For each  $i = 1, \dots, m$ :  $6m$  voters with preference  $(\bar{b}_i, b_i, w, c, \dots)$  and 2 voters with preference  $(\bar{b}_i, d_0, w, c, \dots)$ ;
- For each  $i = 1, \dots, m$ : 1 voter with preference  $(c, b_i, \dots)$  and 1 voter with preference  $(c, \bar{b}_i, \dots)$ ;

Candidate  $c$  will lose this election since reallocated votes will always go to  $w$  before  $c$ . However the last group of voters, and only that group, can make  $c$  a winner by moving him down on their preferences. By lowering  $c$  in the appropriate preferences of the last group of voters,  $w$  can be prevented from gaining any (reallocated) votes until all remaining candidates have at least  $12m$  votes; then  $w$  will be eliminated and  $c$  will win the election. The preferences in which  $c$  must be moved down are exactly those  $(c, b_i, \dots)$  and  $(c, \bar{b}_i, \dots)$  for which the  $\{i\} \subseteq \{1, \dots, m\}$  is a solution to the 3-cover problem.  $\square$

**Corollary 2.** *NON-MONOTONICITY under STV is NP-hard independently of how the excess votes of a winning candidate are reallocated among the remaining contenders.*

A byproduct of our analysis is to show that even a single voter in an STV election can change a candidate from winner to loser by moving him up in his preferences, which slightly strengthens the example of [9]. In the contrived election of Theorem 1, if the manipulative voter constructs an effective preference then his favored candidate  $c$  will win; but if the voter moves  $c$  to the top of his preferences then  $c$  will lose. Thus STV is apparently as sensitive as it can be to changes in preferences, since non-monotonicity can be realized by just a single voter.

Finally, we caution the reader that there does not seem to be a standard use of the term “monotonic”. Fishburn [11] agrees with our use of the term, but Doron and Kronick call the same idea “non-perversity”. Moulin [20] also agrees with our use of the term, but observes that the idea is sometimes called “positive responsiveness”. Elsewhere in the literature the term “monotonicity” is sometimes used to mean “strong positive association” or “Maskin monotonicity” [2], which refers to a similar but stronger property: that a winner remain a winner when gaining support even when the relative rankings of the other candidates are allowed to change in the preferences of each voter. As shown in [22, 27], this stronger sense of monotonicity is equivalent to strategy-proofness; and since we have shown that it is NP-complete to recognize opportunities to vote strategically, it follows that it is also NP-complete to recognize when an election violates Maskin monotonicity.

## 6 Conclusions

There are several ways to measure the vulnerability of a voting procedure to strategic voting (see, for example, [8, 24, 26]). We have suggested measuring the computational effort. This complements the established approach, which is to count the opportunities for manipulation. Both of these measures of manipulability are useful but imperfect. The primary weakness of the former is that it is a worst-case measure. A weakness of the latter is that it is based on some assumption about the distribution of societies and that assumption is difficult

to justify. Frequently it is chosen for convenience of analysis.

We think that even the weakest reading of our theorem makes three contributions. First, we have helped explain Chamberlin’s empirical observation that strategies to manipulate STV are complex and instance-specific [8], since such behavior is the hallmark of NP-complete problems. Also, our results show how a would-be manipulator is constrained to plot his strategy in an STV election: he cannot do significantly better than enumerative search for an effective preference. This is different from other standard voting procedures, for which a natural, “greedy” algorithm will either produce an effective preference or show that none exists [5]. Finally, our result refines the conclusions of impossibility theorems such as those of Gibbard, Satterthwaite, and Gärdenfors: All voting procedures have flaws, but for some it can be difficult to exploit those flaws.

The strongest interpretation of our result is that one cannot in practice expect to vote strategically in an STV election, even given perfect information. This probably overstates the case, for like some of the notions of fairness, computational complexity does not perfectly model our concerns. In particular, the caveats of [5] hold here, the most important of which we briefly summarize. Most immediately, NP-completeness is an asymptotic, worst-case measure of complexity. This means that, for large enough instances of EFFECTIVE PREFERENCE under STV, *some* are difficult. This might be insufficient to provide practical protection against strategic voting. Certainly for small enough elections one can simply generate each of the  $|C|!$  possible preference orders and check each one to see whether it elects the favored candidate. Whether elections with many candidates are in fact difficult to manipulate is an empirical question. Naturally this will depend on the structure of a “typical” instance. It is worth remarking that Chamberlin’s observations on the complexity of manipulating STV were based on small elections (3 candidates and electorates of size 21 and 1000); this suggests that the computational difficulty becomes palpable rather quickly as the size of the election increases.

We close with some additional comments on points that are frequently misunderstood. First, it is important to note that our assumptions of perfect infor-

mation and sincere voting by most of the voters are *intentionally* unrealistic—we are not describing voter behavior, but making a conservative assumption to show that even when the situation is unrealistically favorable it can still be too difficult to determine how to vote strategically. Furthermore, we have shown that it is difficult for even a single manipulator to determine how to vote. One expects it to be only harder, at least in an informal sense, for a group of people to coördinate their votes to achieve a preferred outcome.

A second possible misunderstanding is that one might be tempted to conclude, by the contrived nature of the election in the proof of Theorem 1, that only pathological instances are difficult. This would be overinterpreting the proof. The contrived nature of our election is an artifact of human proof-building: We needed only to show that there exist some instances that are as difficult as another hard problem; we constructed an instance that made our proof as simple as possible, but not necessarily representative. In short, our theorem does not say whether any particular election is difficult to manipulate.

We also point out that the issues of computational difficulty are only intensified when the computational device is a relatively slow one, such as a person with pencil and paper. For example, in [18] it was reported that the site of the 1996 Summer Olympics was chosen by successive elimination, with six cities voted on by the 88 members of the International Olympic Committee. Our results imply that, even with perfect information, a strategic voter would have had to search for an effective preference over  $6! = 840$  possibilities. Even if 90% of the possible preferences could have been eliminated *a priori*, the prospect of evaluating the remaining 84 elections by hand might have been sufficiently discouraging to prevent strategic voting.

Finally, we speculate on some broader suggestions of our results. For STV elections we have established the formal difficulty of recognizing when strategic voting is possible and of recognizing instances of non-monotonicity. Most immediately, this suggests that STV might rise somewhat in our estimation, since two of its weaknesses—susceptibility to strategic voting and non-monotonicity—are perhaps less threatening than previously thought.

More generally, we have illustrated that even when voting procedures can fail to satisfy fairness criteria, nevertheless it can be hard to verify when this occurs. This might be interpreted as offering an explanation for an observation made by Nurmi and others of the “...the differences in difficulty of devising counterexamples showing the violations of criteria by various procedures. For some procedures and criteria the productions of a counterexample may be rather obvious while for other procedures it may be exceedingly difficult” [25].

## Acknowledgements

The authors thank Jeffrey Banks, I. D. Hill, Craig Tovey, and an anonymous referee for helpful comments.

J. Bartholdi was supported in part by a Presidential Young Investigator Award from the National Science Foundation (ECS-8351313) and by the Office of Naval Research (N00014-85-K-0147). J. Orlin was supported in part by a Presidential Young Investigator Award from the National Science Foundation (ECS-8451517).

## References

- [1] ANONYMOUS (1989). *Proportional Representation: Ballot Counting, Tally Sheets, and Background Information*, prepared for the municipal election, Cambridge, Massachusetts, November 1989, FLM Cambridge, Research Division. Available from the Election Commission of the City of Cambridge, 362 Green Street, Cambridge, Massachusetts 02139.
- [2] AUSTEN-SMITH, D. AND J. BANKS (1990). “Monotonicity in electoral systems”, manuscript, Department of Economics, University of Rochester, Rochester, NY, to appear in *The American Political Science Review*

- [3] BARTHOLDI, J. J. III, C. A. TOVEY, AND L. NARASIMHAN (1990). “Recognizing majority-rule equilibrium in spatial voting games”, to appear in *Social Choice and Welfare*.
- [4] BARTHOLDI, J. J. III, C. A. TOVEY, AND M. A. TRICK (1989). “Voting schemes for which it can be difficult to tell who won the election”, *Social Choice and Welfare* 6:157–165.
- [5] BARTHOLDI, J. J. III, C. A. TOVEY, AND M. A. TRICK (1989). “The computational difficulty of manipulating an election”, *Social Choice and Welfare* 6:227–241.
- [6] BRAMS, S. J. (1982). “The AMS nominating system is vulnerable to truncation of preferences”, *Notices of the American Mathematical Society* 29:136–138.
- [7] BRAMS, S. J. AND P. C. FISHBURN (1990). “Alternative voting systems”, preprint from the Department of Politics, New York University, New York, NY 10003; to appear in the *Encyclopedia of American Political Parties and Elections*, ed. L. S. Maisel, Garland Publishing, New York, to appear 1991.
- [8] CHAMBERLIN, J. R. (1985). “An investigation into the relative manipulability of four voting systems”, *Behavioral Science* 30:195–203.
- [9] DORON, G. AND R. KRONICK (1977). “Single transferable vote: an example of a perverse social choice function”, *American Journal of Political Science*, XXI(2):303–311.
- [10] DUMMETT, M. (1984). *Voting Procedures*, Clarendon Press, Oxford.
- [11] FISHBURN, P. C. (1982). “Monotonicity paradoxes in the theory of elections”, *Discrete Applied Mathematics* 4:119–134.
- [12] FISHBURN, P. C. AND S. J. BRAMS (1983). “Paradoxes of preferential voting”, *Mathematics Magazine* 56(4):207–214.

- [13] GÄRDENFORS, P. (1976). “Manipulation of social choice functions”, *Journal of Economic Theory* 13:217–228.
- [14] GAREY, M. R. AND D. S. JOHNSON (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman and Co., San Francisco.
- [15] GIBBARD, A. (1973). “Manipulation of voting schemes”, *Econometrica* 41:587–601.
- [16] HILL, I. D., B. A. WICHMANN, AND D. R. WOODALL (1987). “Single transferable vote by Meek’s method”, *The Computer Journal* 30(3):277–281.
- [17] HOLZMAN, R. (1989). “To vote or not to vote: what is the quota?”, *Discrete Applied Mathematics* 22:133–141.
- [18] JANOFSKY, M. (1990). “No clear favorite in search for site”, *The New York Times*, Sports Section, September 16, 1990.
- [19] MILL, J. S. (1861). *Considerations on Representative Government*, 1962 reprinting by Henry Regnery Company, Chicago of the original edition published in London by Parker, Son, and Bourn in 1861.
- [20] MOULIN, H. (1988). *Axioms of Cooperative Decision Making*, Cambridge University Press.
- [21] MOULIN, H. (1988). “Condorcet’s principle implies the no-show paradox”, *Journal of Economic Theory* 45:53–64.
- [22] MULLER, E. AND M. A. SATTERTHWAITTE (1977). “The equivalence of strong positive associate and strategy-proofness”, *Journal of Economic Theory* 14:412–418.
- [23] NEWLAND, R. A. (1972). *Only Half a Democracy*, London: The Electoral Reform Society of Great Britain and Ireland.

- [24] NITZAN, S. (1985). “The vulnerability of point-voting schemes to preference variation and strategic manipulation”, *Public Choice* 47:349–370.
- [25] NURMI, H. (1990). “Probability models in constitutional choice”, *European Journal of Political Economy* 6:107–117.
- [26] SAARI, D. G. (1990). “Susceptibility to manipulation”, *Public Choice* 64:21–41.
- [27] SAIJO, T. (1987). “On constant Maskin monotonic social choice functions”, *Journal of Economic Theory* 42:382–386.
- [28] SATTERTHWAITTE, M. A. (1975). “Strategy-proofness and Arrow’s conditions”, *Journal of Economic Theory* 10:187–217.