Manipulation of Second-Order Copeland Elections: Heuristic and Experiment

Ramoni O. Lasisi
Department of Computer and Information Sciences
Virginia Military Institute, USA
LasisiRO@vmi.edu

1 Introduction
Preference aggregation is used in a variety of applications, including artificial intelligence and multiagent systems - to ease group decision-making when agents are faced with a number of alternatives to make a choice. Consider the following. Ephrati and Rosenschein (1997) use virtual elections for preference aggregation in multiagent systems planning where agents vote on the next step of a plan under consideration. Also, according to (Bartholdi, Tovey, and Trick 1989), The Federation Internationale Des Echecs and the United States Chess Federation (USCF) implement tie-breaking rules that are either identical to, or are minor variants of, the second-order Copeland scheme to determine winners in competitions.

Voting protocols, such as the Copeland and second-order Copeland schemes are appropriate candidates, among others, for modeling such preference aggregation. Bartholdi, Tovey, and Trick (1989) define a voting scheme as an algorithm that takes as input a set \( C \) of candidates and a set \( V \) of preference orders that are strict (irreflexive and antisymmetric), transitive, and complete on \( C \). The algorithm outputs a subset of \( C \), who are the winners (allowing for ties). The ideal of a society is that a candidate emerging as a winner in an election be as widely and socially acceptable as possible.

Strategic manipulation of elections by agents remains a bane of voting protocols. Thus, the inability to limit or understand the effects of this menace may undermine the confidence agents have in decisions made via such protocols. The famous Gibbard-Satterthwaite theorem states: Every voting scheme with at least three outcomes is either dictatorial or manipulable (Gibbard 1973; Satterthwaite 1975). This implies that in any non-dictatorial voting protocol with at least three candidates, there exist some preferences of the voters such that some voters achieve better outcomes voting strategically i.e., not truthfully representing their preferences.

While the Copeland voting protocol can be efficiently manipulated in polynomial time, the second-order Copeland voting scheme is NP-complete to manipulate even if a manipulator has free and complete information about the preferences of other voters in an election (Bartholdi, Tovey, and Trick 1989). Although this complexity result is daunting to deter a would-be strategic voter from manipulation, NP-completeness is a worst case measure. And only shows that at least one instance of the problem requires such complexity (Conitzer and Sandholm 2006). Thus, real-life instances of an election that we care about may be easy to manipulate. Bartholdi, Tovey, and Trick’s hardness result rests on an assumption that the number of candidates in elections is unbounded. Contrary to this, Conitzer and Sandholm (2006) show that such hardness results lose relevance when the number of candidates is small, because manipulation algorithms that are exponential only in the number of candidates might be available. In line with this, and in an attempt to determine how few the candidates of elections can be for voting procedures to be hard, they show that at least four candidates are needed in any Copeland election for the manipulation to become hard in the second-order Copeland scheme (Conitzer, Sandholm, and Lang 2007).

We note that the problem of manipulation in elections and voting systems are pervasive in human societies and multiagent systems, and has received attention of many researchers in recent years. See (Mossel and Racz 2012; Faliszewski, Hemaspaandra, and Schnoor 2012; Hemaspaandra, Rothe, and Hemaspaandra 2013; Rothe and Schend 2013; Narodytska and Walsh 2014; Lasisi and Allan 2014).
The following is our working hypothesis: There are instances of second-order Copeland elections that may be efficiently manipulated using heuristics. We propose a branch-and-bound heuristic for manipulation of the second-order Copeland elections, and investigate the performance of the heuristic using randomly generated data. The data, which are the preference orders for each voter about candidates, were generated using three distributions, including uniform, normal, and Poisson. We provide empirical evaluation of the effects of manipulation of an election by a strategic voter for a constant number, \( 4 \leq k \leq 7 \), of candidates, and an unbounded number, \( v \in \{2000, \ldots, 10000\} \), of voters. Experiments suggest that there are instances of the elections that may be efficiently manipulated using the proposed heuristic when a voter has perfect information about the preferences of other voters. However, the second-order Copeland scheme becomes resistant to manipulation for fairly large number of candidates.

2 Copeland and Second-order Copeland Voting

The Copeland voting scheme (also known as the first-order Copeland method) is a protocol in which all candidates in an election engage in the same number of pairwise contests. A winner is a candidate that maximizes her Copeland score; the difference between her number of victories and defeats in all pairwise contests (Niemi and Riker 1976; Nurmi 1983). In the case of a tie, the eventual winner is the candidate whose defeated competitors have the largest sum of Copeland score. This tie-breaking rule is the second-order Copeland voting (Bartholdi, Tovey, and Trick 1989).

Let \( C = \{c_1, \ldots, c_n\} \) be a set of candidates in an election, with \( k \geq 4 \). Let \( V = \{v_1, \ldots, v_n\} \) be the preference orders of \( n \) voters over \( C \). We define a relation, \( \succ \), for each \( v \in V \) over \( C \). We say that a voter \( k \) ranks candidate \( i \) over candidate \( j \) denoted by, \( c_i \succ c_j \), if \( k \) prefers \( i \) to \( j \) in her preference order \( v_k \in V \). Since all candidates must engage in the same number of contests, the preference orders are required to be complete on \( C \). Thus, we assume that candidates that are not ranked in a voter’s preference order will be defeated competing with other candidates that are ranked. Finally, denote by \( C_S(c_i) \), the Copeland score of candidate \( c_i \in C \) in an election.

Example 1. Copeland Voting

Consider the following truthful preferences by seven agents for three candidates, \( a, b, \) and \( c \), in an election:

3 agents: \( a \succ c \succ b \)
2 agent: \( c \succ b \succ a \)
1 agent: \( c \succ a \succ b \)
1 agent: \( b \succ a \succ c \)

In the pairwise contests:

\[ a \text{ vs } b : 4 \text{ vs } 3 \Rightarrow a \text{ wins and } b \text{ loses} \]

\[ a \text{ vs } c : 4 \text{ vs } 3 \Rightarrow a \text{ wins and } c \text{ loses} \]

\[ b \text{ vs } c : 1 \text{ vs } 6 \Rightarrow b \text{ loses and } c \text{ wins} \]

The Copeland scores for the three candidates are:

\[ C_S(a) = 2 - 0 = 2 \]
\[ C_S(b) = 0 - 2 = -2 \]
\[ C_S(c) = 1 - 1 = 0 \]

Thus, candidate \( a \) is the overall winner in this election.

Example 2. Manipulation of Copeland Voting

We consider the same set of candidates and voters as in Example 1. However, one of the voters is using a different preference order. Let \( s \) be the strategic voter in this election with the preference \( b \succ a \succ c \). Let \( c \) be the candidate that \( s \) would like to manipulate her preference for. Suppose \( s \) employs the polynomial manipulation algorithm of Bartholdi, Tovey, and Trick, designed to place favored candidate at the top of strategic agent’s preference, to uncover a manipulative preference, \( c \succ a \succ b \), and then participates in the election. The new Copeland scores for the three candidates are:

\[ C_S(a) = 0 - 2 = -2 \]
\[ C_S(b) = 1 - 1 = 0 \]
\[ C_S(c) = 2 - 0 = 2 \]

Thus, candidate \( c \) is the overall winner in this election where the strategic agent \( s \) fails to report her preference order truthfully. Note that the preference orders of the remaining six agents remain the same as before.

Example 3. Second-order Copeland Voting

Consider the following preferences by seven agents for four candidates, \( a, b, c, \) and \( d \), in an election:

3 agents: \( a \succ b \succ c \succ d \)
2 agents: \( d \succ b \succ c \succ a \)
1 agent: \( d \succ b \succ a \succ c \)
1 agent: \( c \succ d \succ a \succ b \)

In the pairwise contests:

\[ a \text{ vs } b : 4 \text{ vs } 3 \Rightarrow a \text{ wins and } b \text{ loses} \]
\[ a \text{ vs } c : 4 \text{ vs } 3 \Rightarrow a \text{ wins and } c \text{ loses} \]
\[ a \text{ vs } d : 3 \text{ vs } 4 \Rightarrow a \text{ loses and } d \text{ wins} \]
\[ b \text{ vs } c : 6 \text{ vs } 1 \Rightarrow b \text{ wins and } c \text{ loses} \]
\[ b \text{ vs } d : 3 \text{ vs } 4 \Rightarrow b \text{ loses and } d \text{ wins} \]
\[ c \text{ vs } d : 4 \text{ vs } 3 \Rightarrow c \text{ wins and } d \text{ loses} \]

The Copeland scores for the four candidates are:

\[ C_S(a) = 2 - 1 = 1 \]
\[ C_S(b) = 1 - 2 = -1 \]
\[ C_S(c) = 1 - 2 = -1 \]
\[ C_S(d) = 2 - 1 = 1 \]
Thus, candidates $a$ and $d$ tied in this first-order Copeland voting. This tie is broken by computing the sum of the Copeland scores for the defeated competitors for both $a$ and $d$. The sum of the Copeland scores for the defeated competitors of $a$ is $C_S(b) + C_S(c) = -2$ and that of the defeated competitors of $d$ is $C_S(a) + C_S(b) = 0$. Thus, candidate $d$ emerge as the overall winner in this election after the application of the tie-breaking rule.

Winners in both the Copeland and second-order Copeland voting can be efficiently computed in polynomial-time. However, while a strategic voter can efficiently manipulate the Copeland voting in polynomial time, the second-order Copeland voting is NP-complete to manipulate (Bartholdi, Tovey, and Trick 1989). According to Bartholdi, Tovey, and Trick: Intuitively, it is difficult to construct a manipulative preference under Second-Order Copeland because it is difficult to know where to place candidates in the preference. For example, placing a favored candidate at the top can unintentionally improve the scores of rivals because of second order effects in the scoring.

3 Proposed Branch-and-Bound Heuristic

The previous section makes it clear that it is difficult for a strategic voter to determine a preference order that allows a distinguished candidate to win in a second-order Copeland scheme. This is because an attempt by a strategic voter to alter her truthful preference in anticipation of making a particular candidate win may improve the Copeland scores of some other candidates in the election. Hence, the manipulator is constrained to examine an exponential number of her preference orderings while being cognizance of other candidates’ Copeland and second-order Copeland scores in each of the orderings.

Heuristics are known to provide preferable solutions to instances of hard problems in most practical situations without having to examine all the possible choices. We propose the use of branch-and-bound technique in the design of our heuristic. The approach is to compute an estimated score of the current problem instance, called the bound. This bound is then used to determine how to branch or bypass (without examining of) large instances of the problem whose scores cannot be better than the bound found so far.

Let $C$ and $V$ be as defined in Section 2. Let $s$ be a strategic voter with preference order $v_s \in V$. Let $c_d \in C$ be a distinguished candidate that $s$ would like to manipulate her preference order for so that $c_d$ wins in an election. Suppose $m \leq k$ of the candidates, including $c_d$ tied under the Copeland voting protocol, then the difficulty is to manipulate the tie-breaking rule (under the second-order Copeland scheme) such that candidate $c_d$ wins. Hence, our interest is to find all such preference orderings $v_s$ of voter $s$ that elicits wins for $c_d$.

We assume that the preference orders of all the voters are fixed except that of $s$ i.e., $v_s$, and these other orderings are known to $s$. Altogether, there are \( \binom{k}{2} \) pairwise contests among the $k$ candidates and each of the candidate participates in exactly $k - 1$ contests. For every new preference order $v_s$ of $s$, dynamically compute the Copeland scores of the $k$ candidates in \( \left[ \frac{k - 1}{2} \right] \) contests and set the bound for the heuristic as the maximum score in the contests, denoted by \( \max c_b \). Let the Copeland score, $C_S(c_d)$ of $c_d$ in the \( \left[ \frac{k - 1}{2} \right] \) contests be $x$. We estimate the overall score of $c_d$ in the expected $k - 1$ contests as $C_{S,v_s}(c_d) = x + \left[ \frac{k - 1}{2} \right]$. Note that this score gives an estimate of the upper bound on the Copeland score that is attainable by $c_d$ since we have assumed that $c_d$ will win in all of the remaining \( \left[ \frac{k - 1}{2} \right] \) contests. If this estimated score is less than or equal to the bound value, i.e., $C_{S,v_s}(c_d) \leq \max c_b$, then bypass this and other similar preferences, since this is the best score attainable by $c_d$ in these $k - 1$ contests using this particular preference order. Otherwise, compute the Copeland scores of the $k$ candidates in the remaining \( \left[ \frac{k - 1}{2} \right] \) contests and determine the winner. If candidate $c_d$ wins, then we have found an instance of the second-order Copeland voting that is manipulable.

The salient point worthy of note in this heuristic is that it allows the distinguished candidate to circumvent election scenarios that lead to a tie with other candidates, thus ensuring a win in all situations where the heuristic reports a win.

4 Experiment and Analysis

This section provides a description of the experimental design for the conduct and manipulation of second-order Copeland elections using the proposed branch-and-bound heuristic of the previous section. Our decision on the choice of parameters in our experiments, i.e., the constant number, $4 \leq k \leq 7$, of candidates, and a voter’s perfect information about the preferences of other voters are influenced by earlier results from the literature as discussed in Section 1. This assumption of free and perfect information about voters’ preferences is a theoretical import that is used to show worst case situations. See for example, the introduction of (Bartholdi, Tovey, and Trick 1989).

4.1 Simulation

In all our simulations, we let $4 \leq k \leq 7$ be the number of candidates and $1000 \leq n \leq 10000$ be the number of voters. Thus, we define a set $C$ of candidates such that $|C| = k$ and a set $V$ of preference orders of voters, where $|V| = n$. We define a strategic voter $s$ with preference order $v_s \in V$ and a distinguished candidate $c_d \in C$ that $s$ would like to manipulate $v_s$ for in order to elicits a win for $c_d$.

We represent the set $V$ by a corresponding set $\Pi$ of permutations of the $k$ candidates in an election. Consider an election with three candidates, 1, 2, and 3. If a voter’s preference order in $V$ is, $2 \succ 1 \succ 3$, then the corresponding permutation $\pi \in \Pi$ of the preference order is given as 213. Since there are $k!$ possible permutations for each $k$, we identify a preference order with a unique number $i$ such that $1 \leq i \leq k!$. The initial preference orders assigned to all voters are randomly generated by assigning preference orders identified by corresponding numbers from $[1, k!]$. We also randomly select from $[1, n]$ and $[1, k]$, the strategic voter and the distinguished candidate respectively. The initial preference order assigned to a strategic voter is the identity permutation i.e., $1, 2, \ldots, k$. Observe that any preference order of
the strategic voter than the identity permutation that results in a win for the distinguished candidate is a manipulative preference order. Details of the different statistical distributions that are used to randomly generate preferences, determines strategic agents, and the distinguished candidates in elections are provided in Subsection 4.2.

When starting a new experiment, we perform the steps described above and freeze the preference orders of all the voters except that of the strategic voter. The strategic voter then examines her $k!$ possible preference orders in relation to other voters’ fixed preferences to determine preference orders that result in wins for the distinguished candidate. Recall that the strategic voter has complete information about other voters preference orders. The original election is the election where the manipulator starts with the identity preference order. The remaining $k! - 1$ elections are attempts by the strategic voter to determine potential manipulative preference orders.

We conduct elections among the $k$ candidates and use the branch-and-bound heuristic as described in Section 3 to engage in manipulation for strategic agents. We determine the percentage of the total elections that results in wins for the distinguished candidate. We repeat the experiment 100 times and compute the average value of the results to minimize deviations. For different runs of the experiments, we vary the number, $k$, of candidates and the number, $n$, of voters appropriately, within the ranges specified above while keeping one of the other two parameters constant alternatively.

### 4.2 Distributions to Generate Preference Orders

Each possible preference order in an election with $k$ candidates has a unique identifier between 1 and $k!$. The integer number representing the unique identifier for each of the preference orders of voters are generated using uniform, normal, and Poisson distributions. We have used these distributions to model different perceptions of candidates by the electorates. We give descriptions of how the preference orders are generated using the distributions.

**Preference Orders Using Uniform Distribution** The preference orders for all voters in an election are chosen based on their corresponding unique identifiers and such that they are integers drawn from a uniform distribution, $U(0, k! + 1)$. $U(a, b)$ defines a uniform distribution over the interval $(a, b)$, where both $a$ and $b$ are finite.

**Preference Orders Using Normal Distribution** The preference orders for all voters in an election are chosen based on their corresponding unique identifiers and such that they are integers drawn from a normal distribution, $N(\mu, \sigma^2)$, where $\mu$ and $\sigma^2$ are the mean and variance. We use a mean of $\mu = \frac{k!}{2}$ and values of standard deviation $\sigma = \frac{k!}{2}$ and $\sigma = \frac{3k!}{4}$, where $k$ is the number of candidates in an election.

**Preference Orders Using Poisson Distribution** The preference orders for all voters in an election are chosen based on their corresponding unique identifiers and such that they are integers drawn from a Poisson distribution, $Poisson(\lambda)$, where parameter $\lambda$ is the mean of the distribution. We use means of $\lambda = \frac{k}{2}$ and $\lambda = \frac{3k}{4}$, where $k$ is the number of candidates in an election.

We note that the parameters in these distributions are sufficient to see patterns of behaviors in our experiments and as well provide some generalization on the evaluation of the effects of manipulations in the second-order Copeland elections.

## 5 Results and Discussion

We present the results of our extensive set of simulations. For our study, we generate different numbers of candidates and voters with a distinguished candidate and a manipulator in each of the elections.

We show a summary of the extent of vulnerability of the second-order Copeland scheme to manipulation by a single voter in Figures 1 - 5, for elections where the preference orders were generated using uniform (Figure 1), normal (Figures 2 & 3), and Poisson (Figures 4 & 5) distributions. The $x$-axes in each of the figures indicate the number of voters in elections while the $y$-axes are the average percentage of manipulations achieved by strategic voters in the elections. We categorized the elections based on the number of candidates involved, so we have a 4-Candidate election for an election involving only four candidates, and so on. There are four types of elections shown in the figure.

![Figure 1: Extent of vulnerability of the second-order Copeland scheme to manipulation (uniform distribution).](image)

Figure 1 suggests that higher percentage of manipulations are achieved by the strategic voter when the number of candidates are few. In particular, consider the 4-Candidate election, the percentage of manipulations fluctuates between 14% and 22% of the elections for all cases of the number of voters. We can easily observe a trend in the figure as follows. As the number of candidates in the elections increases, the range between which the percentage of manipulations achieved by the voter oscillates also reduces. Consider the 7-Candidate election, this value fluctuates between 4% and 8% of the elections for all cases of the number of voters.

One would have expect that the percentage of manipulations achieved by the strategic voter be very small when the number of voters is large even for smaller size candidate elections like 4-Candidate, but this is not so. Our par-
preferences across all voters in the elections. Also, observe
noticed from these figures the effect of the spread of the
distribution; we however, is similar to that when the preference orders of the voters
are generated using the uniform distribution, we however,
tent of vulnerability of the second-order Copeland scheme
to manipulation when the preferences of the voters
number of candidates.

![Figure 2: Extent of vulnerability of the second-order Copeland scheme to manipulation (normal distribution with \( \mu = \frac{k_1}{2} \) and \( \sigma = \frac{k_1}{4} \)).](image2)

![Figure 3: Extent of vulnerability of the second-order Copeland scheme to manipulation (normal distribution with \( \mu = \frac{k_1}{2} \) and \( \sigma = \frac{k_1}{4} \)).](image3)

Figure 2: Extent of vulnerability of the second-order Copeland scheme to manipulation (normal distribution with \( \mu = \frac{k_1}{2} \) and \( \sigma = \frac{k_1}{4} \)).

Figure 3: Extent of vulnerability of the second-order Copeland scheme to manipulation (normal distribution with \( \mu = \frac{k_1}{2} \) and \( \sigma = \frac{k_1}{4} \)).

Figure 4: Extent of vulnerability of the second-order Copeland scheme to manipulation (Poisson distribution with \( \lambda = \frac{k_1}{2} \)).

Figure 5: Extent of vulnerability of the second-order Copeland scheme to manipulation (Poisson distribution with \( \lambda = \frac{k_1}{4} \)).

tial explanation for this unexpected oscillations of the per-
centage of manipulation even when the number of voters is
very large is the following. The more the number of can-
didates, the more independently distributed the preferences
of the voters are. So for the 4 and 5-Candidate elections,
the range of the distributions for the preference orders of
the voters are small, [1, 24] for 4-Candidate and [1, 120] for
5-Candidate elections. Hence, voters preference orders are
very close, making the distinguished candidate also popu-
lar among other voters, so the manipulator has better chance
to achieve more manipulations even for larger values of the
number of voters. On the other hand, the range of the dis-
tribution of the preference orders is large for fairly large
number of candidates.

Similar results of the extent of vulnerability of the second-
order Copeland scheme to manipulation when the prefe-
rence orders of the voters are generated using the normal
distribution are shown in Figures 2 and 3. Although the ex-
tent of vulnerability of the second-order Copeland scheme
is similar to that when the preference orders of the voters
were generated using the uniform distribution, we however,
noticed from these figures the effect of the spread of the
preferences across all voters in the elections. Also, observe
that unlike Figure 1, the percentages of manipulation for
the 7-Candidate elections are higher when the preferences
were generated using the normal distribution (Figures 2 and
3). These values fluctuate between 6% and 15%.

Finally, Figures 4 and 5 show the extent of vulnerability of
the second-order Copeland scheme when preferences of the
voters are generated using the Poisson distribution for dif-
ferent values of the mean. These results differ significantly
from those of the normal and uniform distributions, in part,
because Poisson distribution is used to simulate balls and
bins distributions. Thus, most of the unique identifiers for
the permutations obtained using this distribution concentrate
more around a region. The percentages of manipulation for
the 7-Candidate elections, when \( \lambda = \frac{k_1}{4} \), fluctuates between
2% and 4%, representing the worst values in all three cases.

6 Conclusions and Future Work

Strategic manipulation of elections by agents remains a bane
of voting protocols. Thus, the inability to limit or under-
stand the effects of this menace may undermine the confi-
dence agents have in decisions made via such protocols. The
well-known second-order Copeland voting scheme, that
we consider in this paper, has been shown to be NP-
complete to manipulate. This complexity is daunting enough
to deter a strategic voter from manipulation, however, NP-
completeness is a worst case measure, and only shows that at least an instance of the problem requires such complexity. We provide empirical evidence to show that despite this complexity, finding beneficial manipulation is relatively easy in practice, at least for the elections used in this work.

We propose a branch-and-bound heuristic for manipulation of the second-order Copeland elections, and investigate the performance of the heuristic using randomly generated data. The data, which are the preference orders for each voter about candidates, were randomly generated using three distributions, including uniform, normal, and Poisson. We provide empirical evaluation of the effects of manipulation of an election by a strategic voter for a constant number, $4 \leq k \leq 7$, of candidates, and an unbounded number, $v \in \{2000, \ldots, 10000\}$, of voters. We also evaluate the extent of ease of vulnerability of the second-order Copeland scheme to manipulation using the heuristic.

The results of our extensive experiments suggest that there are instances of the elections that may be efficiently manipulated using the proposed heuristic when a voter has perfect information about the preferences of other voters, thus validating our earlier stated hypothesis that: **There are instances of second-order Copeland elections that may be efficiently manipulated using heuristics.** It is, however, not difficult to infer that second-order Copeland elections involving large number of candidates will be highly resistant to manipulation of the type considered in this work. This is because the extent of vulnerability of the scheme when the number of candidates is seven is lower on the average for the three distributions, especially when the number of voters is very large. In particular, strategic voters achieve less than 5% of manipulations in all elections for the Poisson distribution when the mean, $\mu = \frac{3k}{2}$.

Our argument here is that, if it is difficult for strategic agents to achieve high values of manipulation when they have access to perfect information about the preferences of other voters, then, there is a strong reason to believe that the manipulators will achieve little or nothing in reality where the preferences of other voters are rarely available and the number of candidates is large. At best, only partial information or probabilities of whether or not a candidate will defeat another candidate may be available to strategic agents.

There are several areas of ongoing research on this problem. Here are some directions for future work. An extension of the number of candidates beyond seven that was considered in this paper is planned. Also, based on the outcomes of our experiments, we conjecture that the distributions of other voters’ preferences are key to manipulation. We plan to characterize distributions of voters’ preferences in which manipulation is impossible. An obvious and interesting step after this is to then determine which distributions are efficiently manipulable. Another interesting view of this type of manipulation is to consider the effect of when a coalition of voters is the strategic manipulator for a distinguished candidate. Furthermore, we plan to consider situations where the strategic voter or coalition of strategic voters have imperfect information about the preferences of other voters in the election. Finally, we plan to seek real-life data sets that we can use to evaluate and improve the proposed heuristic.

Acknowledgement

This research is partially supported by the Virginia Military Institute’s Professional Travel Funds.

References


