Awesome Work in Computer Science

Russell Emerine

CMU-CS-25-129

August 2025

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Daniel Sleator, Chair
Fei Fang
Daniel Anderson

*Submitted in partial fulfillment of the requirements
for the degree of Masters in Computer Science.*

Abstract

Much of the past work on voting systems focuses on *ranked voting systems*, which have a number of limitations such as Arrow's Theorem [1]. In this paper we consider ranked voting systems as well as the less commonly used class of *rated voting systems*. The systems differ in that ranked voting systems only allow the voter to order the candidates, while in rated voting systems the voter can score each candidate independently. In 2000, Warren Smith [3] evaluated ranked and rated voting systems under a Monte Carlo simulation model of voter utilities and behaviors. We replicate Smith's results with a wider selection of voting systems and voter utility distributions and conclude that range voting, a rated system where each candidate can be independently scored in the range [0, 1], has the best performance.

This project's source code is freely available online at `https://github.com/RussellEmerine/voting_simulation`.

iv

Acknowledgments

My advisor is cool.

July 28, 2025
DRAFT

Contents

July 28, 2025

DRAFT

List of Figures

July 28, 2025
DRAFT

List of Tables

xii

Chapter 1

Voting Systems

A *voting system* is a system where a (usually large) number of *voters* cast *votes* of some kind, and the votes are used in some process to select one of a (usually small, but greater than one) number of *candidates*.

Some formulations allow the output to be a set of candidates or an ordering of candidates. For simplicity, we will only consider systems that output a single winner. We will also focus on cases with three or more candidates, as with two candidates, all reasonable voting systems exhibit reasonable behavior.

We first describe several types of voting systems, then list and describe the particular voting systems we consider in our simulation.

1.1 Ranked and Rated Voting Systems

1.1.1 Ranked Voting Systems

In a ranked voting system, vote information can be encoded into an ordering of the candidates. For instance, plurality voting uses the most favored candidate from each ordering, while Borda voting uses the whole ordering. Ranked voting systems traditionally output an ordering of candidates (which in this section we refer to as the "outcome"); to choose our single winner we can simply choose the most favored candidate in the outcome.

Arrow's theorem states that, if there are at least three candidates, no ranked voting system may satisfy all of the following properties [1]:

- Pareto efficiency: If every voter places A before B, then the outcome also places A before B. (This can be weakened to non-imposition, i.e. that for any A and B, there is some cast of votes such that the output places A before B [4].)

- Non-dictatorship: There is no voter whose ordering is always the same as the outcome.

- Independence of irrelevant alternatives: In two elections with the same number of voters where each pair of corresponding voters has the same relative ordering of A and B, the outcomes of both elections have the same relative ordering of A and B.

The fact that it is impossible to have all three of these very reasonable conditions is a severe limitation of ranked voting systems.

July 28, 2025

DRAFT

1.1.2 Rated Voting Systems

Votes need not be restricted to orderings of candidates. Rated voting systems allow a voter to provide independent scores for each candidate, which can be more informative than an ordering of the candidates. For instance, range voting allows a voter to give each candidate their own score in the range $[0, 1]$. thereby allowing a voter to express how much they prefer their most favored candidate over their second most favored candidate.

Since these are not ranked voting systems, Arrow's theorem does not apply, and it is completely possible to have Pareto efficiency, non-dictatorship, and independence of irrelevant alternatives. However, they are still subject to some restrictions, notably Gibbard's theorem, that no deterministic process of collective decision may satisfy all of the following properties [2]:

- The process has more than two possible outcomes.
- There is no voter who singlehandedly determines the outcome.
- The game-theoretically optimal vote for a voter will not depend on the voter's beliefs of what other voters will vote.

When applied to voting systems, this implies that non-dictatorial voting systems with three or more possible outcomes (ranked, rated, or otherwise) require some kind of strategic voting that is not completely honest.

> *An aside about determinism:*
> Most voting systems do not have a good way of handling ties without using randomness or allowing multiple winners. However, when the number of voters is reasonably large, the chance of a tie is negligible. We will consider voting processes that are deterministic when there are no ties "good enough."

1.1.3 Other Voting Systems

There are a few "obviously bad" voting systems that we consider, such as "random winner" and "worst candidate". These are only useful as a frame of reference, and are not expected to have any of the properties discussed for ranked and rated systems.

1.2 COAF and Non-COAF Systems

1.2.1 COAF Systems

Smith's specification of compact set based, one-vote, additive, fair voting systems describes many common ranked and rated voting systems [3]. In a COAF system with C candidates there is a compact set $S \subseteq \mathbb{R}^C$ of allowed votes, where S is symmetric across permutations of candidates (fair). Each voter chooses one vote in S to submit. Then, the votes are added, and the candidate with the greatest sum (or equivalently, average, referred to as "score" for the following proof) is selected as the winner. For instance, plurality voting is when $S = \{(1, 0, 0, \dots), (0, 1, 0, \dots), (0, 0, 1, \dots), \dots\}$, and range voting is when $S = [0, 1]^C$.

Smith provides a formal proof of the optimality of the "moving average" strategy. This strategy generates the game theoretically optimal vote in any COAF system when given poll data in the form of a predicted ordering of candidates by likelihood to win the election (assumed to be produced by a random sample of honest pollees). Consider a COAF system with vote set S and without loss of generality label the candidates as $c_1, c_2, \ldots c_C$ in poll order. Let $U_1, U_2, \ldots U_C$ be the utilities of the voter for each candidate, under the utility model discussed later. The vote is generated as follows:

- Let the set X_0 representing the set of potential votes start as $X = S$.
- If $U_1 > U_2$, let X_1 be the subset of X_0 that maximizes the 1st component. Otherwise, let X_1 to the subset of X_0 that minimizes the 1st component.
- For each candidate c_i in poll order starting from c_2, if $U_i > \sum_{j=1}^{i-1} U_j$, let X_i be the subset of X_{i-1} that maximizes the ith component. Otherwise, let X_i be the subset of X_{i-1} that minimizes the ith component.

The final set X_C will consist of exactly one vector, which will be the vote.

We present an alternate interpretation of Smith's proof.

Let us say there are C candidates, V votes, and P pollees, where V and P are reasonably large. Let us say that across all votes, the ith component of the vote vector lies in a distribution with mean μ_i and variance σ_i^2. μ_i is the actual score the candidate will receive. σ_i^2 has no direct effect on the election result but is useful for strategic analysis.

In this strategy, the only information voters have on the candidates is the polling data and their own personal associated utilities. To model this, we will assume a voter's belief of the distribution of a candidate c_i's score μ_i follows a normal distribution with the same variance as would be expected from the sample distribution, which by the central limit theorem is $\frac{\sigma_i^2}{P}$. These belief distributions have means we will call $x_1, x_2, \ldots x_C$ — the numerical values of these are not important, but they are known to be in the polling order, as shown in this plot of the believed distributions of μ_4, μ_3, μ_2, and μ_1:
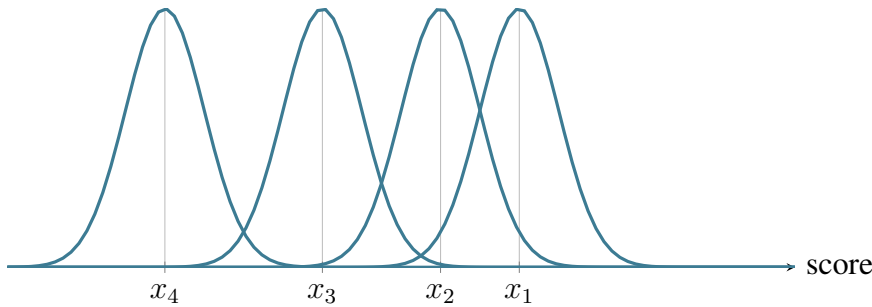
Since the P is large, we can assume the variance $\frac{\sigma_i^2}{P}$ is very small:
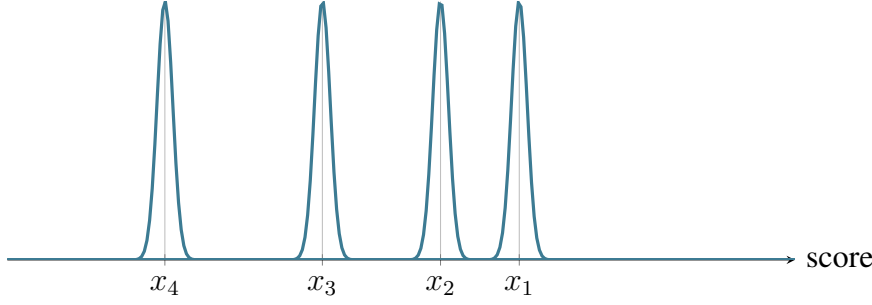
The voter's vote can only change the outcome of the election if the values of the actual two highest scores are a "near tie", within $\frac{1}{V}$ of each other. Otherwise, the vote cannot change the outcome of the election — this case can be considered a fixed component of the expected utility, and so can be ignored for reasoning about the optimal vote. The following reasoning will assume there is some near tie.

Consider some fixed μ_1 near x_1. This can be within, say, 1000 standard deviations. With large P, the standard deviation is small enough that this is nowhere close to x_2. Since the probability that μ_1 is near x_1 is large, its contribution dominates the expected utility, and we can ignore the case that μ_1 is not near x_1. Likewise, the probability that every other candidate c_i has $\mu_i \leq \mu_1 + \frac{1}{V}$ is large, and we can ignore the other case. Now, the only result-determining near ties that can happen are between c_1 and some other c_i. The chance of such a near tie is proportional to $\frac{1}{V} \exp\left(-\frac{P(x_1 - x_i)^2}{2\sigma^2}\right)$. Since P is large, this chance is much larger for c_2 than for any other c_i. Therefore, the voter must optimize the vote for the case that c_1 and c_2 have a near tie, and so will choose some vote that maximizes or minimizes the 1st and 2nd components according to whether the voter prefers c_1 over c_2. This determines the vote subsets X_1 and X_2.

After narrowing down the potential votes to X_2 by considering when c_1 is a member of the near tie, there may still be many possible votes to make. It is already a very small chance for the optimization to X_2 to affect the outcome of the election, and it is even less likely for any remaining optimization to do so — and so the restriction to X_2 is already "good enough" for most purposes (which will be useful for non-COAF systems). However, further optimizations can still be determined using similar reasoning.

1.2.2 Non-COAF Systems

Properties of non-COAF systems, and description of voter behavior.

1.3 A List of Voting Systems Considered

A list of voting systems we simulate.

Chapter 2

Utility Distributions

Discussion of utility distributions, including uniform vs. normal vs. multimodal, and random vs. issue-based.

July 28, 2025
DRAFT

Chapter 3

Simulation

Discussion of implementation of simulation.

Chapter 4

Analysis

Analysis of some plots.

Chapter 5

Conclusions

Conclusions of the project.

Chapter 6

Appendix A: Plots

All the plot images.

Chapter 7

Introduction

Chapter 8

Conclusion

July 28, 2025
DRAFT

Bibliography

[1] Kenneth J. Arrow. A difficulty in the concept of social welfare. *Journal of Political Economy*, 58(4):328–346, Aug 1950. doi: 10.1086/256963. (document), 1.1.1

[2] Allan Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41(4): 587–601, 1973. ISSN 00129682, 14680262. URL `http://www.jstor.org/stable/1914083`. 1.1.2

[3] Smith. Range voting. 2000. URL `https://rangevoting.org/WarrenSmithPages/homepage/rangevote.pdf`. (document), 1.2.1

[4] Robert Wilson. Social choice theory without the Pareto Principle. *Journal of Economic Theory*, 5(3):478–486, 1972. ISSN 0022-0531. doi: https://doi.org/10.1016/0022-0531(72)90051-8. URL `https://www.sciencedirect.com/science/article/pii/0022053172900518`. 1.1.1

July 28, 2025

DRAFT