Quantitative Analysis of Strategic Voting in Anonymous Voting Systems

Tiance Wang

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Electrical Engineering
Adviser: Sanjeev Kulkarni and Paul Cuff

January 2016
Abstract

Democratically choosing a single preference from three or more candidate options is not a straightforward matter. There are many competing ideas on how to aggregate rankings of candidates. However, the Gibbard-Satterthwaite theorem implies that no fair voting system (equality among voters and equality among candidates) is immune to strategic voting, also known as manipulation. This dissertation is a quantitative analysis of strategic voting from a geometric perspective.

Anonymous voting rules, where all voters are equal, can be viewed as a partition of a high dimensional simplex, where different distributions of votes correspond to different points in the simplex, and each particular way of partitioning the simplex corresponds to a voting rule. It is revealed that the orientation, instead of the location, of the boundary determines manipulability. A boundary that separates two winning candidates is not manipulable if and only if the boundary is parallel to all vote changes that does not switch the order of the candidate pair.

We analyze the vulnerability to strategic voting of several popular voting systems, including plurality, Borda count and Kemeny-Young, under various vote distributions. When there are three candidates, we show that the Kemeny-Young method, and Condorcet methods in general, are categorically more resistant to strategic voting than many other common voting systems, due to the existence of non-manipulable boundaries. We verify our results on voting data that we collected through an online survey on the 2012 US President Election.

Finally, we explore the collective behaviors of manipulative voters. Assume every voter can change their vote for an infinite number of times. They formulate strategies based on their observations on the preference of the population. The observations, which contain noise, are generated by some distribution conditioned on the current vote status. We show that the plurality rule almost always elects the instant run-
off winner, while Borda count almost always elects the Condorcet winner (when one exists) as the number of voters grow.
Acknowledgements

I am deeply indebted to my advisors Prof. Sanjeev Kulkarni and Prof. Paul Cuff for their constant guidance, wholehearted support and endless patience at every stage of my Ph.D. study. Being extraordinary researchers and great mentors, they are always ready to help and encourage me to overcome difficulties and frustrations in research. It is them who made this six-year intellectual expedition possible for me.

Prof. Sanjeev Kulkarni is a prolific scholar and a caring mentor. In addition to his research insights, I am also amazed by his meticulousness. He proofreads our write-ups with attention to fine details, in order to help us perfect our works. He offered me two summer internships at the Deutsche Telekom T-Lab in Germany. During the two internships, I gained exposure to different research environment and research methodologies, as well as living experience in a different culture.

Prof. Paul Cuff is a passionate, energetic and incredibly intelligent researcher and problem solver. His brilliant insights, mathematical elegance and clarity of presentation always brought about delightful progress to puzzling problems. He is also very generous in sharing his time with students, discussing all kinds of interesting problems.

I would like to express gratitude to Prof. Pan (Ben) Hui, who was my supervisor when I did two summer internships in Deutsche Telekom Innovation Labs in Berlin. His multitasking ability are fast-paced working style is very influential on me. Outside of work, he befriended his students and interns. Without him my stay in Berlin would not be as fruitful and fun.

I thank Prof. Peter Ramadge for being my thesis reader and providing valuable feedbacks. I thank Prof. Mung Chiang for serving as committee of my general exam. I would like to thank the faculty of the Department of Electrical Engineering, Department of Mathematics and Department of Operational Research and Financial Engineering for their teaching, including Prof. Sergio Verdú, Prof. Robert Calder-
bank, Prof. Peter Ramadge, Prof. Yakov Sinai, Prof. Patrick Cheridito and Prof. Ramon van Handel for their teaching.

Special thank goes to Shang (Stella) Shang. Both of us are co-advised by Prof. Kulkarni and Prof. Cuff. During the five years she was here, we went to research meetings together, took courses together, and also had great teamwork experience collaborating in a research project. Her optimistic spirit, proactive working attitude and love for life has always inspired me. I am especially thankful of Shucheng (Charlie) Zhu. We were friends since undergraduate and came to Princeton as graduates together. He is a fun and a generous person, and a supportive and caring friend. He offered me tremendous help in my job search. I am very fortunate to have a friend like you. I am honored to have many friends who made my six years in Princeton an unforgettable experience: Yue Tian, Tao Han, Aoxiang Tang, Yingzhe Hu, Junchao Xia, Jingwan Lu, Eve Wang, Tracy Ke, Yunlai Zha, Mengjie Wang, Yihong Wu, James Xiang, Liechao Huang, Pingmei Xu and Mia Chen. I would like to thank everyone who together with me worked for our advisors: Haipeng Zheng, Jieqi Yu, Arvid Wang, Curt Schieler, Eva Song, Zhuo Zhang, Sai Satpathy and Jingbo Liu.

Finally, I would like to express my deepest gratitude to my parents, for their unconditional love. They offered my guidance and support in each step of my life, and taught me the value of kindness, enterprise, honor and integrity. I would not be where I am today without their love and sacrifice. To them I dedicate this dissertation.
To my parents.
## Contents

Abstract .................................................. iii  
Acknowledgements ...................................... v  
List of Tables ........................................... xi  
List of Figures ........................................... xii  

1 Introduction ............................................. 1  
1.1 Background ........................................... 1  
1.2 Properties of a voting system ................. 3  
  1.2.1 Anonymity and Dictatorship .................. 3  
  1.2.2 Neutrality ....................................... 4  
  1.2.3 Resoluteness .................................... 4  
  1.2.4 Deterministic vs Stochastic .................. 4  
1.3 Strategic voting ..................................... 4  

2 Common voting systems .............................. 8  
2.1 Scoring rules ....................................... 8  
  2.1.1 Plurality ......................................... 9  
  2.1.2 Anti-plurality ................................... 9  
  2.1.3 Borda count .................................... 10  
2.2 Condorcet methods .................................. 10  
2.3 Generalized scoring rules ....................... 12
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Example: Plurality vote</td>
<td>9</td>
</tr>
<tr>
<td>2.2</td>
<td>Instant run-off, second round</td>
<td>14</td>
</tr>
<tr>
<td>2.3</td>
<td>Instant run-off, third round</td>
<td>14</td>
</tr>
<tr>
<td>3.1</td>
<td>Vote changes that constitute manipulation</td>
<td>19</td>
</tr>
<tr>
<td>3.2</td>
<td>Approximate Number of Manipulable and Pivotal Profiles for Several Voting Method</td>
<td>43</td>
</tr>
<tr>
<td>4.1</td>
<td>Strategic voting table for Borda count. Grey cells are the strategic voters</td>
<td>48</td>
</tr>
<tr>
<td>4.2</td>
<td>Strategic voting table for plurality voting</td>
<td>49</td>
</tr>
<tr>
<td>4.3</td>
<td>An example of Kemeny-Young</td>
<td>50</td>
</tr>
<tr>
<td>4.4</td>
<td>Pivotal and Strategic voters for Kemeny-Young</td>
<td>50</td>
</tr>
<tr>
<td>4.5</td>
<td>Conditional incentive for Borda count with $N = 10,000$ voters</td>
<td>61</td>
</tr>
<tr>
<td>4.6</td>
<td>Conditional incentive for plurality voting with $N = 10000$ voters</td>
<td>61</td>
</tr>
<tr>
<td>4.7</td>
<td>Conditional incentive for Kemeny-Young with $N = 10000$ voters</td>
<td>63</td>
</tr>
<tr>
<td>4.8</td>
<td>Voting outcome of different voting systems based on survey data based</td>
<td>67</td>
</tr>
<tr>
<td>5.1</td>
<td>The order of rankings throughout the paper</td>
<td>74</td>
</tr>
<tr>
<td>5.2</td>
<td>Optimal strategy for plurality</td>
<td>80</td>
</tr>
<tr>
<td>5.3</td>
<td>Optimal strategy for Borda count</td>
<td>90</td>
</tr>
<tr>
<td>5.4</td>
<td>Non-zero entries in the limit distribution of $V^T$</td>
<td>94</td>
</tr>
</tbody>
</table>
List of Figures

3.1 Figure for Lemma 3.8 ............................................................... 23
3.2 Figure for Lemma 3.9 for $f(v_1) = \{b, K\}$ .......................... 25
3.3 Figure for Lemma 3.10 for the case $a \in K_1 \cap K_2$ ............... 26
3.4 Pairwise comparison space $U$ ................................................. 37
3.5 Profile space not fully partitioned by non-manipulable boundaries . 39
3.6 Winning region of $a$ when $\beta_1 = \beta_2 = 1$, $\beta_3 = 1.5$ .......... 40
3.7 Boundaries of Borda count, shown in pairwise comparison space $U$ . 41
3.8 Counting manipulable profiles for $B(a, b|f_{Borda})$ .................. 42
4.1 How Borda count partitions $U$ .................................................. 47
4.2 Strategic and non-manipulable boundaries of the Kemeny-Young method 51
4.3 Score profiles with pivotal and strategic voters for Borda count ........ 55
4.4 The outcome of vote changes by two voter types, one with strategic incentive (bac) and one without (abc), in Borda count. ................. 55
4.5 Hyperplanes parallel to $S_{ab}^B = S_{ba}^B$ ..................................... 60
4.6 The information projection falls on the intersection of $S_{abc}^K = S_{bac}^K$ and $S_{acb}^K = S_{bac}^K$ .............................................................. 62
4.7 Condorcet paradox never occurs for our collected voting data .......... 67
4.8 Kemeny-Young is mostly immune to strategic voting for our collected voting data ................................................................. 68
4.9 Election survey data for Borda count and plurality ...................... 69
5.1 Observation $p^f$ and its projection to the plurality boundary
Chapter 1

Introduction

1.1 Background

Social choice theory is, according to William H. Riker, “the description and analysis of the way that the preferences of individual members of a group are amalgamated into a decision of the group as a whole” [35].

When there are only two items to choose from, merging the preferences of members is a straightforward process. If we allow the members to express their opinions in the form of a binary vote, indicating which item they prefer, then we pick the candidate with more support. However, when there are three or more candidates, there is not an obvious way of merging conflicting opinions.

We start by inspecting what information is conveyed by the ballots. Ballots can be broadly divided into two categories: ranked ballots and range ballots. For ranked ballots, voters rank candidates in a hierarchy on ordinal scale. The rank implies, in particular, which candidate is preferred out of every pair. In some cases indifference is allowed. Rankings are transitive: if a voter prefers $a$ no less than $b$ and prefers $b$ no less than $c$, then she must prefer $a$ no less than $c$. Range ballots associate each candidate with a numerical rating within a certain range. Voting methods that use the first
type of ballots are collectively known as ranked voting methods. Although defined on very different domains, range ballots can be easily mapped to ranked ballots, by sorting candidates according to their scores. In this dissertation, we focus on ranked voting systems that use complete rankings as ballots. We will briefly discuss range voting, a typical voting system based on ranged ballots in Chapter 4.

We use $C$ to denote the set of candidates, use $N$ to denote the number of voters, and use $L$ to denote the set of complete rankings, or all permutations over $C$. Thus $|L| = |C|!$. The ordered collection of all votes is known as the voting profile or simply profile. The set of all profiles is denoted by $L^N$.

There has been some inconsistency in the literature on the use of terminology such as voting rule, social choice function, social welfare function and so on. Our definitions essentially follow [40].

**Definition 1.1.** Given the set of all profiles $L^N$ and the set of candidates $C$, where each element of $L^N$ is an ordered collection of $N$ permutations of the elements in $C$,

1. A function $f : L^N \mapsto 2^C \setminus \emptyset$ is called a voting rule.
2. A function $f : L^N \mapsto 2^L \setminus \emptyset$ is called a social welfare function.
3. A function $f : L^N \times 2^C \mapsto 2^C$ is called a social choice function if $f(v, S) \subseteq S$ for any $S \subseteq C$.

A voting rule is applied when we are only concerned with “who’s the best”. We use $2^C \setminus \emptyset$ instead of $C$ to allow ties. The social welfare function is used when we want to know how good each candidate is compared to others. A social choice function allows “local” elections over a subset of candidates (as if all other candidates dropped out).

We reserve the term “voting system” as a general term that incorporate all of the aggregation functions above.
1.2 Properties of a voting system

1.2.1 Anonymity and Dictatorship

A voting rule is *anonymous* if it treats all voters equally. That is, the voting outcome does not change even if the votes are reshuffled. Formally, $f$ is anonymous if for any $v \in \mathcal{L}^N$ and any permutation $\sigma$ of $N$ objects,

$$f(v_1, \ldots, v_N) = f(\sigma(v_1), \ldots, \sigma(v_N)) \quad (1.1)$$

The anonymity condition guarantees fairness among voters in the sense that “my vote has the same impact as yours.” Since the identities of voters are eliminated, we can represent the profile simply with the vote count for each ranking. Denote the $i$th preference order (or type-$i$ vote) $L_i \in \mathcal{L}$ by a basis vector $e_i \in \mathbb{R}^{K!}$, whose $i$th entry is 1 and all other entries are 0:

$$e_i = (0, \ldots, 0, 1, 0, \ldots, 0) \quad (1.2)$$

**Definition 1.2** (Anonymous Profile). *An anonymous vote profile is represented by a vector: $v = \sum_i N_i e_i = (N_1, N_2, \ldots, N_{K!}) \in \mathcal{V}_N$, where

$$\mathcal{V}_N = \{ v \in \mathbb{Z}^{K!} : \sum_{i=1}^{K!} v_i = N, v_i \geq 0 \} \quad (1.3)$$

We call $\mathcal{V}_N$ the anonymous profile space.

A *dictatorship*, in some sense, is the exact opposite of anonymous voting system. A voting rule is a dictatorship if there exists one voter who controls the voting outcome while all other voters have no influence at all. That voter is called the dictator.
1.2.2 Neutrality

A voting rule \( f \) is neutral if it treats candidates equally. More formally, it commutes with permutations on \( C \):

\[
f(\tau \circ v_1, \ldots, \tau \circ v_N) = \tau \circ f(v_1, \ldots, v_N)
\]

for all \((v_1, \ldots, v_N) \in V\) and all \(\tau \in \mathcal{L}\). If the output contains several candidates, then the permutation \(\tau\) is applied to each one of them. Intuitively, a neutral voting system is not biased for or against any candidate.

1.2.3 Resoluteness

A voting rule \( f \) is resolute if it always outputs a single winner. An irresolute voting system, in contrast, sometimes outputs a set of more than one candidate. The outputs are considered as tied winners.

1.2.4 Deterministic vs Stochastic

A voting system is stochastic if the output is a probability distribution over candidates, rather than a fixed candidate or set of candidates. A voting system is deterministic if it always outputs a fixed candidate or set of candidates.

1.3 Strategic voting

A voter can sometimes achieve a more desirable election result by casting a ballot that does not reflect her honest preference. This phenomenon is known as strategic voting or manipulation of the voting system. We refer to such a voter as a strategic voter or manipulator. Assume each voter holds a sincere preference in the form of a complete ranking of candidates. Manipulation is the scenario in which a voter
reports a ranking different from her sincere preference, thus changing the outcome. In principle, strategic voting is something that designers of voting systems would like to avoid. According to C. L. Dodgson (better known as Lewis Carroll), it would be “better for elections to be decided according to the wishes of the majority than of those who happen to have most skill at the game” [7].

More formally, a voting rule $f$ is manipulable if there are two profiles $v$ and $v'$ and a voter $i \in \{1, \ldots, N\}$ such that $v_i \neq v'_i$ and $v_j = v'_j$ for all $j \neq i$, and the $i$th voter prefers $f(v')$ to $f(v)$. For resolute $f$, this means that $f(v')$ ranks higher than $f(v)$ by voter $i$’s sincere preference. When $f$ is irresolute and $f(v)$ or $f(v')$ contains multiple candidates, the meaning of “prefer” is more ambiguous. There will be detailed discussion on this issue in Chapter 3.

The Gibbard-Satterthwaite theorem [24], [38] (referred to as the G-S theorem) asserts that a resolute, deterministic voting rule with three or more candidates has to satisfy one of three undesirable properties:

1. The voting system is dictatorial.

2. There are at most two candidates who can possibly win under the voting system.

3. The voting system is susceptible to strategic voting (manipulable).

In other words, in order for the voting system to be free from manipulation at all, it either has to be extremely unfair to voters (dictatorship) or extremely unfair candidates (some candidate can never win). However, the Gibbard-Satterthwaite theorem focuses on the worst case rather than the “usual case”, i.e., the kind of voting profiles that are likely to occur for real world elections. Also, it does not quantify which voting systems are less manipulable than others. In order to answer questions such as “how easily is this property violated”, or “how fair is this voting system”, it is necessary to quantify the degree to which a property does or does not hold. We address this problem by introducing probability distributions over vote profiles.
that are biased toward certain profiles, rather than uniformly distributed across the simplex.

We investigate strategic voting of anonymous voting rules from a geometric perspective. Each vote profile (the collection of all votes from every voter) is represented by a point in a high-dimensional space, and a voting rule is represented as a partition of the space into different regions, with each region corresponding to a winner. Strategic voting, then, becomes a property of the boundaries separating these regions. More precisely, any case of strategic voting is a case where a single vote change pushes the original profile across a boundary. We establish a proof of the Gibbard-Satterthwaite theorem for anonymous voting systems with a geometric argument that not only proves the theorem, but also provides additional information on the locations of manipulable profiles and the property of manipulable decision boundaries.

The Gibbard-Satterthwaite theorem concerns the consequence of a single vote change. We extend the exploration to the scenario where all voters may change their votes. Each voter holds a sincere preference of candidates, yet the vote they announce to the public is not necessarily their sincere preference, but the one that they think will bring the best result. A random voter is chosen, obtains a noisy observation on the social preference and adjusts her vote if her current vote is no longer her best strategy. The adjustment of her vote causes a minor shift of the social preference. The accumulation of the feedback might cause the strategy of other voters to change as well. Due to the noise in the observation and the randomness in the sequence of chosen voters, the profile follows a random path. We argue that despite the randomness, this voting mechanism may yield an almost deterministic result that depends only on the sincere preference of the voters. We show that, in effect, the manipulative behavior of voters convert the voting rule to a different one. In particular, we show that under this model, plurality voting almost always elects the instant run-off winner, while Borda count almost always elects the Condorcet winner when one exists.
The rest of the thesis is organized as follows. Chapter 2 gives a brief introduction to some common voting systems that will recur throughout the thesis. Chapter 3 analyzes anonymous voting systems as a function defined on a high dimensional simplex and shows properties of non-manipulable boundaries. Chapter 4 introduces a probabilistic model for the distribution of profiles over the simplex and compares manipulability in a probabilistic context. Also, we present our 2012 US Presidential Election survey result. Chapter 5 explores a different aspect of strategic voting. Given that all voters would want to manipulate, we investigate the collective behavior of voters assuming they have a certain degree of knowledge about the social preference.
Chapter 2

Common voting systems

This chapter provides a brief description of several common voting systems. This includes plurality, Borda count, Condorcet methods (in particular, the Kemeny-Young method), instant run-off voting (IRV) and two-round systems. They may be used either as social welfare functions or voting rules. We focus on their roles as voting rules. Voting systems mentioned here will recur in later chapters of the thesis.

2.1 Scoring rules

Definition 2.1 (Scoring rules). A scoring rule \( f \) is determined by a vector \( w \in \mathbb{R}^K \) with

\[
1 = w_1 \geq w_2 \geq \cdots \geq w_{K-1} \geq w_K = 0 \tag{2.1}
\]

The winner is chosen by assigning a score of \( w_i \) to the candidate in the \( i \)th position on each ballot, summing the scores for each candidate, and declaring the one with the highest score the winner. If several candidates tie as the top scorer, they are all declared as winners.
Remark: Any scoring rule can also be defined as a social welfare function by sorting all candidates according to their scores. Scoring rules are also known as positional voting rules.

2.1.1 Plurality

Definition 2.2 (Plurality). Plurality rule is a scoring rule with the score vector \( w = (1, 0, \ldots, 0) \).

2.1.2 Anti-plurality

Definition 2.3 (Anti-plurality). Anti-plurality rule is the scoring rule with the score vector \( w = (1, \ldots, 1, 0) \).

In words, plurality rule elects the candidate that is the top choice of the greatest number of voters, while anti-plurality rule elects the candidate that is the bottom choice of the least number of voters.

Example 2.4. Suppose there is an election with four candidates \( C = \{a, b, c, d\} \). The votes are distributed as follows:

<table>
<thead>
<tr>
<th>Number of voters</th>
<th>26</th>
<th>26</th>
<th>25</th>
<th>23</th>
</tr>
</thead>
<tbody>
<tr>
<td>First choice</td>
<td>a</td>
<td>c</td>
<td>d</td>
<td>c</td>
</tr>
<tr>
<td>Second choice</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td>Third choice</td>
<td>c</td>
<td>a</td>
<td>a</td>
<td>d</td>
</tr>
<tr>
<td>Fourth choice</td>
<td>d</td>
<td>d</td>
<td>c</td>
<td>b</td>
</tr>
</tbody>
</table>

Table 2.1: Example: Plurality vote

If we use plurality rule as the voting rule, only the top choice of each voter is taken into consideration. Since c wins 49% of the votes, more than a (26%), b (0%) and d (25%), c is the unique winner of this profile. If the anti-plurality rule was used, then
a becomes the unique winner since no one puts a at the bottom, while b, c and d have all been put at the bottom by some voters.

2.1.3 Borda count

Definition 2.5 (Borda count). Borda count is the scoring rule with the score vector

\[ w = \left(1, \frac{K-2}{K-1}, \ldots, \frac{1}{K-1}, 0\right) \]  

(2.2)

People often use \((K, K_1, \ldots, 1)\) or \((K-1, \ldots, 0)\) as the score vector to avoid fractional scores. These two vectors define the same voting rule as (2.2). To make the mapping between score vectors and voting rules bijective, we impose the requirement that the first and last entry be 1 and 0 respectively.

Example 2.6. Look again at the profile in Table 2.1. If Borda count was used, then we first calculate the Borda score for each candidate. Use \(S_i\) to denote the Borda score for candidate \(i\). Since there are 4 candidates, the score vector is \(w = (1, 2/3, 1/3, 0)\). Thus \(S_a = 1.95\), \(S_b = 1.54\), \(S_c = 1.83\) and \(S_d = 0.98\). Thus \(a\) is the unique winner for Borda count.

2.2 Condorcet methods

Denote by \(P(a, b|v)\) the fraction of voters who rank \(a\) above \(b\) on their ballots in \(v\). That is,

\[ P(a > b|v) = \frac{\text{number of voters who rank } a \text{ higher than } b}{N} \]  

(2.3)

\(P(a > b|v)\) can be interpreted as the probability that a uniformly randomly chosen voter prefers \(a\) to \(b\) given profile \(v\). In this dissertation, we are concerned with ballots that disallow ties. Therefore \(P(a > b|v) = 1 - P(b > a|v)\) for any two distinct
candidates $a$ and $b$. We say that $a$ is preferred to $b$ by the majority under profile $v$ if $P(a > b|v) > P(b > a|v)$.

If there exists a candidate $a$ such that $P(a > b|v) \geq P(b > a|v)$ for all $b \in C \setminus \{a\}$, this candidate is called a Condorcet winner.

**Definition 2.7** (Condorcet method). A social choice function $f$ is a Condorcet method if it always elects the Condorcet winner, whenever one exists:

$$a \in f(v) \text{ iff } \forall b \in C \setminus \{a\}, P(a > b|v) \geq P(b > a|v) \quad (2.4)$$

Thus, different Condorcet methods only differ for profiles without a Condorcet winner.

**Example 2.8.** Look again at the profile in Table 2.1. Candidate $b$ is the Condorcet winner for this profile, since $P(b > a) = 0.51, P(b > c) = 0.51$ and $P(b > d) = 0.52$, although it beats other candidates only with a small margin.

Examples of Condorcet methods include Kemeny-Young method [26] [46], Ranked pairs [41], Dodgeson’s method [32],

One may wonder whether there exists a scoring rule that is also a Condorcet method. Unfortunately, this has been shown to be impossible [47]. For any choice of the score vector, there exists a voter number $N$ and a profile $v$ that produces a Condorcet winner who is not the winner under the scoring rule. Newenhizen, et al. [44] argue that among all positional rules, Borda count maximizes the probability of the Condorcet winner being the top scorer in the positional rule assuming the profile is uniformly distributed over the simplex.

However, one may generalize the scoring rule to incorporate a much broader class of voting rules, including many Condorcet methods.
2.3 Generalized scoring rules

We follow the definition of generalized scoring rules in [45] and [30].

**Definition 2.9.** A generalized scoring rule is defined by

1. a score vector $w \in \mathbb{R}^{K'}$ (where $K'$ is dependent on the number of candidates $K$, but not necessarily equal to $K$), and

2. a function $g : \mathbb{R}^{K'} \mapsto 2^\mathbb{C} \setminus \emptyset$, such that $g(x) = g(y)$ if $x_i \geq x_j \leftrightarrow y_i \geq y_j$ for any $i, j \in \{1, \ldots, K'\}$.

The input to $g$ is the sum score of all ballots. The output of $g$ is the winner or set of winners. Note that it is enough to know the ordering of the $K'$ entries of the sum score to find out the winner.

By definition, generalized scoring rules are anonymous. Generalized scoring rules are equivalent to hyperplane rules, which are the voting rules that partition the profile space with a finite number of affine hyperplanes [30].

### 2.3.1 Kemeny-Young method

One particular Condorcet method we focus on is the **Kemeny-Young method**. Denote by $d_K(L_i, L_j)$ the **Kemeny distance** or **Kendall-tau distance** between two rankings $L_i$ and $L_j$, which is defined as the number of disagreements in the ordering of all pairs of candidates between $L_i$ and $L_j$. For $L_i \neq L_j$, $1 \leq d_K(L_i, L_j) \leq \binom{K}{2}$.

**Definition 2.10** (Kemeny score). For a profile $v$ and a ranking $L$, the **Kemeny score** $S^K(L|v)$ of ranking $L$ is defined as the total Kemeny distance between $L$ and each ranking in $v$:

$$S^K(L|v) = \sum_{i=1}^{n} \sum d_K(L, v_i)$$ (2.5)
Definition 2.11 (Kemeny-Young method). The Kemeny-Young method may refer to either of the following:

1. A social welfare function that outputs the rankings with the minimum total Kemeny distance to all rankings in \( v \):

\[
L_i \in f(v) \iff \forall L_j \in \mathcal{L} \setminus \{L_i\}, \sum_{L \in v} d_K(L_i, L) \leq \sum_{L \in v} d_K(L_j, L) \quad (2.6)
\]

2. A voting rule that outputs all top candidates of the output rankings of the social welfare function defined above.

The Kemeny-Young method is both a Condorcet method and a generalized scoring rule with \( K' = K! \).

One drawback with Kemeny-Young method is the high time complexity when the number of candidates is large. If the profile is free of “pairwise preference loops”, i.e. for any subset of candidates there is always a “local” Condorcet winner among them, then finding the optimal ranking (the one with the lowest Kemeny score) is easy. Otherwise, finding the optimal ranking can be difficult. Bartholdi et al. [4] proved that Kemeny-Young method is NP-hard. Dwork et al. [19] and Biedl et al. [6] proved that even with 4 voters the problem is still NP-Hard. There has been numerous literature tackling with this problem, for example, by finding local optimalities [19], fixed-parameter algorithms [5] [39] [22], providing bounds on Kemeny scores [10], [14], efficient approximation algorithms [2] and so on. There is also work concerning with the aggregation of partial rankings [1].

2.3.2 Instant run-off and two-round system

Definition 2.12 (Instant run-off voting, IRV). Instant run-off voting is a voting rule that consists of up to \( K - 1 \) rounds of plurality vote. In each round the candidate
with the least plurality votes is eliminated. In the next round, remove the eliminated
candidate from all votes, run another plurality vote and eliminate another candidate.
The procedure is repeated until only one candidate is left or several candidates are tied
for winning in the last round.

Example 2.13. Look at the profile in Table 2.1. After the first round, candidate \( b \) is
eliminated. The profile becomes:

\[
\begin{array}{ccc}
\text{Number of voters} & 26 & 49 & 25 \\
\text{First choice} & a & c & d \\
\text{Second choice} & c & a & a \\
\text{Third choice} & d & d & c \\
\end{array}
\]

Table 2.2: Instant run-off, second round

After the second round, \( d \) is eliminated. The profile becomes

\[
\begin{array}{cc}
\text{Number of voters} & 51 & 49 \\
\text{First choice} & a & c \\
\text{Second choice} & c & a \\
\end{array}
\]

Table 2.3: Instant run-off, third round

Therefore \( a \) is the unique instant run-off winner for this profile.

Finally, we introduce two-round system.

Definition 2.14 (Two-round system). In a two-round system, a candidate is declared
the winner if she wins 50 % or more of the plurality votes. If no candidate wins more
than half of the votes, then all but the two candidates receiving the most votes are
eliminated, and a second round of voting occurs.

Example 2.15. Look at the profile in Table 2.1. Candidate \( c \) and \( a \) gets the most and
second most plurality vote, respectively. Therefore all candidates except for \( a \) and \( c \)
are eliminated. In the second round, 51 people prefer \( a \) to \( c \) and only 49 people prefer
\( c \) to \( a \), making \( a \) the unique winner under the two-round system.
Two-round system and instant run-off voting are identical when there are only 3 candidates.

These voting systems have a wide range of applications, not only in political science but also in artificial intelligence and pattern recognition. Erp [42] [43] provides an overview of voting methods for Multiple-classifier combination in pattern recognition. For a comprehensive introduction of voting systems of either practical significance or theoretical interest, please refer to Chapter 1.4 of [40].
Chapter 3

Geometric view of strategic voting

3.1 Anonymity and scale invariance

A voting rule $f$ can be viewed as a way of labelling $|C| = K$ subsets of $\mathcal{L}^N$ corresponding to $K$ candidates. If the voting rule is irresolute, then the subsets may not be disjoint. The intersection of two subsets, if non-empty, consists of the profiles for which the voting system outputs a set of winners, including the two candidates. Let $V_i = \{v : v \in f(v)\}$. For each $i \in C$, We refer to $V_i$ as the winning region of $i$. This perspective will be helpful later when we use boundaries to partition the profile space.

We can further simplify the profile by normalizing the anonymous profile by the total number of voters: $p = v/N$. The normalized profile $p$, which is the empirical distribution of the votes, is a probability distribution in the regular $K! - 1$-simplex $\mathcal{P}_{K!-1}$. Therefore, any anonymous voting system is also a function of the empirical distribution of votes $p$ and the number of voters $N$: $f : \mathcal{P}_{K!-1} \times \mathbb{Z}^+ \mapsto 2^C \setminus \emptyset$. Throughout the paper, we use $\mathcal{P}$ instead of $\mathcal{P}_{K!-1}$ when the number of candidates $K$ is clear from context. To remove the dependency on the number of voters, we need another property which we call scale invariance (also known as homogeneity in some literature):
Definition 3.1 (Scale invariance). An anonymous voting system $f$ is scale invariant if it depends only on the empirical distribution of votes: $f : \mathcal{P} \mapsto 2^\mathcal{C}$.

We can simply use $\mathcal{P}$ as the domain of a scale invariant voting system: $f : \mathcal{P} \mapsto \mathcal{C}$. Hence a profile is identified with the empirical distribution it generates. For this reason, we use $p$ instead of $v$ to represent a profile under a scale invariant voting system. Technically, all irrational points should be excluded from the domain, since they are not “real” profiles for any integer number of voters. Thus $f$ is not necessarily defined on the irrational points. However, for simplicity we use $\mathcal{P}$ as the domain since whether we define $f$ on the irrational points is immaterial.

We state our main result before showing the technical details:

Theorem 3.2 (Gibbard-Satterthwaite theorem for scale invariant voting systems). Suppose $f : \mathcal{P} \mapsto 2^\mathcal{C} \setminus \emptyset$ is a scale invariant voting system. If the union of all outcomes of $f$ contains at least 3 candidates, then $f$ is manipulable.

Although it looks just like a small modification the Gibbard-Satterthwaite theorem (occasionally allowing for ties), the structure of our proof is simpler and visually intuitive. The profile space $\mathcal{P}$ is partitioned into cells by boundaries. Cells correspond to outcomes with a unique winner, and boundaries correspond to ties of two or more candidates. It will be shown that only linear boundaries with a certain orientation (determined by the winning candidates on either side of the surface) are immune to manipulation. However, the profile space cannot be partitioned only using such linear boundaries when there are at least 3 winning candidates.

If we do not allow for ties (as in the original Gibbard-Satterthwaite theorem), the same idea of proof applies, and the proof can be made even simpler.

For an anonymous voting system, a vote change is defined as a vector

$$\Delta_{i,j} = e_j - e_i$$

(3.1)
It corresponds to the change applied to a profile if a voter changes her vote from $L_i$ to $L_j$. If $j = i$, then there is no change. We define manipulability as follows:

**Definition 3.3.** Given an anonymous voting system $f$, a profile $v$ is manipulable if there exists a vote change $\Delta_{i,j}$ such that $v > 0$ and $f(v + \Delta_{i,j}) > f(v)$, where $C_1 > C_2$ means that a voter with sincere preference $L_i$ prefers outcome $C_1$ to $C_2$.

This definition of manipulability applies to voting systems with a fixed number of voters. We will develop notions of manipulability for scale invariant voting systems when the number of voters approaches infinity. The following section shows that all boundaries immune to strategy must be linear with certain orientations.

### 3.2 Geometric features of non-manipulable boundaries

For any pair of candidates $a$ and $b$, we define $\mathcal{L}_{ab} = \{L_i \in \mathcal{L} : a > b\}$. Thus the set of rankings $\mathcal{L}$ can be partitioned into $\mathcal{L}_{ab}$ and $\mathcal{L}_{ba}$. Take a pivotal profile $v$. Assume $f(v) = a$, and by a single vote change we get a new profile $v' = v + \Delta_{ij}$ such that $f(v') = b$ or $f(v') = \{a, b\}$. A voter whose honest vote is $L_i \in \mathcal{L}_{ab}$ does not want to see that change happen and therefore has no incentive for strategic voting. It is only when $L_i \in \mathcal{L}_{ba}$ that this voter has an incentive to vote strategically with respect to $a$ and $b$.

Now we look at the possible choices of $L_j$. Consider a positional voting rule (such as Borda count), which adds scores to each candidate according to their position in each vote. A voter who prefers $b$ to $a$ may want to exaggerate the difference between $b$ and $a$ by choosing a different ranking that places $b$ higher above $a$. Although this move also affects the scores of some other candidates, it does not hurt as long as their scores do not surpass $b$’s. Other voting systems may not follow the same
procedure, but a similar argument applies: the new vote gives \( b \) a bigger advantage over \( a \). Therefore, it sounds reasonable that \( L_j \) should be an element of \( \mathcal{L}_{ba} \). The case \( L_j \in \mathcal{L}_{ab} \) sounds less reasonable. However, it is logically possible since we allow arbitrary voting systems. For example, we can define a voting system that elects \( a \) when all votes rank \( b \) ahead of \( a \), and elects \( b \) in all other cases.

The following table summarizes whether or not a vote change is a case of strategic voting, assuming the original winner is \( a \), and the winner after a vote change from \( L_i \) to \( L_j \) is \( b \) or \( \{a, b\} \). Although the case \( L_i, L_j \in \mathcal{L}_{ab} \) itself is not a case of manipulation, the backward vote change from \( L_j \) to \( L_i \) is a case of manipulation. Therefore, both profiles (before and after the vote change) are manipulable.

<table>
<thead>
<tr>
<th>( L_i \in \mathcal{L}_{ab} )</th>
<th>( L_i \in \mathcal{L}_{ba} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not manipulation, but profiles are manipulable</td>
<td>Manipulation</td>
</tr>
<tr>
<td>Not manipulation</td>
<td>Manipulation</td>
</tr>
</tbody>
</table>

Table 3.1: Vote changes that constitute manipulation.

Two profiles \( v_a \) and \( v_b \), with the same number of voters, are \textbf{adjacent} if they differ by only one vote. In other words, there exists a vote change \( \Delta_{i,j} = v_b - v_a \). Two winning regions \( V_a \) and \( V_b \) are adjacent if either

1. \( V_a \cap V_b = \emptyset \), but there exist two adjacent profiles \( v_a \) and \( v_b \), such that \( v_a \in V_a \), \( v_b \in V_b \), or

2. \( V_a \cap V_b \neq \emptyset \).

We say that a boundary exists between two winning regions if they are adjacent. Denote the boundary between \( a \) and \( b \) induced by voting system \( f \) as \( B_{a,b|f} \). To be precise, we define the boundary to be the set of profiles adjacent to profiles in the other winning region:

\textbf{Definition 3.4.} \textit{Let} \( \Delta \) \textit{denote a vote change. The boundary between candidate} \( a \) \textit{and} \( b \) \textit{induced by voting system} \( f \) \textit{consists of all pivotal profiles in} \( V_a \) \textit{and} \( V_b \) \textit{adjacent to}
the other winning region:

\[ B_{a,b|f} = \{v \in V_a : \exists \Delta \text{ such that } f(v+\Delta) \in V_b\} \cup \{v \in V_b : \exists \Delta \text{ such that } f(v+\Delta) \in V_a\} \]

(3.2)

Note that if \( V_a \cap V_b \neq \emptyset \), then \( B_{a,b|f} \) consists of all profiles in \( V_a \cap V_b \) and all their adjacent profiles in \( V_a \cup V_b \). The order of the two candidates in the subscript does not matter: \( B_{a,b|f} = B_{b,a|f} \). We say that a boundary is manipulable if it contains one or more manipulable profiles. Otherwise, the boundary is non-manipulable. If for any two profiles \( v, v' \in B_{a,b|f} \), there is a sequence of vote changes forming a path from \( v \) to \( v' \), such that the entire path is contained in \( B_{a,b|f} \), then we say that \( B_{a,b|f} \) is connected. With relevant terminology defined, we now state our first result regarding strategic voting:

**Lemma 3.5.** Suppose \( v_a, v_b \in B_{a,b|f} \) are adjacent pivotal profiles, connected by a vote change \( \Delta_{i,j} = v_b - v_a \). Suppose \( f(v_a) = a \), \( f(v_b) = b \) or \( f(v_b) = \{a,b\} \), and both \( v_a \) and \( v_b \) are non-manipulable. Then \( a >_i b \) and \( b >_j a \). Moreover, if \( v_a + \Delta \) is non-manipulable and \( f(v_a + \Delta) \subseteq \{a,b\} \) for any vote change \( \Delta \), then \( f(v_a + \Delta_{g,h}) = f(v_b) \) for any \( g, h \) with \( a >_g b \) and \( b >_h a \).

**Proof.** From the non-manipulability of \( v_a \) we can infer that \( L_i \in \mathcal{L}_{ab} \). Otherwise, a supporter of \( L_i \), who would prefer \( b \) to win instead of \( a \), has the incentive to change their vote to \( L_j \). By the same argument, \( v_b \) being non-manipulable implies that \( L_j \in \mathcal{L}_{ba} \).

We prove the second proposition by contradiction. Suppose \( L_g \in \mathcal{L}_{ab} \) and \( L_h \in \mathcal{L}_{ba} \) for some \( g, h \in \{1, \ldots, K!\} \). Let \( v'_a = v_a + \Delta_{g,h} \), with \( f(v'_a) = a \). We have

\[ v_b = v_a + \Delta_{i,j} = v'_a + \Delta_{h,g} + \Delta_{i,j} = v'_a + \Delta_{i,g} + \Delta_{h,j} \]

(3.3)
If \( f(v'_a + \Delta_{i,g}) = a \), then \( v'_a + \Delta_{i,g} \) is a manipulable profile since \( \Delta_{h,j} \) does not switch the preference order between \( a \) and \( b \). If \( f(v'_a + \Delta_{i,g}) = b \) or \( \{a, b\} \), then \( v'_a \) is a manipulable profile, since \( \Delta_{i,g} \) does not switch the preference order between \( a \) and \( b \).

**Corollary 3.6 (Monotonicity).** Suppose \( f \) is non-manipulable and \( f(v) = a \) for some \( v \in \mathcal{V} \). Take any vote in \( v \) that does not rank \( a \) first, and change it by switching \( a \) with the candidate one place higher. The ranking of all other candidates remain unchanged. Call the new profile \( v' \). Then \( f(v') = a \).

**Proof.** Let \( L_i \) and \( L_j \) be the rankings before and after the change. Suppose \( f(v') = C \), where \( C \) is a set (might a be singleton set) of candidates. Assume \( C \neq \{a\} \),

1. If \( C \) contains a candidate who is ranked above \( a \) by \( L_i \), then the change from \( v \) to \( v' \) is a case of manipulation, since a voter with sincere preference \( L_i \) prefers the outcome \( C \) to \( a \).

2. If all elements of \( C \) are ranked lower than \( a \) by \( L_i \), then the change from \( v' \) to \( v \) is a case of manipulation. Since \( a \)'s position is improved in \( L_j \), \( L_j \) also ranks all elements of \( C \) lower than \( a \). Thus a voter with sincere preference \( L_j \) prefers \( a \) to \( C \).

**Corollary 3.7 (Anonymity and non-manipulability implies unanimity).** If a voting system \( f \) is non-manipulable and \( a \in f(\mathcal{V}) \), then \( f(v) = a \) for any profile \( v \) that unanimously rank \( a \) first in all votes.

**Proof.** We project the profiles onto the unit simplex \( \mathcal{P} \) by dividing by the number of voters. All profiles that unanimously rank \( a \) first form a \((K - 1)! - 1\)-face of \( \mathcal{P} \). Call this face \( S_a \). This face is the convex hull of \((K - 1)!\) rankings in \( \mathcal{L} \) that rank \( a \) first and is itself a \((K - 1)! - 1\)-simplex.
Since \( a \in f(V) \), there exists \( v_0 \in P \) such that \( f(v_0) = a \). If \( v_0 \notin S_a \), pick any vote in \( v_0 \) that does not rank \( a \) first, and change the vote by moving \( a \) to the top. By Corollary 3.6, the new profile still elects \( a \). After a number of changes, we end up with a profile on the face \( S_a \), which elects \( a \). Now, any vote change that does not remove \( a \) from the top position should not stop \( a \) from being the unique winner. Otherwise, suppose there is such a vote change \( \Delta_{ij} \) and the profile after change is \( v \), then given the profile \( v \), someone with sincere preference \( L_j \) has the opportunity for strategic voting. Therefore we conclude that \( f(S_a) = \{a\} \). \hfill \Box

For any profile \( v \in V \), denote by \( P(a > b|v) \) the fraction of votes that rank \( a \) higher than \( b \). It can also be interpreted as the conditional probability that \( a \) is preferred to \( b \) by a randomly selected voter given the profile \( v \). The following lemma goes from local to “global” structure of non-manipulable boundaries. It shows that a non-manipulable boundary has a fixed orientation that is determined by the two winning candidates it separates:

**Lemma 3.8.** Let \( f \) be an anonymous voting system and \( N \) be the number of voters. Let \( V_a \) and \( V_b \) be the winning regions of \( a \) and \( b \) respectively. Assume \( B_{a,b|f} \neq \emptyset \) and \( B_{a,b|f} \cap V_i = \emptyset \) for any candidate \( i \in C \setminus \{a,b\} \). Suppose \( B_{a,b|f} \) is connected and non-manipulable. If \( V_a \cap V_b \neq \emptyset \), then

\[
P(a > b|v) = \begin{cases} 
\alpha + 1/N & \text{for all } v \in B_{a,b|f} \text{ such that } f(v) = a \\
\alpha & \text{for all } v \in B_{a,b|f} \text{ such that } f(v) = \{a,b\} \\
\alpha - 1/N & \text{for all } v \in B_{a,b|f} \text{ such that } f(v) = b 
\end{cases}
\]  
(3.4)

where \( \alpha \) is a constant. If \( V_a \cap V_b = \emptyset \), then

\[
P(a > b|v) = \begin{cases} 
\alpha + 1/N & \text{for all } v \in B_{a,b|f} \text{ such that } f(v) = a \\
\alpha & \text{for all } v \in B_{a,b|f} \text{ such that } f(v) = b 
\end{cases}
\]  
(3.5)
Finally, if $B_{a,b|f}$ is disconnected, then each connected non-manipulable part of the boundary satisfies the above conditions.

Proof. Assume $B(a,b|f)$ is connected. Pick two profiles $v_a, v_{a'} \in B_{a,b|f} \cap V_a$. By connectedness, there exists a sequence of vote changes from $v_a$ to $v_{a'}$ such that each profile on the path is in $B(a,b|f)$. By Lemma 3.5, any vote change (along the path) from $L_{ab}$ to $L_{ba}$ changes the winner from $a$ to $b$, since $B_{a,b|f}$ does not intersect with other boundaries. Similarly, any vote change from $L_{ba}$ to $L_{ab}$ changes the winner from $b$ to $a$. Therefore, along the path from $v_a$ to $v_{a'}$, the number of vote changes from $L_{ab}$ to $L_{ba}$ must equal to the number of vote changes from $L_{ba}$ to $L_{ab}$. Hence $P(a > b|v_a) = P(a > b|v_{a'})$.

By the same argument, for any two profiles $v_b, v_{b'} \in B_{a,b|f} \cap V_b$, $P(a > b|v_b) = P(a > b|v_{b'})$. Since the boundary profiles on the $V_b$ side differ from those on the $V_a$ side by a single vote change from $L_{ab}$ to $L_{ba}$, $P(a > b|v_a) - P(a > b|v_b) = 1/N$.

\[ \]

If there are three or more winning regions in the neighborhood of $v_a$, then strategic voting becomes inevitable. We show a result that extends the Gibbard-Satterthwaite Theorem by specifying the locations of manipulable profiles. Specifically, a manipulable profile always exists within two vote changes away from $v_a$. We assume the profiles are in the interior of the simplex $\mathcal{P}$: each ranking receives at least 2 votes.
This requirement is reasonable when the number of voters is large compared to the number of candidates.

**Lemma 3.9.** Let \( f : \mathcal{V} \mapsto 2^C \setminus \emptyset \) be an anonymous voting system with at least 3 winning candidates. Suppose \( v \) and \( v_1 \) are two adjacent profiles, \( v_1 = v + \Delta_{ij} \), \( f(v) = a \) and \( f(v_1) \neq f(v) \). Therefore \( f(v_1) \) contains at least one candidate other than \( a \). Call her \( b \). Let \( L_g \) be a ranking that prefers \( a \) to \( b \), and prefers \( b \) to any other candidate. Let \( L_h \) be a ranking that prefers \( b \) to \( a \), and prefers \( a \) to any other candidate. Let \( v_2 = v + \Delta_{gh} \) and \( v_3 = v_1 + \Delta_{jh} = v_2 + \Delta_{ig} \). Suppose \( v, v_1, v_2 \) and \( v_3 \) are not manipulable. Then,

1. If \( a \in f(v_1) \), then \( f(v_2) = \{a, b\} \).
2. If \( a \notin f(v_1) \), then \( f(v_2) = b \).

**Proof.** From the non-manipulability of \( v_2 \) we can infer that \( f(v_2) \subseteq \{a, b\} \). If \( f(v_2) \) contains any candidate other than \( a \) and \( b \), then a type-\( h \) pessimist voter would consider \( f(v) = a \) to be a better result than \( f(v_2) \).

Assume \( f(v_1) = \{a, b, K\} \) \((K \subseteq C \setminus \{a, b\} \text{ might be empty})\). Then \( f(v_3) \subseteq \{a, b\} \). Otherwise, \( v_3 \) is manipulable since a type-\( g \) pessimist voter would change her vote to \( L_i \). Moreover \( f(v_3) \neq a \). Otherwise, an optimist type-\( h \) voter would change her vote to \( L_j \) and add \( b \) to the list of winners. Finally, if \( f(v_3) = b \) then \( v_1 \) is manipulable because a type-\( j \) pessimist has the incentive to change her vote to \( L_h \). Therefore, \( f(v_3) = \{a, b\} \). This excludes the possibility that \( f(v_2) = a \) and \( f(v_2) = b \). Therefore, \( f(v_2) = \{a, b\} \).

If \( f(v_1) = \{b, K\} \), we can use the same argument to prove that \( f(v_2) = b \). \( \square \)

**Lemma 3.10.** Let \( f : \mathcal{V} \mapsto 2^C \setminus \emptyset \) be an anonymous voting system with at least 3 winning candidates. Suppose \( v_0 \) is adjacent to \( v_1 \) and \( v_2 \) \((v_1 \text{ and } v_2 \text{ need not be adjacent})\).
\[ f(v) = a \quad L_i \in \mathcal{L}_{ab} \quad f(v_2) = b \]

\[ L_g : \quad a > b > \ldots \quad L_h : \quad b > a > \ldots \]

\[ f(v_1) = \{b, K\} \quad L_j \in \mathcal{L}_{ba} \quad f(v_3) = b \]

\[ L_j \in \mathcal{L}_{ba} \quad L_h : \quad b > a > \ldots \]

Figure 3.2: Figure for Lemma 3.9 for \( f(v_1) = \{b, K\} \)

\( f(v_0) = a, \quad f(v_1) = \{b, K_1\}, \quad f(v_2) = \{c, K_2\} \). Then there is a manipulable profile at most two steps away from \( v_a \).

**Proof.** First, consider the case where at least one of \( K_1 \) and \( K_2 \) does not contain \( a \). Without loss of generality, assume \( a \notin K_1 \). Consider vote changes \( \Delta_{ij} \) and \( \Delta_{gh} \) defined by

\[
L_i = (a, b, R) \\
L_j = (b, a, R) \\
L_g = (a, c, R') \\
L_h = (c, a, R')
\]  

(3.6)

where \( R \) and \( R' \) are arbitrary rankings of the remaining candidates. That is, \( L_i \) prefers \( a \) to \( b \) and prefers \( b \) to other candidates, which is ranked according to \( R \). Let \( v_b = v_a + \Delta_{ij} \) and \( v_c = v_a + \Delta_{gh} \). Then by Lemma 3.9 \( f(v_b) = b \) and \( f(v_c) = c \) or \( \{a, c\} \), depending on whether \( a \in K_2 \). Let \( v = v_b + \Delta_{gh} = v_c + \Delta_{ij} \). If \( v_b \) is not manipulable, then \( a, c \notin f(v) \). If \( v \) is not manipulable, then \( d \notin f(v) \) for any candidate \( d \in C - \{a, b, c\} \). Therefore \( f(v) = b \). However, this makes \( v_c \) manipulable since a type-\( g \) pessimist voter has the incentive to vote \( L_h \) since she prefers \( b \) to \( c \).
Now consider the case where \( a \in K_1 \cap K_2 \). Define \( v_b, v_c \) and \( v \) the same as above. By Lemma 3.9, \( f(v_b = \{a, b\}) \) and \( f(v_c) = \{a, c\} \). We claim that \( f(v) = \{a, b, c\} \) or \( \{b, c\} \) if none of \( v_b, v_c \) and \( v \) is manipulable. If \( d \in f(v) \) for some \( d \neq a, b, c \), then the “worst candidate” of \( v_b \) is better than \( d \) according to a type-\( j \) voter, thus making \( v \) manipulable.

We look at two more profiles: \( v_b' = v_b + \Delta_{ij} \) and \( v' = v_b' + \Delta_{gh} = v + \Delta_{ij} \). By Lemma 3.9 \( f(v_b') = b \) if both \( v_b \) and \( v_b' \) are not manipulable. Also, \( d \notin f(v') \) for the same reason as \( d \notin f(v) \). Therefore \( f(v') \subseteq \{a, b, c\} \). From this we can infer that \( c \in f(v') \). Otherwise \( v \) is manipulable since a type-\( i \) voter can improve the “worst case” candidate by switching to \( L_j \). However, this makes \( f(v_b) \) manipulable since a pessimist type-\( g \) voter has the incentive to switch to \( L_h \).

In conclusion, we can always find a manipulable profile. Note that \( f(v') = f(v) - 2\Delta_{ij} + \Delta_{gh} \). This is why we require that each ranking of \( v \) receive at least 2 votes.

\[
\begin{align*}
f(v) &= a & f(v_b) = \{a, b\} & f(v_b') = b \\
\Delta_{ij} & & \Delta_{ij} & \\
\Delta_{gh} & & \Delta_{gh} & \\
f(v_c) = \{a, c\} & f(v) = \{b, c\} & f(v') = ? \\
\Delta_{ij} \quad /\{a, b, c\} & \Delta_{ij} & \\
\end{align*}
\]

Figure 3.3: Figure for Lemma 3.10 for the case \( a \in K_1 \cap K_2 \)

For voting system without ties, the proof is simpler.

**Lemma 3.11.** Let \( f : \mathcal{P} \times \mathbb{N} \mapsto \mathcal{C} \) be an anonymous voting system. Suppose two profiles \( v_b \) and \( v_c \) are both adjacent to \( v_a \) (\( v_b \) and \( v_c \) need not be adjacent), with \( f(v_a) = a \), \( f(v_b) = b \) and \( f(v_c) = c \). Suppose \( v_a, v_b \) and \( v_c \) are non-manipulable.
Also, suppose each entry of \(v_a, v_b\) and \(v_c\) is positive. Then there exists a manipulable profile \(v\) which is at most two vote changes away from \(v_a\).

**Proof.** Let \(\Delta_{i,j} = v_b - v_a\) and \(\Delta_{g,h} = v_c - v_a\). Since these profiles are non-manipulable, we have \(L_i \in \mathcal{L}_{ab}, L_j \in \mathcal{L}_{ba}, L_g \in \mathcal{L}_{ac}\) and \(L_h \in \mathcal{L}_{ca}\). Consider vote changes \(\Delta_{1,2}\) and \(\Delta_{3,4}\) defined by

\[
L_1 = (a, b, c, R) \\
L_2 = (b, a, c, R) \\
L_3 = (a, c, b, R') \\
L_4 = (c, a, b, R')
\] (3.7)

where \(R\) and \(R'\) are arbitrary permutations on the set \(C - \{a, b, c\}\). \(R\) and \(R'\) need not be different. Let \(v'_b = v_a + \Delta_{1,2}, v'_c = v_a + \Delta_{3,4}\).

1. Assume \(f(v'_b) \neq b\). Since \(\Delta_{1,2}\) does not switch the pairwise preference order between \(a\) and any candidate except \(b\), we have \(f(v'_b) = a\). Note that

\[
v_b = v_a + \Delta_{i,j} = v'_b + \Delta_{2,1} + \Delta_{i,j} = v'_b + \Delta_{i,1} + \Delta_{2,j}
\] (3.8)

Since \(L_1\) puts \(a\) at the top, \(f(v'_b + \Delta_{i,1}) = a\) by Lemma 3.5. Therefore, if the honest profile is \(v'_b + \Delta_{i,1}\), then someone with honest preference \(L_2\) can help \(b\) win by switching their vote to \(L_j\), a case of strategic voting. By the same argument, if \(f(v'_b) \neq c\), \(L_4\)-supporters have strategic incentive.

2. Assume \(f(v'_b) = b, f(v'_c) = c\). Apply two vote changes to \(v_a\):

\[
v = v_a + \Delta_{1,2} + \Delta_{3,4} = v'_b + \Delta_{3,4} = v'_c + \Delta_{1,2}
\] (3.9)

Is \(f(v)\) equal to \(a, b, c\) or some other candidate? We shall discuss each case:
(a) Suppose \( f(v) = a \). If the honest profile is \( v'_a \), then someone with honest preference \( L_3 \) can help \( a \) defeat \( b \) by voting \( L_4 \), thus making \( v_b \) manipulable.

(b) Suppose \( f(v) = b \). If the honest profile is \( v'_c \), then someone with honest preference \( L_1 \) can help \( b \) defeat \( c \) by voting \( L_2 \), thus making \( v_c \) manipulable.

(c) Suppose \( f(v) = c \). This is the same as case 2.

(d) Suppose \( f(v) = d \) for some other candidate \( d \in C \). Since \( v = v'_c + \Delta_{1,2} \), and \( \Delta_{1,2} \) does not switch the pairwise preference order between \( c \) and \( d \), \( v'_c \) is manipulable.

Therefore, we have shown that at least one of \( v_a, v_b, v_c, v'_b, v'_c \) and \( v \) must be a manipulable profile, under the mild assumption that each ranking has a positive number of supporters.

With just a short additional argument, this result can be used to prove the Gibbard-Satterthwaite Theorem for anonymous voting systems. Although it is just a special case (with anonymity), our proof takes a different approach and reveals more about the relationship between strategic voting and decision boundaries. The proof is essentially a combination of Corollary 3.7 and Lemma 3.10.

**Theorem 3.12.** Let \( f : \mathcal{P} \times \mathbb{N} \mapsto C \) be an anonymous voting system with \( M \geq 3 \) candidates and \( N \geq 2M \) voters, and suppose there are at least 3 candidates who can potentially win. Then there exists a manipulable profile \( v \in \mathcal{V} \).

**Proof.** We can always find two candidates with adjacent winning regions. Call them \( a \) and \( b \). Then there exists two adjacent profiles \( v_a, v_b \) such that \( f(v_a) = a \) and \( b \in f(v_b) \).

Let \( c \in f(\mathcal{V}) \) be another winning candidate. By Corollary 3.7, \( f(S_c) = \{c\} \), where \( S_c \) is the set of profiles that unanimously rank \( c \) at the top. It follows that \( v_a, v_b \notin S_c \).

Take an arbitrary vote present in both \( v_a \) and \( v_b \) that does not rank \( c \) first. Change this vote by moving \( c \) to the top position. Denote the vote change by \( \Delta_1 \) and apply it...
to both \(v_a\) and \(v_b\). Two scenarios might occur: If either \(v_a + \Delta_1 \in V_c\) or \(v_b + \Delta_1 \in V_c\), then \(f\) is manipulable by Lemma 3.10 or Lemma 3.11. If either \(f(v_a + \Delta_1)\) or \(v_b + \Delta_1\) contains some candidate other than \(a, b\) or \(c\), then \(f\) is manipulable by Lemma 3.5.

If neither scenario happens, re-label \(v_a + \Delta_1\) and \(v_b + \Delta_1\) as \(v_a\) and \(v_b\), and move \(c\) to the top in another vote present in two new profiles. Since \(v_a\) and \(v_b\) differ by only one vote, we can repeat the process until either \(v_a\) or \(v_b\) is only one vote change away from \(S_c\). If the winner ever changes, then \(f\) is manipulable by the same argument. Otherwise, at the end of the path, one of \(v_a\) and \(v_b\) is adjacent to a profile in \(S_c\), which elects \(c\). Thus \(f\) is manipulable by Lemma 3.10 or Lemma 3.11.

\[\square\]

### 3.3 Scale invariant voting systems with hyperplanar boundaries

Scale invariant voting systems follow the same partition rule of the vote simplex for any number of voters. Therefore, it is desirable to use a definition of boundary that is independent of the number of voters. Define \(B_P(a, b|f) \subseteq P\) to be the set of profiles such that \(p \in B_P(a, b|f)\) if any open ball centered at \(p\) has non-empty intersections with both \(V_a\) and \(V_b\). We say that the boundary is non-manipulable if for any number of voters \(N\), the corresponding discrete boundary contains no manipulable profiles. The definition of connectedness follows the usual topological definition of connectedness in \(\mathbb{R}^{K_1}\).

It is possible to design a pathological voting system by assigning winners irregularly on the simplex \(P\). But as we will see in the following corollary, a non-manipulable boundary must be a hyperplane. The orientations of the hyperplanes (the direction of their normal vectors) are determined by the winning regions they separate.
Corollary 3.13. Let $f$ be a scale invariant voting system with $K \geq 3$ candidates. Let $V_a$ and $V_b$ be two adjacent winning regions induced by $f$. If $B_{a,b|f}$ is connected and non-manipulable, then $B_{a,b|f}$ is a subset of the hyperplane $P(a > b) = \alpha$ for some constant $\alpha \in (0,1)$. Moreover, $V_a$ is in the halfspace $P(a > b) > \alpha$ and $V_b$ is in the halfspace $P(a > b) < \alpha$. If $B_{a,b|f}$ is disconnected, then each connected and non-manipulable component satisfies the above property.

Proof. Assume $B_{a,b|f}$ is connected. Suppose $B_{a,b|f}$ is not a subset of the hyperplane $P(a > b) = \alpha$ for any $\alpha$. Then there exists two profiles $v_1 \in B_{a,b|f}$, $v_2 \in B_{a,b|f}$ such that $P(a > b|v_1) = \alpha_1$, $P(a > b|v_2) = \alpha_2$, $\alpha_1 \neq \alpha_2$. Since $v_1, v_2 \in \mathcal{P}$, there exists a large enough integer $N$ such that $v_1, v_2$ are valid profiles when there are $N$ voters. Thus we can apply Lemma 3.8 to show that a manipulable profile exists.

Suppose the boundary that separates $V_a$ and $V_b$ is the hyperplane:

$$m \cdot p = \sum_{i=1}^{K!} m_i p_i = C$$

such that $f(p) = a$ if $mp > C$ and $f(p) = b$ if $mp < C$. If $mp = C$, $f(p) = \{a,b\}$. $m \in \mathbb{R}^{K!}$ is normalized by:

1. $\sum m_i = 0$
2. $\|m\|_2 = 1$

If $\sum m_i \neq 0$, then we can replace $m$ with $m' = m - (1, \ldots, 1) \sum m_i / K!$ and then normalize $m'$. The intersection of the new hyperplane with the simplex $\mathcal{P}$ is the same as the old one.

Example 3.14. Consider the plurality system for three candidates. Suppose the six rankings are listed in the order:

$$\mathcal{L} = \{abc, acb, cab, bac, bca, cba\}$$
Then the boundary between $a$ and $b$ is defined by

$$\begin{align*}
\left[\frac{1}{2}, \frac{1}{2}, 0, -\frac{1}{2}, -\frac{1}{2}, 0\right] \cdot p &= 0 \quad (3.12) \\
\left[\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, 0, 0, -\frac{1}{2}\right] \cdot p &> 0 \quad (3.13) \\
\left[0, 0, -\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right] \cdot p &< 0 \quad (3.14)
\end{align*}$$

Equation (3.12) states that an equal proportion of voters put $a$ or $b$ on the top of their votes. The inequalities (3.13) and (3.14) states that the vote count of $c$ is lower than $a$ and $b$.

Define the following notation:

- $m = \max_i m_i$
- $m = \min_i m_i$
- $\overline{m}_{ab} = \max_{i: L_i \in L_{ab}} m_i$
- $\underline{m}_{ab} = \min_{i: L_i \in L_{ab}} m_i$

**Lemma 3.15.** If $mp = C$ is a non-manipulable boundary between candidate $a$ and $b$, and $V_a$ is located on the $mp > C$ side, then the entries of $m$ take only two values:

$${m_i} = \sqrt{\frac{1}{K!}}$$

for $i : L_i \in L_{ab}$, and

$${m_i} = -\sqrt{\frac{1}{K!}}$$

for $i : L_i \in L_{ba}$.

**Proof.** Any equation $mp = C$ satisfying the condition above is equivalent to $P(a > b) = \frac{1}{2} + C$, and is therefore a non-manipulable boundary. Otherwise, $m$ has two unequal entries $m_i \neq m_j$ where $L_i$ and $L_j$ rank $a$ and $b$ in the same order. Suppose $m_i - m_j = \epsilon > 0$. For a sufficiently large number of voters $N$, we can find a profile $v \in V$ satisfying $0 < mv/N - C < \epsilon$ and $v_i > 0$. Therefore a vote change $\Delta_{ij}$ brings the profile to the other side of the boundary, making it a case of manipulation.
Since we have \( m_i = m_j \) for all \( i, j \) such that \( L_i, L_j \) rank \( a, b \) in the same order, it follows from the normalizing conditions that \( m_i = \sqrt{\frac{1}{K!}} \) for \( i : a >_i b \), and \( m_i = -\sqrt{\frac{1}{K!}} \) for \( i : a <_i b \).

We define the blowup of the boundary, as in [30]:

**Definition 3.16.** For \( \epsilon > 0 \), the blowup of the set \( B \) by \( \epsilon \) is

\[
B^\epsilon = \{ p \in P : \exists p' \in B \text{ such that } ||p - p'|| \leq \epsilon \}
\]  

**Lemma 3.17.** Suppose there are \( N \) voters. Let \( \epsilon = \frac{1}{N} \max\{m - m_{ab}, m_{ba} - m_{ba}\} \). Then \( v \) is a manipulable profile in \( B(a,b|f) \) iff \( v/N \in B(a,b|f)^\epsilon \).

We will give a constructive proof showing that non-manipulable boundaries can partition the space into at most two cells. The following theorem is parallel to Lemma 3.10 for a continuous profile space.

**Theorem 3.18.** Suppose \( f \) is a scale invariant voting system with \( K \) candidates. Let \( X \) be an intersection of non-manipulable hyperplanar boundaries. Then a neighborhood of \( X \) intersects with at most two winning regions.

**Proof.** Let \( C = \{c_1, \ldots, c_K\} \), and the corresponding winning regions be \( \{V_1, \ldots, V_K\} \). There are \( \binom{K}{2} \) different pairs of candidates, hence at most \( \binom{K}{2} \) classes of non-manipulable boundaries. Some of them are used to form the intersection \( X \). Sort the candidate pairs as follows:

\[
(1,2), \ldots, (1,K), (2,3), \ldots, (K - 1, K)
\]  

This will also be the order when we enumerate boundary classes. If all \( \binom{K}{2} \) classes of boundaries were used, they would (locally) divide the profile space into \( 2^{\binom{K}{2}} \) convex cones. Label the convex cones with vectors in \( \{0,1\}^{\binom{K}{2}} \). Each entry of the vector
corresponds to a boundary class. The $n$th entry takes the value of 0 or 1 depending on which side of the boundary the convex cone locates on. Define a mapping from the convex cones to the candidate set $h : \{0, 1\}^{K \choose 2} \mapsto C$. If changing the $n$th entry of $s \in \{0, 1\}^{K \choose 2}$ never changes $h(s)$ for any $s$, then the $n$th boundary class is not used. Moreover, the boundary class $(i, j)$ is only non-manipulable if it is used to separate $V_i$ and $V_j$ but not other winning regions.

Suppose $c_1$ and $c_2$ are among the winners, and $V_1$ and $V_2$ share a boundary. By properly reordering the candidates and labelling the convex cones, we can always make $h((0, \ldots, 0)) = c_1$ and $h((1, 0, \ldots, 0)) = c_2$. We will show that these two initial values determine $h$, and the image of $h$ is $\{c_1, c_2\}$.

1. Since $h((0, \ldots, 0)) = c_1$, we have

$$h((0, \ldots, 0, *, \ldots, *)) = c_1 \quad (3.17)$$

where a * stands for either 0 or 1. Otherwise, $V_1$ is separated from another winning region by a manipulable boundary.

2. Since $h((1, 0, \ldots, 0)) = c_2$, by the same argument we have

$$h((1, *, \ldots, 0, 0, *, \ldots, *)) = c_2 \quad (3.18)$$

3. From (3.17) and (3.18) we can conclude that

$$h((0, 0, \ldots, 0, 1, 0, \ldots, 0, *, \ldots, *)) = c_1 \quad (3.19)$$

If the 1 is between the $K$th and the $2K - 3$th bit (both ends included), then this vector is one of the vectors in (3.17). So it maps to $c_1$. 

33
If the 1 is between the second and the $K-1$th bit (both ends included), then this vector is neighboring to one of the vectors in (3.18), and the differing bit corresponds to the (1,2) boundary. Therefore $h(\cdot) = c_1$ or $c_2$. However, it is also neighboring to one of the vectors in (3.17), and the differing bit corresponds to a boundary between $c_1$ and one of $c_3, \ldots, c_K$, so $c_2$ is impossible.

4. By repeating the above arguments, we can increase the number of 1’s between the second and the $2K-3$th bit. Eventually we get the mapping of every vector in $\{0,1\}^{(K)}_{(2)}$:

$$h((0,*,\ldots,*)) = c_1 \text{ and, by the same argument, } (3.20)$$

$$h((1,*,\ldots,*)) = c_2 \text{ (3.21)}$$

Therefore $h$ is fixed, and the image of $h$ is $\{c_1, c_2\}$.

Finally, if $X$ is on the face of a simplex, then some of the convex cones are outside the simplex and can’t be assigned a winner. However, this does not affect the conclusion that a neighborhood of $X$ intersects with at most two winning regions. This completes the proof.

We are now ready to state the Gibbard-Satterthwaite Theorem for scalable voting systems. This theorem a special case of Theorem 3.12 for scale invariant voting systems. However, it offers a different perspective.

**Theorem 3.19** (Gibbard-Satterthwaite Theorem for scalable voting systems). Suppose $f$ is a scale invariant voting system with $K$ candidates, and there are at least 3 winning candidates. Then $f$ is manipulable.

**Proof.** Suppose $f$ is non-manipulable. Then all of its boundaries are hyperplanar. There exist two candidates $a$ and $b$ with adjacent winning regions. Consider a con-
nected component of \( B(a, b|f) \). It is part of a hyperplane \( P(a > b) = \alpha \) for some \( 0 < \alpha < 1 \). It either (a) slices through the vote simplex, dividing it into two polytopes \( P_1 \) and \( P_2 \), or (b) intersects with a boundary of a different class.

First suppose case (a). Since there are at least 3 winning candidates, one of the two polytopes contains at least two of the candidates, one of which is neither \( a \) nor \( b \). Suppose the two winners are \( a \) and \( c \), and the boundary \( B(a, c|f) \) exists. (Note that \( V_a \) cannot be separated from \( V_c \) by \( V_b \), otherwise the boundary between \( a \) and \( b \) put the two candidates “on the wrong side” and becomes manipulable.) Take one of its connected component. It is part of a hyperplane \( P(a > c) = \beta \) for some \( 0 < \beta < 1 \). It is easy to see that the two hyperplanes \( P(a > b) = \alpha \) and \( P(a > c) = \beta \) have a non-empty intersection inside \( \mathcal{P} \). Therefore \( B(a, c|f) \) either intersects with \( B(a, b|f) \) or intersects with another boundary. In both cases we are transferred to case (b).

Suppose case (b). We can directly conclude that a manipulable boundary exists based on Theorem 3.18.

\[ \square \]

### 3.4 Pairwise comparison methods with minimal manipulability

The transformation \( g : \mathcal{P} \rightarrow \mathcal{U} \subseteq [0, 1]^{\binom{K}{2}} \) from the profile space to the pairwise profile space \( \mathcal{U} \) is defined by:

\[
u_{ij} = \sum_{i > k > j} p_k \quad \text{for any } i, j \text{ in the candidate set} \tag{3.22}\]

Since the space \( \mathcal{U} \) has lower dimension than \( \mathcal{P} \), not all voting systems can be represented losslessly in the pairwise comparison space. Nevertheless, there are several good reasons to investigate voting systems in the pairwise comparison space. First, when there are 3 candidates, \( \mathcal{U} \) has only 3 dimensions, making it easy to visualize the
voting system. Secondly, non-manipulable boundaries can be represented easily in the pairwise comparison space. Finally, a wide range of voting systems, including Borda count (sum up the number of pairwise victories of one candidate against all other candidates and use it as the “Borda score” of this candidate) and several Condorcet methods can be represented this way.

When there are 3 candidates, we only need three axes: $u_{ab}$, $u_{bc}$ and $u_{ca}$ to span the space $U$. For example, a preference profile with 50% voters for $abc$, 30% for $bac$ and 20% for $acb$ has 70% of the population preferring $a$ to $b$, 80% preferring $b$ to $c$, and 0% preferring $c$ to $a$. Thus it maps to the point $(0.7, 0.8, 0)$. The impartial anonymous culture, where every ranking is equally likely, maps to the center $(0.5, 0.5, 0.5)$. The mapping $V \mapsto U$ has a two-dimensional null space. Therefore, different profiles may map to the same pairwise comparison profile. For example, the two profiles $v_1 = (1/3, 0, 1/3, 0, 1/3, 0)$ and $v_2 = (0, 1/3, 0, 1/3, 0, 1/3)$ both project to $(0.5, 0.5, 0.5)$. This “lossy” representation of the profile space has the advantage that non-manipulable boundaries are very easy to identify (hyperplanes perpendicular to the axes).

Note that $U$ does not fill up the unit cube $[0, 1]^3$. It is the convex hull of 6 of its vertices. The two excluded vertices are $(0, 0, 0)$ and $(1, 1, 1)$, which correspond to cyclic pairwise preference orders ($a < b, b < c, c < a$ or $a > b, b > c, c > a$). The pairwise profile space $U$ is shown in Figure 3.4.

Suppose the set $B_f^U \subseteq U$ is the boundary in $U$ under voting system $f$. The preimage $B_f^P = g^{-1}(B_f^U)$ is the corresponding boundary set of $P$. Lemma 3.20 states that $B_f^P$ is a valid boundary set for $P$. To show this, we prove that two points in $P$ are separated by $g^{-1}(B_f^U)$ if their mapping in $U$ are separated by $B_f^U$.

**Lemma 3.20.** If a set $B_u \subseteq U$ partitions $U$ into $K$ cells, then its preimage in $P$: $B_P = g^{-1}(B_u)$ partitions $P$ into $K$ cells as well.
Proof. If two points $u_1, u_2 \in \mathcal{U}$ belong to different cells, then any path connecting them must cross a boundary. Take any $p_1 \in g^{-1}(u_1), p_2 \in g^{-1}(u_2)$, connect $v_1, v_2$ with any path in $\mathcal{P}$, and map the path using $g$. Since $g$ is a continuous mapping, the mapped path is also continuous and must cross the boundary in $\mathcal{U}$. Therefore $f^{-1}(u_1)$ and $f^{-1}(u_2)$ also belong to two different winning regions in $\mathcal{P}$.

On the other hand, if $v_1, v_2 \in \mathcal{P}$ belong to the same cell, then their images $f(v_1), f(v_2)$ also belong to the same cell in $\mathcal{U}$, since a connected path between $v_1$ and $v_2$ maps to a connected path between $f(v_1)$ and $f(v_2)$ in $\mathcal{U}$ without crossing a boundary.

The above lemma ensures us that we can work safely in the space of pairwise comparison.

Let $Cond(a) \subseteq \mathcal{V}$ denote the set of profiles with Condorcet winner $a$ for any $a \in \mathcal{C}$. If there exists a profile $v_b \in Cond(a)$ such that $f(v_b) = b$, then by neutrality, there has to exist another profile $v_a \in Cond(b)$ with $f(v_a) = a$, and $v_a$ is just the result of coordinate permutation of $v_b$. Similarly, for any connected set $V'_b \subseteq Cond(a)$ such
that \( f(V'_a) = b \), there is a corresponding set of the same volume \( V'_a \subseteq Cond(b) \) such that \( f(V'_a) = a \). By reassigning the winner of \( V'_a \) as \( a \) and \( V'_b \) as \( b \), we monotonically decreases the number of manipulable profiles.

After all possible reassignments, we are left with a partition with no boundary in the interior of the Condorcet region, but the hyperplanes that separate the Condorcet region from the non-Condorcet region might be a boundary.

If \( P(i > j) = c, c \neq 1/2 \) is the boundary between \( i \) and \( j \), then by neutrality \( P(i > j) = 1 - c \) must also be a boundary. However, at least one of them is manipulable by Lemma 3.5. Therefore, the choice of non-manipulable boundaries is reduced to \( P(i > j) = 1/2 \) for all \( i, j \in C \).

For a neutral voting method with 3 candidates, the non-manipulable boundary pieces cannot fully partition the profile space, as shown in the last section. The unpartitioned regions are the two tetrahedrons in Figure 3.5. This is exactly the region of the “Condorcet paradox”, where no Condorcet winner exists for these profiles. The tetrahedron in the octant with \( u_{ab} > 0, u_{bc} > 0 \) and \( u_{ca} > 0 \) corresponds to the loop \( a > b > c > a \), and the other tetrahedron represents the other loop \( a > c > b > a \).

### 3.4.1 Impossibility of domination

While different voting systems have different boundaries (hence different manipulable profiles), it is natural to ask the following questions:

1. Does there exist a scale invariant voting system whose manipulable profiles are a subset of all other voting system’s manipulable profiles? In other words, it dominates other voting systems in terms of manipulability.

2. Does there exist a finite number of voting systems, such that any voting system is dominated by one of these voting systems?
Does there exist a scale invariant voting system that has fewer manipulable profiles than any other voting system for any fixed number of voters?

Unfortunately the answer to both the first and second question is no. We have constructed a counterexample in Theorem 3.21:

**Theorem 3.21.** There exists a set of infinitely many anonymous and neutral voting systems, none of which is dominated by any other voting system in the set.

**Proof.** We construct a class of voting systems with three continuously changing parameters. Assign a score to each ranking in the following way:

\[
S_{abc}(v) = \beta_1 P(a > b|v) + \beta_2 P(b > c|v) + \beta_3 P(a > c|v) \tag{3.23}
\]

The top choice of the highest-scored ranking is the winner, or in the case of a tie, the top choice of each of the highest scoring rankings. This voting system is neutral since the score is assigned in the same manner to any ranking. By continuously changing...
the values of $\beta_1$, $\beta_2$ and $\beta_3$, we get a set of voting systems with continuously changing boundaries. None of these dominate any other.

Remark: If we set $\beta_1 = \beta_2 = \beta_3 > 0$, we actually get the Kemeny-Young method. If we set $\beta_3 = 2\beta_1 = 2\beta_2 > 0$, then we get Borda count.

3.4.2 Counting pivotal and manipulable profiles

Given the non-existence of a finite number of voting systems that dominate others in terms of manipulability, one way to search for strategy resistant voting systems is to count the number of manipulable profiles. Direct counting of the number of profiles is intractable. Instead, we put the profiles in $\mathbb{R}^{K!}$ and find the polytope that packs in all pivotal or manipulable profiles. For hyperplane rules (voting systems that partitions the simplex with hyperplanes), the polytope is formed by a set of linear inequalities $Av \leq b$. Calculating the volume of a high-dimensional convex polytope

Figure 3.6: Winning region of $a$ when $\beta_1 = \beta_2 = 1$, $\beta_3 = 1.5$
is not straightforward. For simplicity, we analyze voting systems with 3 candidates, using the method for computing polytope volumes proposed in [27].

Example: Borda count

Figure 3.7:Boundaries of Borda count, shown in pairwise comparison space $U$

We use Borda Count as an example. By symmetry, we only need to analyze one of the boundaries, say $B(a, b|f_{Borda})$. The boundary is formed by two equations and inequalities:

\[
\begin{align*}
p_1 + 2p_2 + p_3 - p_4 - 2p_5 - p_6 &= 0 \\
p_1 + p_2 + p_3 + p_4 + p_5 + p_6 &= 1 \\
2p_1 + p_2 - p_3 - 2p_4 - p_5 + p_6 &\geq 0 \\
0 &\leq p_i \leq 1, i = 1, \ldots, 6
\end{align*}
\]
where (3.24) is obtained by equalizing the Borda scores of $a$ and $b$. When the number of voter is $N$, (3.24) is replaced by two inequalities:

$$-\frac{3}{2N} \leq p_1 + 2p_2 + p_3 - p_4 - 2p_5 - p_6 \leq \frac{3}{2N}$$

(3.28)

The choice of $\frac{3}{2N}$ is due to the following. The distance from profiles to the boundary are multiples of $\frac{1}{N}$. Profiles right on the boundary or with distance $\frac{1}{N}$ to the boundary are manipulable. Cutting off from the middle between the manipulable profiles and the closest non-manipulable profiles leads to the additional width of $\frac{1}{2N}$. The simplex constraint (3.25) is used to remove a variable, say $v_6$, so that we have a 5-dimensional polytope. The new polytope is defined by

$$-\frac{3}{2N} \leq p_1 + 2p_2 + p_3 - p_4 - 2p_5 - p_6 \leq \frac{3}{2N}$$

(3.29)

$$p_1 + p_2 + p_3 + p_4 + p_5 \leq 1$$

(3.30)

$$p_1 - 2p_3 - 3p_4 - 2p_5 \geq 1$$

(3.31)

The conditions $v_i \leq 1$ is implied by (3.29) and (3.31).

![Figure 3.8: Counting manipulable profiles for $B(a, b|f_{Borda})$](image)

42
Theorem 3.22. Suppose the equation of a boundary hyperplane is \( mp = C \). Let

\[
\epsilon_1 = \frac{1}{N} \arg \min_{1 \leq i < j \leq K} |m_i - m_j| \tag{3.32}
\]

Then the blowup boundary for volume calculation of manipulable profiles inside \( S \) is defined by

\[
C - \epsilon - \epsilon_1/2 \leq mp \leq C + \epsilon + \epsilon_1/2 \tag{3.33}
\]

where \( \epsilon \) is the “width” of the strip, for either manipulable or pivotal profiles. The number of manipulable or pivotal profiles is equal to the volume of the polytope defined by (3.33) and other boundary constraints, divided by \( N^{K!-1} \).

The width \( \epsilon \) depends on the voting system and the number of voters. Using the above results, we calculated the number of manipulable and pivotal profiles for Borda Count, Kemeny-Young method, and Baldwin method, which is another Condorcet method that use Borda count to eliminate candidates one by one. For accurate approximation, the number of voters should be reasonably large. In our simulation we make \( N \geq 1000 \).

<table>
<thead>
<tr>
<th>Voting method</th>
<th># of manipulable profiles</th>
<th># of pivotal profiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Borda count</td>
<td>( 2.604 \times 10^{-2} \times N^4 )</td>
<td>( 7.812 \times 10^{-2} \times N^4 )</td>
</tr>
<tr>
<td>Kemeny-Young</td>
<td>( 1.563 \times 10^{-2} \times N^4 )</td>
<td>( 2.604 \times 10^{-2} \times N^4 )</td>
</tr>
<tr>
<td>Baldwin method</td>
<td>( 1.21 \times 10^{-2} \times N^4 )</td>
<td>( 1.56 \times 10^{-2} \times N^4 )</td>
</tr>
<tr>
<td>Condorcet region</td>
<td>0</td>
<td>( 4.69 \times 10^{-2} \times N^4 )</td>
</tr>
</tbody>
</table>

Table 3.2: Approximate Number of Manipulable and Pivotal Profiles for Several Voting Method

The results agree with the intuition that Condorcet methods are more resistant to strategic voting. Note that for Kemeny-Young method and Baldwin method, the numbers only include the boundaries outside the Condorcet region. The number of pivotal profiles of any Condorcet method inside the Condorcet region are shown in the last row.
Chapter 4

Probabilistic evaluation of manipulability

There are two views on the purpose of voting. One view is that voters have inherently different preferences, while candidates are not inherently “better” or “worse” than each other. Alternatively, one may assume a “ground truth” ranking of candidates, and voters’ preferences are correlated with the ground truth. Thus voting is a method of estimating the truth given many noisy observations. The latter view dates back to the Condorcet Jury Theorem [15], which states that in an election with two candidates, suppose each voter has probability $p > 1/2$ of choosing the correct candidate and votes are i.i.d., then the majority decision converges to the ground truth as the number of voters goes to infinity. Conitzer [11] investigates which voting rules can be interpreted as maximum likelihood estimators under suitable noise models. Conitzer [9] applies this interpretation to voting on social networks.

In practice, we do not expect all profiles to be equally likely. Some are more likely to occur in real settings than others. This is addressed by introducing a probability distribution over all profiles. The motivation for using probability here is also to deviate from the unrealistic assumption that all profiles are deemed to be equally
likely; in real elections voters usually have some information on the expected vote
distribution.

Recall from Chapter 1 that we use $K$ to denote the number of candidates and $N$ to
denote the number of voters. A widely used model [36], [33] assumes that, each voter
randomly draws a vote from a probability distribution $p$ over $K!$ rankings. Thus,
with $K$ candidates, under anonymity, the preference profile is a random vector $v$ with
a multinomial distribution in $K!$ dimensions. For this reason we refer to this model
as the *multinomial model*, and refer to $p$ as the *social bias*. The *impartial anonymous
culture* assumption, where each voter is completely unbiased and picks a vote with
uniform distribution, is just a special case of the multinomial model (uniform social
bias). But it’s not a particularly inspiring one. We find it more interesting to allow
for the model to represent some sort of coherence throughout the population, and this
is made possible through the selection of the bias in the multinomial model. Whether
or not the multinomial model is realistic, it allows us to uncover interesting properties
that are missed by the uniform random votes model and that we believe would hold
true under many realistic probability models. To enrich our discussion, we introduce
another closely related probability model, called the sampling model. For this model,
each voter holds a fixed vote (contrary to the random decision in the multinomial
model), therefore the vote profile is also a fixed point in the simplex. However, the
vote profile is unknown, and can only be estimated from a random sample of votes,
which resembles the information from polls and surveys before an election. It should
be emphasized that the dichotomy of manipulable and non-manipulable boundaries
is independent of the probabilistic model.

The main contribution of this chapter include:

1. A detailed analysis of the voting decision boundaries of three well known voting
systems - plurality, Borda count, and Kemeny Young: Which voting profiles
near the decision boundaries are manipulable; and for each manipulable profile,
which voters would vote strategically. For the Kemeny-Young method (and Condorcet methods in general), part of the decision boundary is immune to strategic voting: No profiles near these boundaries are manipulable. This result does not hold for plurality and Borda count. All boundaries for plurality and Borda count are manipulable.

2. A new measure for the manipulability of voting systems that is compatible with an arbitrary probability distribution on the space of profiles.

3. Large deviations analysis which yields the probability of strategic voting in the limit of large populations or sample sizes for a variety of social biases. In particular, the Kemeny-young method is shown to be immune to strategic voting for most vote distributions.

4. Validation using results from an online survey of the 2012 US Presidential Election. The Kemeny-Young method is largely strategy-proof under the data we collected.

Some earlier work measures the degree of manipulability by counting the number of manipulable voting profiles [25]. It is shown that random voting rules (that is, winners are randomly assigned to each vote profile) have a high degree of manipulability. This approach provides a basis for detailed calculation and theory development. The work in [23] define a distance metric between two voting systems as the probability that two voting systems with the same input elect different winners, with the assumption that each vote is uniformly distributed over all rankings. This gives us a tool to characterize “how dictatorial” a voting system is, by calculating the distance from the voting system to the nearest dictatorial voting system. It is shown that for any voting system with three candidates, the sum of each voter’s probability of strategic voting is lower bounded by the square distance between the voting system and the set of dictatorial voting systems, multiplied by a constant. This also implies that with a
fixed voting system, there always exists a voter whose probability of strategic voting is at least $\Omega(1/N)$.

### 4.1 Strategic Voting in Different Voting Systems

#### 4.1.1 Borda count

Based on the score profile, we can conclude which types of voters have strategic incentive. Consider an example:

**Example 4.1.** In an election using Borda count, the scores of three candidates are $S^B_a = 100, S^B_b = 100$ and $S^B_c = 70$. A voter with preference cba, learning that her top choice c has no chance of winning, is willing to compromise and vote bca instead of cba in order to help her second choice win. Another voter with preference abc, knowing that b threatens her top choice a, will bury b by voting acb. In both cases, voters change the result to their benefit by not voting honestly.
Table 4.1: Strategic voting table for Borda count. Grey cells are the strategic voters.

<table>
<thead>
<tr>
<th>Leading</th>
<th>Runner-up</th>
<th>abc</th>
<th>acb</th>
<th>cab</th>
<th>cba</th>
<th>bca</th>
<th>bac</th>
</tr>
</thead>
<tbody>
<tr>
<td>ab</td>
<td>ab</td>
<td>abc</td>
<td>acb</td>
<td>acb</td>
<td>bca</td>
<td>bca</td>
<td>bca</td>
</tr>
<tr>
<td>ac</td>
<td>ac</td>
<td>abc</td>
<td>acb</td>
<td>cba</td>
<td>cba</td>
<td>abc</td>
<td>abc</td>
</tr>
<tr>
<td>bc</td>
<td>bc</td>
<td>cab</td>
<td>cab</td>
<td>cba</td>
<td>cba</td>
<td>abc</td>
<td>abc</td>
</tr>
<tr>
<td>a</td>
<td>b</td>
<td>abc</td>
<td>acb</td>
<td>cab</td>
<td>bca</td>
<td>bca</td>
<td>bca</td>
</tr>
<tr>
<td>a</td>
<td>c</td>
<td>abc</td>
<td>acb</td>
<td>cba</td>
<td>cba</td>
<td>cba</td>
<td>bac</td>
</tr>
<tr>
<td>b</td>
<td>c</td>
<td>abc</td>
<td>cab</td>
<td>cba</td>
<td>bca</td>
<td>bac</td>
<td>bac</td>
</tr>
<tr>
<td>c</td>
<td>a</td>
<td>abc</td>
<td>acb</td>
<td>cba</td>
<td>bca</td>
<td>abc</td>
<td>abc</td>
</tr>
<tr>
<td>b</td>
<td>a</td>
<td>acb</td>
<td>acb</td>
<td>cba</td>
<td>bca</td>
<td>bac</td>
<td>bac</td>
</tr>
<tr>
<td>c</td>
<td>b</td>
<td>bac</td>
<td>acb</td>
<td>cab</td>
<td>cba</td>
<td>bac</td>
<td>bac</td>
</tr>
</tbody>
</table>

A voter is pivotal if by changing her vote alone, some losing candidate can outscore the winner or become tied with it. However, if the runner-up is lagging behind by 2 points or more, then none of the pivotal voters has the incentive to change his vote. For example, if a leads b by 2 points, then some voter with honest preference bac, bca or cba would like to help b defeat a. But the best they can do is to shrink the difference to 1 point, which is not enough to change the result.

Table 4.1 enumerates all cases of strategic voting. If the leading candidate and the runner-up are both ab, it means a and b are tied as first. Otherwise, we assume the runner-up is only 1 point behind the leading candidate. The 6 columns on the right indicates how a strategic voter would vote given their sincere preferences. Cases of strategic voting are shaded in gray. It can be seen that manipulator exists in all types of boundaries. This is an expected result, since boundaries for Borda count are not pairwise comparison boundaries.

### 4.1.2 Plurality

The analysis of strategic voting for plurality is similar to Borda count. For plurality, the first place gets all the credit and no score is assigned to the second place. The candidate with the most first place votes is declared the winner. In Table 4.2 we
list all cases of strategic voting for plurality where \( a \) and \( b \) are either tied as the winner or differ by 1 point. Other cases are completely symmetric and hence omitted. We denote the case where \( a \) and \( b \) are tied at the top by putting \( ab \) in both the “Leading” and the “Runner-up” column. It can be seen that for every winner/runner-up combination, there is chance for strategic voting for at least one type of voter.

<table>
<thead>
<tr>
<th>Leading</th>
<th>Runner-up</th>
<th>abc</th>
<th>acb</th>
<th>cab</th>
<th>cba</th>
<th>bca</th>
<th>bac</th>
</tr>
</thead>
<tbody>
<tr>
<td>( ab )</td>
<td>( ab )</td>
<td>( a )</td>
<td>( a )</td>
<td>( a )</td>
<td>( b )</td>
<td>( b )</td>
<td>( b )</td>
</tr>
<tr>
<td>( a )</td>
<td>( b )</td>
<td>( a )</td>
<td>( a )</td>
<td>( c )</td>
<td>( b )</td>
<td>( b )</td>
<td>( b )</td>
</tr>
<tr>
<td>( b )</td>
<td>( a )</td>
<td>( a )</td>
<td>( a )</td>
<td>( a )</td>
<td>( c )</td>
<td>( b )</td>
<td>( b )</td>
</tr>
</tbody>
</table>

Table 4.2: Strategic voting table for plurality voting

4.1.3 The Kemeny-Young Method

We start with a lemma that states that the boundaries of the Kemeny-Young method are always between two rankings with Kendall tau distance 1 or 2, but not 3.

**Lemma 4.2.** For a voting profile with 3 candidates, the Kemeny distance between the optimal ranking and the runner-up is 1 or 2 if the optimal ranking is unique.

**Proof:** Without a loss of generality, assume \( abc \) is the optimal ranking. We have \( S^K_{abc} + S^K_{cba} = S^K_{bca} + S^K_{acb} \). (This holds even if the votes are partial rankings.) Therefore, either \( bca \) or \( acb \) has a lower Kemeny score than \( cba \). So a Kemeny distance of 3 is impossible. However, the runner-up might have Kemeny distance 2 to the winner. An example is shown in Table 4.3:

Here the winner is \( abc \), with \( cab \) and \( bca \) being tied as runner-ups. They both have Kemeny distance of 2 to the optimal ranking. Table 4.4 lists all cases where a pivotal voter exists, assuming the leading ranking is \( abc \). Note that

1. A voter who can change the optimal ranking from \( abc \) to \( acb \) is not considered a pivotal voter, since she does not change the overall result.
<table>
<thead>
<tr>
<th>Ranking</th>
<th>Votes Received</th>
<th>Kemeny Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>abc</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>acb</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>cab</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>cba</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>bca</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>bac</td>
<td>0</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 4.3: An example of Kemeny-Young

<table>
<thead>
<tr>
<th>Optimal</th>
<th>Runner-up</th>
<th>abc</th>
<th>acb</th>
<th>cab</th>
<th>cba</th>
<th>bca</th>
<th>bac</th>
</tr>
</thead>
<tbody>
<tr>
<td>abc</td>
<td>acb</td>
<td>abc</td>
<td>acb</td>
<td>cab</td>
<td>cba</td>
<td>bca</td>
<td>bac</td>
</tr>
<tr>
<td>abc</td>
<td>bac</td>
<td>abc</td>
<td>acb</td>
<td>cab</td>
<td>cba</td>
<td>bca</td>
<td>bac</td>
</tr>
<tr>
<td>abc</td>
<td>cab</td>
<td>abc</td>
<td>acb</td>
<td>cab</td>
<td>cba</td>
<td>cab or cba</td>
<td>bac</td>
</tr>
<tr>
<td>abc</td>
<td>bca</td>
<td>abc</td>
<td>acb</td>
<td>cab</td>
<td>cba</td>
<td>bca</td>
<td>bca or cba</td>
</tr>
</tbody>
</table>

Table 4.4: Pivotal and Strategic voters for Kemeny-Young

2. The runner-up is either tied with, or 1 or 2 points higher than, the leading ranking. Otherwise, there is no pivotal voter. This follows from Lemma 1.

3. We do not consider the cases where three rankings are tied or almost tied as these do not come up in the multinomial analysis.

Observe that Kemeny-Young is categorically different from Borda count and plurality voting in the following way: Strategic voting happens only when the boundary separating two “winning” rankings with Kemeny distance 2. For those with Kemeny distance 1, none of the six types of voters can benefit from modifying their vote. For example, suppose $S^K_{abc} = S^K_{bac} - 1$. Three types of voters – bac, bca and cba would like to help b to win. However, no matter how they change their vote, there is simply no way to fill in the gap! We plot the strategic and non-manipulable boundaries in the space of pairwise comparisons in Figure 4.2. The transparent tetrahedrons are the regions of Condorcet paradox, i.e. profiles with a pairwise majority loop. Note that all manipulable boundaries reside within the regions of Condorcet paradox.
This phenomenon can be explained by Lemma 3.5. Since Kemeny-Young method is a Condorcet method, the boundary between \( a \) and \( b \) in the Condorcet region

\[
\text{(4.1)}
\]

![Figure 4.2: Strategic and non-manipulable boundaries of the Kemeny-Young method](image)

We end this section by summarizing the above conclusions with a theorem:

**Theorem 4.3.** The following propositions hold for a voting system with three candidates:

1. For Borda count and plurality, strategic voting is possible for every profile near the boundary: Each pivotal profile is either manipulable, or adjacent to a manipulable profile.

2. For the Kemeny-Young method, strategic voting is possible at a boundary if and only if it separates two rankings with Kemeny distance 2.
4.2 Probabilistic Analysis for Strategic Voting

In this section, we introduce a new metric for measuring the manipulability of voting systems, the *conditional incentive*. It is defined as the probability that a random voter has incentive for strategic voting *given* that she is a pivotal voter. We believe that the conditional incentive more accurately measures the likelihood of strategic voting than simply counting the number of manipulable profiles. We then apply it to two probabilistic models and provide a simple algorithm to calculate the conditional incentive.

4.2.1 Conditional Incentive

We reserve the capital $P$ for profile distributions and use lower case letters for vote profiles and the probability distribution of votes. Let $P_N$ be the profile distribution with $N$ voters. For example, under the *impartial culture* assumption, each voter uniformly randomly picks a vote. In this case $P_N$ is a multinomial distribution, and the probability that we get a particular profile $(v_1, \ldots, v_6) \in V$ is:

$$P_N((v_1, \ldots, v_6)) = \frac{N!}{6^N(Nv_1)! \cdots (Nv_6)!} \quad (4.2)$$

**Definition 4.4.** Define the mean incentive $I(P_N)$ as the probability that a random voter has the incentive to manipulate:

$$I(P_N) = \sum_{v \in V} P_N(v) \sum_{i=1}^{M!} v_i 1\{\text{type-}i \text{ voter is strategic}\} \quad (4.3)$$

For infinitely many voters, let

$$I(P) = \lim_{N \to \infty} I(P_N) \quad (4.4)$$
if the limit exists.

**Definition 4.5.** Define the mean effect $E_m(P_N)$ as the probability that a random voter is a pivotal voter. It is obvious that $I(P_N) \leq E_m(P_N)$.

\[
E_m(P_N) = \sum_{v \in V} P_N(v) \sum_{i=1}^{M!} v_i 1\{\text{type-}i \text{ voter is pivotal}\}
\] (4.5)

For infinitely many voters, let

\[
E_m(P) = \lim_{N \to \infty} E_m(P_N)
\] (4.6)

if the limit exists.

**Definition 4.6.** The conditional incentive $I_c(P_N)$ is defined as the probability that a random voter is a strategic voter, given that she is pivotal.

\[
I_c(P_N) = \frac{I(P_N)}{E_m(P_N)}
\] (4.7)

For infinitely many voters, let

\[
I_c(P) = \lim_{N \to \infty} I_c(P_N)
\] (4.8)

if the limit exists.

For a scale invariant voting system, we expect $E_m(P) = 0$. We do not show a formal proof, but intuitively the boundary should grow “thinner” as the number of voters grows. In other words, the voting result stabilizes after more and more people have voted. However, the conditional incentive $I_c$ does not necessarily converge to 0, as will be analyzed in the following section. A voting system may have a very low mean incentive simply because there is a small fraction of pivotal profiles. But what we are really after is a voting system with a low conditional incentive. After all, in
the settings of Gibbard-Satterthwaite Theorem, everyone votes as if their vote makes the difference.

We illustrate the concept using the example of Borda count. We show the score profiles instead of the vote profiles because 1) the pattern of strategic voting for Borda count can be directly read from the score profile, and 2) score profiles are easily shown on a two-dimensional plot. Figure 4.3 depicts a microscopic view of the score profiles of Borda count near the boundary $S_a^B = S_b^B$. Here each dot represents a score profile $(S_a^B, S_b^B, S_c^B)$. All score profiles lie on a plane $S_a^B + S_b^B + S_c^B = 3N$. The vertical line in the middle represents the boundary $S_a^B = S_b^B$. To the left of the boundary are the score profiles such that $S_a^B > S_b^B > S_c^B$ and to the right are the profiles satisfying $S_b^B > S_a^B > S_c^B$. $S_c^B$ is the same for all profiles on the same horizontal level, and increases when moving upwards. The set of manipulable score profiles is surrounded by the rectangle. The fractions over each dot denotes the number of types of strategic voter (numerator) and types of pivotal voter (denominator). For example, 2/5 means that if a voting profile (among many) maps to this score profile, then 5 out of 6 types of voter are pivotal, but only two types are strategic voters. The procedure of counting follows from Table 4.1. Figure 4.4 shows how a vote change translates to the change of the score profile. In the left figure, a bac-type voter benefits from strategic voting by changing her vote to bca (arrow number 3). In the right figure, an abc-type voter have three options (arrow 2, 3 and 5) to make b the winner by changing her vote, but she has no incentive to do so.

It is easy to count the number of pivotal and manipulable score profiles. Unfortunately, not all score profiles are equiprobable. Even if they were, the ratio of voters supporting each ranking given the score profile is non-uniform. Calculating the conditional incentive $I_c(P_N)$ with large $N$ seems to be a formidable task due to the very large number of pivotal profiles. However, given the specific profile distribution
\( (x + 2, x - 2, y) \cdot \frac{4}{6} \bullet 0/3 \cdot 0/1 (x - 2, x + 2, y) \cdot \\
\frac{2}{5} \bullet 0/3 \cdot 0/1 (x - 1, x + 2, y - 1) \bullet \)

\( S^B_a = S^B_b \)

Set of manipulable profiles

Figure 4.3: Score profiles with pivotal and strategic voters for Borda count

\( S^B_a = S^B_b \)

\( S^B_a = S^B_b \)

(a) Possible vote changes for voter bac. The arrows 1-5 denotes the resulting score profile if the vote is changed to abc, acb, bca, cab or cba, respectively.

(b) Possible vote changes for voter abc. The arrows 1-5 denotes the resulting score profile if the vote is changed to ach, bac, bca, cab or cba, respectively.

Figure 4.4: The outcome of vote changes by two voter types, one with strategic incentive (bac) and one without (abc), in Borda count.

\( \{P_N\} \), the calculation of conditional incentive is much simplified using large deviation analysis.

### 4.2.2 Multinomial Model

In real elections, people usually have some prior knowledge on the possible vote distribution, which they could get from polls, surveys, social networks, etc. However, there is still quite a lot of randomness remaining due to sampling error, faulty or biased sampling method, mishandling of data among others. Both models we are about to introduce simulate the reality with a unimodal distribution that represents some sort
of coherence throughout the population. The first probabilistic model, known as the 

multinomial model, assumes that the behavior of each individual voter is intrinsically random: Each voter draws a random vote the (normalized) voting profile $v$ is subject to a multinomial distribution: $P_N \sim \text{Multi}(N, p)$, where $p_0$, called the social bias, is a probability distribution over 6 rankings. For example, $p = [0.3, 0.2, 0.5, 0, 0, 0]$ means that the probability that any individual voter votes for $abc$ is 0.3.

However, the problem can be much simplified by using large deviation analysis. To find the conditional incentive as $N \to \infty$, we first find the information projection of $p$ onto the set of all pivotal profiles $T$:

$$q^* = \arg\min_{q \in T} D(q||p) \tag{4.9}$$

Here $D(q||p) = \sum_i q_i \log(q_i/p_i)$, the Kullback-Leibler divergence or relative entropy from $q$ to $p$, is a metric of the information lost when $p$ is used to approximate $q$ [12]. Let $v$ be the empirical vote distribution drawn with bias $p$. For any $\epsilon > 0$, denote by $q^*$ the neighborhood of $q^*$ with radius $\epsilon$ (we do not specify a distance metric). By Sanov’s Theorem [37],

$$\lim_{N \to \infty} P[v \in q^*|v \in T] = 1 \tag{4.10}$$

In other words, if strategic voting happens at all, then with high probability the profile is inside a small neighborhood of $q^*$ as the number of voters grow. Therefore, $I_c(P_N)$ can be calculated by just looking at the few pivotal profiles near $q^*$. $D(q||p)$ is a convex function of $q$, but $T$ may not be a convex set: The convex combination of two pivotal profiles is not necessarily pivotal. However, if $T$ is “piecewise linear”, i.e. $T$ can be written as a finite union of subsets, each subset being a hyperplane bounded by linear inequalities, then convex optimization can be applied to minimize $D(q||p)$ over each linear segment. Most voting systems we encounter in practice have piecewise linear boundaries. To focus on the central idea, assume the optimal solution
$q^*$ is located on the boundary $T$ defined by the linear equation $L \cdot q = 0$ and a set of linear inequalities $L_i q \leq 0$.

Minimize $D(q||p) = \sum_{i=1}^{6} q_i \log \frac{q_i}{p_i}$ \hspace{1cm} (4.11)

subject to $Lq = 0$ \hspace{1cm} (4.12)

$0 \leq q_i \leq 1$ for $i = 1 \ldots 6$ \hspace{1cm} (4.13)

$\sum_{i=1}^{6} q_i = 1$ \hspace{1cm} (4.14)

The solution can be found using the KKT conditions:

$q_i^* = \frac{p_i e^{\lambda L_i}}{\sum_i p_i e^{\lambda L_i}}$ \hspace{1cm} for $i = 1, \ldots, 6$ \hspace{1cm} (4.15)

where $\lambda$ is the Lagrange multiplier for the constraint (4.12). $\lambda$ can be solved by plugging (4.15) into (4.12).

The following result further simplifies the calculation of conditional incentive.

**Theorem 4.7.** Suppose the social bias $p$ satisfies $p_i > 0$ for all $i$ and $p \notin T$. Let $q^* = \arg\min_{q \in T} D(q||p)$. For a profile $v$ satisfying $||v - q^*||_1 < \epsilon$ for some $\epsilon > 0$ a vote change $\Delta_{ij}$ satisfying $L \cdot \Delta_{ij} = 0$,

$$\left| \frac{P_N(v + \Delta_{ij})}{P_N(v)} - 1 \right| \leq \epsilon \cdot \frac{\max_i p_i}{\min_i p_i} + \frac{1}{N}$$ \hspace{1cm} (4.16)

**Proof.** By the probability mass function of multinomial random variables,

$$P_N(v) = \frac{N!}{(N v_1)! \ldots (N v_6)!} p_1^{N v_1} \ldots p_6^{N v_6}$$ \hspace{1cm} (4.17)
Therefore

\[
P_N(v + \Delta_{ij}) \quad \frac{v_i}{v_j + 1/N} \cdot \frac{p_j}{p_i}
\]

(4.18)

Therefore, for an arbitrary profile \(v\), the ratio is roughly \((p_i v_j)/(p_j v_i)\). However, due to the proximity of \(v\) to the projection point \(q\) and the fact that \(L_i = L_j\), we can show this ratio is actually close to 1. Let \(r_i = v_i - q_i\) and \(r_j = v_j - q_j\). By \(||v - q^*||_1 < \epsilon\) we have \(|r_i| < \epsilon\) and \(|r_j| < \epsilon\). By plugging in (4.15) to (4.18),

\[
\left| \frac{P_N(v + \Delta_{ij})}{P_N(v)} - 1 \right| \\
\leq \left| \frac{r_i p_j / p_i - r_j - 1/N}{v_j + 1/N} \right| \\
\leq |r_i p_j / p_i - r_j - 1/N| \\
\leq \epsilon \cdot \frac{\max_i p_i}{\min_i p_i} + 1/N
\]

(4.19)

This theorem implies the following. Consider any profile \(v\) near \(q^*\), and draw a hyperplane orthogonal to \(L\). Other profiles on this hyperplane are almost equiprobable to \(v\), and the difference in probability vanishes as \(N\) goes to infinity. Therefore, a single profile can be used to represent the entire hyperplane in (4.7). Of course, as \(v\) moves further and further away from \(q^*\) the probability would decrease, but with a fairly large \(N\) the approximation can be good enough.
Theorem 4.8. Assume \( p \not\in T \), \( p_i > 0 \) for \( i = 1, \ldots, 6 \), and the information projection \( q^* \) is on the boundary defined by \( Lq = C \). The conditional incentive \( I_c(p) \) is given by

\[
I(p) = \sum_{d=-D}^{D} \sum_{i=1}^{6} e^{\lambda k} q_i \mathbb{1}\{\text{type-}i\text{ voters are manipulators on the hyperplane } Lq = C + \frac{d}{N}\}
\]

(4.20)

\[
EF(p) = \sum_{k=-1}^{1} \sum_{i=1}^{6} e^{\lambda k} q_i \mathbb{1}\{\text{type-}i\text{ voters are pivotal on the hyperplane } Lq = C + \frac{d}{N}\}
\]

(4.21)

\[
I_c(p) = \frac{I(p)}{EF(p)}
\]

(4.22)

where \( D \) is chosen appropriately such that all pivotal profiles are included. Whether type-\( i \) voters are pivotal or strategic can be verified by enumerating all possible vote changes.

Application to Borda count

Without a loss of generality, let us assume the social bias \( p \) is closest to the boundary between \( a \) and \( b \). Lemma 4.9 says that all pivotal profiles are such that the score difference between \( a \) and \( b \) is at most 4:

Lemma 4.9. For Borda count, suppose \( v \in \{v : |S^B_a - S^B_b| \leq 4, \min(S^B_a, S^B_b) > S_c\} \), and \( v_i > 0 \) for \( i = 1, \ldots, 6 \), then there exists at least one pivotal voter in \( v \).

Proof. Since a voter contributes 2 points to her top choice and 0 point to her bottom choice, by reversing the order she could increase or decrease the point difference of two candidates by at most 4. Therefore, if \( a \) and \( b \) are 5 points or more apart no single voter can change the outcome. Q.E.D.

We explicitly write the convex program for Borda count. \( B_a \) and \( B_b \) denote the Borda scores of \( a \) and \( b \).
Minimize \( D(q||p) = \sum_{i=1}^{6} q_i \log \frac{q_i}{p_i} \) \hspace{1cm} (4.23)

subject to \((B_a - B_b)q = 0\) \hspace{1cm} (4.24)

\[ 0 \leq q_i \leq 1 \text{ for } i = 1 \ldots 6 \] \hspace{1cm} (4.25)

\[ \sum_{i=1}^{6} q_i = 1 \] \hspace{1cm} (4.26)

Figure 4.5 plots what happens near \( q^* \). These hyperplanes are actually 4-dimensional, but are plotted as 2-dimensional for the sake of visualization. \( I_c(p) \) is given by Theorem 3. The sign of \( \lambda \) does not matter. If \( \lambda > 0 \), then the hyperplanes with \( k > 0 \) are closer to \( p \), and vice versa. Choosing \( \lambda > 0 \) or \( \lambda < 0 \) results in the same value of \( I_c(p) \).

Figure 4.5: Hyperplanes parallel to \( S_a^B = S_b^B \)

For illustration, we pick 8 different social biases, and set the number of voters to \( N = 10000 \). The conditional incentives computed analytically using Theorem 3 are very close to the simulation results (Table 4.5), except for the uniform distribution. This is expected since the uniform distribution is at the intersection of 3 boundaries.

Are there any profiles with zero conditional incentive? The answer is yes, but such profiles require \( p_i = 0 \) for some \( i \). One example is \( p = [.01, .49, 0, 0, .49, .01] \). Though all 4 types of voters can be pivotal voters, only \( abc \) and \( bac \) can potentially vote
strategically (refer to Table 1). But nobody holds that preference anyway. In real elections with three candidates, it is very unlikely for a ranking to have no support at all, as suggested by historical voting data in [33].

**Application to plurality**

Calculation of $I_c(p)$ for plurality involves the same procedure as Borda count. We solve for the information projection of $p$ on the boundary set $T^P$, find the pivotal and strategic voters on hyperplanes parallel to the boundaries, and calculate the mean incentive and mean effect separately. Table 4.6 compares the analytical and simulation results of the conditional incentive with $N = 10000$ voters. The analytical and simulation results are very close except when $p$ is uniform.

<table>
<thead>
<tr>
<th>$p$</th>
<th>$I_c(P_N)$ (simulation)</th>
<th>$I_c(p)$ (analytical limit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>uniform</td>
<td>0.2177</td>
<td>0.1829</td>
</tr>
<tr>
<td>[0.2 0.2 0.18 0.18 0.12 0.12]</td>
<td>0.1598</td>
<td>0.1631</td>
</tr>
<tr>
<td>[0.2 0.15 0.15 0.15 0.2 0.15]</td>
<td>0.2010</td>
<td>0.1983</td>
</tr>
<tr>
<td>[0.18 0.18 0.17 0.17 0.15 0.15]</td>
<td>0.2005</td>
<td>0.1988</td>
</tr>
<tr>
<td>[0.2 0.2 0.1 0.2 0.2]</td>
<td>0.1335</td>
<td>0.1342</td>
</tr>
<tr>
<td>[0.19 0.18 0.17 0.16 0.15 0.15]</td>
<td>0.2006</td>
<td>0.1993</td>
</tr>
<tr>
<td>[0.2 0.18 0.17 0.16 0.15 0.14]</td>
<td>0.1941</td>
<td>0.1933</td>
</tr>
<tr>
<td>[0.2 0.15 0.18 0.17 0.15 0.15]</td>
<td>0.2007</td>
<td>0.2000</td>
</tr>
</tbody>
</table>

Table 4.6: Conditional incentive for plurality voting with $N = 10000$ voters
Application to Kemeny-Young

Finding the information projection on the boundary set for Kemeny-Young is trickier than Borda count. This is because the Kemeny-Young method partitions the probability simplex $\mathcal{V}$ into non-convex cells. Therefore the information projection $q^*$ may not always fall on the interior of a boundary, but instead on the intersection of two linear pieces. Figure 4.6 illustrates an example, where $q^*$ is at the intersection of two boundaries $S^K_{abc} = S^K_{bac}$ and $S^K_{acb} = S^K_{bac}$. If the social bias favors $a$, then the information projection may fall on the intersection of the two boundaries. Once that happens, we need to look at the microscopic structure near $q^*$ to determine which pivotal profiles are manipulable and which are not. Again, for visualization purpose we are showing the space of profiles in a lower dimension. The two intersecting boundary pieces are actually 4-dimensional hyperplanes, and the intersecting “line” is actually a 3-dimensional hyperplane.

![Figure 4.6: The information projection falls on the intersection of $S^K_{abc} = S^K_{bac}$ and $S^K_{acb} = S^K_{bac}$](image)

Again, we compare the simulation and analytical results of the conditional incentive for the Kemeny-Young method in Table 4.7. The last distribution is adjusted slightly because we want to show an example where the social bias is not on the boundary and is closer to a manipulable boundary.
<table>
<thead>
<tr>
<th>$p$</th>
<th>$I_c(P_N)$ (simulation)</th>
<th>$I_c(p)$ (analytical limit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>uniform</td>
<td>0.0830</td>
<td>0</td>
</tr>
<tr>
<td>[0.2 0.2 0.18 0.18 0.12 0.12]</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>[0.2 0.15 0.15 0.15 0.2 0.15]</td>
<td>0.0939</td>
<td>0</td>
</tr>
<tr>
<td>[0.18 0.18 0.17 0.17 0.15 0.15]</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td>[0.2 0.2 0.1 0.1 0.2 0.2]</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td>[0.19 0.18 0.17 0.16 0.15 0.15]</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td>[0.2 0.18 0.17 0.16 0.15 0.14]</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td>[0.205 0.15 0.18 0.17 0.15 0.145]</td>
<td>0.162</td>
<td>0.196</td>
</tr>
</tbody>
</table>

Table 4.7: Conditional incentive for Kemeny-Young with $N = 10000$ voters

### 4.2.3 Sampling Model

The sampling model assumes a finite number of voters $N$, and each individual voter holds a fixed preference. Therefore, unlike the multinomial model, the vote distribution is exactly the social bias $p \in \mathcal{P}_N$. However, voters do not know $p$ completely. Instead, they observe $p^s$, the profile of a small group of randomly chosen voters. Suppose the group size is $N'$, $N' \ll N$. The sample follows a multinomial distribution $\text{Multi}(p, N_0)$. A rational voter would then estimate which boundary $p$ most probably falls on and would then determine her strategy of manipulation accordingly.

To find the closest boundary to $p^s$, we are faced with the problem of choosing a prior distribution for $p$. A prior that give more weight to the center (equiprobable for all votes) might be preferred (in real elections, people do often see close matches rather than landslide victories), but for now we assume uniform prior for generality and lack of good model for non-uniform priors. Thus maximizing $P(p|p^s)$ is equivalent to maximizing $P(p^s|p) \propto 2^{-N'D(p^s||p)}$. We would like to find a distribution $\bar{p}$ that maximizes the KL-Divergence $D(p^s||p)$ among all $p \in T$:

$$
\bar{p} = \arg\max_{p \in T} D(p^s||p) = \arg\min_{p \in T} - \sum_{i=1}^{6} p^s_i \log p_i \tag{4.27}
$$
We first consider the general case. The boundary $T$ is defined by an equation $h(p) = 0$. For now, put the inequality constraints aside and assume that the information projection is on the interior of a boundary piece. Since log is a concave function, we are solving the constrained convex program for $p$:

$$\text{minimize} \quad - \sum_{i=1}^{6} q_i \log p_i$$  \hspace{1cm} (4.28)

subject to \( h(p) = 0 \)  \hspace{1cm} (4.29)

$$\sum_{i=1}^{6} p_i = 1$$  \hspace{1cm} (4.30)

We can use the KKT condition to get a necessary condition for the optimal solution $\bar{p}$. Let $\lambda_b$ and $\lambda_p$ be the Lagrange multipliers corresponding to the boundary restriction (4.29) and the probability simplex restriction (4.30). To get rid of the minus sign, we flip the sign of the Lagrange multipliers.

$$\frac{q_i}{p_i} - \frac{\partial h}{\partial p_i} \cdot \lambda_b - \lambda_p = 0, \quad i = 1, \ldots, 6$$  \hspace{1cm} (4.31)

or, after rearrangement,

$$p_i = \frac{q_i}{h' \lambda_b + \lambda_p}$$  \hspace{1cm} (4.32)

where $h' = \partial h/\partial p_i$. The 6 equations of (4.31) together with (4.29) and (4.30) form a system of 8 equations and 8 unknowns. However, with a general $h$ it is difficult to find the analytical solution. The values of $\lambda_b$ and $\lambda_p$ could be found by applying numerical methods. Using (4.32), we will show that a small shift of the social bias parallel to the boundary defined by $h(p)$ does not change the KL-Divergence $D(p||q)$. 

64
\[\frac{P(p + \Delta_{ij}|p^s)}{P(p|p^s)} = \frac{P(p^s|p + \Delta_{ij})}{P(p^s|p)} = \left(\frac{p_i + 1/N}{p_i}\right)^{p_i^K} \left(\frac{p_j - 1/N}{p_j}\right)^{p_j^K} \approx 1 \quad (4.33)\]

For the sampling model, the change of probability is much less sensitive to perturbations on the social bias. Therefore, the conditional incentive can be calculated by sampling at \(\bar{p}\) alone.

### 4.3 2012 Presidential Election Survey

In our 2012 US Presidential Election Survey, we asked each participant to evaluate 11 Republican, Democrat and independent candidates through an online survey form. Although Hilary Clinton was not a candidate for the 2012 election, she was also included because of her popularity. The 11 candidates were presented to participants in randomized order. Each participant was asked to assign a numerical score (on a scale from 0 to 100) to each candidate. A 100 indicates the strongest support for that candidate to be president, and 0 indicates the strongest opposition. The participants had the choice not to rate candidates they did not know or did not want to evaluate.

The participants also, on a voluntary basis, provided basic demographic information (age, gender, level of education, state of residence, party affiliation and so on) for better analysis of the data. We recruited participants through three channels:

1. Google Adwords and FiveThirtyEight.com. Our link was shown to random Google users and NYTimes (FiveThirtyEight blog) readers in the US. Users who clicked on our ads were taken to our online survey.
2. Mercer County Panel. The respondents were residents of Mercer County, New Jersey.

3. Amazon Mechanical Turk. The respondents were Amazon Mechanical Turk users living in the US.

We collected 1,650 valid responses in total (responses that rated at least one candidate were considered valid). Among the respondents, 359 identified themselves as Republicans, 391 as Democrats, and 537 as independent.

<table>
<thead>
<tr>
<th>Method</th>
<th>Valid Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google/FiveThirtyeight</td>
<td>887</td>
</tr>
<tr>
<td>Mercer County Panel</td>
<td>291</td>
</tr>
<tr>
<td>Mechanical Turk</td>
<td>472</td>
</tr>
<tr>
<td>Total</td>
<td>1650</td>
</tr>
</tbody>
</table>

Based on the collected data, we ranked the candidates using five different ranking methods: Kemeny-Young, Borda count, range voting, instant run-off voting, and plurality voting. Except for range voting, all scores are first converted to rankings. Kemeny-Young and single run-off voting are compatible with partial rankings, whereas for Borda count, we assign all unranked candidates the “average score”. For example, if someone rates Obama 80, Romney 75, Ron Paul 72 and leaves all other candidates unrated, then a score of 3 is assigned to Obama, 1 is assigned to Ron Paul and 2 is assigned to everyone else (including Romney). Thus a voter rating more candidates contributes more information, which hopefully gives voters a reason to rate as many candidates as possible. The outcomes with different methods are very similar, but not exactly the same. Only Borda count puts Mitt Romney at the top, while all other methods decide on Barack Obama as the winner.

As in previous sections, we plot the results in the pairwise comparison space. In Figure 4.7 each dot represents the vote profile for one subset of three candidates.
<table>
<thead>
<tr>
<th>Rank</th>
<th>Kemeny-Young</th>
<th>plurality</th>
<th>Borda count</th>
<th>Run-off</th>
<th>Range Voting</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Barack Obama</td>
<td>Barack Obama</td>
<td>Mitt Romney</td>
<td>Barack Obama</td>
<td>Barack Obama</td>
</tr>
<tr>
<td>2</td>
<td>Hillary Clinton</td>
<td>Mitt Romney</td>
<td>Barack Obama</td>
<td>Mitt Romney</td>
<td>Hillary Clinton</td>
</tr>
<tr>
<td>3</td>
<td>Mitt Romney</td>
<td>Ron Paul</td>
<td>Hillary Clinton</td>
<td>Ron Paul</td>
<td>Mitt Romney</td>
</tr>
<tr>
<td>4</td>
<td>Ron Paul</td>
<td>Hillary Clinton</td>
<td>Ron Paul</td>
<td>Hillary Clinton</td>
<td>Ron Paul</td>
</tr>
<tr>
<td>5</td>
<td>Jon Huntsman</td>
<td>Newt Gingrich</td>
<td>Jon Huntsman</td>
<td>Newt Gingrich</td>
<td>Newt Gingrich</td>
</tr>
<tr>
<td>6</td>
<td>Rick Santorum</td>
<td>Rick Santorum</td>
<td>Rick Santorum</td>
<td>Rick Santorum</td>
<td>Jon Huntsman</td>
</tr>
<tr>
<td>7</td>
<td>Newt Gingrich</td>
<td>Jon Huntsman</td>
<td>Newt Gingrich</td>
<td>Jon Huntsman</td>
<td>Rick Santorum</td>
</tr>
<tr>
<td>8</td>
<td>Rick Perry</td>
<td>Rick Perry</td>
<td>Rick Perry</td>
<td>Rick Perry</td>
<td>Rick Perry</td>
</tr>
<tr>
<td>9</td>
<td>Michele Bachmann</td>
<td>Michele Bachmann</td>
<td>Gary Johnson</td>
<td>Michele Bachmann</td>
<td>Gary Johnson</td>
</tr>
<tr>
<td>10</td>
<td>Gary Johnson</td>
<td>Gary Johnson</td>
<td>R. Lee Wrights</td>
<td>Gary Johnson</td>
<td>Michele Bachmann</td>
</tr>
<tr>
<td>11</td>
<td>R. Lee Wrights</td>
<td>R. Lee Wrights</td>
<td>Michele Bachmann</td>
<td>R. Lee Wrights</td>
<td>R. Lee Wrights</td>
</tr>
</tbody>
</table>

Table 4.8: Voting outcome of different voting systems based on survey data.

Figure 4.7: Condorcet paradox never occurs for our collected voting data.

There are 11 candidates, hence \( \binom{11}{3} = 165 \) different subsets. The two dark tetrahedrons denote the regions of Condorcet paradox. We reordered the candidates so that all dots fall in the same sub-cube \( u_{ab} > 0.5, u_{bc} > 0.5, u_{ca} < 0.5 \). Therefore, the same candidate, e.g. Mitt Romney, may be attached to the name \( a \) in one dot but attached to the name \( b \) in another dot. This, however, does not affect the cycle-free property of the voting profile. If we had not done so, the dots would scatter around the space, but still would not be inside the Condorcet paradox region. By sorting it is easier to see that no dots falls inside the “Condorcet paradox” region.
Some historical voting data report the same finding that real elections are free of Condorcet paradoxes.

Figure 4.8: Kemeny-Young is mostly immune to strategic voting for our collected voting data

What can we read from these data pertaining to strategic voting? In the last section we showed that for the Kemeny-Young method, all manipulable boundaries are inside the region of Condorcet paradox. It is reasonable to suspect that most data points are closer to non-manipulable boundaries than manipulable boundaries. In Figure 4.8, we plotted the log-distance (using the KL-divergence as the distance metric) of each sample to its maximum likelihood prior on the manipulable and non-manipulable boundaries, respectively. The log-scale is used to prevent the data points from clustering near the origin. The figure shows that all but a few profiles are closer to a non-manipulable boundary than a manipulable boundary. This suggests that for a real world election using the Kemeny-Young method, the vote count is more likely to be on a non-manipulable boundary than on a manipulable boundary.

For comparison, we also show the data points with the Borda count and plurality in Figure 10(a) and 10(b), respectively. We plot against the boundary set of Borda count $T^B$ in Figure 4.9(a). Most data points locate in the winning region of $a,$
but there are also some data points located along the boundary between \( a \) and \( b \). Therefore, for Borda count not only is the conditional incentive high, but the absolute probability (measured by the mean effect) is also likely to be high in real elections. The data points are plotted in a different space for plurality in Figure 10(b). The three axes denote the portion of the plurality votes that go to \( a \), \( b \) or \( c \), respectively. The polytope inside the cube is the vote simplex, and all data points reside on the simplex. Different colors (or shades of gray) denote different winning regions.

![Diagram of Borda count and plurality](image)

(a) Borda count  
(b) Plurality

Figure 4.9: Election survey data for Borda count and plurality

### 4.4 Concluding Remarks

It is difficult to verify whether in real elections votes are subject to a multinomial distribution. However, for Condorcet methods, the categorical difference between “manipulable boundary” and “non-manipulable boundary” applies universally. Given the theoretical calculation and simulation results under the multinomial model, we believe that the Kemeny-Young method, and Condorcet methods in general are less manipulable than competing methods.
The concept of strategic voting is used in different contexts. First, ballots may not always be in the form of a (partial or complete) rank-ordering. For example, Range voting and Majority judgement requires a voter to assign an “evaluation”, in the form of a numerical value, to each candidate. Balinski and Laraki [3] argue that Majority judgement is “strategy-proof” in the sense that each voter cannot improve the status of a candidate if that candidate’s score (median of every voter’s evaluation) is lower than her sincere evaluation. However, if we convert the scores of each candidate into a ranking (or pick the top scorer), then a voter can still help her favourite by lowering the score of other candidates. Range voting, in contrast is very susceptible to strategic voting. Voters will exaggerate the difference between their top choice and the bottom choice, to the point that it virtually becomes approval voting. For approval voting, each voter converts a preferential order into a binary opinion for each candidate: Approve or disapprove. The choice of the approval cut-off can be tactical. But it is debatable whether this should be considered as strategic voting. In sum, the exact meaning of strategic voting is highly dependent on the voting rules.
Chapter 5

Dynamics of strategic voting

5.1 Introduction and Motivation

Given the voting rule, a voter knows the outcome her vote change leads to provided that she knows how everyone else’s voted. This “perfect knowledge” assumption is essential to Gibbard-Satterthwaite theorem and many other related theorems on strategic voting. Alternatively, one may apply game-theoretic considerations to strategic voting and model the election as a dynamic process, where each voter acts to maximize their gain, which in this scenario is the chance to elect their favorite candidates.

In this chapter, we investigate the collective behavior of manipulative voters when they strategize based on imperfect knowledge of the social preference. All voters may update their votes based on their noisy perception of the current social preference. This causes the temporary vote profile of the population to change as voters make strategic adjustments. Vote changes are asynchronous and without collaboration. Broadly speaking there are two types of outcome: 1) the profile keeps drifting without reaching an equilibrium 2) the social preference eventually reaches a Nash equilibrium, at which point no voter benefits from unilaterally changing her own vote.
We investigate plurality rule and Borda count in this chapter because they are widely used and have relatively simple geometric representation. We make a few assumptions to enable mathematical analysis:

1. Each voter holds a sincere preference order of the candidates, in the form of a ranking. No partial ranking or indifference is allowed.

2. Voters make independent, noisy observations of the vote profile.

3. Voters do not share their information with other voters and do not form coalitions.

Voters’ information on the social preference may be obtained from public data, such as surveys, polls and social networks. Consider a plurality vote with three candidates $a$, $b$ and $c$. A survey shows that the current support rate for each candidate is 40% for $a$, 38% for $b$ and 22% for $c$. The credibility of this survey might suffer from small poll size, biased sampling method, etc. However, if a voter with honest preference $c > b > a$ believes this polling result to be fairly accurate, then she has the incentive to vote for $b$ instead of $c$. Given the polling result, a tie between $a$ and $c$ is very unlikely. If she votes for $b$, then her vote has a much greater likelihood to be the tie-breaker between $a$ and $b$. A voter whose top choice is $a$ or $b$ will, in contrast, stick to their honest vote because the strategic choice coincides with their honest preference. Eventually, the unfortunate candidate $c$ loses almost all votes to $a$ and $b$, and the winner is the pairwise majority between $a$ and $b$.

The above intuitive argument is essentially in agreement with what Duverger’s Law [17] suggests: plurality system marginalizes smaller political parties and results in a two-party system. However, Duverger’s Law is applicable to the scenarios where the gap of support levels between the middle and bottom-ranked candidate is large. When $b$ and $c$ share similar popularity, people are less certain about the advantage of being dishonest.
If such a process ever stabilizes, then no unilateral change of vote leads to a better outcome for any voter. In other words, the points of stabilizations are Nash equilibria. Do such Nash equilibria exist in all voting systems and in all distributions of honest preferences? Moreover, can they always be reached regardless of the sequence of voters? We will apply this analysis to several common voting systems, including plurality and Borda Count.

5.2 Related work

5.2.1 Works on strategic voting

Maurice Duverger suggested that plurality voting system favors a two-party system in the 1950s [16] [17]. This effect is referred to as Duverger’s law nowadays. However, Duverger’s law is more a prediction on the tendency towards bipartisanship rather than a precisely stated mathematical problem. There are formulations of the problem where Duverger’s law fails. For example, David Myatt [31] shows that stable multicandidate support may arise when voters observe private information and act without complete coordination.

Applying Nash equilibrium to the analysis of strategic voting sounds appealing at first. However, due to the very low ratio of pivotal profiles (which tends to zero as the number of voters increases), a single-voter vote change most likely would not change the outcome. Therefore, a naive application results in Nash equilibrium almost everywhere in the profile simplex. Meir et al. [28] addresses this issue by adding some uncertainty to each voter’s belief. A voter seeks a way to vote that dominates all other votes within a certain radius to the current score profile that the voter believes. The vote status is said to be at equilibrium if no voter has an action that dominates her current action. It is shown that if voters act one by one (no simultaneous changes),
then a voting equilibrium will always be reached in any sequence of actions starting from the honest profile.

Other works exploring the dynamics of strategic voting for plurality include [29][8]. There are also works considering coordination or collaboration among voters [34][20]. Pierre Favardin et al. [21] considers the scenario where there are communications and collective strategies either among all voters, or among voters with the same preference. They concluded that Borda Count is the least manipulable among scoring rules.

5.2.2 Relationship with statistical physics

In our analysis of the plurality vote, the least popular candidate loses all votes in the long run, at a rate dependent on both the current vote distribution and the observation noise. A similar analysis can be found in [13].

5.3 Modelling for 3 candidates

We design the model with the equality of voters in mind, both in terms their influence on the outcome and their information source. Consider an election with $N$ voters and 3 candidates. Since we exclusively investigates scale invariant voting rules, we treat a profile as a point on the 5-simplex $\mathcal{P} \subseteq \mathbb{R}^6$. Any deterministic scale invariant voting system can be represented by a partition of the simplex $\mathcal{P}$, as outlined in Chapter 3. List the 6 rankings in the following order:

<table>
<thead>
<tr>
<th>vote</th>
<th>$i$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>abc</td>
<td>1</td>
<td>abc</td>
<td>acb</td>
<td>cab</td>
<td>cba</td>
<td>bca</td>
<td>bac</td>
</tr>
<tr>
<td>acb</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cab</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cba</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bca</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bac</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1: The order of rankings throughout the paper

For example, out of 1000 voters, 450 hold preference order $abc$, 350 hold $bac$ and 200 $cab$. The profile is represented with $v = [0.45, 0, 0.2, 0, 0, 0.35] \in \mathcal{P}$. However,
if we consider voters’ sincere and declared (strategic) preferences, we need to divide voters into 36 types. To fully represent the vote status, we define a $6 \times 6$ matrix $V \in \mathbb{R}^{6 \times 6}$, where $V_{ij} \in [0, 1]$ is the ratio of voters whose honest vote is $i$ and current vote is $j$. Therefore, the row sum of $V$ is the honest profile $v^h$, and the column sum is the current vote profile $v$. All entries of $V$ sum up to 1. If we choose a voter randomly, then $V_{ij}$ is the probability of choosing a voter with honest preference $i$ and current vote $j$. The current vote status $V$ and current profile $v$ changes with time, but the honest profile does not change. We use the superscript $t$ as the time index.

Let $1 \in \mathbb{R}^{6 \times 1}$ denote the all-one vector, then $v^h$ and $v^t$ can be represented by:

\begin{align*}
    v^h &= V^t \cdot 1 \text{ for any } t \quad (5.1) \\
    v^t &= 1^T \cdot V^t \quad (5.2) \\
    \sum_{i,j} V^t_{ij} &= 1 \text{ for all } t \quad (5.3)
\end{align*}

Recall from Chapter 3 that a profile is pivotal if there exists a single vote change $\delta$ such that $f(v + \delta) \neq f(v)$, where $f(v)$ is the winner under voting rule $f$ given profile $v$. Denote the set of all pivotal profiles by $B$, and the subset of the boundary that separates the winning regions of $a$ and $b$ by $B_{ab}$.

The process of evolution can be summarized in 4 steps:

\subsection*{5.3.1 Initialization}

Suppose the vote status is initialized with everyone voting honestly: At time $T = 0$, $V^0 = D(v^h)$, where $D : \mathbb{R}^6 \mapsto \mathbb{R}^{6 \times 6}$ is a diagonal matrix with the input as its diagonal. It then follows that $v^0 = v^h$.  

75
5.3.2 Forming strategies

At time $t$, a voter is chosen uniformly randomly. We use two random variables $h^t \in \{1, \ldots, 6\}$ to denote her honest preference and $c^t \in \{1, \ldots, 6\}$ to denote her current strategic vote. Therefore $V^t$ is the joint distribution for $(h^t, c^t)$. The voter receives a signal $p^t$, which is a noisy observation of the current profile $v^t$. Conditioned on $v^t$, the observation is independent of the random voter’s honest preference or current vote.

The voter then strategically updates her vote based on this observation. Since her vote only makes a difference when the profile happens to locate on a boundary, the voter’s strategic vote will depend on the conditional distribution of the vote profile restricted to the boundaries. We assume a uniform prior and that the observation is dependent on the current vote status through a distribution $P(p|v)$. Some examples of $P(p|v)$ are:

1. **Unbiased sampling model.** The observation is a small unbiased sample of votes. This model corresponds to observations obtained by surveys and polls. Suppose the number of samples is $N_s$, $N_s < N$, then the empirical distribution is $p \sim \frac{1}{N_s} \cdot \text{Multi}(v, N_s)$. Note that this result applies to independent sampling without replacement. However, when $N \gg N_s$, the difference between sampling with and without replacement is small.

2. **Laplace noise model.** Independent and identically distributed Laplace random variables are added to each entry of the profile. The Laplace distribution is a continuous distribution with probability density function $p(x) = \frac{1}{2b} \exp\left(-|x - \mu|/b\right)$. For our problem we set $\mu = 0$. The noise is independent of $v^t$.

The rationale for using Laplace noise is as follows: Suppose the party conducting the poll has collected every voter’s ballot, and wants to reveal the result to the
public. However, by doing so they run the risk of revealing the individual selections of those surveyed. One popular notion of privacy is the differential privacy [18], which aims to disclose general non-specific information while hiding information about any specific voter. It is shown that $\epsilon$-differential privacy can be achieved by adding Laplacian noise to the query results [18].

3. **Gaussian noise model.** Gaussian noise is a natural choice for modelling noises. In the context of differential privacy, the slightly relaxed $\epsilon, \delta$-differential privacy can be achieved by adding Gaussian noise [18].

*Remark:* By adding Laplace or Gaussian noise, we get a “profile” with all entries not adding up to 1. However, many voting rules can be extended to admit such profiles, including all scoring rules.

Recall that in Definition [3.4] we use $B_{a,b|f}$ to denote the boundary set that separates the winning region of $a$ and $b$. We omit the voting rule $f$ in the subscript when the voting rule is clear from context. The knowledge of $P(p'|v^t)$ and the prior $P(v)$ allows us to find $P(v^t|p^t)$. Therefore, the voter finds the boundary profile that maximizes $P(p'|v)$ over all $v$ on the boundary, which, under the uniform prior assumption, is the same as maximizing $P(v|p^t)$:

$$ v' = \arg\max_{v' \in B} P(v|p^t) \quad (5.4) $$

The voter then form the strategy as if $v'$ is the actual profile.

Denote by $d$ a distance function that measures the distance from one profile to another: $d: \mathcal{P} \times \mathcal{P} \mapsto \mathbb{R}$. We do not require $d$ to be symmetric (thus $d$ is not necessarily a distance metric). We say that a distance function $d(p, v)$ is compatible with a distribution $P(p|v)$ if (5.5) is satisfied:

$$ P(p|v) \leq P(q|v) \text{ if and only if } d(p, v) \geq d(q, v) \quad (5.5) $$
Lemma 5.1. 1. The Laplace noise model is compatible with the $l_1$ distance.

2. The Gaussian noise model is compatible with the Euclidean distance.

3. As the number of voters increases, The unbiased sampling model is asymptotically compatible with the KL-divergence distance function as the number of samples grow.

Proof. First consider the Laplace noise model. Let $X_1, \ldots, X_6$ be iid Laplace random variables with mean 0 and variance $2\sigma^2$. Then the density function $f(p|v)$ is given by

$$f(p|v) = \prod_{i=1}^{6} f(X_i) = (2\sigma)^{-6} \exp\{-\frac{||p - v||_1}{\sigma}\}$$  \hspace{1cm} (5.6)

Therefore, the Laplace noise model is compatible with the $l_1$ distance function $d(p, v) = ||p - v||_1$.

If $X_1, \ldots, X_6$ are i.i.d. Gaussian random variables with mean 0 and variance $\sigma^2$, then the density function $f(p|v)$ is

$$f(p|v) = (2\pi\sigma^2)^{-3} \exp\{-\frac{||p - v||_2^2}{2\sigma^2}\}$$  \hspace{1cm} (5.7)

Therefore, the Gaussian noise model is compatible with the $l_2$ distance function $d(p, v) = ||p - v||_2$.

Now consider the unbiased sampling model. By Sanov’s theorem \cite{37, 12}, $P(p|v) \approx 2^{-N_s D(p||v)}$ for large $N_s$, where $N_s$ is the number of samples.

Remark: KL-Divergence is not a metric since it is not symmetric: in general $D(p||q) \neq D(q||p)$. However, there is a statistical meaning to this quantity: $2^{-ND(p||v)}$ gives the order of the probability that an empirical distribution $p$ is observed when $N$ samples are drawn independently with distribution $v$. This is a good approximation.
of polling with a large population base. Another important property of the KL-
Divergence is that $D(p||q) = \infty$ if $p_i > 0$, $q_i = 0$ for some $i = 1, \ldots, 6$. The $l_1$
distance also has a voting theoretical interpretation: $N/2 \cdot ||p - q||_1$ is the minimum
number of vote changes required to change the profile $p$ to $q$.

$$D(p||q) = \infty$$

$$||p - q||_1$$

$$\frac{N}{2} \cdot \text{minimum number of vote changes}$$

Figure 5.1: Observation $p^t$ and its projection to the plurality boundary

Denote the vote change as $\Delta(h_t, c_t, p^t) \in \mathbb{R}^{6 \times 6}$, where $(h_t, c_t)$ are as defined before.
If, for example, a voter with honest preference $abc$ changes her vote from $bac$ to $abc$
based on her observation, and $N$ is the total number of voters, then $\Delta^t(1, 6, p^t)$ is a
matrix with $\Delta_{1,6}^t = -1/N$, $\Delta_{1,1}^t = 1/N$ and all other entries 0. The **expected vote change** is the expectation of $\Delta^t$ over all voters, which is given by

$$F(V^t, p^t) = \sum_{i,j} V_{i,j} \Delta(h_t, c_t, p^t)$$ (5.8)

Table 5.2 shows the strategic vote given the closest boundary for plurality. When
a voter has more than one strategic choice, she chooses the one closest to her honest
preference. For example, if a voter with sincere preference $abc$ actually votes for $b$,
she reports her vote as $bac$ instead of $bca$. 

79
5.3.3 Feedback to the social expectation

We have the update equations:

\[ V^0 = D(v^h) \]
\[ V^{t+1} = V^t + \Delta(h^t, c^t, p^t) \quad (5.9) \]

At time \( t + 1 \), another voter of type \((h^{t+1}, c^{t+1})\) is drawn with distribution \( V^{t+1} \), makes an observation \( p^{t+1} \sim P(p|v^{t+1}) \) and updates her vote accordingly.

5.3.4 Equilibrium

**Definition 5.2.** We say that the current vote status \( V^t \) is at an equilibrium for the noise model \( P(p|v) \) if \( \Delta(i, j, p^t) = 0 \) for all \( i, j \) and \( p^t \) such that \( P(p^t|v^t) > 0 \)

The vote status \( V^t \) stops changing once reaching an equilibrium. When the vote status reaches such a point, no voter finds it beneficial to change her vote unilaterally under any observation with a positive probability. This is a very strong notion of equilibrium. However, with unbounded noise models like Laplace or Gaussian, no equilibrium can be ever achieved since any profile has a positive probability of being observed given any current vote status. For such noise models we need a weaker notion of equilibrium. The exact meaning of “weaker” equilibrium will be discussed in later sections.
5.4 Plurality

For plurality vote with 3 candidates, the boundary can be divided into three convex subsets: $B_{ab}$, the boundary between $a$ and $b$, $B_{bc}$, the boundary between $b$ and $c$, and $B_{ac}$, the boundary between $a$ and $c$. We can plot the voting system on a 2-simplex, with each point on the simplex corresponding to a plurality profile, defined as follows:

**Definition 5.3.** Given the profile $p$, the **plurality profile** is a vector $(p_a, p_b, p_c) \in \Delta^2$, where $p_a = p_1 + p_2$, $p_b = p_5 + p_6$ and $p_c = p_3 + p_4$ are the percentage of votes that rank $a$, $b$ and $c$ first respectively, and $\Delta^2$ is the 2-simplex:

\[
\Delta^2 = \{ (x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_1 + x_2 + x_3 = 1 \} \quad (5.10)
\]

Using the plurality profile, the boundaries can be represented succinctly:

\[
B_{ab} = \{ p : p_a = p_b \geq \frac{1}{3} \} \quad (5.11)
\]

where $B_{ac}$ and $B_{bc}$ are similarly defined.

5.4.1 Laplace noise model

**Differential Privacy**

We begin the discussion by introducing some basic notions of differential privacy. This discussion will help us choose a meaningful noise power. Let $h$ be a query function that maps a dataset to a real vector. Define the $l_1$- and $l_2$-sensitivity:

**Definition 5.4.** For $h : \mathcal{D} \mapsto \mathbb{R}^d$, the $l_1$-sensitivity of $f$ is

\[
\Delta h = \max_{D_1, D_2} || f(D_1), f(D_2) ||_1 \quad (5.12)
\]
for all $D_1, D_2$ differing in at most one element. Similarly, the $l_2$-sensitivity of $f$ is

$$\Delta h = \max_{D_1, D_2} ||f(D_1), f(D_2)||_2$$

(5.13) for all $D_1, D_2$ differing in at most one element.

Suppose the query function $h$ returns the anonymous profile. The $l_1$ sensitivity of $h$ is the maximum $l_1$ distance between any two neighboring profiles, which is $2/N$. Similarly, the $l_2$-sensitivity of $h$ is $\sqrt{2}/N$. It is shown in [18] that the $\epsilon$-privacy, defined below could be achieved by adding Laplace noise of variance $2\sigma^2$ to the query function $h$, such that $\sigma \geq \Delta h/\epsilon$.

**Definition 5.5.** [18] A random algorithm $\mathcal{M}$ satisfies $\epsilon$-differential privacy if for any neighboring datasets $D_1$ and $D_2$, and any subset $S$ of possible outcomes $\text{Range}(\mathcal{M})$,

$$\Pr[\mathcal{M}(A) \in S] \leq \exp(\epsilon) \times \Pr[\mathcal{M}(A') \in S].$$

(5.14)

Intuitively, smaller $\epsilon$ means smaller difference in the distribution before and after the vote change. To maintain the level of privacy, the algorithm only needs to add noise with standard deviation inversely proportional to the number of voters.

**Error Analysis**

We derive the probability that a voter makes a wrong judgement on the closest boundary given the current vote status and the noise distribution. Note that the error probability does not depend on the honest preference of the voter, but the voter’s action upon getting the observation does depend on her honest preference.
Lemma 5.6. The $l_1$ distances from a profile $p$ to the three boundaries can be expressed in terms of the plurality profile. Assume $p_a \geq p_b \geq p_c$, we have

\begin{align*}
d_1(p, B_{ab}) &= p_a - p_b \\
d_1(p, B_{ac}) &= \begin{cases} 
    p_a - p_c & \text{if } p_b < \frac{1}{3} \\
    \frac{2}{3} - 2p_c & \text{if } p_b \geq \frac{1}{3}
\end{cases} \\
d_1(p, B_{bc}) &= \begin{cases} 
    2p_a - \frac{2}{3} & \text{if } p_b < \frac{1}{3} \\
    \frac{2}{3} - 2p_c & \text{if } p_b \geq \frac{1}{3}
\end{cases}
\end{align*}

The proof is straightforward. From Lemma 5.6 we can conclude that for plurality profile satisfying $p_a \geq p_b \geq p_c$, $B_{ab}$ is the closest boundary while $B_{bc}$ is the farthest one:

Lemma 5.7. A plurality profile $p = (p_a, p_b, p_c)$ with $p_a \geq p_b \geq p_c$ satisfies

\begin{align*}
d_1(p, B_{ab}) &\leq d_1(p, B_{ac}) \leq d_1(p, B_{bc}) \tag{5.19}
\end{align*}

Proof.

\begin{align*}
d_1(p, B_{ac}) - d_1(p, B_{ab}) &= \begin{cases} 
    p_b - p_c \geq 0 & \text{for } p_b < \frac{1}{3} \\
    p_b - p_c + (p_b - \frac{1}{3}) \geq 0 & \text{for } p_b \geq \frac{1}{3}
\end{cases} \\
d_1(p, B_{bc}) - d_1(p, B_{ac}) &= \begin{cases} 
    \frac{1}{3} - p_b \geq 0 & \text{for } p_b < \frac{1}{3} \\
    0 & \text{for } p_b \geq \frac{1}{3}
\end{cases}
\end{align*}
From the Lemma it follows that $B_{ab}$ is the closest boundary iff $p_c \leq \min(p_a, p_b)$.

Now we add Laplace noise to the profile $p$. Denote by $X = (X_1, \ldots, X_6)$ the noises added to the profile $v$, such that $X_i$'s are independent Laplace random variables with zero mean and variance $2\sigma^2$.

**Theorem 5.8.** Let $(v_a, v_b, v_c)$ denote the plurality profile of $v$. Suppose $v_c < \min(v_a, v_b)$. Let $\Delta v = \min(v_a, v_b) - v_c$ and $p = v + X$. Also, suppose $\sigma$ is a function of $N$ with $\lim_{N \to \infty} \sigma = 0$. Then there exists a large enough $N_1$ such that for all $N > N_1$,

$$P(p \sim B_{ab}) \geq 1 - C \cdot \sigma^5 (\Delta v)^3 e^{-\Delta v / \sigma}$$

(5.22)

for $C > \frac{1}{144}$. Specifically, if we choose $\sigma = 1/\epsilon N$, then the bound becomes

$$P(p \sim B_{ab}) \geq 1 - C \cdot (\Delta v)^3 (\epsilon N)^{-5} e^{-\Delta v \epsilon N}$$

(5.23)

**Proof.** The plurality profile of the observation is

$$p_a = v_a + X_1 + X_2$$
$$p_b = v_b + X_5 + X_6$$
$$p_c = v_c + X_3 + X_4$$

(5.24)
Therefore

\[ P(p \sim B_{ab}) = P(v_c + X_3 + X_4 < \min(v_a + X_1 + X_2, v_b + X_5 + X_6)) \geq 1 - P(v_c + X_3 + X_4 \geq v_b + X_5 + X_6) - P(v_c + X_3 + X_4 \geq v_a + X_1 + X_2) \geq 1 - 2P(X_3 + X_4 - X_5 - X_6 \geq \Delta v) \] (5.25)

Since a zero mean Laplace distribution with variance \(2\sigma^2\) is the difference of two iid exponential distribution with mean \(\sigma^{-1}\), the sum of \(n\) iid Laplace random variables with mean 0 and variance \(2\sigma^2\) can be treated as the difference of two iid Erlang random variables with probability density function \(\frac{1}{6\sigma^2}x^3e^{-x/\sigma}\). Therefore the probability density function of \(Y = X_3 + X_4 - X_5 - X_6\) is

\[
f_Y(x) = \frac{1}{36} \int_0^{+\infty} (u + |x|)^3 e^{(-u - |x|)/\sigma} u^3 e^{-u/\sigma} du = \frac{1}{36e|x|/\sigma} \int_0^{+\infty} (u^2 + u|x|)^3 e^{-2u/\sigma} du = \frac{1}{36e|x|/\sigma} g_3(|x|) \] (5.26)

where \(g_3\) is a cubic polynomial whose coefficient of the cubic term is \(\int_0^{+\infty} u^3 e^{-2u/\sigma} du = \sigma^4/8\). Therefore,

\[
P(Y \geq \Delta v) = \int_{\Delta v}^{\infty} f_Y(x) dx = \frac{1}{36} \int_{\Delta v}^{\infty} e^{-x/\sigma} g_3(x) dx = \frac{1}{288} \sigma^5 (\Delta v)^3 e^{-\Delta v/\sigma} + (\Delta v)^3 O(\sigma^5 e^{-\Delta v/\sigma}) \] (5.27)
We can then derive (5.22) and (5.23) from (5.25) and (5.27).

To simplify the expression, we define
\[
g(\Delta v, \sigma) = \sigma^5 (\Delta v)^3 e^{-\Delta v/\sigma}
\]
for \(\Delta v > 0, \sigma > 0\). By taking the partial derivatives with respect to \(\sigma\) and \(\Delta v\), we can see that it is an increasing function of \(\sigma\), and attains maximum at \(\Delta v = 3\sigma\) when \(\sigma\) is fixed.

Recall the update equations
\[
V^0 = D(v^h)
\]
\[
V^{t+1} = V^t + \Delta(h^t, c^t, p^t)
\]
(5.28)

Let \(\{\mathcal{F}_t\}\) be the filtration generated by \(\{V^t\}\). We will show that for large enough \(N\), the profile converges to the instant run-off voting outcome after \(O(Nv_c \log(Nv_c))\) vote changes, where \(v_c\) is the ratio of the plurality votes of the least popular candidate in the initial profile.

**Theorem 5.9.** Suppose candidate \(c\) wins the least plurality vote in the honest profile \(v^h\). Denote by \(\mathcal{P}_c(\epsilon) \subseteq \mathcal{P}\) the part of the simplex \(\mathcal{P}\) such that \(p_c \leq \epsilon\). The hitting time of \(\mathcal{P}_c(\epsilon)\) is \(O(Nv_c \log(Nv_c))\).

**Sketch of proof:** The proof technique is similar to [13]. We show that for a large enough number of voters, the plurality vote of the least popular candidate is a supermartingale with a strictly negative drift.

\[
\begin{align*}
- \mathbb{E}[v_c^{t+1} - v_c^t | \mathcal{F}^t] &= \frac{1}{N} \left[ v_c^t P(p^t \sim B_{a,b}) - v_c^t P(p^t \sim B_{b,c}) - v_c^t P(p^t \sim B_{a,c}) \right] \\
&\geq \frac{1}{N} \left[ v_c^t (1 - g(v_c^t, \sigma)) - (1 - v_c^t)g(v_c^t, \sigma) \right] \\
&= \frac{1}{N} \left[ v_c^t - g(v_c^t, \sigma) \right]
\end{align*}
\]
(5.29)
where $v^t_c \gg g(v^t, \sigma)$ for $v^t_c \geq \epsilon$. We then apply Doob decomposition and use martingale concentration results to bound the hitting time of $\mathcal{P}_c(\epsilon)$.

### 5.4.2 Four candidates or more

Suppose there are four candidates $a, b, c$ and $d$ in a plurality vote. There are $4! = 24$ types of preferences. Thus the profile space is a $23$-simplex, and the plurality profile is a four dimensional vector $p = (p_a, p_b, p_c, p_d)$. Without loss of generality we assume the honest profile satisfies $p_a \geq p_b \geq p_c \geq p_d$. That is, $a$ is the leading candidate and $b$ is the runner up, with $d$ at the bottom.

Now any voter who likes $a$ or $b$ most will stick to their honest vote. However, voters who like $c$ or $d$ best will find their favorite candidate very unlikely to win. Thus they will shift their support to either $a$ or $b$. For example, a voter with sincere preference $c > b > d > a$ probably wants to vote for $b$ instead if she learns that $a$ and $b$ are the top two candidates. However, if a voter has sincere preference $c > d > b > a$, it is less certain that she would like to vote for $b$, since $b$ is only her third most preferred candidate. In our analysis, we insist that voters always use the closest boundary to the estimated voting result.

For a plurality vote with 4 candidates, the boundary between $V_a$ and $V_b$ is represented by

$$B_{a,b} = \{p \in \mathcal{P} : p_a = p_b \geq 0.25\}$$

(5.30)

If a profile $p$ satisfies $p_a \geq p_b \geq p_c \geq p_d$, then the closest boundary is $B_{ab}$. There is a one-way flow of votes from $c$ and $d$ to $a$ and $b$. In expectation, once $a$ and $b$ are in lead, they will remain in lead until all supports for $c$ and $d$ are drained. The outcome is determined by the pairwise preference between $a$ and $b$. Therefore, strategic voting in effect turns instant run-off to a two-round run-off.
When there are only two candidates left, we just pick the one with more supporters. In fact, instant run-off and two-round system are equivalent when there are only 3 candidates. We do not perform explicit analysis but only makes a conceptual argument, as the analysis is very similar to the three candidate case.

5.5 Borda count

5.5.1 Direction of drift

We defined scoring rules in Definition 2.1. It is known that none of the scoring rules is also a Condorcet method [47]. However, as we show in this section, collective strategic behavior turns Borda count (Definition 2.5) into a Condorcet method. As shown in the last section, noise may sometimes leads to the wrong decision but overall does not affect the destination of the drift. In this section we do not incorporate the effect of noise in our analysis and focus on the expected direction of the drift of the profile.

We look at Borda count with 3 candidates. For each vote, Borda count assigns a score of 2 to the top candidate, a score of 1 to the middle-ranked candidate and no score to the bottom candidate. The total score for candidate $a$, $b$ and $c$, known as Borda score, are denoted by $S^B_a$, $S^B_b$ and $S^B_c$ respectively. To simplify the expressions, we normalize the Borda score by the number of voters. Thus $S^B_a + S^B_b + S^B_c = 3$. Denote the vote profile by $v = (v_1, \ldots, v_6)$ and the score of candidate $i$ given profile $v$ as $S^B_i(v)$. Then

$$S^B_a(v) = (2, 2, 1, 0, 0, 1) \cdot v$$  \hspace{1cm} (5.31)

$$S^B_b(v) = (1, 0, 0, 1, 2, 2) \cdot v$$  \hspace{1cm} (5.32)

$$S^B_c(v) = (0, 1, 2, 1, 0) \cdot v$$  \hspace{1cm} (5.33)
We omit the profile and simply write $S_i^B$ as the Borda score of candidate $i$ when the profile is clear from context. The boundary between $V_a$ and $V_b$ is

$$B_{ab} = \{ v \in V : S_a^B(v) = S_b^B(v) \geq S_c^B(v) \}$$

$$= \{ v \in V : (1, 2, 1, -1, -2, -1) \cdot v = 0,$$

$$(2, 1, -1, -2, -1, 1) \cdot v \geq 0 \}$$ (5.34)

The boundaries between $b$ and $c$, $c$ and $a$ are similarly defined. Note that we do not put non-negativity constraints on $v_1, \ldots, v_6$ since a noisy profile may contain negative values.

The $l_1$ distance from a profile to the closest boundary can be expressed in terms of Borda score. Assume that a profile $v$ satisfies $\min(S_a^B, S_b^B) \geq S_c^B$, and the number of voters is $N$. If we want to equalize $a$ and $b$’s Borda score using as few vote changes as possible, we should change $acb$ votes to $bca$ votes. Since each change decreases $S_a^B$ by $2/N$ and increases $S_b^B$ by $2/N$, the number of vote changes required is $N|S_a^B - S_b^B|/4$, where $N$ is the number of voters. Thus the $l_1$ distance from $v$ to the $ab$ boundary is

$$d_1(v, B_{ab}) = |S_a^B - S_b^B|/2$$ (5.35)

The projection of $v$ on the $ab$ boundary is

$$v^* = v + (0, -1, 0, 0, 1, 0) \cdot \frac{|S_a^B - S_b^B|}{4}$$

And the score profile of $v^*$ satisfies $S_a^B(v^*) = S_b^B(v^*) = (S_a^B(v) + S_b^B(v))/2$, $S_c^B(v^*) = S_c^B(v)$. It is easy to show that the distance from $v$ to $B_{ab}$ is smaller than the distance from $v$ to $B_{ac}$ and $B_{bc}$.

Table 5.3 shows the optimal strategic vote for Borda count given the closest boundary.
Same as the last section, we use a process \( \{V^t\} \), where \( V^t \in [0,1]^{6 \times 6}, t \in \mathbb{N} \), to represent the evolution of the vote status. The row sum of \( V^t \) is the honest profile \( v^h \), which is invariable with time, and the column sum of \( V^t \) is the current vote status \( v^t \) (see equations 5.3).

From the above table we can conclude the following:

**Lemma 5.10.** In the absence of noise, the bottom scorer’s Borda score always improves in expectation as a result of strategic voting. That is, if \( v^t \) satisfies \( S^B_c(v^t) \leq \min(S^B_b(v^t), S^B_a(v^t)) \), then

\[
\mathbb{E}[S^B_c(v^{t+1})|V^t] \geq S^B_c(v^t)
\]  

(5.36)

**Proof.** In the absence of noise, only one source of randomness remains: which voter is selected to update her vote. From Table 5.3 we can see that for a profile \( v^t \) (which is also the observation) that satisfies \( S^B_c(v^t) \leq \min(S^B_b(v^t), S^B_a(v^t)) \), voters with sincere preference \( abc \) or \( cab \) strategically change their votes to \( cab \) (if their current vote is not \( cab \)), and voters with sincere preference \( cba \) or \( bac \) strategically change their votes...
to bca if their current vote is not bca. Thus

\[
\mathbb{E}[S^B_c(v^{t+1})|V^t] - S^B_c(v^t)
\]

\[
= \sum_{i=1}^{6} (V^t_{i1} - V^t_{i3} - V^t_{i4} + V^t_{i6})
\]

\[
= \sum_{i=1}^{6} (v^t_1 - v^t_3 - v^t_4 + v^t_6)
\]

\[
=(S^B_a(v^t) + S^B_b(v^t) - 2S^B_c(v^t))/3
\]

\[
\geq 0 \quad (5.37)
\]

\[
\Box
\]

However, no simple expression exists for the expected change in a and b’s Borda score:

**Lemma 5.11.** *In the absence of noise, given any vote status V^t that satisfies S^B_c(v^t) \leq\min(S^B_b(v^t), S^B_a(v^t)), the expected score change for candidate a, b and c at time t + 1 are respectively:*

\[
\mathbb{E}[S^B_a(v^{t+1})|V^t] - S^B_a(v^t)
\]

\[
= \sum_{i=1}^{3} (V^t_{i3} + 2V^t_{i4} + 2V^t_{i5} - V^t_{i6}) - \sum_{i=4}^{6} (V^t_{i1} + 2V^t_{i2} + 2V^t_{i3} - V^t_{i6}) \quad (5.38)
\]

\[
\mathbb{E}[S^B_b(v^{t+1})|V^t] - S^B_b(v^t)
\]

\[
= -\sum_{i=1}^{3} (V^t_{i1} + 2V^t_{i4} + 2V^t_{i5} - V^t_{i6}) + \sum_{i=4}^{6} (V^t_{i1} + 2V^t_{i2} + 2V^t_{i3} - V^t_{i4}) \quad (5.39)
\]

So how do their scores change in the long run? Note that at time 0, the probability that a type i, j voter gets picked is V^0_{ij}. If type i, j voters find a better option than voting for j, then the expected decrement of V^0_{ij} is exactly V^0_{ij}/N. If we extrapolate \( \mathbb{E}[V^t|V^0] \) to a continuous process, \( V^t_{ij} \) would decrease at an exponential rate in expec-
With a suitable stretch of the time axis, the remaining votes of type $i, j$ at time $t$ is
\[
\mathbb{E}[V_{ij}^t | V^0] = V_{ij}^0 e^{-t}
\] (5.40)

Using this result, we can get the expression of the expected score after time 0 (which can be replaced by any starting time) for $a$, $b$, and $c$:

\[
\begin{align*}
\mathbb{E}[S_a^B(v^t)|V^0] &= A_0 + A_1 e^{-t} \\
\mathbb{E}[S_b^B(v^t)|V^0] &= B_0 + B_1 e^{-t} \\
\mathbb{E}[S_c^B(v^t)|V^0] &= C_0 + C_1 e^{-t}
\end{align*}
\] (5.41)

where $A_0$, $A_1$, $B_0$, $B_1$, $C_0$ and $C_1$ can be calculated from the honest profile $v^h$ and current vote status $v^0$:

\[
\begin{align*}
A_0 &= 2(v_1^h + v_2^h + v_3^h) \\
A_1 &= S_a^B(0) - A_0 \\
B_0 &= 2(v_4^h + v_5^h + v_6^h) = 2 - A_0 \\
B_1 &= S_b^B(0) - B_0 \\
C_0 &= 1 \\
C_1 &= S_c^B(0) - C_0
\end{align*}
\] (5.42)

Let the profile follows the same direction of drift. As $t$ approach infinity, eventually all voters who sincerely prefer $a$ to $b$ will switch their vote to $acb$, and all voters who sincerely prefer $b$ to $a$ will vote for $bca$. But if this happened, then $c$ would have a Borda score of 1, surpassing either $a$ or $b$. This contradicts with the assumption
that $c$ is the bottom scorer. Thus the direction of the drift must have changed before reaching the end point.

We divide the problem into two cases by whether a Condorcet winner exists in the honest profile.

5.5.2 Condorcet winner exists

A Condorcet winner is preferred to any other candidate by more than half of the voters. If the initial profile is honest, then the Condorcet winner cannot be ranked at the bottom since she has a Borda score higher than 1. However, if the vote status is initialized with arbitrary values, then the Condorcet winner may be ranked at any position.

Assume $a$ is the Condorcet winner in the honest profile. This means that

\[
\begin{align*}
 v^h_1 + v^h_2 + v^h_3 &> 0.5 \\
 v^h_1 + v^h_5 + v^h_6 &> 0.5
\end{align*}
\] (5.43)

Case 1: $a$ initially ranks first.

Without loss of generality, assume $S^B_a(0) \geq S^B_b(0) \geq S^B_c(0)$. Then $A_0 > 1$, $B_0 < 1$ and $a$’s Borda score always dominate $b$’s and $c$’s in expectation. This is guaranteed by the following lemma:

**Lemma 5.12.** Let $f(x) = f_0 + (f_1 - f_0)e^{-x}$ and $g(x) = g_0 + (g_1 - g_0)e^{-x}$, where $f_0 > g_0$ and $f_1 > g_1$. Then $f(x) > g(x)$ for all $x \geq 0$.

**Proof.**

\[
 f(x) - g(x) = (f_0 - g_0)(1 - e^{-x}) + (f_1 - g_1)e^{-x} > 0 
\] (5.44)
Thus there exists a time $t'$ such that $S^B_b(t') \leq S^B_c(t')$. What happens after time $t'$? If $c$‘s score ever surpasses $b$, then by the same argument we can show that $b$’s score will catch up back again with $c$, because we are in the same scenario as in the time range $t \in [0, t')$ except $b$ and $c$ switch their names. Therefore, after time $t'$ the profile will shift along the hyperplane $S^B_b = S^B_c$. The winner is unambiguously $a$, the Condorcet winner. Table 5.4 shows which entries of $V^t$ are 0 and which entries are non-zero (represented by 1) after sufficiently long time:

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>honest</td>
<td>current</td>
<td>abc</td>
<td>acb</td>
<td>cab</td>
<td>cba</td>
</tr>
<tr>
<td>abc</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>acb</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>cab</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>cba</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>bca</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>bac</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5.4: Non-zero entries in the limit distribution of $V^t$

For each row the votes flow back and forth between the two nonzero entries, depending on whether $b$ or $c$ scores higher.

**Case 2: $a$ initially ranks second.**

If $a$ is ranked second initially, without loss of generality assume $S^B_b(0) \geq S^B_a(0) \geq S^B_c(0)$. Since $A_0 > C_0 = 1 > B_0$ by equations (5.42), by Lemma 5.12, $E[S^B_a(v^t)|V^0]$ will not cross $E[S^B_c(v^t)|V^0]$ before crossing with $E[S^B_b(v^t)|V^0]$. After the crossing, we go back to case 1. From then on the Borda score of $a$ will always dominate $b$ and $c$. Thus we conclude that $a$ is the winner in the long run.

**Case 3: $a$ initially ranks last.**

Assume $S^B_b(0) \geq S^B_c(0) \geq S^B_a(0)$. Similar to case 2, $A_0 > C_0 = 1 > B_0$. Thus $S^B_a(v^t)$ will surpass $b$ and $c$ in the long run. Whether $S^B_b(v^t)$ crosses with $S^B_c(v^t)$ does not change the nature of the problem.

Summarizing the three cases, we can get the following theorem:
Theorem 5.13. In a three party election, if there exists Condorcet winner according to every voter's honest opinion, then the Condorcet winner will always win in the Borda count after sufficiently long time under the effect of strategic voting.

5.5.3 No Condorcet winner

Suppose the initial profile satisfies $S^B_a(0) \geq S^B_b(0) \geq S^B_c(0)$. If there is no Condorcet winner in the honest profile, the pairwise preferences among $a$, $b$ and $c$ form a loop: $p(a > b|v^h)$, $p(b > c|v^h)$ and $p(c > a|v^h)$ are all no smaller than 0.5 or all no greater than 0.5. We denote these two cases with $a > b > c > a$ and $a > c > b > a$.

Case 1: $a > b > c > a$

By equations (5.42) and Lemma 5.12, $E[S^B_b(v^t)]$ will not cross with $E[S^B_c(v^t)]$, but will cross with $E[S^B_a(v^t)]$. After crossing, the ranking becomes $a > c > b$. The problem is now equivalent to case 2.

Case 2: $a > c > b > a$

By equations (5.42) and Lemma 5.12, $E[S^B_b(v^t)]$ will not cross with $E[S^B_a(v^t)]$, but will cross with $E[S^B_c(v^t)]$. After crossing, the Borda ranking becomes $a > c > b$, which turns the problem back to case 1.

In summary, the profile will keep drifting, with $a$, $b$ and $c$ alternatively emerging as the winner but will always be replaced by the candidate that beats her in pairwise comparison. There is no point of equilibrium.

5.6 Conclusion

In this chapter, we analyzed the dynamics of strategic voting for plurality and Borda count. Voters make noisy observations and asynchronously update their vote according to which decision boundary their observation is closest to.
For plurality, we performed probabilistic analysis on the evolution of the profile, using Laplace noise. We showed that strategic voting in effect turns plurality into a two-round system, which is identical to instant run-off when there are three candidates. For Borda count, we performed deterministic analysis on the expected direction of profile drift. It turns out that if there is a Condorcet winner in the honest profile, then this Condorcet winner will eventually stand out as the Borda count winner after a certain time. If there is no Condorcet winner in the honest profile, then the profile will drift on, with different candidates leading at different times, and there is no “eventual” winner. Thus in effect, strategic voting turns Borda count to a Condorcet method.
Bibliography


