

UNIVERSITÉ DE TUNIS
INSTITUT SUPÉRIEUR DE GESTION



MASTER RECHERCHE

SPÉCIALITÉ

SCIENCES ET TECHNIQUES DE L'INFORMATIQUE DECISIONNELLE (STID)

OPTION

INFORMATIQUE ET GESTION DE LA CONNAISSANCE (IGC)

**MULTI-ROUND VOTE ELICITATION FOR MANIPULATION UNDER
CANDIDATE UNCERTAINTY**

MANEL AYADI

SOUTENU LE 09/07/2015, DEVANT LE JURY COMPOSÉ DE:

ELOUADI ZIED	PROFESSEUR, ISG-TUNIS	PRÉSIDENT
ACHOUR ADHEMI	MAÎTRE ASSISTANT, IISG-TUNIS	RAPPORTEUR
BEN AMOR NAHLA	PROFESSEUR, ISG-TUNIS	DIRECTEUR DU MÉMOIRE

LABORATOIRE: **LARODEC**

ANNÉE UNIVERSITAIRE 2014 / 2015

Acknowledgments

I would like to express my sincere gratitude to my advisor Professor Nahla Ben Amor for introducing me to the area of research. I would like to address special thanks for you, for your patience in supervising this work, your support, your advises, your productive comments and for the many hours spent helping this work. Thank you for your endless patience in improving my writing, for your comments on chapter drafts and for your great efforts to explain things clearly and simply.

I owe a debt of gratitude to Professor Saoussen Krichen for spending time discussing my research and for her useful remarks. I am deeply indebted to you for your encouragements and for all the interested discussion that have always been of great interest during the elaboration of this work.

I would like to express my consideration to Mr Didier Dubois and Mr Jerome Lang for their interest to this work and their valuable discussions and comments. This work has certainly benefited from their wisdom and their insightful and thoughtful suggestions.

Thanks to all the members of my laboratory LARODEC in *Institut Suprieur de Gestion de Tunis* for their support. A particular acknowledgment goes to my friends for their encouragement and moral support, and for all good and bad times we get together.

Lastly, and most importantly, I wish to thank my parents for their steadfast support, their love and patience. I dedicate this work to them.

Contents

Introduction	1
Part I : Theoretical Aspects	4
1 Social choice and manipulation	5
1.1 Introduction	5
1.2 Basic concepts and notations	6
1.2.1 Positional scoring rules	8
1.2.2 Condorcet rules	10
1.3 Vote manipulation in an election	11
1.3.1 Manipulation aspects	12
1.3.2 Manipulating an election	14
1.4 Computational problem	15
1.5 Impact of manipulation on social welfare	16
1.6 Conclusion	17
2 Uncertainty in preferences	18
2.1 Introduction	18

2.2	Incomplete preferences	19
2.2.1	Partial votes and profiles	19
2.2.2	Probabilistic preference models	20
2.3	Solution concepts with partial preferences	23
2.3.1	Possible and necessary winners	23
2.3.2	Minimax regret	24
2.3.3	Probabilistic preference models	24
2.4	Preference elicitation	25
2.5	Conclusion	27
Part II	: Contributions	28
3	New manipulation strategy with uncertain knowledge	29
3.1	Introduction	29
3.2	Manipulation strategy under uncertainty	30
3.3	Multi-round elicitation process	32
3.4	New manipulation strategy with uncertain knowledge	37
3.5	Conclusion	41
4	Experimental study	42
4.1	Introduction	42
4.2	Data sets description	43
4.3	Multi-round elicitation experiments	44
4.3.1	Sushi data experiments	44
4.3.2	Gen-Ran data experiments	49
4.4	New manipulation strategy experiments	54
4.5	Discussion	59

CONTENTS

iv

4.6 Conclusion	60
--------------------------	----

Conclusion	61
-------------------	-----------

References	63
-------------------	-----------

List of Figures

1.1	Single peaked preferences	12
2.1	Partial preference profile	20
3.1	New manipulation strategy with uncertain knowledge	39
4.1	Sushi data format	43
4.2	Sushi data histograms with <i>SP-IC</i> model	46
4.3	Sushi data histograms with <i>IC</i> model.	47
4.4	(a) Summary of Sushi results with <i>IC</i> model, (b) Summary of Sushi results with <i>SP-IC</i> , (c) Probability of the approximate winner quality.	48
4.5	Gen-Ran data histograms with <i>IC</i> model	51
4.6	Gen-Ran data histograms with <i>SP-IC</i> model	52
4.7	(a) Gen-Ran data with <i>IC</i> model, (b) Gen-Ran data results with <i>SP-IC</i> , (c) Probability of the approximate winner quality.	53
4.8	(a) Manipulation strategy with $P=2$ and $M = 750$, (b) Manipulation strategy with $P=3$ and $M = 750$, and (c) Manipulation strategy with $P=4$ and $M = 750$	56
4.9	(a) Manipulation strategy with $P=2$ and $M = 1000$, (b) Manipulation strategy with $P=3$ and $M = 1000$, and (c) Manipulation strategy with $P=4$ and $M = 1000$	57

4.10 (a) Probability of optimal manipulation with $M = 500$, (b) Probability of optimal manipulation with $M = 750$, and (c) Probability of optimal manipulation with $M = 1000$	58
--	----

List of Tables

1.1	Preference profile's example	7
1.2	Preference profile's example	8
1.3	Example due to Fishburn	10
2.1	Impartial culture model	20
2.2	Impartial anonymous culture model	21
2.3	Mallows model's Kendall tau distance	21
2.4	Mallows model's probabilities	22
2.5	Single peaked impartial culture model	23
2.6	Partial preference with a probability distribution	25
2.7	Probability table with probabilistic preferences	25

Introduction

Preferences are a common feature of everyday decision making generally used in collective decision when multiple agents need to choose one out of a set of possible decisions. In fact, each agent can express its preferences over the possible decisions, and a centralized system aggregates such preferences to determine the winning decision.

A very similar aggregation problem has been studied for a long time in the framework of *voting theory*. It consists in searching a 'reasonable' mechanism, typically in the form of voting rule, aggregating the opinions expressed by several voters on the candidates in an election, in order to determine a winner or to rank all candidates. The diversity of voting systems actually used in the world shows that this problem is still important. In the 1950s, the works of (Arrow, 1951; Black, Newing, McLean, McMillan, & Monroe, 1958; May, 1952) have initiated a huge literature (Kelly, 1991) forming what is today called *social choice theory*, concerned with the design and analysis of methods for collective decision making.

Results obtained in social choice theory are valuable for multicriteria decision aiding. Arrow (Arrow & Raynaud, 1986) proposed to go from one to the other by replacing the words 'action', 'criterion', 'partial preference' and 'overall preference' by 'candidate', 'voter', 'individual preference' and 'collective preference'. In other words, in social choice, each voter expresses an individual preference ordering over the set of candidates, and an election is held to compute the winner and obtain a collective decision. In this context, computational social choice appears as an interdisciplinary field linking social choice theory and computer science, promoting an exchange of ideas in both directions.

One of the most challenging topics in computational social choice is the study of strategic voting known also as *manipulation* consisting for a given voter or coalition of voters, in expressing an insincere preference profile so as to give more chance to a preferred candidate to be elected. For example, if an agent prefers Nader to Kerry to Bush, but knows that Nader has too few other supporters to win, while Kerry and Bush are close to each other, the agent would be

better off by declaring Kerry as its top candidate. The cheating party who misreport their preferences is called the manipulator while voters who reveal their true preferences are the honest party.

Clearly, strategic voting problem has a strong impact on elections. It represents an undesirable phenomenon because social choice schemes are tailored to aggregate preferences in a socially desirable way, and if the agents reveal their preferences insincerely, a socially undesirable candidate may be chosen. The issue of strategic voting has been studied extensively. A seminal negative result stated by Gibbard and Satterthwaite (Gibbard, 1973; Satterthwaite, 1975) show that voting rules that are not manipulable do not exist. A well known fact in strategic voting is the knowledge held by the manipulators on the preferences of the honest voters. Standard approaches to manipulation in social choice theory as well as in computational social choice is often confined to cases in which the manipulators have complete knowledge of the preferences of sincere voters (Conitzer, Sandholm, & Lang, 2007; Gibbard, 1973; Bartholdi III, Tovey, & Trick, 1989; Satterthwaite, 1975). This assumption does not cope with real world situation where manipulators are generally uncertain about the votes of sincere voters or even completely ignorant about their voters.

This work studies the connection between manipulation problem and vote elicitation. We investigate unweighted coalitional manipulation problem in the setting where the manipulators have identical preferences and are uncertain about the non-manipulators votes. In this context, elicitation may reveal information about sincere voters' votes allowing the manipulators to ensure the victory of their most preferred candidates. To this end, we propose to solve the problem of uncertainty in manipulators knowledge, regarding the non-manipulators votes, using an efficient top-k elicitation process. Given an incomplete voters preferences, our first contribution consists of a multi-round elicitation process aiming to determine the minimal amount of information, with the minimal number of rounds; in order to predict the right outcome of the election. Our second contribution is a *new manipulation strategy with uncertain knowledge*, where the restricted knowledge of the manipulators is solved using the proposed multi-round elicitation process. Using this approach, our goal is to determine an optimal manipulation strategy consists of a successful manipulation with less damage on social welfare and to answer the question: *How many candidates, per sincere voter, are needed to be known for an optimal manipulation?*

In order to deal with partial profile, two probabilistic models will be used, namely: *Impartial Culture* (Guilbaud, 1952) and *SP-IC*. In order to test the effectiveness of our proposals, we use two data sets namely: *Sushi data* (Kamishima, Kazawa, & Akaho, 2005) and *Gen-Ran data*.

This dissertation is organized as follows: Part I composed of two chapters presents the theoretical aspects of this report: Chapter 1 is an introduction to voting theory where it introduces the main research topics in computational social choice, and the most important voting rules used. Also, this chapter addresses the problem of strategic manipulation. Chapter 2 deals with uncer-

tainty in preferences. It provides an overview of a number of models and techniques developed to make decision with incomplete information about voter preferences. Part II is dedicated to our contributions and is composed of two chapters: Chapter 3 describes our *new manipulation strategy with uncertain knowledge*. In this chapter, we define our manipulation problem precisely and we propose a multi-round elicitation process to deal with incomplete preferences in order to solve the incomplete knowledge held by the manipulators. Finally, Chapter 4 is dedicated to the experimental study.

Part I

Theoretical Aspects

Part I presents the theoretical aspects of this report. It provides an overview of social choice theory. Chapter 1 introduces voting theory and the most important voting rules used to select the winner based on the voters' preferences. Also, this chapter addresses the problem of strategic manipulation. Chapter 2 deals with uncertainty in preferences. It provides an overview of a number of models and techniques developed to make decision with incomplete information about voter preferences.

Social choice and manipulation

1.1 Introduction

Social choice theory concerns the design and formal analysis of methods for aggregating the preferences of multiple voters, that is, to make a socially desirable decision as to which a winner is chosen from a set of candidates. In this context, voting rules are used to aggregate the preferences of voters over a set of candidates standing for election to determine which candidate should win the election. As an example of voting rule, *majority rule* selects candidates which have a majority, that is, more than half the votes.

The first controversy regarding social choice rules was presented by Marquis de Condorcet who noted that the concept of a social preference relation can be problematic. Thus, as proved by the *Condorcet paradox* (de Caritat et al., 1785), the majority rule can result in cycles where there are more than two candidates. For instance, if a majority of voters prefers a to b . Another majority prefers b to c and yet another one c to a . Clearly, the pairwise majority relation is cyclic. Hence, the majority rule does not constitute a social welfare function.

The best known incompatibility result is *Arrow's impossibility theorem* (Arrow, 1951) showing that this difficulty in the concept of social welfare is not specific to the majority rule. Arrow's impossibility theorem states that a set of axioms could not be simultaneously met when aggregating preferences whenever there are more than 2 candidates. In other words, *Arrow's impossibility theorem* shows that at least one of the required conditions has to be omitted or relaxed in order to obtain a positive result. Given two candidates a and b , these axioms are defined as follows:

- *Pareto*: if all individuals prefer candidate a to b , so does the collectivity, i.e. a will be ranked at least as high as b in the social preference relation.

- *Independence of irrelevant alternatives (IIA)*: the social preference between a and b depends on the individual preferences between a and b only.
- *Non-dictatorship*: there is no individual whose preference determines the social preference between all pairs of candidates.

One of the most challenging topics in computational social choice is the study of *manipulation* which refers to a coalition of one or more voters obtaining a more desirable outcome by misreporting their preferences. A voter is said to vote strategically when it does not rank the candidate according to its true preferences, but rather so to make the candidate most favorable to itself. Manipulation controversy the social choice goal, since if the voters reveal their preferences insincerely, a socially undesirable candidate may be chosen.

This chapter is organized as follows: Section 1.2 defines the basic notions of voting model used in preference aggregation. Section 1.3 presents the most important voting rules which are as important determinants of the voting outcomes as the individual opinions expressed in voting. Section 1.4 introduces manipulation problem in social choice. Finally, Section 1.5 deals with computational problems related to manipulation.

1.2 Basic concepts and notations

In this section we investigate social choice domain which represents the voting model used to aggregate voters' preferences into a collective decision. Formally, a voting model is defined by an election $E = (N, A, >^N)$ where:

- $N = \{1, \dots, n\}$, is the set of *voters (agents)*,
- $A = \{a, b, \dots\}$, is the set of *candidates (alternatives)* such that $|A| = m$, and
- $>^N = (>_1, \dots, >_n)$, is the *preference profile* of voters in N

For each voter $i \in N$, let $>_i \in >^N$ denote the *preference order* (or *vote*) of voter i over A . When $a >_i b$ for some $a, b \in A$, said that voter i prefers a to b .

Example 1.1. Let us consider a voting model with 3 candidates and 2 voters i.e. $A = \{a, b, c\}$ and $N = 2$. Table 1.1 represents the preference profile of the voters 1 and 2 where the preference order of each voter is represented by a one column.

In this case, voter 1 prefers a to b to c , his preference order is represented as follows: $a >_1 b >_1 c$.

Voter 1	Voter 2
a	b
b	c
c	a

Table 1.1: Preference profile's example

The use of voting rules to aggregate preferences has become a topic of intense study, and one of great importance in ranking, recommender systems, resource allocation, and other applications of social choice to computational systems. Voting rule is a very common way of resolving disagreements, determining common opinions and choosing public policies. It represents a procedure for making a choice from the set of candidates.

Formally, given an election $E = (N, A, \succ^N)$ as input, a *voting rule* f is a function $f : E \rightarrow S$, outputs a non-empty subset $S \subseteq A$. The elements of S are called the winners of the election E under f . If $|f(E)| > 1$ for any election E , the mapping f is called a *voting correspondence*.

In an election, the decision to choose which candidate to elect is no doubt important, but so are the questions related to the way in which the vote is taken. In other words, the voting procedure to be applied plays an important role as well. In fact, voting rules are as important determinants of the voting outcomes as the individual opinions expressed in voting. Voting rules can be classified in several ways generally:

- *Positional scoring rules*: Each voter ranks the candidates in order of preference. This ballot is represented by a vector where the first candidate is the most preferred candidate; the second one is the second preferred candidate and so on. e.g. IRV, Borda and plurality with run-off.
- *Condorcet rules*: Based on pairwise comparisons of candidates. They are a class of ranked voting systems that meet the Condorcet criterion. That is, the candidate who, when compared in turn with each of the other candidates, is preferred over the other candidate is always declared to be the winner, if such a candidate exists. e.g. copeland and maximin.

The following example will be used in the remaining to illustrate different voting rules.

Example 1.2. Let us consider a situation with five candidates $A = \{a, b, c, d, e\}$ and $N = 100$ voters. Table 1.2 presents the preference profile of the N voters. Each column represents the preference order of a subset of voters, where the 1st cell is the number of people who voted in that order, and the following is the order of the vote. For example, 33 voters have the preference order $a > b > c > d > e$.

33 voters	16 voters	3 voters	8 voters	18 voters	22 voters
a	b	c	c	d	e
b	d	d	e	e	c
c	c	b	b	c	b
d	e	a	d	b	d
e	a	e	a	a	a

Table 1.2: Preference profile's example

1.2.1 Positional scoring rules

The plurality rule selects the candidate who was ranked first by the most of voters. Plurality is the voting rule most often used in real-world elections, but note that it completely disregards all the information provided by the voter preferences except for the top ranking. Under a positional scoring rule, each time a candidate is ranked i^{th} by some voter, it gets a particular score s_i . The scores of each candidate are then added and the candidate with the highest cumulative score is selected. Scoring rules are also known as ranked voting systems or preferential voting systems. Formally, for a fixed number of candidates m , a positional scoring voting rule is defined by a non-negative vector called a *scoring vector* $s = (s_1, \dots, s_m) \in \mathbb{R}^m$ such that $s_1 \geq \dots \geq s_m$ defines a scoring rule over a set of candidates of size m . A candidate receives s_j points from each voter who ranks him in the j^{th} position, and the score of a candidate is the total number of points he receives from all voters. Three well known example of scoring rules are *Borda rule*, *plurality rule*, and *anti-plurality rule*.

Borda's rule: In the Borda voting rule, each voter gives $(m - 1)$ points to the candidate she ranked first, $(m - 2)$ points to the candidate she ranked second, or in general $(m - k)$ points to the candidate she ranked k^{th} . The score vector for the Borda rule is $s = (m - 1, m - 2, \dots, 1, 0)$. The winner is the candidate who amasses the highest total number of points. Borda's rule takes a special place within the class of scoring rules as it chooses those candidates with the highest average rank in individual rankings.

Example 1.3. Let us consider data in Example 1.3, then from Table 1.2, the Borda vector is $(4; 3; 2; 1; 0)$ and the Borda scores are computed as follows:

- $s(a) = (33 \times 4) + (16 \times 0) + (3 \times 1) + (8 \times 0) + (18 \times 0) + (22 \times 0) = 135$
- $s(b) = (33 \times 3) + (16 \times 4) + (3 \times 2) + (8 \times 2) + (18 \times 1) + (22 \times 2) = 247$
- $s(c) = (33 \times 2) + (16 \times 2) + (3 \times 4) + (8 \times 4) + (18 \times 2) + (22 \times 3) = 244$

- $s(d) = (33 \times 1) + (16 \times 3) + (3 \times 3) + (8 \times 1) + (18 \times 4) + (22 \times 1) = 192$
- $s(e) = (33 \times 0) + (16 \times 1) + (3 \times 0) + (8 \times 3) + (18 \times 3) + (22 \times 4) = 182$

This means that the Borda winner is b, since he has the highest Borda score.

Plurality rule: In the plurality voting rule, each voter gives 1 point to the candidate she ranked first, and the winner is the candidate who receives the highest total number of points. The score vector for the plurality rule is $s = (1, 0, \dots, 0)$. Hence, the cumulative score of a candidate equals the number of voters by which it is ranked first.

Example 1.4. *The plurality winner is a, since she is the candidate ranked first most often (33 votes).*

Anti-plurality rule: The anti-plurality rule selects the candidate who was ranked last by the least number of voters. The score vector for the anti-plurality rule (which is sometimes also called veto) is $s = (1, \dots, 1, 0)$. As a consequence, it chooses those candidates that are least-preferred by the lowest number of voters.

Example 1.5. *In this voting system, each voter would mark a vote against his or her fifth preference. The anti-plurality vector is $(1; 1; 1; 1; 0)$. The anti-plurality scores are computed as follows:*

- $s(a) = (33 \times 1) + (16 \times 0) + (3 \times 1) + (8 \times 0) + (18 \times 0) + (22 \times 0) = 36$
- $s(b) = (33 \times 1) + (16 \times 1) + (3 \times 1) + (8 \times 1) + (18 \times 1) + (22 \times 1) = 100$
- $s(c) = (33 \times 1) + (16 \times 1) + (3 \times 1) + (8 \times 1) + (18 \times 1) + (22 \times 1) = 100$
- $s(d) = (33 \times 1) + (16 \times 1) + (3 \times 1) + (8 \times 1) + (18 \times 1) + (22 \times 1) = 100$
- $s(e) = (33 \times 0) + (16 \times 1) + (3 \times 0) + (8 \times 1) + (18 \times 1) + (22 \times 1) = 42$

This means that candidates a and e are eliminated, and there is a tie between b, c and d.

Due to their simplicity, scoring rules are among the most-used voting rules in the real world. Borda rule is used in the National Assembly of Slovenia, and is similar to that used in the Euro-vision song contest.

1.2.2 Condorcet rules

This category of voting rules is based on pairwise comparisons of candidates that meet the Condorcet criterion. The *Condorcet winner*, for a given preference profile, is the candidate who beats every other candidate in pairwise elections. The Condorcet rule outputs the Condorcet winner if it exists; otherwise, it outputs the set of all candidates. As already seen in the Condorcet Paradox, there are preference profiles that do not admit a Condorcet winner. However, whenever a Condorcet winner does exist, it obviously has to be unique.

Many social choice theorists consider the existence of Condorcet winners to be of great significance and therefore call any voting rule that picks a Condorcet winner whenever it exists a *Condorcet extension* such as copeland and maximin. Under copeland rule, the winner is the candidate that wins the most pairwise contests (in a pairwise contest, a candidate wins if it is preferred over the other candidate by more than half of the voters), the score for every candidate is 1 point when it wins, -1 when it loses and 0 if the pairwise contest ends with a draw. The candidate with the most points wins. Under the Maximin rule, every candidate is evaluated by its worst pairwise defeat by another candidate; the winners are those who lose by the lowest margin in their worst pairwise defeats. If no candidate is undefeated, the candidate that is defeated by the fewest votes in its worst defeat, wins.

However, the voting rules presented above (positional scoring rules) thus far do not satisfy the Condorcet criterion: every scoring rule fails to select the Condorcet winner for some preference profile (Fishburn, 1973). This is shown by using one universal example given in Table 1.3.

Example 1.6. In Table 1.3, it is easily verified that candidate a is a Condorcet winner as 9 out of 17 voters prefer a to b and 10 out of 17 voters prefer a to c . Now, consider an arbitrary scoring rule with score vector (s_1, s_2, s_3) . Candidate b is the winner under plurality rule with 8 voters against 6 votes for a and 3 votes for c . Even with Borda rule, b is the winner with a score of 22 against 20 points for candidate a and 10 points for c . The score of candidate b always exceeds that of candidates a and c . In other words, b is the unique winner in any scoring rule, even though a is the Condorcet winner.

6 voters	3 voters	4 voters	4 voters
a	c	b	b
b	a	a	c
c	b	c	a

Table 1.3: Example due to Fishburn

In order to gain more insight into the huge zoo of voting rules, various axioms that may or

may not be satisfied by a voting rule have been put forward. Sometimes a certain set of axioms completely characterizes a single voting rule or an interesting class of voting rules (such as the class of scoring rules).

1.3 Vote manipulation in an election

A significant problem in social choice is that there is generally one or more voters that can obtain a more desirable outcome by misreporting their preferences. This is called *manipulation* or *strategic voting*. For example, consider a plurality election between three candidates, a , b , and c . Consider voter i with preferences $a \succ_i b \succ_i c$. Moreover, suppose that voter i believes that almost nobody else will rank a first, but it will be a close race between b and c . Then, i may be best off casting a vote in which b is ranked first: he has little hope of getting a to win, so he may be better off focusing on ensuring that at least b will win.

Manipulation is an undesirable phenomenon leading to fairness issues. Also, energy and resources are wasted on determining how best to manipulate. It follows from the negative result of *Arrow's impossibility theorem* (Arrow, 1951), that of *Gibbard-Satterthwaite theorem* (Gibbard, 1973; Satterthwaite, 1975) which states that if there are three or more candidates, then in any non-dictatorial voting scheme, there are preferences under which an agent is better off voting strategically. Consequently, this theorem shows that a dictatorship is the only non-manipulable voting mechanism. A voting rule that is *strategy proof* is also called non-manipulable where a voter cannot improve the outcome by misreporting his preferences.

One possible escape from the *Gibbard-Satterthwaite theorem* is to apply very specific restrictions on the class of allowed preferences. By relaxing the axiom of universal domain in which all preferences are possible, this theorem does not hold any more. For example, one such restriction is to *single-peaked preferences* (Black, 1948). *Black's median voter theorem* implies that if an odd number of voters have preferences that are single-peaked, then the median-voter rule is *strategy-proof*. That is, if the distribution of preference orders is such that they are single peaked, the Gibbard-Satterthwaite theorem does not apply and there is no chance for strategic voting to succeed (Mackie, 2003). Suppose that there is some predetermined linear ordering of the alternative set, a voter's preference is single-peaked, with respect to this ordering; if he has some special place that he likes best along that line, and his dislike for an alternative grows larger as the alternative goes further away from that spot. That is, put an order on the candidates, every voter has a most preferred candidate and likes a candidate less as he moves away from their preferred candidate. It is possible to sort the most preferred candidates according to this order. With single-peaked preferences, the median voter is the voter whose most preferred candidate is the median in this order amongst the voters.

Example 1.7. Lets us consider 5 candidates $A = \{a, b, c, d, e\}$ and 3 voters having the following preferences: $a \succ_1 b \succ_1 c \succ_1 d \succ_1 e$, $b \succ_2 c \succ_2 d \succ_2 a \succ_2 e$ and $c \succ_3 b \succ_3 a \succ_3 d \succ_3 e$. These preferences can be represented by Figure 1.1. On the vertical axis, some measure of the voters' utility is graphed. Utility = 1 represents the most preferred candidate by the voter. Each of the three plots shows one of the sets of individual preferences: In the plots, there is an oval for each candidate, and its height corresponds to its position in the list. As drawn, the single peak in an individual's ranking emerges visually as a peak in the plot.

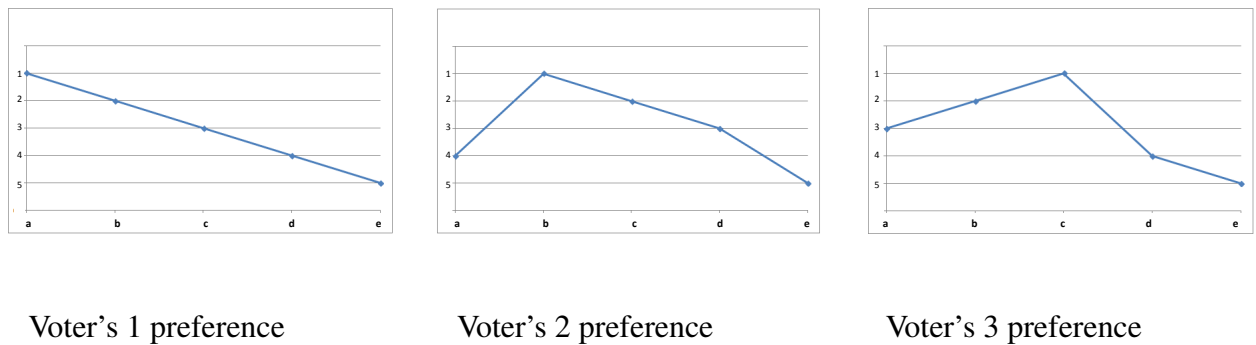


Figure 1.1: Single peaked preferences

Applying the single peaked preference model to the three voters would identify b as the median individual favorite, and place it first in the group ranking (the sorted list of individual favorites would be a, b, c and so the median is b). Once b is removed, three single-peaked rankings on the candidates a, c, d and e remain. The individual favorites in this reduced set are a, c , and c , so c is the new median individual favorite, and ranked second in the group ranking. Proceeding in this way, the resulting group ranking is: $b \succ c \succ a \succ d \succ e$, and the winner is candidate a .

1.3.1 Manipulation aspects

Manipulation can be characterized via three aspects:

Manipulators' knowledge: The knowledge that the manipulators have about the non-manipulators can differ according to these three situations:

- *Complete information* setting, the manipulators know the non-manipulators' votes exactly.
- *Incomplete information* setting, the manipulators are uncertain about the non-manipulators' votes.

- *No information* setting, the manipulators are ignorant about the non-manipulators' votes.

A well-known fact in social choice theory is that strategic voting, becomes harder when voters know less about the preferences or votes of other voters. Standard approaches to manipulation in social choice theory as well as in computational social choice (Bartholdi III et al., 1989) assume that the manipulating voter or the manipulating coalition have complete information about the other voters' votes, i.e. they know perfectly how the other voters will vote. Some approaches (Barbera, Bogomolnaia, & Van Der Stel, 1998) assume that voters have a probabilistic prior belief on the outcome of the vote, which encompasses the case where each voter has a probability distribution over the set of profiles. A recent paper (Conitzer, Walsh, & Xia, 2011) extends coalitional manipulation to incomplete knowledge, by distinguishing manipulating from non-manipulating voters and by considering that the manipulating coalition has, for each voter outside the coalition, a set of possible votes encoded in the form of a partial order over candidates (top-k favorite candidates). An extreme case of uncertainty is when a voter is completely ignorant about other votes.

Manipulators' number: Regarding this point, we can face two situations:

- *Coalitional manipulation* refers to a situation where a group of voters cast votes not according to their true preferences, but rather to obtain some goal.
- *Single manipulation* refers to a situation where a single manipulator casts votes not according to their true preferences, but rather to obtain some goal.

For any constant number of candidates, manipulation by an individual in the complete information setting is computationally easy because the manipulator can enumerate and evaluate all its possible votes (rankings of candidates) in polynomial time (Conitzer & Sandholm, 2002a). For example, with Borda rule, it is polynomial for a single voter to compute how to manipulate the result but NP-hard for two or more voters. In (Conitzer & Sandholm, 2002a), authors show that hardness results for manipulation by coalitions in the complete information setting, imply hardness of manipulation by individuals in the incomplete information setting.

Manipulators' goal: Manipulators can have two different goals, namely:

- *Constructive* manipulation, where manipulators try to make a given candidate win,
- *Destructive* manipulation, where manipulators try to prevent one particular candidate from winning.

Destructive manipulation is easier to compute than constructive manipulation (Conitzer et al., 2007). However, there are also rules where both destructive and constructive manipulation are in the same complexity class (e.g., both constructive and destructive manipulation of plurality are polynomial to compute, whilst both constructive and destructive manipulation of plurality with runoff for weighted votes are both NP-hard (Conitzer et al., 2007)).

1.3.2 Manipulating an election

Formally, in a manipulation problem, the N voters are divided into two groups, the manipulators and the non-manipulators known also as *honest* or *sincere* voters. For any voting rule f , an instance $I = (H, M, A, >^H, p)$ is given by an election $E = (N, A, >^H)$ where:

- $H = \{1, \dots, h\}$, is the set of *sincere voters*,
- $M = \{1, \dots, (n - h)\}$, is the set of *manipulators*,
- $A = \{a, b, \dots, m\}$, is the set of *candidates* where $|A| = m$, and
- $>^H = (>_1, \dots, >_h)$, is the *preference profile* of voters in H
- a distinguished candidate $p \in A$

The goal is to answer whether the M manipulators, knowing the preferences of the H sincere voters, can provide a preference such that p will be chosen by f . In other words, *Is there a way for the manipulators to cast their votes so that p wins?*

Example 1.8. Let us consider an election $E = (3, 4, >^3)$ using the Borda voting rule with the following preference profile: $b >_1 a >_1 c >_1 d$, $b >_2 a >_2 c >_2 d$ and $a >_3 b >_3 c >_3 d$. Thus, Borda scores are: $s(a) = 7$, $s(b) = 8$, $s(c) = 3$, $s(d) = 0$. The Borda winner is b , beating candidate a 8 points to 7.

However, if voter 3 wants to manipulate the election and make a wins, he will change his vote and ranks candidate b last. The manipulation model will be as follows: $I = (2, 1, 4, >^2, a)$, 2 sincere voters, one manipulator how want to make candidate a wins. The new preference profile is: $b >_1 a >_1 c >_1 d$, $b >_2 a >_2 c >_2 d$ and $a >_3 c >_3 d >_3 b$. New Borda scores are: $s(a) = 7$, $s(b) = 6$, $s(c) = 4$, $s(d) = 1$. Candidate b 's score goes down to 6, and a (voter 3's favorite candidate) wins!

1.4 Computational problem

The problem of manipulation is defined more strongly than the simple definition of manipulation. The question that generally raised is not whether the manipulator can guarantee a victory of a more preferred candidate, but whether he can guarantee the victory of a specific candidate by misreporting his preferences and casting a false preference order. A greedy algorithm deciding the problem under specific voting rules exists. This algorithm generates the preference of the voter i (the manipulator) such that the candidate p will be declared the winner, or returns that this is not possible. The algorithm does the following:

1. Rank p in 1st place
2. While there are candidates that were not ranked yet:
 - If there exists a candidate that can be ranked in the next spot without preventing p from winning, rank that candidate in the next spot,
 - otherwise, declare that the desired preference does not exist.

Example 1.9. *Let us consider the following example with 4 candidates $A = \{a, b, c, p\}$ and $N = 3$ voters. The Borda voting rule will be used. The preference order of sincere voters 1 and 2 is $a > p > b > c$. The 3rd voter is the manipulator who wants to make candidate p wins the election. The result of the algorithm for the 3rd voter is:*

1. Rank p in 1st place, and find that with the two other voters p has 7 points.
2. Notice that a can't be ranked next, since he will end with 8 points. Rank b in 2nd place, leaving him with 4 points.
3. Notice that a can't be ranked next, since he'll end with 7 points. Rank c in 3rd place, leaving him with 1 point.
4. Rank a in 4th place, leaving him with 6 points.

According to f , p wins with 7 points (while a , b , and c have 6, 4, and 1 points respectively). The preference order of the manipulator is: $p > a > b > c$.

It is obvious that if the algorithm found a preference resulting in p winning, the decision that the manipulator can provide a preference such that p wins is correct. However, in some cases, the algorithm could not find such a preference. Authors in (Bartholdi III et al., 1989; Bartholdi III & Orlin, 1991), show that it can be computationally difficult to manipulate certain rules. For

example, with some rules like STV, it is NP-hard to compute how a single voter needs to vote to manipulate the result. The possibility of manipulation depends on the voting protocol, the number of manipulators and whether we are interested in constructive or destructive manipulation. For example, authors in (Conitzer & Sandholm, 2002a) show that the plurality protocol is easy to manipulate constructively and destructively for any number of candidates, however, the Borda protocol becomes hard to manipulate constructively with 3 candidates already, but is easy to manipulate destructively for any number of candidates.

1.5 Impact of manipulation on social welfare

Social welfare designs the degree of satisfaction of the society. It considers welfare across different possible sets of individual preferences. Employing a voting rule to choose a more desirable alternative by the society (optimal) leads to a possible gap between the social welfare of the optimal alternative and the social welfare of the one that is ultimately elected using the voting rule.

It is often natural to consider the utility that a voter derives from the choice or decision that is made for the group as a whole. When it is related to the manipulation problem, social welfare provides a different analysis of manipulability of various voting rules. While the probability of manipulation can inform us on the chances given to manipulators to change the outcome of the election, this does not tell the whole story, since alternatives with higher success probability can have an undesirable impact on social welfare causing less societal dissatisfaction. Intuitively, societal satisfaction is guaranteed when the more preferred alternative is to the sincere voters, the more likely it is that manipulators can succeed in causing the alternative p to be selected. In other word, if candidates that are "closer to winning" are those that are generally ranked more highly by a coalition of manipulators, this means that such candidates are generally more desirable by the society. As a consequence, social welfare for alternative p chosen by the manipulators must be close to that of the optimal (non-manipulated) alternative if p has a reasonable chance of winning, which in turn means that the *damage*, or *loss* in social welfare, caused by manipulation will itself be limited.

Characterizing the impact of a manipulating coalition's action only in terms of its probability of succeeding can sometimes be ambiguous. A deficiency of current manipulation analyses gains prominence on success probability and ignoring the impact of manipulation on social welfare. To the best of our knowledge, only one work considers the impact of social welfare under manipulation (Lu, Tang, Procaccia, & Boutilier, 2012). Authors shed light on the stark difference between the loss in social welfare and the probability of manipulation by showing that even when manipulation is likely, impact to social welfare is slight.

1.6 Conclusion

In this chapter we have examined social choice theory, by introducing basic notions of voting model used in preference aggregation. We have also presented the most important voting rules to select the winner based on the voters' preferences. Then, we have focused on manipulation problem, where a coalition of manipulators want to guarantee the victory of a specific candidate by misreporting their preferences; based on the knowledge held of sincere votes. Manipulation becomes harder when the manipulators have partial information about the preferences of sincere voters. In this case it may be worth having an idea of the possible outcomes without waiting for the preferences to be completed. To this end, the next chapter introduces existing solutions to determine the winners with partial information about voter preferences.

Uncertainty in preferences

2.1 Introduction

Voting rules are used to aggregate the preferences of voters over a set of candidates standing for election to determine which candidate should win the election. However, many voting schemes make stringent assumptions about the preferences provided by voters. For instance, it is usually assumed that each voter provides a *complete* preference ranking of all candidates under consideration. In practice, however, voting systems often permit voters to declare a partial order over a subset of the candidates. For example, in the 1992 General Election for Dublin West, 9 candidates ran, but 29988 voters ranked only a median of 4 candidates over the set of candidates, and only 12.7% of the voters cast a complete vote.

Partial voting can have a significant effect on elections (Emerson, 2013). For example, in elections for the Tasmanian Parliament, voters are forced to rank a minimum number of candidates to prevent certain types of strategic voting (for example, when three candidates are running, voters must rank at least two candidates, whilst when four or more candidates are running, at least three candidates must be ranked). Also, incomplete vote has a significant impact on computational issues surrounding strategic voting, by modeling the uncertainty of the manipulators about the preferences of sincere voters. A well-known fact in strategic voting, becomes harder when voters know less about the preferences or votes of other voters. In this context, uncertainty regarding the non-manipulators' votes can be solved using different solution concepts used to determine the winner with incomplete preferences.

This chapter is organized as follows: Section 2.2 presents different forms of partial vote and probabilistic approaches to deal with incomplete profiles. Section 2.3 introduces solution concepts to determine the winners with partial or stochastic information about voter preferences,

this includes: possible and necessary winners, regret-based winner computation and probabilistic models. Finally, Section 2.4 provides practical techniques to elicit voter's preferences.

2.2 Incomplete preferences

Incomplete information about voter preferences represents a challenging problem in preference aggregation while communicating voter preferences to the mechanism is probably the most critical communication in voting. Uncertainty appeals when the votes are incompletely specified. For example during preference aggregation, it is possible to have an incomplete preference profile in which some of the votes are incomplete orders.

2.2.1 Partial votes and profiles

An incomplete vote is defined by an election $E' = (N, A, \Pi)$ where:

- $N = \{1, \dots, n\}$, the set of *voters*,
- $A = \{a, b, \dots\}$, the set of *candidates* such that $|A| = m$, and
- $\Pi = \{\pi_1, \dots, \pi_n\}$, the *partial preference profile* of voters in N .

Let π_i denote the partial preference order of voter i . A partial vote π_i can represent information about voter i 's responses to queries. The most natural queries that can be represented as a partial preference order are:

- *Pairwise comparisons*: Voter i is asked to state which of two alternatives, a or b , is preferred to the other.
- *Top- k queries*: Voter i provides the *top- k* (ranked) alternatives from \succ_i for some $1 \leq k \leq m$.

A *completion* of π_i is any vote \succ_i that extends π_i . Let $C(\pi_i)$ denote the set of *completions* of π_i , that is, the set of all complete votes \succ_i that extend π_i . Let $C(\Pi) = C(\pi_1) \times \dots \times C(\pi_n)$ be the set of completions of Π .

Example 2.1. Let us consider three candidates c_1, c_2 and c_3 with the three partial preference orders illustrated in Figure 2.1. Let $\Pi = \{O_1, O_2, O_3\}$ be the partial preference profile, such that in O_1 , the voter prefers $c_1 \succ c_2$ and $c_1 \succ c_3$, but there is no relation between candidates c_2 and c_3 . In this case, O_1 (resp. O_2, O_3) can be completed by adding $c_2 \succ c_3$ (resp. $c_1 \succ c_2, c_1 \succ c_2$ and $c_1 \succ c_3$). As a result, $C(\pi_i) = \{c_2 \succ c_3, c_1 \succ c_2, c_1 \succ c_2, c_1 \succ c_3\}$.

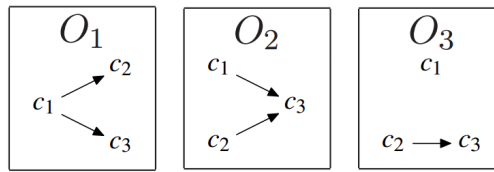


Figure 2.1: Partial preference profile

2.2.2 Probabilistic preference models

If one is given partial information about the preferences of voters, but is forced to make a decision and select the winner of an election; several styles of approach can be used, including worst-case or probabilistic analyses. Probabilistic analysis requires the specification of some prior distribution over voter preferences or preference profiles, so we briefly review several key probabilistic models.

Probabilistic analysis has been used primarily to study the likelihood that various phenomena (e.g., Condorcet cycles, manipulability) occur in randomly drawn voter populations, than as a basis for decision making or elicitation with incomplete information. Abstractly, a distribution over preference order $>$ can be viewed as a "culture" indicating the probability that a random voter will hold a particular preference ranking. By far, there are two widely adapted probability models in the social choice literature which are used for sampling voters' preferences.

- *Impartial culture model (IC)* (Guilbaud, 1952), in which each ranking is equally likely to be a voter's preference, and all voter's preferences are independent. For totally ordered m alternatives chosen by N voters, *IC* assumes that each voter independently selects her preference ranking according to a uniform probability distribution over preference orders.

Example 2.2. Table 2.1 presents a preference profile over 3 candidates $A = \{a, b, c\}$. For any preference order, the *IC* assumes that the probability of observing any preference order is uniform.

$a > b > c$	$a > c > b$	$b > a > c$	$b > c > a$	$c > a > b$	$c > b > a$
$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Table 2.1: Impartial culture model

- *Impartial anonymous culture (IAC)* (Kuga & Nagatani, 1974), which provides a uniform distribution over all preference profiles. The *IAC* model relies also on an equiprobability assumption, but this time without taking the identity of the voters into account. Hence,

the preferences of the electorate are generated by using anonymous profiles in which the names of the voters are neglected. IAC assumes that each anonymous profile is equally likely.

Example 2.3. Table 2.2 presents a preference profile over 3 candidates $A = \{a, b, c\}$. The IAC assumes that the probability of observing any distribution over preference orders is equally likely. That is, any vector where $X_1 + X_2 + X_3 + X_4 + X_5 + X_6 = 1.0$

$a > b > c$	$a > c > b$	$b > a > c$	$b > c > a$	$c > a > b$	$c > b > a$
X_1	X_2	X_3	X_4	X_5	X_6

Table 2.2: Impartial anonymous culture model

More realistic probabilistic models of preferences, or parametrized families of distributions over rankings, have been proposed in statistics, econometrics and psychometrics. These models typically reflect some process by which people rank, judge or compare alternatives. A commonly used model, adopted widely in voting theory is the *Mallows ϕ -model* (Mallows, 1957).

- *Mallows ϕ model* is parametrized by a modal or reference ranking σ and a dispersion parameter $\phi \in (0, 1]$; and for any ranking r : $P(r; \sigma, \phi) = \frac{1}{Z} \phi^{d(r, \sigma)}$, where:
 - d is the Kendall tau distance and
 - $Z = \sum_r \phi^{d(r, \sigma)} = 1 \times (1 + \phi) \times (1 + \phi + \phi^2) \times \dots \times (1 + \phi^{m-1})$ is a normalization constant.

When $\phi = 1$ the uniform distribution over rankings is obtained (i.e., impartial culture).

Example 2.4. Let us consider the following preference orders presented in Table 2.3 over 3 candidates $A = \{a, b, c\}$ and the corresponding Kendall tau distance.

$a > b > c$	$a > c > b$	$b > a > c$	$b > c > a$	$c > a > b$	$c > b > a$
0	1	1	2	2	3

Table 2.3: Mallows model's Kendall tau distance

- Let us consider the ranking $r : a > b > c$, and compute the probability of the preferences ranking. The normalization constant is equal to: $Z = 1 \times (1 + \phi) \times (1 + \phi + \phi^2)$.

- Let us consider the case where $\phi = 0.5$. In this case, $Z = 1 \times (1 + 0.5) \times (1 + 0.5 + 0.5^2) = 2.625$. The probability of different preference ranking is shown in Table 2.4 and obtained as follows:

$$\begin{aligned}
 * P(a > b > c) &= \frac{1}{Z} \phi^{d(a>b>c, a>b>c)} = \frac{1}{2.625} \times 0.5^{(0)} = 0.381 \\
 * P(a > c > b) &= \frac{1}{Z} \phi^{d(a>b>c, a>c>b)} = \frac{1}{2.625} \times 0.5^{(3-0)} = 0.1905 \\
 * P(b > a > c) &= \frac{1}{Z} \phi^{d(a>b>c, b>a>c)} = \frac{1}{2.625} \times 0.5^{(1-0)} = 0.1905 \\
 * P(b > c > a) &= \frac{1}{Z} \phi^{d(a>b>c, b>c>a)} = \frac{1}{2.625} \times 0.5^{(2-0)} = 0.095 \\
 * P(c > a > b) &= \frac{1}{Z} \phi^{d(a>b>c, c>a>b)} = \frac{1}{2.625} \times 0.5^{(2-0)} = 0.095 \\
 * P(c > b > a) &= \frac{1}{Z} \phi^{d(a>b>c, c>b>a)} = \frac{1}{2.625} \times 0.5^{(3-0)} = 0.04762
 \end{aligned}$$

$a > b > c$	$a > c > b$	$b > a > c$	$b > c > a$	$c > a > b$	$c > b > a$
0.381	0.1905	0.1905	0.095	0.095	0.04762

Table 2.4: Mallows model's probabilities

Among probabilistic analysis used in social choice, there is also an increasing focus on identifying tractable special cases using tools like fixed parameter tractability (Hemaspaandra, Lavaee, & Menton, 2013) and domain restrictions (Walsh, 2007). One of the most common domain restrictions considered in social choice theory is that of *single peaked preferences*.

- *Single peaked impartial culture (SP-IC)* generates singled peaked votes uniformly from the IC model. The model draws uniformly at random from the set of orders that are consistent with a given social axis. In general, given m candidates at integer points on the axis choice form 1 to m , there are 2^{m-1} different singled peaked votes. In (Walsh, 2015) discusses how to generate singled peaked votes uniformly from the IC model. He shows that half of all these single peaked votes end in m and are made up of all the single peaked votes from 1 to $m - 1$ augmented with m at their end. The other half of these single peaked votes end in 1 and are made up of all the single peaked votes from 2 to m augmented with 1 at their end.

Example 2.5. Let us consider the following preferences order over candidates in $A = \{a, b, c\}$. If the social axis is $a > b > c$ then positive probability is obtained only on the $2^{m-1} = 2^2 = 4$ consistent orders with an uniform distribution of $\frac{1}{4}$ on each order.

Table 2.5 shows the distribution of probabilities on consistent orders based on (Walsh, 2015). Given the social axis $a > b > c$, let us consider the half of all these single peaked votes made up of all the single peaked votes containing candidates a and b , and augmented with c at their end. This consideration concludes to two preferences order: $a > b > c$ and

$b > a > c$ having a positive probability of $\frac{1}{4}$. Now, let us consider the other half of these single peaked votes end in a and are made up of all the single peaked votes containing b and c . The preferences order obtained are $b > c > a$ and $c > b > a$ with probability of $\frac{1}{4}$.

$a > b > c$	$a > c > b$	$b > a > c$	$b > c > a$	$c > a > b$	$c > b > a$
$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{4}$

Table 2.5: Single peaked impartial culture model

A variety of other models have been proposed that reflect different interpretations of the ranking process (e.g., Plackett-Luce, Bradley-Terry, Thurstonian, etc.) and many of these have found application in computer science, machine learning, and recommender systems.

2.3 Solution concepts with partial preferences

Two fundamental questions arise in voting situations related to incomplete preference profiles. First, *is the information provided sufficient to determine the winner (or have certain alternatives have been ruled out as possible winners)?* Second, *if required to choose a winning alternative under specific conditions, what are the right criteria for doing so?* These questions are discussed in this section.

2.3.1 Possible and necessary winners

Suppose it is given a voting rule f but have only a partial profile Π . This partial information may or may not be sufficient to determine the true winner under f . However, if the winner cannot be determined, Π may still contain enough information to rule out certain alternatives as potential winners. (Konczak & Lang, 2005) introduce the notions of possible and necessary winners to reflect these possibilities. A candidate is a *possible winner* if there exists a voting order such that this candidate is a plurality winner over the cast votes. Similarly, a candidate is a *necessary winner* if he is a plurality winner over the cast votes for each voting order. The possible and necessary winner problems are interesting in their own right. In addition, they provide insight into several related and interesting problems.

Example 2.6. In Figure 2.1, candidate c_1 is a possible winner of Π with respect to plurality rule. In the first place, O_1 (resp. O_2, O_3) can be completed by adding $c_2 > c_3$ (resp. $c_1 > c_2, c_1 > c_2$ and $c_1 > c_3$); then, c_1 is a necessary winner. Secondly, however, c_1 is not the only winner, because

we can complete O_1 (resp. O_2, O_3) by adding $c_2 > c_3$ (resp. $c_2 > c_1, c_2 > c_1$ and $c_1 > c_3$); then, c_2 is the only winner.

The applications of these concepts are many. For instance, if a is a necessary winner under Π , no further elicitation or assessment of voter preferences is needed to make a decision. Knowing whether a is a possible winner is tightly related to the coalitional manipulation problem as well. Assume a partial profile Π in which the votes of the non-manipulators are fully specified, while no information is given for the manipulators: if a is not a possible winner with respect to Π , the manipulating coalition cannot (constructively) manipulate the election to install a ; conversely if a is a possible winner, then a is susceptible to (constructive) manipulation.

2.3.2 Minimax regret

While possible and necessary winners are valuable concepts for providing constraints on the decisions one might make given a partial profile, they do not provide a general means for selecting winners given arbitrary partial profiles. In circumstances when one is required to make a decision without complete information the following questions remain: *which of the possible winners should one select? and is it reasonable to select an alternative that is not a possible winner?*

In recent work, authors in (Lu & Boutilier, 2011a) proposed a solution using the notion of *minimax regret (MMR)* to determine winners given partial profile, and also to guide elicitation. Assume a voting rule f defined using some scoring function. Intuitively, one measures the quality of a proposed winner a given Π by considering how far from optimal solution (true winner) could be in the worst case, given any completion of Π . *MMR* is defined to be the difference between the score of an alternative a and the score of the winner (i.e., alternative with the optimal score), in the worst case given any completion of Π . The *MMR* optimal solution is any alternative that is nearest to optimal in the worst case. This gives us a form of robustness in the face of vote uncertainty.

2.3.3 Probabilistic preference models

Probabilistic analysis of partial profiles can be used to answer several important questions. (Hazon, Aumann, Kraus, & Wooldridge, 2012) study the possibility of computing the probability of a particular candidate winning an election, given incomplete information. The incomplete preferences are represented in the form of a probability distribution over a set of preference orderings. Authors in (Hazon et al., 2012), proposed a model that allows to compute the probability of candidate's victory under a variety of voting rules. For example, given a constant bounded number of alternatives, they show that computing the probability of an alternative winning is polytime

solvable for any voting rule with polytime winner determination. However, when the number of candidates is not bounded, they prove that the problem becomes P-hard for the Plurality, Borda and copeland voting rule.

Example 2.7. *Let us consider the following probability distribution over a set of preference order of 3 candidates $A = \{a, b, c\}$ and 2 voters. Each voter i 's distribution is given in explicit form; that is, a probability is specified for each possible ranking the voter may hold. The voters' preferences are summarized in Table 2.6.*

Voter 1	Voter 2
$\frac{1}{2} a$	$\frac{1}{4} a$
$\frac{1}{3} b$	$\frac{3}{4} b$
$\frac{1}{6} c$	

Table 2.6: Partial preference with a probability distribution

Table 2.7 is an example of how authors in (Hazon et al., 2012), builds the probability table given a set of probability distribution over candidates. The first row represents a voting result which is a vector of votes. For example vector 2, 0, 0 means that 2 votes are given to candidate a and 0 vote for candidate b and c . The second column shows the probabilities for every possible voting result with the two voters. Thus, the probability that candidate a is the winner, is: $\frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \left(\frac{1}{3} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{3}{4} \right) + \frac{1}{2} \cdot \left(\frac{1}{6} \cdot \frac{1}{4} \right)$.

Voting results (a, b, c)	Probability
2, 0, 0	$\frac{1}{2} \cdot \frac{1}{4}$
1, 1, 0	$\frac{1}{3} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{3}{4}$
1, 0, 1	$\frac{1}{6} \cdot \frac{1}{4}$
0, 2, 0	$\frac{1}{3} \cdot \frac{3}{4}$
0, 1, 1	$\frac{1}{6} \cdot \frac{3}{4}$
0, 0, 2	0

Table 2.7: Probability table with probabilistic preferences

2.4 Preference elicitation

Preference elicitation refers to elicit a partial profile from voters with the minimal amount of information to determine a winning outcome of high quality. In this context, one of the main

reasons to implement a social choice function with partial profile, is to minimize the amount of information needed to reach a group decision. Several natural forms of queries can be used to communicate information about voter preferences:

- *Full ranking queries*: Voter i provides her entire ranked list of alternatives \succ_i .
- *Top- k queries*: Voter i provides the *top- k* alternatives from \succ_i for some $1 \leq k \leq m$.
- *Pairwise comparisons*: Voter i is asked to state which of two alternatives, a or b , is preferred to the other.
- *Set choice*: Voter i is presented with a subset $S \subseteq A$, and asked which alternative in S is most preferred.

In order to deal with preference elicitation, theoretical analyses that provide upper and lower bounds of information elicited to determine the outcome of various voting rules have been studied (Conitzer & Sandholm, 2005). Also, approaches to preference elicitation that are driven by more practical considerations exist. These techniques use the different solution concepts (Section 2.3) to deal with partial profiles and consider different measures for the quality of the resulting outcome selected such as winner quality, the amount of information elicited and the number of rounds of elicitation.

To the best of our knowledge, (Kalech, Kraus, Kaminka, & Goldman, 2011) were among the first to develop and evaluate a practical vote elicitation scheme that allowed the selection of winners using partial profiles using scoring rules. Their technique dubbed *greedy voting scheme* uses possible and necessary winners notion to determine a true winner and proceeds for a predetermined number of rounds, asking each voter at each stage for fixed number of positional rankings (e.g., the *top- k* candidates). Since termination is predetermined, necessary winners may not be discovered; instead possible winners are returned. One especially attractive feature of this approach is that voters are only queried a fixed number of times, allowing to minimize interruption and waiting time.

(Lu & Boutilier, 2011b) propose an elicitation scheme that exploits minimax regret solution to determine winners given a partial profile. Two different forms of queries are considered: simple pairwise comparisons (e.g., *do you prefer a to b ?*), and *top- k* queries (e.g., *who are your top- k candidates?*). (Lu & Boutilier, 2011b)'s approach proceeds in rounds, and simultaneously reduce the amount of information elicited from voters and the number of rounds of elicitation. Furthermore, their model is able to measure the quality of the approximate winner.

Eliciting only the pertinent information from voters, allows one to design mechanisms for preference aggregation that have less stringent informational requirements; and more robust to missing information.

2.5 Conclusion

In this chapter we have relaxed the assumption that each voter provides a complete preference ranking of all candidates under consideration, by dealing with incomplete preferences of voters over a set of candidates. We have described the basic notations and the different models of partial preferences. Then, we have introduced solution concepts for winner determination with uncertain preferences. Finally, we have described techniques for eliciting the necessary preferences needed to implement a voting rule. In this context, the main reason to implement social choice functions with partial profiles, is to minimize the amount of information needed to reach a group decision. In the next chapters, we will show how to use these techniques when considering a manipulation problem.

Part II

Contributions

Part II presents the contributions of this report. Chapter 3 describes our new manipulation strategy with uncertain knowledge. In this chapter we define our manipulation problem precisely and we propose a multi-round elicitation process to deal with incomplete preferences in order to solve the incomplete knowledge held by the manipulators. Chapter 4 is dedicated to the experimental results performed to test the effectiveness of our proposals.

New manipulation strategy with uncertain knowledge

3.1 Introduction

Previous chapters introduce a number of interesting and challenging issues relative to manipulation and the implementation of voting schemes with incomplete information, conceptually and computationally. Analysis of manipulation usually assumed that the manipulators have full knowledge of the votes of the non-manipulating voters. However, this assumption is unrealistic in the real world setting where manipulators have rarely access to such information. To this end, we relax this assumption by proposing a *new manipulation strategy with uncertain knowledge*, in the setting in which manipulators have incomplete information about the non-manipulators votes. In order to deal with uncertainty regarding the non-manipulators votes, a *multi-round elicitation process*, of sincere voters preferences, is derived that yields to an optimal manipulation with minimal information elicited.

This chapter is organized as follows: Section 3.2 defines our manipulation problem precisely by discussing its different dimensions and the reasons behind our choices. Section 3.3 discusses deficiencies in preference elicitation and proposes a multi-round elicitation process to deal with incomplete preferences. Section 3.4 describes the *new manipulation strategy with uncertain knowledge*.

3.2 Manipulation strategy under uncertainty

In this section, we define our manipulation problem precisely by detailing its different dimensions. To this end, we try to summarize our manipulation strategy through a series of the most important questions characterizing our problem:

1. *Who is manipulating: an individual voter or a coalition of voters?*

Both of these are important variants, but we focus on coalitional manipulation for the following reason: In elections with many voters, it is very unlikely that an individual voter can affect the outcome by misreporting his preferences even with unlimited computational power. Indeed, a more significant problem is that of coalitional manipulation, where a group of voters coordinates their actions in order to affect the election's outcome.

2. *Are the voters weighted or unweighted?*

Weighted votes occur in a number of real-world settings (e.g., shareholder elections). Weights are typically integers, and a voter with weight k counts as k voters voting identically. However, the *unweighted* version of the coalitional manipulation problem is more natural in most election settings, where the manipulators' votes are unweighted. For this reason, this work focuses on unweighted coalitional manipulation.

3. *What is the goal of manipulation?*

In constructive manipulation, a coalition attempts to cause a single preferred candidate to win. Destructive manipulation, attempts to prevent a specific candidate from winning. In this work, we focus on constructive manipulation variation.

4. *Do manipulators have identical preferences or different preferences?*

If all the manipulators have identical preferences, this is exactly the problem that they want to be able to solve. Manipulators would try to ensure their most preferred candidate's victory. In the case where manipulators do not have identical preferences, they may still work together, but it is much less clear which candidate they should try to promote. While they may all agree that they would prefer a different winner than the truthful one, however, deciding which candidate to support is a whole new game that they need to play among themselves. To push the scenario to the limit, this work considers a situation where all manipulators have identical preferences.

5. *What information do the manipulators have about the non-manipulators' votes?*

In this work we focus on the incomplete information held by the manipulators for the following reasons: Much work in social choice assumes that manipulators know perfectly how the other voters will vote (Narodytska & Walsh, 2014; Conitzer, Lang, & Sandholm, 2003; Conitzer et al., 2007; Gibbard, 1973; Bartholdi III et al., 1989; Satterthwaite, 1975).

Analysis of the impact of manipulation under more realistic incomplete knowledge assumptions has been sparse, but is clearly of vital importance. Recently, there has been some work relaxing this assumption. For example, (Ditmarsch, Lang, & Saffidine, 2013; Lu et al., 2012) extends coalitional manipulation to incomplete knowledge. Still, the study of strategic voting under incompleteness or uncertainty regarding the votes of sincere voters has received little attention.

6. *Which solution concept is considered with partial preferences?*

Several solution concepts discussed in Chapter 2 deal with partial preferences such as: possible and necessary winners, minimax regret and probabilistic preference models. To deal with uncertainty in preferences, this work uses probabilistic information about voter preferences. Notice that possible and necessary winners do not prescribe a general method for choosing winners under partial profiles. Also, minimax regret quantifies the lack of information using a worst-case analysis, ignoring the probability of specific completions. By contrast, at least in cases where probabilistic preference models are available, a more refined analysis of the likelihood of various completions can be used to make decisions.

7. *Which voting rule is considered?*

While plurality is considered as the most voting rule used in real-world elections, but it completely disregards all the information provided by the voter preferences except for the top ranking. Under positional scoring rules like Borda, a scoring vector is provided, ranking the candidate set from the highest cumulative score to the lowest one. Also these category of rules seems especially robust to evaluate the impact of manipulation on social welfare. For this reason, our work considers positional scoring rules to evaluate the outcome of the election.

In summary, this work focuses on manipulation problem when manipulators may declare preferences that are not their true ones, with the aim of obtaining a better outcome for themselves. In this context, the goal is to study unweighted constructive coalitional manipulation problem in the setting in which all manipulators have identical preferences and they are uncertain about the non-manipulators' votes. This uncertainty is represented by a top-k most preferred candidates over the sincere voters' votes. To deal with partial knowledge, probabilistic preference models are adopted. The scoring rule considered in this work is the Borda rule.

Manipulation analysis gains prominence on success probability and ignoring the impact of manipulation on social welfare (see Section 1.5). However, characterizing the impact of a manipulating coalition's action only in terms of its probability of succeeding can sometimes be ambiguous. Formally, we refer to $SW(s, s^*)$ as the difference between the true score s^* and the approximate score s associated to a set of candidates under a given scoring rule and we propose analyzing rules in this light. In this study, we will consider both the probability of manipula-

tion and its impact on social welfare, by computing the loss in welfare under the manipulation strategy where maximizing social welfare is the main objective.

3.3 Multi-round elicitation process

Effective elicitation concerns eliciting intelligently pertinent preference information from the voter because determining and expressing preferences can be arduous. Authors in (Conitzer & Sandholm, 2002b) show that eliciting partial order preferences from the voters can allow the voting protocol to determine the outcome well before all of the preferences have been elicited. The requirement of eliciting only the pertinent preferences is desirable for any of several reasons: First, it can be costly for an agent to determine its own full preferences. Second, in practice, voters may wish to rank candidates only as necessary, in cases where there are many available options or it will be time-consuming than the complete ranking. Third, bandwidth and communications can introduces overhead (e.g. network traffic) and make it impractical to send the entire set of preferences at once to the voting center if the candidate set is large. Finally, less preference revelation is desirable due to privacy reasons.

In this context, much work has mainly focused on theoretical analyses of preference elicitation, reporting the upper and lower bounds of the required communication with the voters (Conitzer & Sandholm, 2005). While these analyses provide the limits of what is possible and are efficient to determine the right outcome, but theoretical approaches that minimize the amount of information may sometimes elicit more preferences than necessary in practice which controversy the effective elicitation goal. Therefore, preference elicitation that are driven by more practical considerations are required. However, to the best of our knowledge, only two studies propose practical algorithms for effective preference elicitation (Kalech et al., 2011; Lu & Boutilier, 2011b).

- (Kalech et al., 2011) assume that each voter holds a predefined decreasing order of the preferences in the initial call for votes, then the voters are requested to submit their highest preferences in an iterative process. Requiring the users to predefine their preferences can be inconvenient to the users for the several reasons mentioned above. Therefore, we require voters to respond and rate a specific preference only when necessary.
- In the second study (Lu & Boutilier, 2011b), a practical elicitation process is proposed for the Borda voting protocol. Authors develop an approach to vote elicitation that exploits partial preference order over voter preferences. The iterative process simultaneously reduce the amount of information elicited from voters and the number of rounds. However, in the underlying process, one round vote elicitation protocols was analyzed. It will

be interesting to generate practical results for optimal multi-round instantiations which will allow to further reduce the amount of information elicited with an optimal number of rounds.

In order to overcome these weaknesses, we propose a practical multi-round elicitation process that yields to approximate the winner of the election with high quality, based on the partial information elicited from voters in multiple rounds. The key idea is that the voters incrementally send their preferences in rounds, one preference each round, in a decreasing order of their preferred candidates. After each round, the voters wait for the voting center to decide whether sufficient information has been received to determine a winner, or additional voters' preferred candidates need to be sent in the next round. In each round the response query of each voter is added to all his previous votes and the center evaluates the outcome of the election using a specific voting rule and based on the available information. The multi-round process stop eliciting preferences once the approximate winner is determiner with high probability, and from that point the voters no longer send preference values.

Example 3.1. *For instance, if the full preference order of the voter is $a > b > c > d$, initially, in the first round he will send only his most preferred candidate a , then the voting center will evaluate the quality of the winner using the Borda rule. If the quality of the approximate winner is far from the true one, then the second preferred candidate of the voter is required at the second round. The vote sent to the voting center will be $a > b$ and so on, until a high quality of the winner is detected.*

Now, let us introduce our multi-round elicitation process formally. Given a set of voters $N = \{1, \dots, n\}$, a set of candidates $|A| = m$, and a partial preference profile $\Pi = \{\pi_1, \dots, \pi_n\}$ of voters in N (see Section 2.2). In this context, a *query* refers to a single request for information from a voter. Type of query is in the form of *top-k* most preferred candidates (e.g., *who is your top-k preferred candidate?*). The elicitation of *top-k* preferences is simulated for $k \in \{1, \dots, (m - 1)\}$. The voter is required to rank only $m - 1$ candidates in the worst case. The complete order ranking is not considered, since the goal is to minimize the amount of information elicited from voters. Given a particular class of queries Q , informally, a *multi-round* voting protocol selects, at each round, a subset of voters, and one query per selected voter. The voter response query at round t is added to all his previous responses.

Formally, let I_{t-1} be the *information set* available at round t . I_{t-1} represents the voters' responses to queries for rounds $t \in \{1, \dots, (t - 1)\}$ allowing them to rank $k = m - 1$ candidates in the worst case. This information set is represented as a partial profile π^{t-1} of a top-k preferred candidates for each voter. Then, we propose a protocol based on three functions:

1. *Querying function* ψ , represents a sequence of mappings $\psi_t : \Pi \mapsto (N \mapsto Q)$, selecting

for each voter a single query at stage t given the current information set I_{t-1} .

2. *Completion function* $C : \Pi \mapsto V$, given a partial profile Π , $C(\Pi)$ is the set of consistent extensions of Π to obtain a full ranking profile. The output of C , is a complete ranking preference profile V over the set of candidates given the partial preference profile Π . Probabilistic models are used to deal with uncertainty regarding the voter preference and to complete the profile.
3. *Winner selection function* $\omega : V \mapsto A$, where $\omega(V)$ denotes the approximate winner under a given complete profile V . Borda scoring rule is used to evaluate the outcome given the voters' votes V and to determine the approximate winner over the set of candidates A . If the approximate winner is far from the true one, the protocol proceeds to round $t + 1$, otherwise the protocol terminates with the chosen winner at round t .

Given a protocol $\Omega = (\psi, C, \omega)$, in order to evaluate the effectiveness of our multi-round elicitation process in our experimental study presented in Chapter 4; we propose to use three properties, namely:

- *Quality of the winner*: Determine the quality of the approximate winner of the election, using partial information elicited in each round. While the Borda rule is used, the output of Ω is a scoring vector over the set of candidates in a decreasing order. The most preferred candidate have the highest score. The approximate winner is evaluated by considering how far from the true winner (given the complete preferences) could be.
- *Amount of elicited information*: Determine the necessary amount of information elicited from voters in each round in order to predict the outcome of the election. The amount of information will be measured by counting the number of k preference elicited in each round.
- *Number of rounds*: Determine the number of rounds needed, in order to obtain a high quality of the winner.

Clearly, the proposed metrics are traded off against one another. At one extreme, higher quality outcomes and fewer rounds of elicitation can usually be achieved at the cost of additional preference information where we have to ask each voter for full rankings. At the other extreme, minimizing elicited information can be constructed by eliciting one query per voter in each round, at a dramatic cost in terms of number of rounds. In this work, our goal is to find the optimal protocol which will allow us to reduce the number of rounds, as well as the amount of information elicited in each round; in order to reach a high quality of approximate winner. We answer the question: *How many candidates are needed to be elicited with the minimal number of*

interactions, in order to determine the true winner ? Uncertainty regarding the voter preferences is facilitated by using probabilistic preference models such as *IC* and *SP-IC* to complete the voters' votes.

Algorithm 3.1 Multi-round elicitation process

Input: N voters, m candidates, Top- k votes: $(\pi_1^k, \dots, \pi_n^k)$

Output: $\omega(V)$ (approximate winner)

Initialization: $t=1, k=1, I_t = \emptyset, \omega^*$ (true winner)

```

1: while (  $(\omega(V) \neq \omega^*)$  OR  $(t < m)$  ) do
2:   for  $i \in N$  do
3:      $\psi_t.add(\pi_i^k)$ 
4:   end for
5:    $I_t.add(\psi_t)$ 
6:    $V \leftarrow C(I_t)$ 
7:   if  $(\omega(V) \neq \omega^*)$  then
8:      $t = t + 1$ 
9:      $k = k + 1$ 
10:  else
11:    return  $\omega(V)$ 
12:  end if
13: end while

```

Algorithm 3.1 presents our multi-round elicitation process. The *querying function* is performed in lines (1-5) where ψ_t contains the partial profile elicited from voters in each round. This profile is a mapping of preference orders in π_i^k which denotes the *top-k* ranking of voter i over a set of m candidates. The response query of each voter is added to all his previous votes in I_t . The *completion function* is presented in line 6 where the partial profile is extended to a complete one V . Then, the *winner selection function* is introduced in lines (7-10) where the Borda rule is applied. The approximate winner $\omega(V)$, obtained with the predicted profile, is compared to the true one; by considering the complete ranking. If the predicted winner is far from the true one, the algorithm proceeds to the next round (line 8) by eliciting the next most preferred candidates of each voter (line 9); otherwise, the approximate winner is declared the winner of the election (line 10).

Example 3.2. Let us consider the following example to illustrate the application of the multi-round elicitation process. We use 5 candidates and 20 voters i.e. $m = 5$ and $N = 20$, having the following preference profile: 6 voters: $1 > 2 > 3 > 4 > 5$, 3 voters: $2 > 4 > 3 > 5 > 1$, 4 voters: $3 > 4 > 2 > 1 > 5$, 4 voters: $5 > 3 > 2 > 4 > 1$, 3 voters: $4 > 5 > 3 > 2 > 1$.

Applying the Borda rule (see Section 1.2) on this profile, will determine the scores of each

candidate as follows: $s(1) = 28$, $s(2) = 49$, $s(3) = 52$, $s(4) = 43$, $s(5) = 28$; and the true winner ω^* is candidate 3. Now let us apply the multi-round elicitation process on this profile, which will be performed in three steps:

- *Querying function:* we consider at the beginning $t = 1$ and $k = 1$ i.e. in the first round the process will elicit top-1 queries of each voter over the set of candidates, π_i^1 ; which will be added to the querying function ψ_1 , that is, $\psi_1 = \{1, 2, 3, 5, 4\}$ and the information set $I_1 = \{1; 2; 3; 5; 4\}$.
- *Completion function:* the partial preference profile in I_1 will be completed using the IC probabilistic model ($C(I_1)$); the output of this step is the following predicted preference profile V : 6 voters: $1 > 2 > 3 > 4 > 5$, 3 voters: $2 > 4 > 5 > 3 > 1$, 4 voters: $3 > 1 > 5 > 4 > 2$, 4 voters: $5 > 2 > 4 > 1 > 3$, 3 voters: $4 > 1 > 5 > 2 > 3$.
- *Winner selection function:* Borda rule is applied to determine the approximate winner $\omega(V)$ of the election given the predicted complete profile V . Borda scores are: $s(1) = 49$, $s(2) = 45$, $s(3) = 31$, $s(4) = 39$, $s(5) = 36$ and the approximate winner is $\omega(V) = 1$ which is different from the true one $\omega^* = 3$. In this case, top-1 queries are not sufficient to determine the right outcome, therefor, the process proceeds to round 2.

In the second round, the 3 functions are performed as follows:

- *Querying function:* in the second round ($t = 2$), the process will elicit top-2 queries of each voter over the set of candidates, π_i^2 ; which will be added to the querying function ψ_2 , that is, $\psi_2 = \{2, 4, 4, 3, 5\}$ and the information set $I_2 = \{1, 2; 2, 4; 3, 4; 5, 3; 4, 5\}$, where the response query of each voter is added to all his previous votes from $t = 1$ (separated by ;)
- *Completion function:* the partial preferences order in I_2 will be completed using the IC probabilistic model ($C(I_2)$); the output of this step is the following predicted preference profile V : 6 voters: $1 > 2 > 4 > 3 > 5$, 3 voters: $2 > 4 > 5 > 3 > 1$, 4 voters: $3 > 4 > 1 > 5 > 2$, 4 voters: $5 > 3 > 2 > 1 > 4$, 3 voters: $4 > 5 > 2 > 1 > 3$.
- *Winner selection function:* Borda rule is applied to determine the approximate winner $\omega(V)$ of the second round given the predicted complete profile V . Borda scores are: $s(1) = 39$, $s(2) = 44$, $s(3) = 37$, $s(4) = 45$, $s(5) = 35$ and the approximate winner is $\omega(V) = 4$ which is also different from the true one $\omega^* = 3$. In this case, top-2 queries are not sufficient to determine the right outcome, and the process proceeds to round 3.

In the third round, the 3 functions are performed as follows:

- *Querying function:* in the third round ($t = 3$), the process will elicit top-3 queries of each voter over the set of candidates, π_i^3 ; which will be added to the querying function ψ_3 , that is, $\psi_3 = \{3, 3, 2, 2, 3\}$ and the information set $I_3 = \{1, 2, 3; 2, 4, 3; 3, 4, 2; 5, 3, 2; 4, 5, 3\}$, where the response query of each voter is added to all his previous votes from $t = 1$ and $t = 2$.
- *Completion function:* the partial preferences order in I_3 will be completed using the IC probabilistic model ($C(I_3)$); the output of this step is the following preference profile V : 6 voters: $1 > 2 > 3 > 4 > 5$, 3 voters: $2 > 4 > 3 > 1 > 5$, 4 voters: $3 > 4 > 2 > 1 > 5$, 4 voters: $5 > 3 > 2 > 4 > 1$, 3 voters: $4 > 5 > 3 > 1 > 2$.
- *Winner selection function:* Borda rule is applied to determine the approximate winner $\omega(V)$ of the third round given the predicted complete profile V . Borda scores are: $s(1) = 34$, $s(2) = 46$, $s(3) = 52$, $s(4) = 43$, $s(5) = 25$ and the approximate winner is $\omega(V) = 3$ which is the true one $\omega^* = 3$. In this case, top-3 queries are sufficient to determine the right outcome, and the process stop eliciting preferences from voters.

To conclude, in this example, eliciting top-3 queries from voters in 3 rounds will allow us to determine the correct winner of the election with the minimal amount of information elicited.

3.4 New manipulation strategy with uncertain knowledge

In this section, we propose to study the connection between strategic voting and preference elicitation. Elicitation may reveal information about honest voters' votes to a manipulator, which allows him to vote strategically. To this end, we introduce a *new manipulation strategy with uncertain knowledge*, to solve restricted manipulators' knowledge, regarding the non-manipulators votes (see Section 3.2), using our proposed multi-round elicitation process (see Section 3.3). In this context, the multi-round elicitation process will elicit the *top-k* preferences of non-manipulators. This amount of information elicited will represent the manipulators' knowledge, which will allow them to ensure the victory of their most desired candidate with the minimal amount of information. Using this new approach, we aim to:

- Determine the minimal value of k needed to be known by the manipulators to achieve their goal, with the minimal number of rounds.
- Decide when to stop eliciting sincere voters' preferences, as an optimal manipulation is guaranteed with less damage on social welfare (see Section 3.2).

While manipulation and elicitation are two closely related fields, unfortunately, previous investigations have mainly focused on each field separately. To the best of our knowledge, only two studies introduce the connection between them, but no practical process has been proposed.

1. (Conitzer & Sandholm, 2002b) show that elicitation introduces additional opportunities for strategic manipulation of the election by the voters. They consider manipulation as an undesirable problem that must be avoided by curtailing the space of elicitation schemes so that no such additional strategic issues arise. However, no framework was proposed to emphasize this view.
2. (Walsh, 2008) studies the connection between the complexity of manipulation and the termination of eliciting preferences. He shows that the complexity of manipulation by a coalition of agents is closely related to the complexity of deciding if preference elicitation can be terminated, which may create a tension since we want it to be computationally hard to manipulate an election but computationally easy to decide when to terminate elicitation. However, strategic issues that may be introduced into a voting protocol by an elicitation process were not studied.

Our *new manipulation strategy with uncertain knowledge* will allow us to overcome existing weaknesses. This new approach uses the three defined functions in the multi-round elicitation process, namely: *querying function*, *completion function* and *winner selection function*; in order to elicit preferences from sincere voters and to determine the outcome of the election by considering the vote of manipulators. The whole process is illustrated in Figure 3.1 and proceeds as follows:

- In the first round ($t=1$), the *querying function* will elicit top-1 sincere voters' preferences. Each sincere voter expresses his most preferred candidate over the available set of candidates.
- The information set elicited is sent to the manipulators, which constitutes their knowledge at the first round. Based on these incomplete preferences, the manipulators will cast a vote that makes their preferred candidate p wins the election.
- In order to complete the sincere votes, we consider an internal process where the *completion function* is performed. In this step, sincere votes are extended using probabilistic models. A complete preference profile V over the whole set of candidates is provided in output.
- Given the complete predicted profile V and the manipulators desired candidate p , gathered on the *new data set*; the *winner selection function* is performed, where a scoring rule is

applied. The output of this step is an approximate winner $\omega(V)$ which will be compared to the true one in order to evaluate the impact of manipulation i.e. verify if the manipulators have succeeded to change the outcome of the election to their favor.

- In order to evaluate the performance of the protocol to elicit the minimal value of k from the sincere voters, we measure both, the probability of manipulation and its impact on social welfare. If no *optimal manipulation* is found, the process proceeds to the next round by eliciting the next most preferred candidate of the sincere voters, until an optimal strategy is detected. We refer to *optimal manipulation* by a successful manipulation with less damage on social welfare.

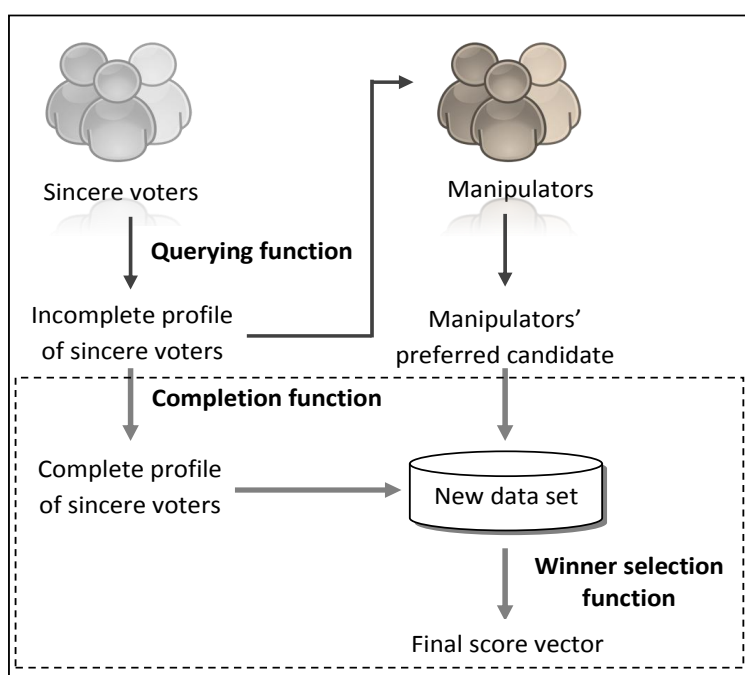


Figure 3.1: New manipulation strategy with uncertain knowledge

The use of the multi-round elicitation process on the new manipulation strategy, will allow us to investigate the relation between the amount of information revealed by sincere voters and the probability of manipulation; by studying how a reduction or increase in uncertainty may change a strategic vote. To this end, our goal is to answer the question: *How many candidates, per sincere voter, are needed to be known for an optimal manipulation?*

Example 3.3. Let us consider the following example to illustrate the application of our new manipulation strategy with uncertain knowledge. In this example, we use 20 sincere voters having the same preference profile as in Example 3.2 and 2 manipulators having a desired candidate, candidate 4, which they want to ensure his victory, i.e. $H=20$, $M=2$ and $p=4$ (see Section 1.3).

The preference profile will be as follows: 6 H: $1 > 2 > 3 > 4 > 5$, 3 H: $2 > 4 > 3 > 5 > 1$, 4 H: $3 > 4 > 2 > 1 > 5$, 4 H: $5 > 3 > 2 > 4 > 1$, 3 H: $4 > 5 > 3 > 2 > 1$ and 2 M: 4.

In the first round, the 3 functions used in the new manipulation strategy (Figure 3.1), are performed as follows:

- The querying function and the completion function are performed in the same way as in Example 3.2 in the first round ($t=1$). The process will elicit top-1 queries of each sincere voter over the set of candidates and I_1 represents the partial manipulators' knowledge in the first round. The output V of the completion function, represents the complete predicted profile obtained by applying IC probabilistic model ($C(I_1)$) on the sincere voters' votes.
- Winner selection function: given the predicted complete profile V of sincere voters in the first round, the manipulators will cast their vote i.e. p = candidate 4, and the Borda rule is applied to determine the approximate winner $\omega(V)$ of the election by considering the manipulators' vote. In this context, Borda scores are: $s(1) = 49$, $s(2) = 45$, $s(3) = 31$, $s(4) = 47$, $s(5) = 36$ and the approximate winner, with manipulation is, $\omega(V) = 1$. While the approximate winner is different from the desired one i.e. candidate 4, however, the manipulators succeeded in raising the chances of their preferred candidate by ranking him in the second place i.e. candidate 4 has the second highest score. If we want to determine a higher quality of approximate winner i.e. candidate 4 in the first place, we proceed to the second round by increasing the manipulators knowledge i.e. elicit top-2 queries of sincere voters.

In the second round, the 3 functions are performed as follows:

- The querying function and the completion function are performed in the same way as in Example 3.2 in the second round ($t=2$). The process will elicit top-2 queries of each sincere voter over the set of candidates and I_2 represents the partial manipulators' knowledge in the second round. The output V of the completion function, represents the complete predicted profile obtained by applying IC probabilistic model ($C(I_2)$).
- Winner selection function: given the predicted complete profile V of sincere voters in the second round, the manipulators will cast their vote, and the Borda rule is applied to determine the approximate winner $\omega(V)$ of the election by considering the manipulators' vote. In this context, Borda scores are: $s(1) = 39$, $s(2) = 44$, $s(3) = 37$, $s(4) = 53$, $s(5) = 35$ and the approximate winner, with manipulation is, $\omega(V) = 4$, which represents the desired candidate of the manipulators. In this case, the manipulators succeeded to ensure the victory of their preferred candidate by knowing only top-2 queries of sincere voters' votes.

To conclude, this example shows the effectiveness of our new manipulation approach, where eliciting only top-2 queries of sincere votes, in two rounds, are sufficient to allow the manipulators to achieve their goal.

3.5 Conclusion

In this chapter we have studied the connection between vote elicitation and manipulation problem. To this end, we have addressed the problem of coalitional manipulation using an efficient top- k elicitation process. Our first contribution is a *multi-round elicitation process* for choosing the ideal threshold k with the minimal number of rounds, in order to obtain a high quality of approximate winner. Then, our second contribution consists of a *new manipulation strategy under candidate uncertainty*, using the proposed multi-round elicitation process. This new approach aims to determine the minimal amount of information needed to be known by the manipulators for an optimal manipulation. The next chapter is dedicated to test the effectiveness of our proposals.

Experimental study

4.1 Introduction

In this chapter, we run several experiments with top-k preference distributions, as well as real voting data, to test the effectiveness of our multi-round elicitation process and the new manipulation strategy proposed. In this study, we are interested in:

- Evaluate the effectiveness of our multi-round elicitation process for determining the optimal top-k needed to make a high quality decision with minimal number of rounds.
- Evaluate the performance of our new manipulation strategy for determining the minimal number of candidates needed to be known by manipulators for an optimal manipulation, by measuring both the probability of manipulation and its impact on social welfare.

To this end, we will use two data sets, *Sushi data* and *Gen-Ran data*, to perform the experiments. We use *Python* programming language to implement the two proposed approaches. To complete profiles, we will consider two probabilistic models *IC* and *SP-IC* (see Section 2.2).

This chapter is organized as follows: Section 4.2 describes the data sets used to perform the different experiments. Section 4.3 is dedicated to investigate the performance of the multi-round elicitation process. Section 4.4 deals with manipulation strategy, where we will test the effectiveness of the *new manipulation strategy with uncertain knowledge*, by measuring the probability of manipulation and its impact on social welfare.

4.2 Data sets description

In this study, two data sets are used:

- *Sushi data* (Kamishima et al., 2005) from *Preflib*¹ (Mattei & Walsh, 2013). This data set contains the results of a series of surveys asking 5000 individuals for their preferences about various kinds of sushi. This work considers Sushi experiments with 5000 voters, where each voter has a complete preference order over a set of 10 alternatives.
- Since Sushi data is not a single peaked data (see Section 1.3), we have randomly generated a data set so called *Gen-Ran data* with 10 alternatives and 1000 voters having a single peaked preferences.

The preference data presented in *Preflib* are expressed in an easily accessible and readable format. Figure 4.1, presents an example of Sushi data format with 4 kinds of sushi and 9 individuals, i.e. $A = \{ebi, anago, maguro, ika\}$ and $N = 9$. The first line presents the number of alternatives. The different kind of Sushi are presented in the four next lines. The fifth line which contains '9,9,3', indicates: the number of individuals, sum of vote count and number of unique orders, respectively. Finally, the remaining lines, present the preference order of the different individuals; where the first number of each line illustrates the number of voters who has this specific preference order.

```

4
1,ebi (shrimp)
2,anago(seaeel)
3,maguro(tuna)
4,ika(squid)
9,9,3
3,8,5,6,2,1,3,9,4,10,7
3,2,5,6,8,3,1,9,7,10,4
3,8,3,6,5,9,2,1,4,10,7

```

Figure 4.1: Sushi data format

¹<http://www.preflib.org/>

4.3 Multi-round elicitation experiments

The multi-round elicitation process aims to minimize the amount of information elicited from voters, as well as the number of rounds. To this end, we run a suit of experiments using Sushi data and Gen-Ran data.

4.3.1 Sushi data experiments

In order to evaluate the performance of our proposed multi-round elicitation process, our first set of experiments uses Sushi data by drawing 10 preference profiles each with 5000 voters. Each profile contains *top-k* preference order $k \in \{1, \dots, 9\}$ with 5000 rankings. We simulate the elicitation of top-k preferences and measure both margin of victory of partial preferences under *IC* and *SP-IC* models for $k \in \{1, \dots, 9\}$ comparing with the complete profile ($k=10$); and the winner quality by measuring the probability of deviation from the true winner. Borda rule is used in order to evaluate the outcome.

Figure 4.2 (resp. Figure 4.3), shows histograms using Sushi data with *SP-IC* (resp. *IC*) model. That is, in each collection of histograms, as defined by a particular top-k queries $k \in \{1, \dots, 10\}$, we generated 10 profiles over a set of 10 alternatives. For each profile and each k , we compute the margin between the scores of different alternatives in each profile, and the true scores by considering the case of complete profile. We use Borda scoring to evaluate the scores in each profile. In each histogram, the horizontal axis corresponds to the different alternatives in a partial profile and the vertical axis represents their scores. The bins of each partial profile in the histogram (blue bins), are compared with those of the complete one (red bins).

Let us start by analyzing Figure 4.2. Clearly, the margin is always zero when $k=9-10$. Also, with top-k queries, where $k \in \{8, 7, 6, 5\}$, the winner is always determined with a slight difference in the scores of the other alternative. While our goal is not only to predict the true winner, but to minimize the impact on social welfare by predicting the correct ranks of the largest number of alternatives; our approach succeeds to achieve this goal. For instance the true ranking of the alternatives with $k=10$ is: 8, 3, 1, 6, 2, 5, 4, 9, 7, 10; with top-k queries $k \in \{7, 8\}$ (resp. $k \in \{5, 6\}$), the approximate ranking ranks 9 alternatives among 10 (resp. 6 among 10 alternatives) in the correct position. Now, let us consider a small number of $k \in \{1, 2, 3, 4\}$; eliciting top-1 queries from voters are not sufficient to determine the true winner. Starting from $k \in \{2, 3, 4\}$, the true winner is always predicted, with 4 correct ranking among 10. To conclude, results with Sushi data using the *SP-IC* model, show that eliciting information from voters in 2 rounds (top-2) is always sufficient to predict the correct winner. Results with 3 rounds (top-3), are a very good approximate for the largest number of correct ranking.

Results with Sushi data in Figure 4.3, show that our multi-round elicitation process performs better using *IC* model than the first model. That is, the true winner is always predicted with all values of $k \in \{1, \dots, 9\}$ i.e. the predicted winner is always alternative 8 which represents the true winner. This is also shown by the slight difference in the scores of different alternatives in the histograms. Results with 2 rounds (top-2), are a very good approximate for the largest number of correct ranking.

Figure 4.4 summarizes the above experiments performed with Sushi data, by showing, in plot (a) (resp. (b)), the margin of victory of the 10 preference profiles, by increasing the number k of alternatives elicited $k \in \{1, \dots, 9\}$ in each profile, using *IC* model (resp. *SP-IC*). The peaks in each preference profile, in a decreasing order, present the order of the winners using the Borda rule. Clearly, under *IC* model, the shape of the different curves is the same, which means that even with partial preferences, the scoring rule ranks the alternatives, almost, in the same order i.e. top-4 queries are usually sufficient to rank all alternatives in the same order comparing with the true ranking ($k=10$). More interestingly, in order to evaluate the impact on social welfare, eliciting preferences from voters in 2 rounds (top-2), guarantee not only a correct outcome, but also a correct ranking of 4/10 alternatives. Sushi results, with *IC* model, show that it is always possible to determine the winner with incomplete information, even with $k = 1$ i.e. alternative 8 has the highest score with different values of k which represents the true winner. Sushi's results with *SP-IC* model (plot b) show that the margin of victory is sparse. For small value of $k = 1$, the approximate winner deviate from the true one i.e. with $k=1$, the approximate winner is alternative 5 while the true winner is alternative 8. Results with *SP-IC* model, suggest that, with top-3 queries, one can usually find the true winner; but top-2 queries are usually enough to obtain the desired outcome.

To conclude, the *IC* model performs well with Sushi data with small values of $k = 1, 2$, this is shown in the 3rd plot where the probability of deviation of the true ranking is of 27% under *IC* model against 34% under *SP-IC* model. With top-3 queries, the true ranking is determined in the most cases with a slight deviation of 17%. Intuitively, while increasing the values of k , the process converges to the correct prediction, and is near perfect with top-3 under the two models (deviation of 15%).

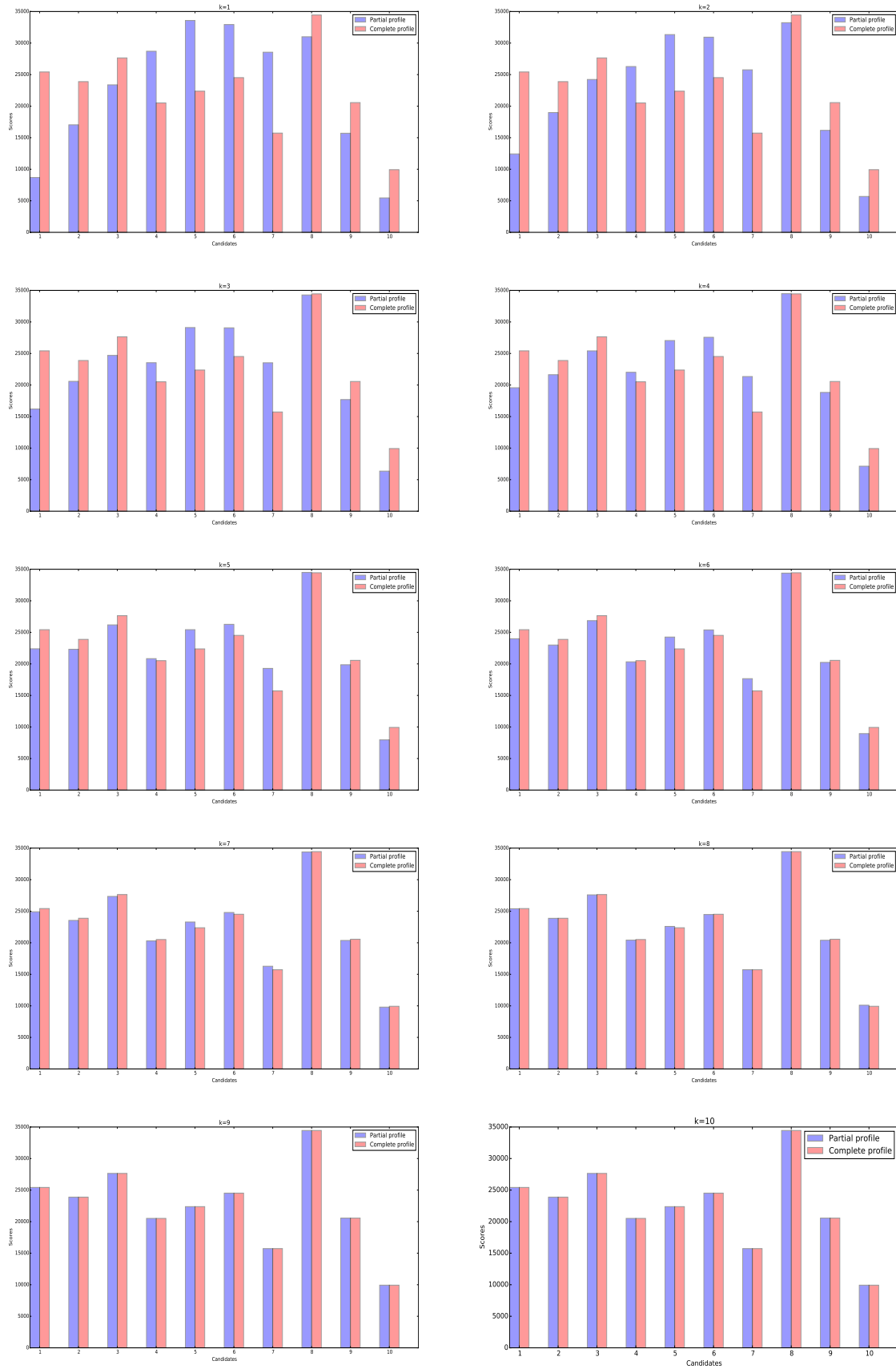
Figure 4.2: Sushi data histograms with $SP-IC$ model

Figure 4.3: Sushi data histograms with IC model.

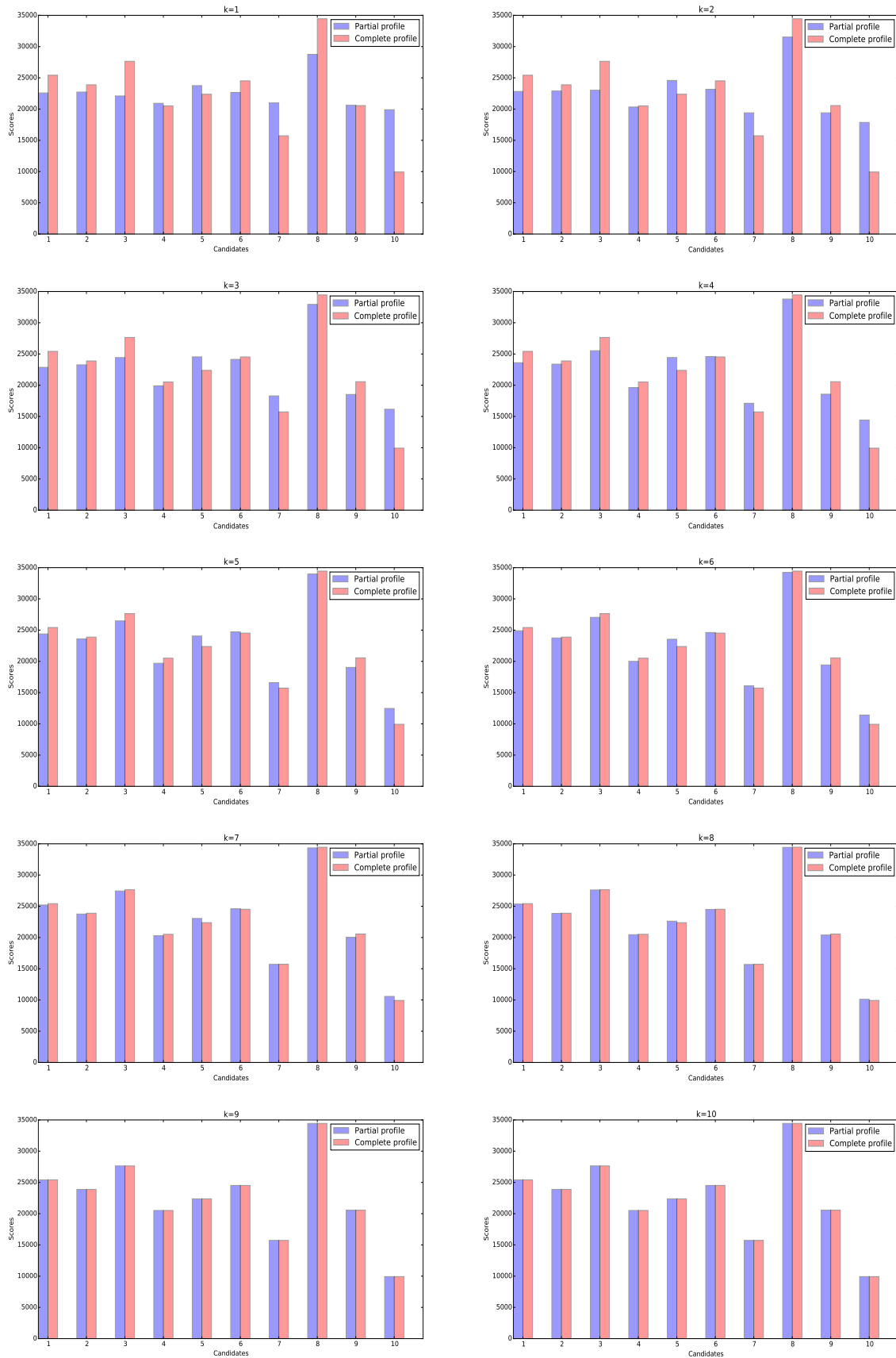
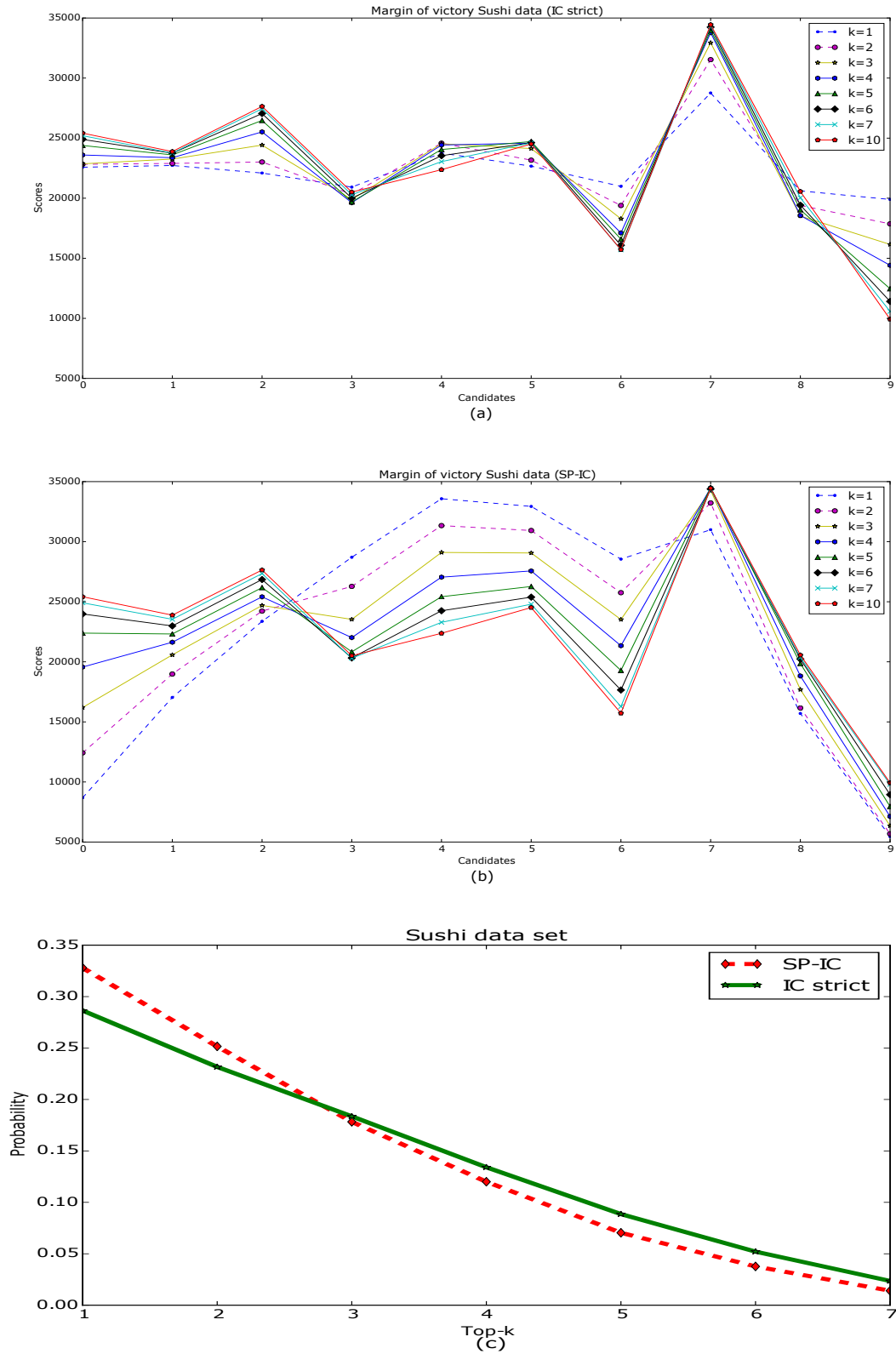


Figure 4.4: (a) Summary of Sushi results with *IC* model, (b) Summary of Sushi results with *SP-IC*, (c) Probability of the approximate winner quality.



4.3.2 Gen-Ran data experiments

Recall that preferences are single peaked if there is some way of lining up all the alternatives, so that the graph of every voters preferences has a single local maximum. To this end, we run a suit of experiments with Gen-Ran data, in order to test the ability of a single peaked data to predict the correct outcome using two models *IC* and *SP-IC*. With Gen-Ran data, we drew 10 profiles each with: $N=1000$ voters and with top- k preferences order $k \in \{1, \dots, 10\}$. We simulate the elicitation of top- k preferences and measure both the margin of victory of partial preferences using the two models for $k \in \{1, \dots, 9\}$ comparing with the complete profile ($k=10$); and the winner quality by measuring the probability of deviation from the true ranking.

Figure 4.5 (resp. Figure 4.6), shows histograms using Gen-Ran data with *IC* (resp. *SP-IC*) model. That is, in each collection of histograms, as defined by a particular top- k queries $k \in \{1, \dots, 10\}$, we generated 10 profiles over a set of 10 alternatives, each with $N=1000$ voters. For each profile and each k , we compute the margin between the scores of different alternatives in each profile, and the true scores by considering the case of complete profile ($k=10$). Similarly to the above experiments, we use Borda rule to evaluate the scores in each profile. In each histogram with a specific value k , the horizontal axis corresponds to the different alternatives in a partial profile and the vertical axis represents their scores. The bins of each partial profile in the histogram (blue bins), are compared with those of the complete one (red bins).

Interesting results are shown when analyzing Gen-Ran data using *IC* model (Figure 4.5), where the margin is almost near to zero when $k \in \{3, 4, 5, 6, 7, 8, 9\}$. Since, not only the true winner is determined, but also the model ranks the alternatives in the correct ranking with a slight difference in the scores. Considering a small values of $k \in \{1, 2\}$, both the correct winner and the correct ranking are predicted, however the bins of partial profiles are not close enough which is due to the lack of information with top-1 and top-2 queries. Results with *SP-IC* are even more illuminating, where the margin between the bins with all values of k is slight and near to zero. That is, *SP-IC* model performs well with Gen-Ran data which represents a single peaked data set. Clearly, top-1 query, with only one round, is always sufficient to determine the true winner with the lowest loss in social welfare.

Figure 4.7 summarizes experiments performed on Gen-Ran data. Clearly, with *IC* model (plot (a)), the curves' shape, with different values of k , is almost the same; except with $k = 1, 2$ the true winner is not determined and the vector score deviates from the true one; but an additional alternative (top-3 queries) is sufficient to determine the true winner with approximately the same score vector. A very good approximate winners are obtained with top-4 queries. Results in plot (c) show that the probability of deviation from true ranking is evaluated to 30% with top-1 and decreases by increasing the number of queries. A very good results are obtained using *SP-IC* where the shape of the curve is the same with all values of k . This is, with only one round

of elicitation, and with top-1 preferences revealed by voters; the correct outcome is determined with a correct ranking over the set of alternatives i.e. a negligible impact on social welfare. The probability of deviation from the true ranking is showed clearly in (plot (c)), which is evaluated to 15% with all the incomplete preferences $k \in \{1, \dots, 7\}$.

Let us turn our attention to the normal form of the distribution generated by Ran-Gen data, constituted of single peaked preferences. Our results suggest that using *IC* and *SP-IC* models, to complete an original single peaked data, generate always a single peaked vote able to determine the correct ranking of alternatives in almost all cases (except $k=1-2$ with *IC* model). For instance, plots (a) and (b), present a local maximum on alternative 4 i.e. alternatives graphed along the horizontal axis have one peak on alternative 4 which represents the winner with higher score, and then the curve goes down with the other alternatives in a decreasing order.

Figure 4.5: Gen-Ran data histograms with IC model

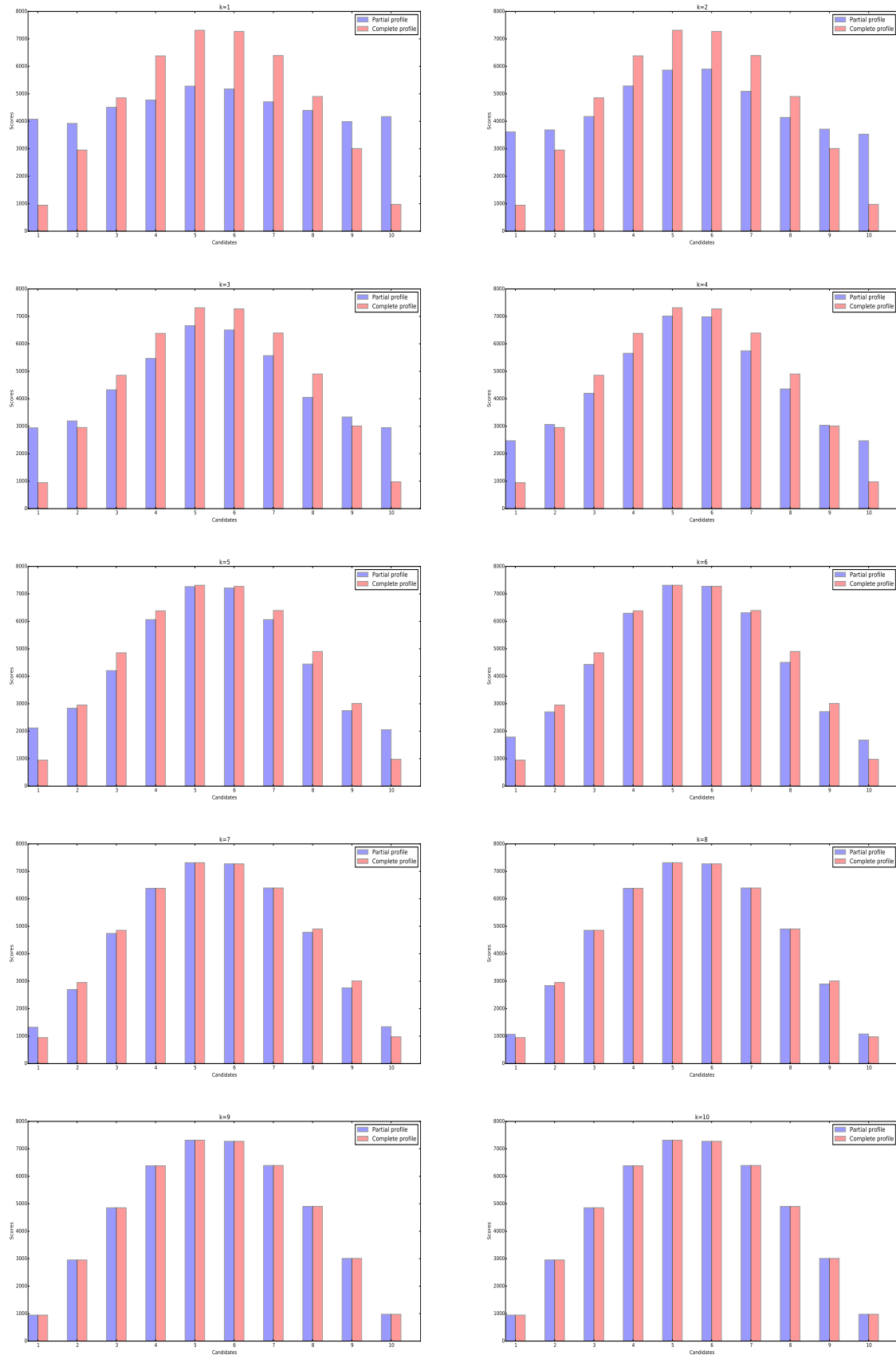


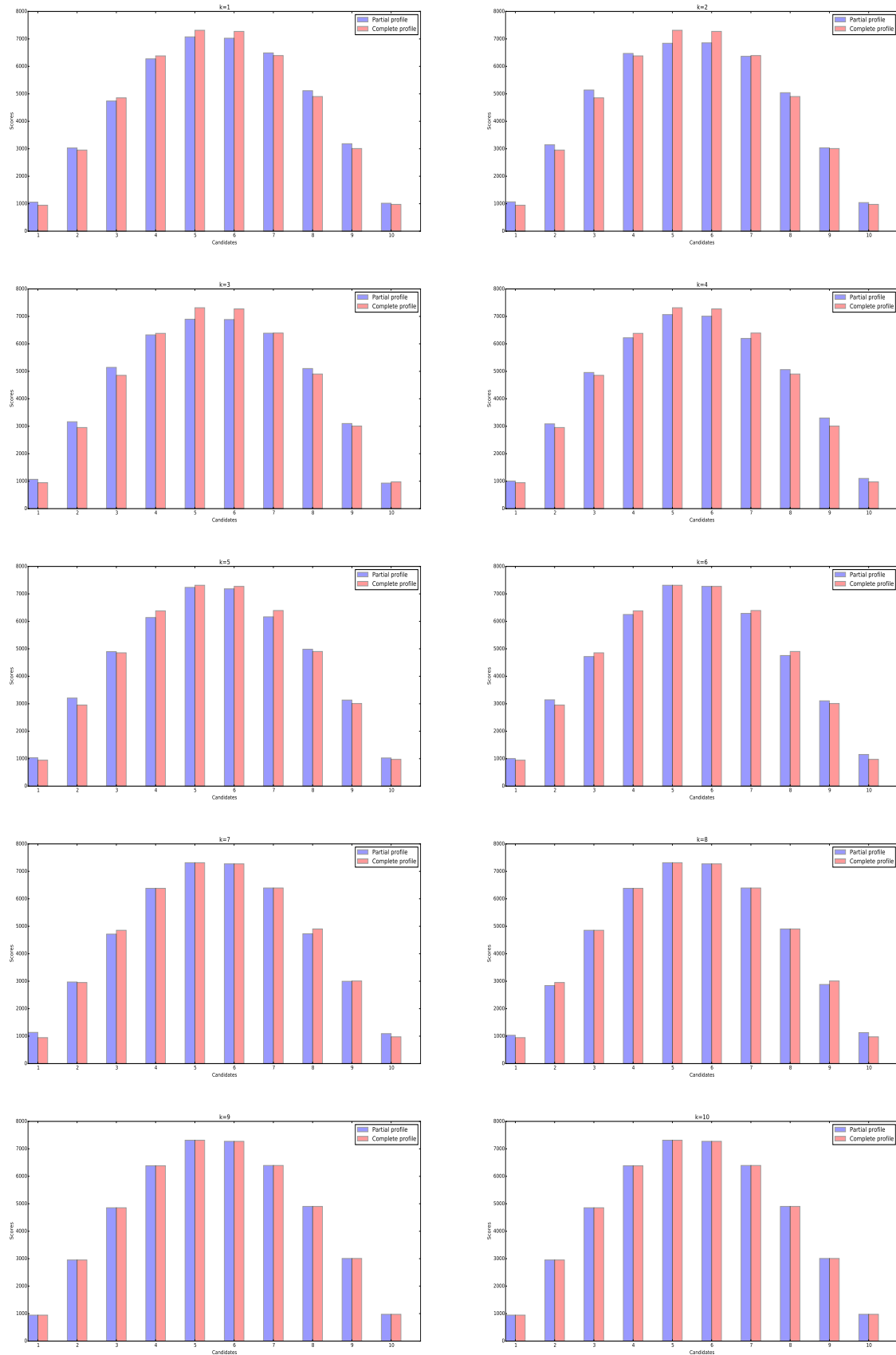
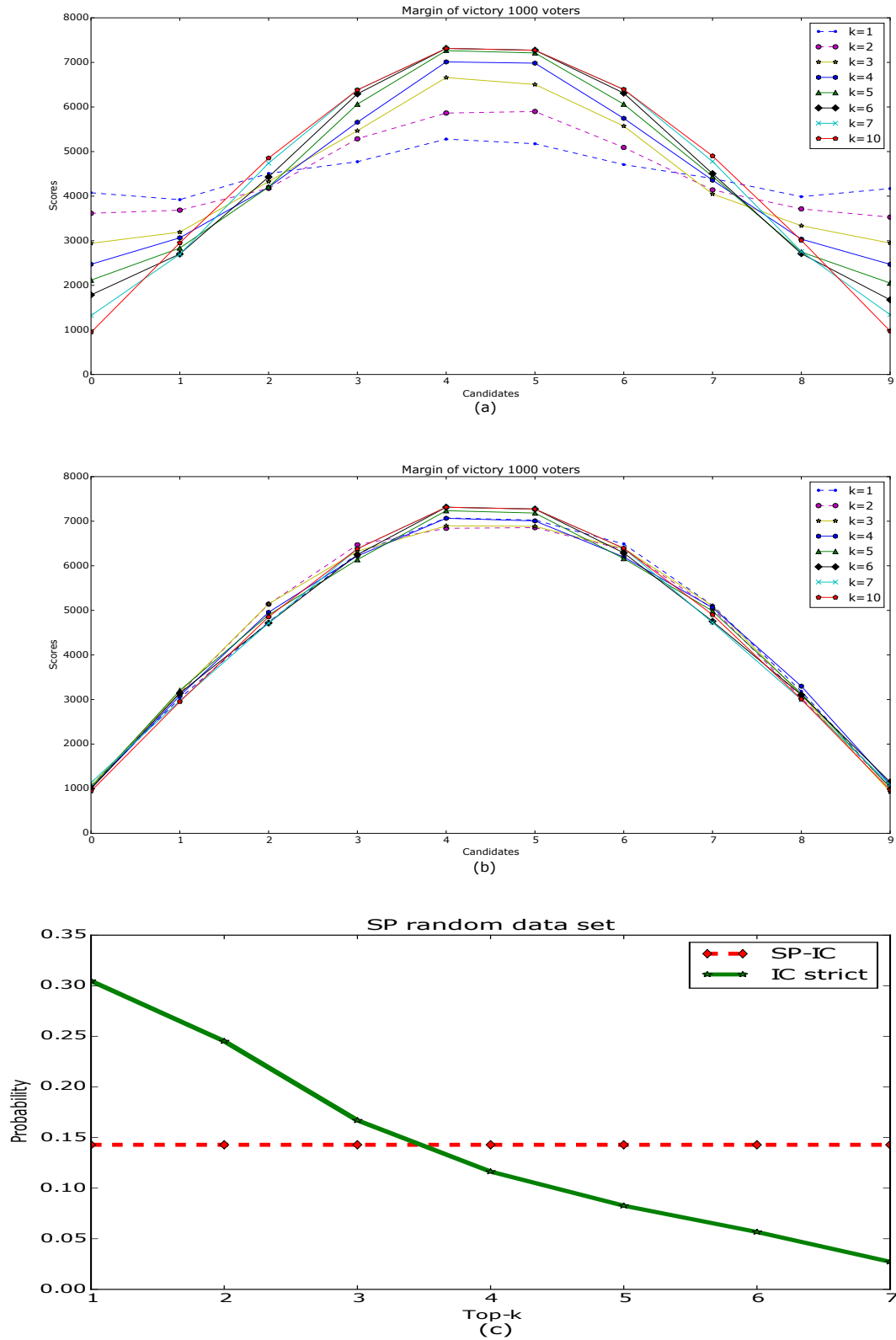
Figure 4.6: Gen-Ran data histograms with $SP-IC$ model

Figure 4.7: (a) Gen-Ran data with *IC* model, (b) Gen-Ran data results with *SP-IC*, (c) Probability of the approximate winner quality.



4.4 New manipulation strategy experiments

In order to test the effectiveness of our manipulation approach, we measure both the probability of manipulation and its impact on social welfare. To this end, we use Sushi data to test the manipulators' strategy. Our choice to this specific data set, and not Gen-Ran data; is due to the Black's median voter theorem (Black, 1948) which implies that Gibbard-Satterthwaite's negative conclusions regarding preference aggregation (Gibbard, 1973; Satterthwaite, 1975) i.e. any voting rule with at least three alternatives, that is non-dictatorial is manipulable; disappear if one assumes that the alternatives can be ordered along a social axis so that every individual's preference over the alternatives is single-peaked (see Section 1.3). In other word, when we consider single peaked preferences, such as Gen-Ran data, the manipulation is not possible.

In this context, we consider Sushi data by drawing 10 profiles such that, each of them contains top- k preferences order $k \in \{1, \dots, 10\}$. We vary the number of manipulators $M \in \{500, 750, 1000\}$ with $N= 5000$ voters in each profile, where $M= 500$ manipulators (resp. 750, 1000), represents 10% (resp. 15% and 20%) of the total number of voters. We use the *IC* model (see Section 2.2) to complete non-manipulators votes, since its effectiveness performed with Sushi data in the above experiments. Since manipulability varies greatly with the preferred alternative chosen by the manipulators, we show results for the alternatives whose expected ranks in different profile are second, third, and fourth i.e. $P=2$, $P=3$, $P=4$.

Figure 4.8 shows results of three different manipulators strategies with $M=750$. In plot (a) (resp. plot (b), plot (c)), the manipulators preferred alternative is the one whose expected rank in different profiles is second, $P=2$ (resp. $P=3$, $P=4$). In the horizontal axis, we represent the 10 preference profiles by varying the value of $k \in \{1, \dots, 5\}$ in each profile. For instance, in plot (a), $K=2$ means that the manipulators know only top-2 preferences of sincere voters; based on these incomplete information, their aim is to ensure the victory of the alternative whose expected rank is second ($P=2$), that is, to make him in the first rank. In order to evaluate the effectiveness of different manipulators' strategies under candidate uncertainty; in each plot, 'opt' refers to the true ranking of alternatives, with complete preferences and without manipulation, which is: 8, 3, 1, 6, 2, 5, 4, 9, 7, 10. The different ranks are presented by the peaks in the graph. Interesting results are presented with top-1 queries, where, with different manipulation strategies and knowing only one preference of non-manipulators votes, the manipulators can ensure the victory of their most preferred alternative. However, an additional knowledge about the non-manipulators votes (top-2 queries), allows the manipulators to only increase their chances of winning, by raising their most preferred alternative to the second rank. For instance, when the manipulators strategy is $P=3$, i.e. manipulators desired alternative is 1; with top-1 queries, the winner is alternative 1; however, with top-2, top-3 and top-4 queries, the winner is always the alternative 8, while alternative 1 is ranked second. In summary, our results with $M=750$ manipulators, suggest that the manipulators

goal can be achieved only if the manipulators know top-1 queries of the non-manipulators votes.

Let us now consider a larger number of manipulators when $M=1000$, presented in Figure 4.9. Results show that, when the manipulators preferred alternative is in the second expected rank ($P=2$), the manipulators are always able to ensure their victory. For instance, in plot (a), when the manipulators desired alternative is 3, this latter is ranked first and declared the winner with all values of $k \in \{1, \dots, 5\}$. However, when $P=3$ (resp. $P=4$), manipulators are able to achieve their goal only with restricted knowledge, that is, top-1 and top-2 queries (resp. top-1, top-2 and top-3 queries); with $k \in \{3, 4, 5\}$ (resp. $k \in \{4, 5\}$), manipulators can only increase the chances of their preferred alternative by ranking him second. While our goal is to determine the minimal amount of information elicited, our results with $M=1000$, suggest that top-1 queries of the manipulators knowledge, are always sufficient to achieve a successful manipulation, that is, with any manipulators' strategy.

Figure 4.8: (a) Manipulation strategy with $P=2$ and $M = 750$, (b) Manipulation strategy with $P=3$ and $M = 750$, and (c) Manipulation strategy with $P=4$ and $M = 750$

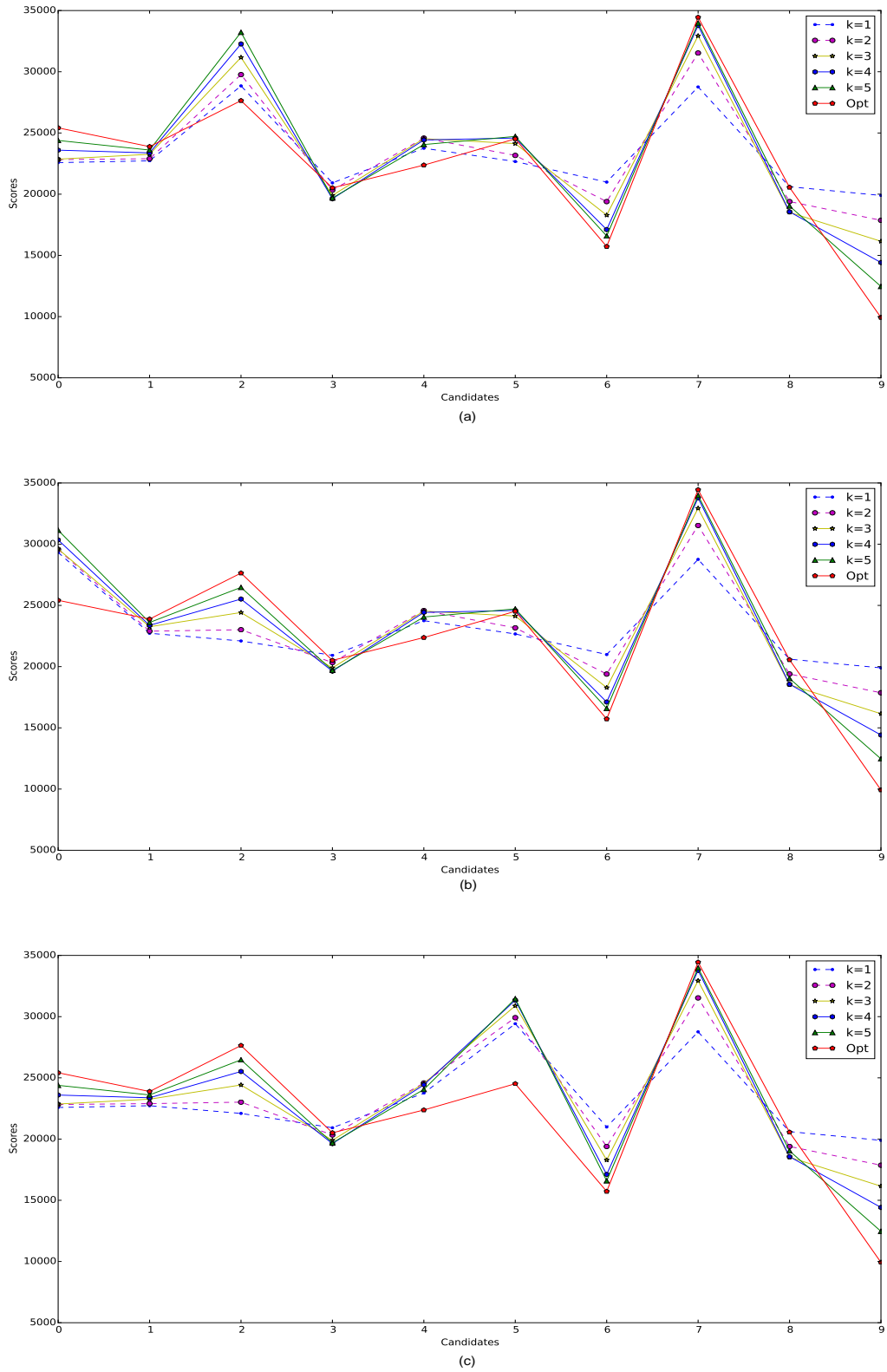


Figure 4.9: (a) Manipulation strategy with $P=2$ and $M = 1000$, (b) Manipulation strategy with $P=3$ and $M = 1000$, and (c) Manipulation strategy with $P=4$ and $M = 1000$

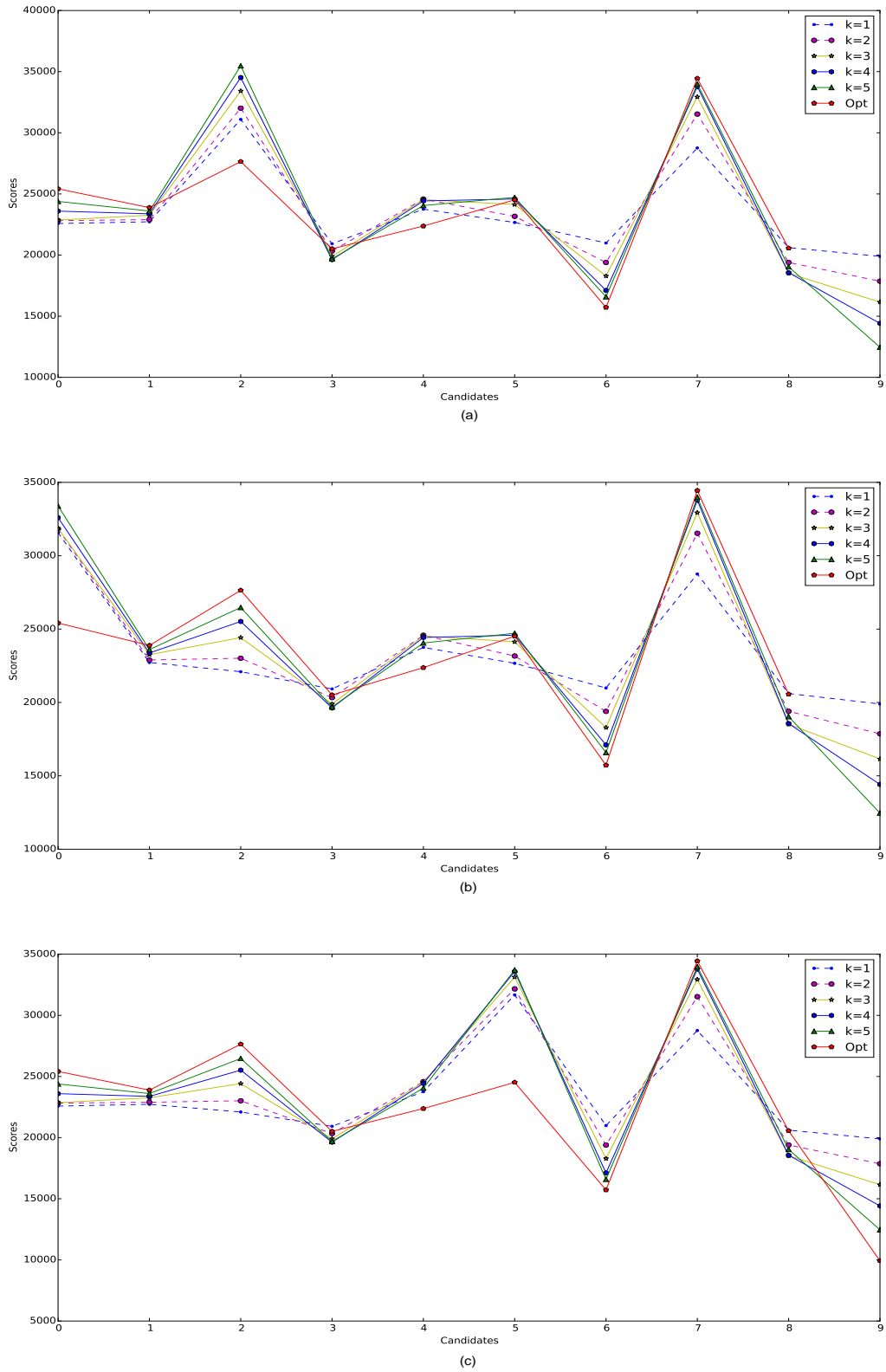
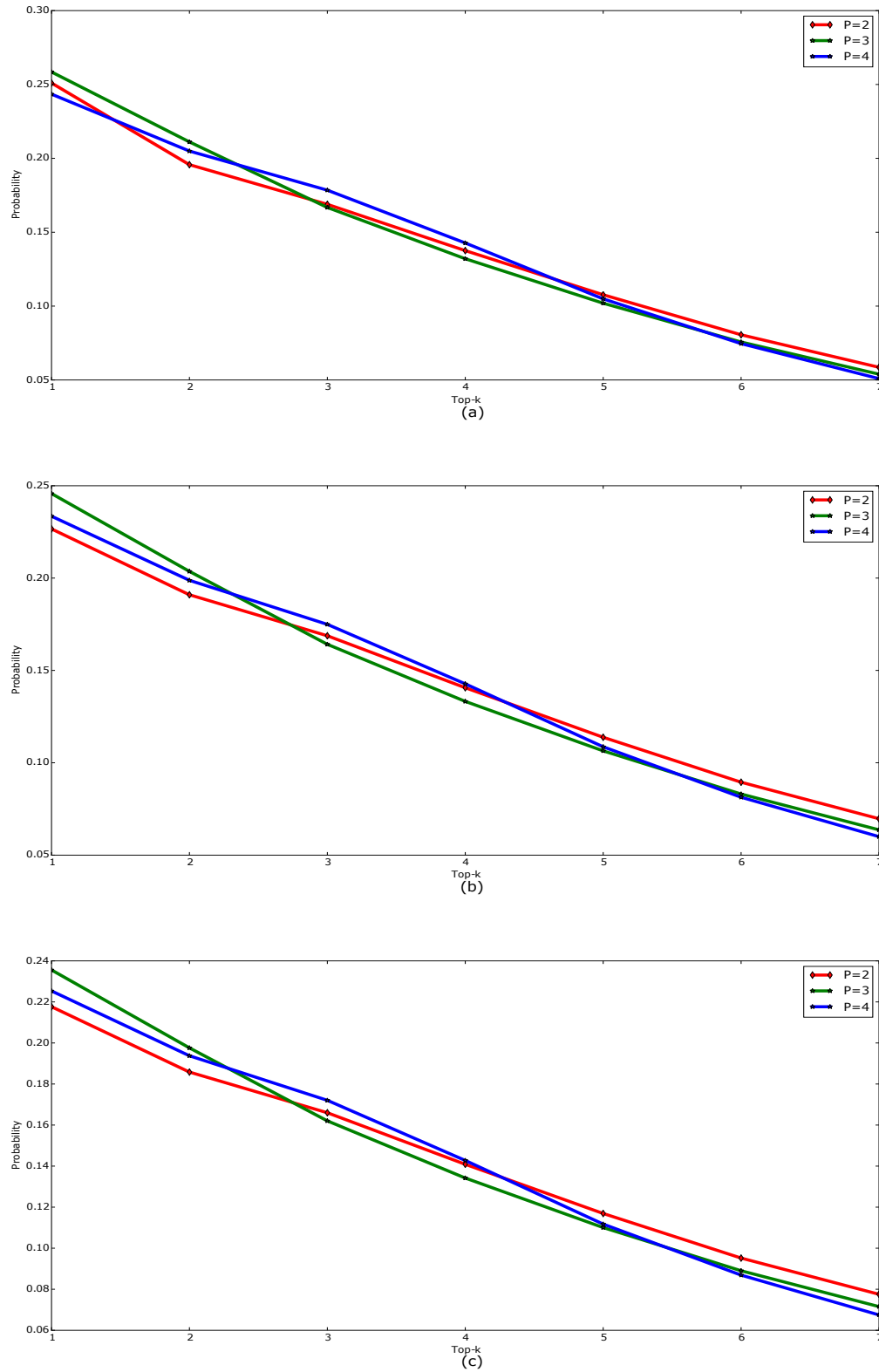


Figure 4.10: (a) Probability of optimal manipulation with $M = 500$, (b) Probability of optimal manipulation with $M = 750$, and (c) Probability of optimal manipulation with $M = 1000$



The impact of manipulation on social welfare is evaluated in Figure 4.10, by measuring the loss on social welfare depending on the knowledge held by the manipulators, as we vary the manipulators number $M \in \{500, 750, 1000\}$ and consider desired alternatives $P=2, P=3$ and $P=4$. Results show that the impact of manipulation on social welfare is 'high' when manipulators know less about the votes of the non-manipulators, and the probabilities gradually drop as k grows. For instance, in plot (b), with 750 manipulators who preferred $P = 3$, loss in social welfare is of 24% (resp. 20%, 16% and 12%) when they know only top-1 (resp. top-2, top-3 and top-4) of non-manipulators votes. Also, the probability of optimal manipulation depends on the chosen alternative by the manipulators under uncertainty. For instance, with top-2 queries, the curve of $P = 2$ is always on the bottom with the different number of manipulators, while, with top-3 queries, manipulators can ensure the victory of $P = 3$ with low loss. Finally, results show that 10% of manipulators ($M=500$), are able to change the outcome with small impact on social welfare by increasing the chances of victory of their most preferred alternative. While, 15% of manipulators are able to make their alternative win the election.

4.5 Discussion

In order to evaluate the effectiveness of our multi-round elicitation process, we propose to compare our results with those obtained in (Lu & Boutilier, 2011b). Authors in (Lu & Boutilier, 2011b) theoretically outlined a general framework for the design of a multi-round elicitation protocol, however they have empirically dealt only with one-round elicitation of top- k candidates. In this context, they have estimated the expected minimax regret using the *MMR* solution (see Section 2.3) to determine winners given partial profile. As in our study, they have used Borda scoring rule to evaluate the outcome of the election. Their results with Sushi data, suggest that with top-5 queries one can usually find the true winner; but top-4 queries are usually enough to obtain low regret and determine a high quality of the winner. While our goal is to minimize the information elicited as well as the number of rounds, our results are more efficient when we consider probabilistic models to complete profiles. Our results suggest that top-1 (resp. top-2) queries, are sufficient to determine the correct winner in one round (resp. 2 rounds), using the *IC* (resp. *SP-IC*) probabilistic model. More interesting results are obtained with single peaked data, where the winner of the election is always determined with any value of k , using the two probabilistic models.

More recent work proposed by authors in (Doucette, Larson, & Cohen, 2014), addressing the problem of determining the outcome of an election with only partial information. In this context, authors propose a novel application with stronger connections to machine learning, where they used classification algorithms to predict the missing preferences of a ballot via latent patterns in the partial information provided. To this end, an imputation based approach to social choice was

derived relying on the ability of machine learning algorithms to provide reasonable imputations of user’s preferences in order to deal with incompleteness. As in our work, authors consider partial elicitation process in the form of top-k style preferences and Borda scoring rule to select the winner. However, our work differs from this recent context insofar as it uses conventional social choice techniques and recommends a particular elicitation strategy for voters i.e. the voter is asked to incrementally send his partial preferences in rounds, one preference each round; but instead works with the partial preferences to achieve the same goal.

While the proposed approach in (Doucette et al., 2014) provides good results with incomplete data expressed by a low error rate using the SVM classifier, however, there are a number of reasons to suppose our multi-round vote elicitation process should be preferred, at least in some situations. Chief among them is that in the case where the ballots have exceptionally high incompleteness (e.g. voters provide only top-1 preferences), classification becomes a difficult problem with the imputation method requiring additional refinement to perform well, thus, the correct winner is not predicted. In the similar situation, our results show that even with high incompleteness; not only the correct winner is determined but also the damage on social welfare is limited, which guarantee a high approximation of the entire ordering using a scoring rule.

Regarding the evaluation of our new manipulation strategy, there is no study in the literature that investigates the relation between the amount of information and the success of manipulation, empirically.

4.6 Conclusion

Experiments with two different data sets (namely *Sushi data* from Preflib and a randomly generated single peaked data *Gen-Ran data*), demonstrate the practical viability and advantages of our multi-round elicitation process to minimize the amount of information elicited; with a high quality of the approximate winner. Also, experiments performed with the new manipulation strategy indicate that our approach is quite tractable. Our results suggest that an optimal manipulation is always obtained when the manipulators have restricted knowledge (top-1 queries). Specifically, the probability of manipulation is sometimes quite significant with fewer manipulators’ knowledge than with full one. Furthermore, our results indicate that a few number of manipulators are able to change the outcome (15% of manipulators).

Conclusion

This work has studied the connection between vote elicitation and manipulation problem. In this context, we have addressed the problem of uncertainty in manipulators' knowledge, using an efficient multi-round elicitation process. Our contribution is of two-fold:

- First, we propose a *multi-round elicitation process* for choosing the ideal threshold k . In this setting, the voters are asked to provide the length k prefix of their ranking; in order to determine the minimal amount of information, with the minimal number of rounds, needed to predict the right outcome of the election.
- Second, we propose a *new manipulation strategy with uncertain knowledge*, where the incomplete knowledge of the manipulators is solved using the proposed multi-round elicitation process. This new approach presents an empirical framework for computational optimal manipulation strategies, when manipulators have incomplete knowledge about non-manipulators votes; by determining the minimal amount of information needed to be known by the manipulators to achieve their goal with less damage on social welfare.

Experiments with two different data sets (namely *Sushi data* from Preflib and a randomly generated single peaked data *Gen-Ran data*), prove the practical viability and advantages of our multi-round elicitation process to predict the right outcome of the election; with the minimal amount of information elicited. Results show that eliciting top-1 preferences from voters is always sufficient to determine the true winner in only one round, with *IC* model. Interesting results are derived when we consider a single peaked data, in particular, the high quality of the approximate winner using only top-1 preferences from voters.

Moreover, experiments performed with the new manipulation strategy indicate that our approach is quite tractable. Specifically, our approach allows the determination of an optimal manipulation strategy using only a small fraction of sincere voters' preferences (top-1) and with less

impact on social welfare. Also, our results show that with restricted knowledge, the manipulators are always able to change the outcome to their favor, however when we increase the number of queries elicited, the manipulators can only raise the rank of their preferred candidates to the second place. Furthermore, our results suggest that a few number of manipulators are able to change the outcome (15% of manipulators). In addition, our experiments demonstrate that manipulation may not be as undesirable a phenomenon as is commonly believed, when the quality of the outcome is the main objective i.e. minimize the impact of manipulation on social welfare.

While our results show the effectiveness of the proposed approaches with incomplete preferences, a number of interesting avenues remain:

Our first interest is to examine the implication of single-peakedness data on our manipulation strategy. That is, study if it is sufficient to eliminate the possibility of profitable manipulation by using single peaked preferences as demonstrated in Black's median voter theorem (Grofman, 2004). Assessing the impact of manipulation on single peaked preferences is a very important step since many intractable voting problems become tractable for single-peaked profiles (Faliszewski, Hemaspaandra, Hemaspaandra, & Rothe, 2009).

Extending our analysis to a richer class of probabilistic models, such as the Mallows model (Mallows, 1957) which is used to learn preferences distributions from voters, allowing inferences to be drawn about their preferences. This model performs well with pairwise preferences form "e.g., I like alternative a better than b ", than with $top-k$ preferences "e.g., I like a better than b , better than c, \dots ". While our data is in the latter form, as a next step, we aim to test Mallows model on our framework by considering the appropriate data set.

Investigate the use of probabilistic conditional preference networks (PCP-nets) as a way to aggregate partial preferences from voters. Since PCP-nets provide a compact representation of a probability distribution over a collection of CP-nets (Cornelio et al., 2014), adapting PCP-nets to our multi-round elicitation process might yield interesting results.

On the theoretical side, deriving bounds on the amount of information required when eliciting voters preferences is a very interesting step to prove the effectiveness of our empirical results.

References

- Arrow, K. J. (1951). Social choice and individual values. *Wiley, New York*.
- Arrow, K. J., & Raynaud, H. (1986). Social choice and multicriterion decision-making. *MIT Press Books, 1*.
- Barbera, S., Bogomolnaia, A., & Van Der Stel, H. (1998). Strategy-proof probabilistic rules for expected utility maximizers. *Mathematical Social Sciences*, 35(2), 89–103.
- Bartholdi III, J. J., & Orlin, J. B. (1991). Single transferable vote resists strategic voting. *Social Choice and Welfare*, 8(4), 341–354.
- Bartholdi III, J. J., Tovey, C. A., & Trick, M. A. (1989). The computational difficulty of manipulating an election. *Social Choice and Welfare*, 6(3), 227–241.
- Black, D. (1948). On the rationale of group decision-making. *The Journal of Political Economy*, 23–34.
- Black, D., Newing, R. A., McLean, I., McMillan, A., & Monroe, B. L. (1958). *The theory of committees and elections*. Springer.
- Conitzer, V., Lang, J., & Sandholm, T. (2003). How many candidates are needed to make elections hard to manipulate? In *Proceedings of the 9th conference on theoretical aspects of rationality and knowledge* (pp. 201–214).
- Conitzer, V., & Sandholm, T. (2002a). Complexity of manipulating elections with few candidates. , 314–319.
- Conitzer, V., & Sandholm, T. (2002b). Vote elicitation: Complexity and strategy-proofness. In *Aaai/iaai* (pp. 392–397).
- Conitzer, V., & Sandholm, T. (2005). Communication complexity of common voting rules. In *Proceedings of the 6th acm conference on electronic commerce* (pp. 78–87).
- Conitzer, V., Sandholm, T., & Lang, J. (2007). When are elections with few candidates hard to manipulate? *Journal of the ACM (JACM)*, 54(3), 14.
- Conitzer, V., Walsh, T., & Xia, L. (2011). Dominating manipulations in voting with partial information. In *Aaai* (Vol. 11, pp. 638–643).

- Cornelio, C., Grandi, U., Goldsmith, J., Mattei, N., Rossi, F., & Venable, K. (2014). Voting with cp-nets using a probabilistic preference structure. In *Proc. of the 5th international workshop on computational social choice (comsoc)*.
- de Caritat, M. J. A. N., et al. (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. L'imprimerie royale.
- Ditmarsch, H., Lang, J., & Saffidine, A. (2013). Strategic voting and the logic of knowledge.
- Doucette, J. A., Larson, K., & Cohen, R. (2014). Approximate winner selection in social choice with partial preferences. In *Proceedings of the 12th international conference on autonomous agents and multiagent systems (aamas)*. international foundation for autonomous agents and multiagent systems.
- Emerson, P. (2013). The original borda count and partial voting. *Social Choice and Welfare*, 40(2), 353–358.
- Faliszewski, P., Hemaspaandra, E., Hemaspaandra, L. A., & Rothe, J. (2009). The shield that never was: Societies with single-peaked preferences are more open to manipulation and control. In *Proceedings of the 12th conference on theoretical aspects of rationality and knowledge* (pp. 118–127).
- Fishburn, P. C. (1973). *The theory of social choice* (Vol. 264). Princeton University Press Princeton.
- Gibbard, A. (1973). Manipulation of voting schemes: a general result. *Econometrica: journal of the Econometric Society*, 587–601.
- Grofman, B. (2004). Black's single-peakedness condition. In *The encyclopedia of public choice* (pp. 367–369). Springer.
- Guilbaud, G. T. (1952). Les théories de l'intérêt général et le problème logique de l'agrégation. *Economie appliquée*, 5, 501–584.
- Hazon, N., Aumann, Y., Kraus, S., & Wooldridge, M. (2012). On the evaluation of election outcomes under uncertainty. *Artificial Intelligence*, 189, 1–18.
- Hemaspaandra, L. A., Lavaee, R., & Menton, C. (2013). Schulze and ranked-pairs voting are fixed-parameter tractable to bribe, manipulate, and control. In *Proceedings of the 2013 international conference on autonomous agents and multi-agent systems* (pp. 1345–1346).
- Kalech, M., Kraus, S., Kaminka, G. A., & Goldman, C. V. (2011). Practical voting rules with partial information. *Autonomous Agents and Multi-Agent Systems*, 22(1), 151–182.
- Kamishima, T., Kazawa, H., & Akaho, S. (2005). Supervised ordering-an empirical survey. In *Data mining, fifth ieee international conference on* (pp. 4–pp).
- Kelly, J. S. (1991). Social choice bibliography. *Social Choice and Welfare*, 8(2), 97–169.
- Konczak, K., & Lang, J. (2005). Voting procedures with incomplete preferences. In *Proc. ijcai-05 multidisciplinary workshop on advances in preference handling* (Vol. 20).
- Kuga, K., & Nagatani, H. (1974). Voter antagonism and the paradox of voting. *Econometrica: Journal of the Econometric Society*, 1045–1067.
- Lu, T., & Boutilier, C. (2011a). Robust approximation and incremental elicitation in voting

- protocols. In *Ijcai proceedings-international joint conference on artificial intelligence* (Vol. 22, p. 287).
- Lu, T., & Boutilier, C. (2011b). Vote elicitation with probabilistic preference models: Empirical estimation and cost tradeoffs. In *Algorithmic decision theory* (pp. 135–149). Springer.
- Lu, T., Tang, P., Procaccia, A. D., & Boutilier, C. (2012). Bayesian vote manipulation: Optimal strategies and impact on welfare. *arXiv preprint arXiv:1210.4895*.
- Mackie, G. (2003). *Democracy defended*. Cambridge University Press.
- Mallows, C. L. (1957). Non-null ranking models. i. *Biometrika*, 114–130.
- Mattei, N., & Walsh, T. (2013). Preflib: A library for preferences <http://www.preflib.org>. In *Algorithmic decision theory* (pp. 259–270). Springer.
- May, K. O. (1952). A set of independent necessary and sufficient conditions for simple majority decision. *Econometrica: Journal of the Econometric Society*, 680–684.
- Narodytska, N., & Walsh, T. (2014). The computational impact of partial votes on strategic voting. *arXiv preprint arXiv:1405.7714*.
- Satterthwaite, M. A. (1975). Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of economic theory*, 10(2), 187–217.
- Walsh, T. (2007). Uncertainty in preference elicitation and aggregation. In *Proceedings of the national conference on artificial intelligence* (Vol. 22, p. 3).
- Walsh, T. (2008). Complexity issues in preference elicitation and manipulation. In *Aamas*.
- Walsh, T. (2015). Generating single peaked votes. *CoRR*, *abs/1503.02766*. Retrieved from <http://arxiv.org/abs/1503.02766>