Theoretical and Practical Advances in Computational Social Choice and Crowdsourcing

by

Yeawon Yoo

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved May 2021 by the
Graduate Supervisory Committee:

Adolfo R. Escobedo, Chair
Pitu B. Mirchandani
Theodore (Ted) P. Pavlic
Erin K. Chiou

ARIZONA STATE UNIVERSITY

August 2021

ABSTRACT

Computational social choice theory is an emerging research area that studies the computational aspects of decision-making. It continues to be relevant in modern society because many people often work as a group and make decisions in a group setting. Among multiple research topics, rank aggregation is a central problem in computational social choice theory. Oftentimes, rankings may involve a large number of alternatives, contain ties, and/or be incomplete, all of which complicate the use of robust aggregation methods.

To address these challenges, firstly, this work introduces a correlation coefficient that is designed to deal with a variety of ranking formats including those containing non-strict (i.e., with-ties) and incomplete (i.e., unknown) preferences. The new measure, which can be regarded as a generalization of the seminal Kendall tau correlation coefficient, is proven to satisfy a set of metric-like axioms and to be equivalent to a recently developed ranking distance function associated with Kemeny aggregation.

Secondly, this work derives an exact binary programming formulation for the generalized Kemeny rank aggregation problem—whose ranking inputs may be complete and incomplete, with and without ties. It leverages the equivalence of minimizing the Kemeny-Snell distance and maximizing the Kendall-tau correlation, to compare the newly introduced binary programming formulation to a modified version of an existing integer programming formulation associated with the Kendall-tau distance.

Thirdly, this work introduces a new social choice property for decomposing large-size problems into smaller subproblems, which allows solving the problem in a distributed fashion. The new property is adequate for handling complete rankings with ties. The property is leveraged to develop a structural decomposition algorithm, through which certain large instances of the NP-hard Kemeny rank aggregation problem can be solved exactly in a practical amount of time.

Lastly, this work applies these rank aggregation mechanisms to novel contexts for extracting collective wisdom in crowdsourcing tasks. Through this crowdsourcing experiment, we assess the capability of aggregation frameworks to recover underlying ground truth and the usefulness of multimodal information in overcoming anchoring effects, which shows its ability to enhance the wisdom of crowds and its practicability to the real-world problem.

To my husband, Dongjin, without whom my journey would not have been possible;

To my little one, Wonjae, whose little giggling and smile make my day;

To my sister, Ye-eun, who has never left my side;

To my parents, who have provided unconditional love;

To my grandparents, who I miss and love:

This dissertation is dedicated to you all with love.

사랑하는 남편 이동진 박사님,

사랑하는 아들 원재,

사랑하는 엄마, 아빠, 언니

보고싶은 할머니, 그리고 하늘에 계신 할아버지께

나의 박사 논문을 바칩니다.

ACKNOWLEDGMENTS

First of all, I gratefully acknowledge funding support from the National Science Foundation (Award 1850355) and the Army Research Office (Award 74113NSII).

I would like to express my deepest appreciation to my advisor, Professor Adolfo Escobedo for his guidance and advice in making this dissertation a reality. He has inspired me to become an independent researcher and helped me realize my potential and strengths throughout my Ph.D. studies. He has supported me not only academically, but also emotionally through the rough road to complete my Ph.D. degree. He is always there for me when I need him and listens to me patiently. I cannot say thank you enough for his tremendous help and support.

My dissertation committee guided me through all these years. I would like to express my sincere thanks to Professor Pitu Mirchandani, who is an exemplary researcher and educator. His active work in research and teaching always inspires me. I would like to thank Professor Ted Pavlic. I have benefited from his course which provided me with another perspective of viewing decision-making problems. I would also like to extend my gratitude to Erin Chiou for being my committee member and supporting me sincerely to get to the next stage of my career.

I am very grateful to several professors who I met during my undergraduate studies. Their encouragement motivates me to pursue my Ph.D. degree. In particular, I was fortunate to take a course from Professor Kenneth Ribet of the University of California Berkeley. His warm words during the office hours change my life. Professor Soyoung Sohn of Yonsei University and Professor Daegon Cho of the Korea Advanced Institute of Science and Technology provided me with the foundations for conducting research. Professor Heecheon You, Professor Byung-In Kim, Professor Kwangjae Kim, and Professor Euiho Suh of Pohang University of Science and Technology provided me with honest advice throughout and beyond my undergraduate and graduate studies.

I would like to acknowledge all my brilliant friends and colleagues at CIDSE and Los Alamos National Laboratory. I am honored to take classes and collaborate with you. Also, I would like to recognize all the assistance and prayers that I received from the fellows in Arizona Korean All Nations Church. To my friends scattered all around the world, thank you for your texts, video chats, well-wishes/prayers, phone calls, and being there whenever I needed someone to talk.

I would also like to say a heartfelt thank you to my mother, father, sister, grandmother, grandfather, and my aunt for always believing in me and praying for me. I can never repay all of you for everything you have done for me. I miss you all so badly. Also, I would like to thank to my mother-in-law and father-in-law for supporting us in whatever way they could.

I'm deeply indebted to my loved one, Dongjin. He has been extremely supportive of me throughout this entire process and has made countless sacrifices to help me get to this point, especially during the last one year which has been particularly challenging. His words kept me going when I wanted to give up. I cherish all the nights we studied side-by-side after laying Wonjae down to sleep—thank you and I love you. To my little treasure, Wonjae, how lucky I am to have the gift of you in my life! You are a great source of relief from scholarly endeavor. Thank you for waiting for mommy to get this done. I did this for you—I hope one day you can be proud of your mom.

Lastly, I would like to praise and thank God, the almighty, who brought me here and granted countless blessings.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Group decision-making has been studied extensively since the shaping of democratic society. To give equal rights to each individual, rather than a selected few, many people have devoted their efforts to develop fair and consistent systems that aggregate the opinions of several individuals to make egalitarian social decisions. Eliciting and/or expressing the preferences over a set of alternatives or candidates as rankings (e.g., candidate i is in first place, candidate j is in second place, etc.) is popular across many decision-making contexts due in part to the scale-free characteristics of these evaluations and their efficient encapsulation of large numbers of pairwise comparisons. Therefore, rank aggregation is a common and widely studied topic in group decision-making. A famous early result is that of Arrow *et al.* (1951) who studied the theoretical implications of the concept of a social welfare function (SWF), which maps individual rankings into a single ranking that should represent the best compromise among the given rankings. Therein, the author provided a set of fundamental conditions that a SWF should satisfy and demonstrated that they could not be satisfied simultaneously by any SWF. Despite this "impossibility" result, rank aggregation has been widely used across a number of practical group decision-making settings. For instance, Fields *et al.* (2013) consider a health care problem of improving nurse triage and patient prioritization in the emergency department of a hospital. When there are more patients waiting in the emergency room than available resources or staff, it is important to order the patients based on the severity of their condition. However, different nurses at times provide differing prioritization of patients, which can be represented by rankings, and thus it is necessary to resolve the conflicts among the multiple rankings. As additional examples, rank aggregation has been applied to evaluate research proposals (Cook *et al.*, 2007a), to judge student paper competitions (Hochbaum and Levin, 2006b; Escobedo *et al.*, 2021), and to improve the annual draft preparation decision-making process of Major League Baseball (Streib *et al.*, 2012). Besides decision-making, rank aggregation has extensive applicability in various other domains, such as information retrieval (Farah and Vanderpooten, 2007; Yilmaz *et al.*, 2008), similarity search (Fagin *et al.*, 2003; Ye *et al.*, 2016; Gao and Xu, 2019), and bioinformatics (Lin, 2010a,b; Marbach *et al.*, 2012; Mandal and Mukhopadhyay, 2017). From an operations research perspective, rank aggregation has been previously connected to the linear ordering problem (Martí and Reinelt, 2011) and the theory of order polytopes owing to its inherent combinatorial nature—i.e., a linear ordering is a permutation of the candidates (Fiorini and Fishburn, 2004; Heiser, 2004). Therefore, the rank aggregation problem is worth discussing in terms of both its potential impact on core methodological aspects and its practical benefits in a wide array of applications.

To solve the rank aggregation problem, Kemeny and Snell (1962) introduced a distance-based framework founded on a set of intuitive metric axioms; its associated SWF has been verified to possess many theoretical and practical benefits. Indeed, the consensus ranking problem based on the Kemeny-Snell distance has competitive advantages over other aggregation frameworks. Known widely as *Kemeny (rank) aggregation*, the objective of this problem is to find a consensus ranking solution, which is defined as a ranking with the minimum number of pairwise reversals to the set of input rankings. Assuming there are no cycles in the majority's pairwise preferences and the input rankings are complete, Kemeny aggregation is guaranteed to return the ranking solution that reflects the majority's pairwise preferences. On the other hand, scoring methods are not guaranteed to do so, for instance, the consensus ranking solution may not place the *Condorcet winner* (see Section 5.1) in first place. Kemeny aggregation has been repeatedly demonstrated to be less vulnerable to manipulation than scoring methods and more robust to outliers (Feld and Grofman, 1988; Favardin *et al.*, 2002; Endriss *et al.*, 2016). The satisfaction of this and other key social choice properties are fundamental reasons that Kemeny aggregation is preferred over various voting methods. The Borda count method (de Borda, 1781) serves as a notable example of the vulnerability to outliers of scoring methods. This method, which assigns a score to each candidate based on the number of opponents it beats in an evaluation and calculates a final score for each candidate by summing the scores earned over all evaluations, is widely employed even though it can yield very inconsistent outcomes (Dummett, 1998; Favardin *et al.*, 2002), especially when the rankings are incomplete (Moreno-Centeno and Escobedo, 2016). We remark that Kemeny aggregation and various extensions of the Kemeny-Snell distance have been developed and applied in the area of decision analysis—e.g., (Dwork *et al.*, 2001; Cook, 2006; Moreno-Centeno and Escobedo, 2016)—to reflect different assumptions about the judges' evaluations.

1.1 Existing Challenges in Rank Aggregation

Yet, there are remaining challenges for rank aggregation: *high dimensionality*, *incompleteness*, and *ties*. Large rank aggregation problems are prevalent in practice. For example, it is not uncommon for a federal funding agency to receive hundreds of submissions to a single program. Cushman *et al.* (2015) mentioned that the number of submitted proposals to the AAG program in the National Science Foundation Astronomical Sciences Division Astronomy was 731 in 2014. As a second example from portfolio decision analysis, Keisler (2004) and Schilling *et al.* (2007) consider problems with 500 options and 173 options, respectively; in the latter, it is mentioned that individual decision-makers had to assess their preferences for more than 50 alternatives. Furthermore, rank aggregation is applicable not only to preferential rankings obtained from human judges but also to a variety of ordinal data encountered in a wide array of non-human contexts such as bioinformatics, web-search engines, and recommendation systems. In bioinformatics, rank aggregation is used to integrate several lists of genes obtained from genomic experiments and find putative genes for specific diseases, where each list may consist of thousands of elements (Lin, 2010a,b; Wald *et al.*, 2012; Kolde *et al.*, 2012; Marbach *et al.*, 2012; Mandal and Mukhopadhyay, 2017). Rank aggregation can be also used in metasearch, where a user query is sent to multiple search engines and then the separate ranked lists returned are aggregated into a representative collective list (Dwork *et al.*, 2001; Desarkar *et al.*, 2016). These and other examples underscore the need to consider large rank aggregation problems in practical decision analysis research.

Incomplete rankings are another common occurrence across various decision-making contexts. When the overall number of alternatives to evaluate is large, it may not be feasible or prudent for any single judge to provide a complete ranking of these alternatives. Indeed, according to Miller's law (Miller, 1956), an average human can hold in short-term memory and judge properly only 7±2 alternatives. In addition to this cognitive limitation, there are various other constraints (e.g., time) that would motivate the evaluation of a smaller subset of the alternatives (i.e., an incomplete ranking). Similarly, having the flexibility to tie some of the alternatives (i.e., a non-strict ranking) may help make an evaluation task more manageable. In practice, it is common for groups of alternatives to be perceived as being indistinguishable from one another and, therefore, it may not be possible for judges to order them strictly (Kendall, 1945). Additionally, in a wide array of contexts, a set of evaluations may have very few distinct values and, hence, the corresponding rankings obtained from them may have many ties (Fagin *et al.*, 2004).

Due to their combinatorial nature, rank aggregation problems with a large number of alternatives are highly computationally demanding. It is known that finding the consensus ranking is NP-hard (Bartholdi *et al.*, 1989; Good, 1975), even when there are only four complete rankings to be aggregated. Considering incomplete-ranking inputs exacerbates these computational difficulties. When solving the problem via the standard branch and bound algorithm, incompleteness increases solution symmetry, which is defined as a permutation of the values of the variables that preserves the set of solutions (Cohen *et al.*, 2005; Liberti, 2008). This has the effect of slowing down pruning of nodes and, consequently, leads to a larger branch and bound tree (Sherali and Smith, 2001). Moreover, incomplete-ranking instances may yield a higher number of alternative optimal solutions than complete-ranking instances, which could also lead to less decisive outcomes (Yoo *et al.*, 2020).

Although a number of methodologies to solve the rank aggregation problem have been proposed, only few works have addressed the difficulties of solving this NP-hard problem exactly (Good, 1975). To date, the vast majority of works have been able to solve only small instances of non-strict complete rank aggregation problems exactly. For example, Emond and Mason (2000) developed and applied a special branch-and-bound algorithm to problems with 15 alternatives, which took an average of one hour to solve. Ali and Meilă (2012) performed an experiment on comparing several methods for Kemeny rank aggregation including both exact and approximate algorithms. They tested the algorithms on the various datasets, the largest of which considered 348 alternatives (the average length of the rankings is 314.86), which took an average of 16 minutes to solve exactly. However, ties in the dataset are broken with an arbitrary alphabetical rule, which means the rankings are transformed to be strict. Betzler *et al.* (2014) also tested real-world as well as synthetic datasets. The largest instance consists of 200 alternatives and it took 100 seconds to solve.

However, their algorithm neither handled ties nor incomplete rankings. Yoo *et al.* (2020) used a customized branch-and-bound algorithm to find all the exact optimal solutions for non-strict and incomplete ranking instances. Many instances with weak collective similarity and more than 16 alternatives could not be solved to completion because of insufficient memory and excessive computing time.

Because the computing time and the number of solutions increase drastically as the number of alternatives increases in the rank aggregation problem (Gross, 1962; Dwork *et al.*, 2001), most works have primarily focused on (meta)heuristics and approximation algorithms. For instance, Mandal and Mukhopadhyay (2017) proposed a metaheuristic rank aggregation approach, called particle swarm optimization-based rank aggregation. However, this approach is not applicable to partial ranked lists (i.e., incomplete rankings) and still returns suboptimal solutions. Davenport and Kalagnanam (2004) developed a greedy heuristic approach and applied to small and medium sized instances (up to 50 alternatives), but their method does not provide performance guarantees. Amodio *et al.* (2016) provided two heuristic algorithms, and the largest instance they considered consists of 50 alternatives. While one of the algorithms obtained three optimal solutions in 20 minutes, the other obtained only one optimal solution within 17 seconds. It is important to highlight, however, that neither algorithm is guaranteed to obtain optimal solutions. Moreno-Centeno and Escobedo (2016) introduced an axiomatic distance to solve the incomplete rank aggregation problem and developed an algorithm that was tested on instances with up to 40 alternatives by adapting the implicit hitting set approach (Moreno-Centeno and Karp, 2013). Their algorithm takes at most 160 seconds to solve these instances, however, the solution cannot contain ties. It is important to note that several approximation algorithms (i.e., heuristics with provable performance guarantees) have been proposed—e.g., (Fagin *et al.*, 2003; Kenyon-Mathieu and Schudy, 2007; Ailon *et al.*, 2008; Ailon, 2010). Most notably, Ailon *et al.* (2008) proposed a 11/7-approximation algorithm for strict complete rankings and Ailon (2010) introduced a 3/2-approximation algorithm for non-strict complete rankings. The objective of this dissertation is to delve into the three above-mentioned characteristics of rank aggregation (high dimensionality, incompleteness, and ties) and to propose new exact approaches for addressing the associated computational challenges.

1.2 Multimodal Judgments in Decision Making: Ranking and Rating

While this introduction has focused on ranking aggregation up to this point, it is important to recognize that ratings are another popular mechanism for eliciting and aggregating preference data. In fact, there is a longstanding debate as to whether rankings or ratings, otherwise known as ordinal and cardinal preferences, should be adopted for opinion elicitation (Ammar and Shah, 2011). Each format has its pros and cons. One of the advantages of using ratings is that they enable the expression of the intensity of preference, while rankings are only able to express preferences in the relative sense. A key disadvantage of ratings is that the rating scale may not be consistent from one person to another (Ammar and Shah, 2012); for example, in conference peer review, certain reviewers are lenient and tend to provide higher scores and others are more stringent and tend to provide lower scores (Wang and Shah, 2018). Contrary to cardinal inputs, ordinal inputs can avoid the issue of inconsistent subjective scales by focusing on pairwise comparisons between items, which can be condensed into a ranking vector (assuming each individual's pairwise preferences are transitive). For these and other reasons, there is no definitive conclusion whether ratings are superior to rankings for preference elicitation or vice versa.

The vast majority of existing works on preference elicitation and collective decision-making focus on only one modality of preference data (cardinal or ordinal). A few works use both types of modalities to arrive at a more comprehensive decision—e.g., (Kim *et al.*, 2015; Wang and Shah, 2018; Li *et al.*, 2018; Wang and Shah, 2020). However, these works tend to aggregate the two types of preference data separately or to convert from one type of information to the other (e.g., induce rankings from ratings). For example, Kim *et al.* (2015) integrates the scores (ratings) of individual genes to determine a prioritized ordered list of a set of genes (ranking), which is induced from the aggregate rating vector. One exception is that Li *et al.* (2018) proposes approaches to aggregating rating and ranking information jointly from a statistical perspective. This dissertation investigates not only the distinctive usefulness of different ranking and rating measures, but also the effectiveness of jointly integrating both types of data via multimodal aggregation (Section 6.5). It is worthwhile to note that concept of multimodality is broadly defined across various applications (e.g., image-text representation (Kruk *et al.*, 2019), visual-acoustical features (Sun *et al.*, 2020)). This dissertation focuses only on integrating cardinal and ordinal inputs (i.e., rankings and ratings) in a context where they are naturally related to each other.

1.3 The Wisdom of Crowds

The essence of eliciting and combining individual preferences into collective preferences can be related to the idea of crowdsourcing, which is a growing paradigm that has proven to be beneficial in a wide range of applications. Certain benefits of crowdsourcing are enabled by the principle commonly referred to as the "wisdom of crowds". The wisdom of crowds theorizes that aggregated information from large groups of people generally results in better outcomes than that from any individual, including experts (Galton, 1907). Many prominent researchers have argued and demonstrated that individual decisions are riddled with biases and/or subjectivity, (e.g., (Tversky and Kahneman, 1974; Kruger and Dunning, 1999; Budescu and Chen, 2015). For example, anchoring is a cognitive bias to be assessed in this dissertation (see Chapter 6), whereby an individual judgment heavily depends on an implicit or explicit reference point (Tversky and Kahneman, 1974). Underestimation and overestimation refer to the tendency to provide relatively small estimates (Hollingsworth *et al.*, 1991; Charras *et al.*, 2012; Au and Watanabe, 2013) and relatively large estimates (Goldstone, 1993; Gebuis and Reynvoet, 2012), respectively. There exist several other cognitive biases in quantitative estimation tasks (Helson, 1964; Kahneman *et al.*, 1982). Utilizing the collective intelligence of crowds can be recommended as a way to attenuate such biases (Jayles and Kurvers, 2020).

However, crowds are not always wise; according to Surowiecki (2005), the following four conditions are required to extract crowd wisdom: *independence, diversity, decentralization,* and *aggregation.* In greater detail, each person in the crowd should be independent, so that they pay attention mostly to their own information. The crowd needs to be diverse, so that people are bringing different pieces of information and not worrying about what everyone around them thinks. Moreover, each crowd member should work in a decentralized way, so that no one is dictating or unduly influencing the collective answer. Lastly, there needs to be a reasonable mechanism for aggregating the separate judgments into one collective verdict. When one of these conditions is violated, the crowd may fail to provide an accurate judgment. As a side note, the four aforementioned conditions help define a specific version of crowdsourcing founded on the wisdom of crowds. Although these two concepts are used interchangeably in many works, they are in fact not equivalent. Specifically, while the four conditions are necessary for achieving crowd wisdom, it is not uncommon for one or more of them to be violated in certain crowdsourcing contexts. For example, TripAdvisor and Yelp are online review platforms that use crowdsourcing but violate the independence and decentralization conditions—each person's ratings can be affected by reviews from other users and other external factors. Hence, one must be cautious when utilizing these terminologies.

Given the importance of utilizing crowd wisdom in decision-making, it is worthwhile to sample a few recent works that utilize this concept to address a variety of problems. As an example, Galton's experiment use a demonstrated that the average of 787 individual estimates of the weight of an ox provided a nearly perfect guess of the true value (Galton, 1907). Da and Huang (2020) describes how collective intelligence can accurately forecast corporate earnings in an open web-based platform and offers empirical evidence that encouraging independent voices among individuals improves the accuracy of the forecast consensus. Steyvers *et al.* (2009) demonstrates the idea of the wisdom of crowds effect in various complex settings. In particular, this work asks participants to recall the correct ordering of specific sets of items or events—e.g., ordering the U.S. states from east to west, sorting U.S. presidents based on the time they served in office, and ordering U.S. cities from largest to smallest populations. The featured activities are more complicated than estimating single numerical point estimates or answering multiple-choice questions. Yi *et al.* (2010) demonstrates that the wisdom of crowds effect can be leveraged to solve combinatorial optimization problems including instances of minimum spanning tree problem and of the traveling salesman problem. The authors explain that the aggregated solution outperforms the solution from the best individual. In contrast to guessing facts and predicting future events, Müller-Trede *et al.* (2018) investigates whether the benefits of crowd wisdom can be extended to deal with judgments where there is no formal ground truth (i.e., preferences, level of satisfaction). The work supports the notion that combining divergent perspectives can provide wise advice even in subjective decision-making contexts—e.g., predicting the enjoyment of musical pieces and short films.

As the preceding paragraphs explain, the majority of the wisdom of crowds literature deals with quantitative judgments in large part because it is more straightforward than aggregating qualitative judgments. The average and median are the most popular methodologies to aggregate quantitative judgments (Mannes, 2009; Müller-Trede *et al.*, 2018; Winkler *et al.*, 2019). For example, Galton (1907), Da and Huang (2020) and Müller-Trede *et al.* (2018) use a simple arithmetic average to compute the collective estimates. A few works implement more advanced aggregation methods. Mao *et al.* (2012, 2013) assess the accuracy of computational social choice approaches (e.g., the Borda rule) to derive the top alternative and the correct ordering of all alternatives and compare the results with traditional approaches. This dissertation seeks to compare the performances of nine different aggregation methods, including the average and the median,to obtain accurate collective estimate on a crowdsourcing activity.

Although the wisdom of crowds is generally beneficial, collecting information costs time and money, and the quality of the resulting collective judgment may have a practical ceiling. In other words, the additional cost to increase the crowd size may be outweighed by the marginal improvement in collective estimation accuracy. Finding the optimal trade-off point at which good crowd wisdom can be obtained with fewer resources is an intriguing topic. A small number of studies have investigated the power of the wisdom of crowds under limited crowd sizes (Goldstein *et al.*, 2014; Siddharthan *et al.*, 2016; Navajas *et al.*, 2018). For example, Mannes (2009) discovered that as few as five selected crowd members can outperform the best member or the average of the whole crowd, when these members are selected based on historical performance. Goldstein *et al.* (2014) showed theoretically that it is worthwhile to increase the size of the crowd when the added individuals are at least half as good as the crowd average. However, the historical performance of individual participants or workers may not be available, and it may be difficult to assess how well new participants will perform in a particular real-world context. This dissertation contributes to this research direction by finding the adequate crowd size for different aggregation methods with respect to a standard dot-counting crowdsourcing experiment which does not require any information regarding the historical performance of the participants.

1.4 Contributions and Overview of the Dissertation

This dissertation makes both theoretical and practical contributions to the fields of computational social choice and crowdsourcing. In summary, we introduce mathematical frameworks for handling heterogeneous ranking data with three practical characteristics that have been historically overlooked: high dimensionality, incompleteness, and ties. More specifically, we introduce a new correlation coefficient measure for handling non-strict and incomplete rankings and prove that this measure satisfies a set of metric-like axioms. Moreover, we develop a generalized binary programming formulation for high-dimensional non-strict incomplete rank aggregation. To handle even larger ranking instances, we also develop a social choice property to solve large-scale rank aggregation problems that allows for the decomposition of certain large instances into smaller subproblems. Moreover, we adapt these mathematical frameworks in a crowdsourcing application that seeks to reconcile multiple modalities of information to derive better collective judgments. The following paragraphs contain the overview of the dissertation and descriptions of the contents of each chapter.

Chapter 2 begins with notation and preliminary conventions used throughout the dissertation. It provides a literature review on the aggregation frameworks, including axiomatic distances, correlation coefficients, and voting rules. This discussion includes an overview of the underlying axioms of some of these measures along with the respective optimization models.

Chapter 3 introduces a correlation coefficient that is designed to deal with a variety of ranking formats including those containing non-strict (i.e., with-ties) and incomplete (i.e., unknown) preferences. The new measure, which can be regarded as a generalization of the seminal Kendall-τ correlation coefficient, is proven to satisfy a set of metric-like axioms and to be equivalent to a recently developed ranking distance function associated with Kemeny aggregation. In an effort to further unify and enhance both robust ranking methodologies, this chapter proves the equivalence of an additional distance and correlation-coefficient pairing in the space of non-strict incomplete rankings. The bridging of these complementary theories reinforces the singular suitability of the featured correlation coefficient to solve the general consensus ranking problem.[1]

Chapter 4 introduces an exact binary programming formulation for the generalized Kemeny rank aggregation problem. The formulation can deal with complete and incomplete rankings with and without ties, and it has a special connection with the weak-order polytope. This formulation provides an exact optimal solution of Kemeny rank aggregation problems with up to 210 alternatives within 10 minutes. As such, it differentiates itself from the vast majority of existing rank aggregation approaches, which focus on approximation algorithms and heuristics. To assess the practical implications of the binary programming formulation, we conduct a set of computational experiments on benchmark datasets as well as on instances drawn from probabilistic distributions.

Chapter 5 derives a new social choice property for expediting the solution process to Kemeny aggregation with ties, which we refer to as NXCC (Non-strict Extended Condorcet Criterion). This property is leveraged to develop a structural decomposition algorithm that decomposes large-size problems into smaller subproblems, while guaranteeing that the optimal solutions to the subproblems can be joined to provide

[1] An expanded version of Chapter 3 has been published in Yoo *et al.* (2020).

the overall optimal solution. To test the effectiveness of NXCC, we compare the computing time of solving non-decomposed (i.e., full) instances and decomposed instances using the binary programming formulation introduced in Chapter 4.[2]

Chapter 6 applies the featured binary programming formulation and other computational social choice mechanisms in a novel context, namely a cognitive crowsourcing task. In more detail, we develop a human subject study and implement it in a popular online crowdsourcing platform. In the experiment, we assess the capability of various aggregation frameworks to recover an underlying ground truth, highlighting the ability of the proposed methodologies to enhance the wisdom of crowds. We also investigate whether multimodal input elicitation can cause an anchoring and other cognitive biases, and whether these biases negatively affect the collective estimation quality. Expressly, we discover that eliciting multimodal inputs interdependently (i.e., asking cardinal estimates based on the ordinal estimates) can create anchoring, which negatively affects the collective estimation accuracy. This experiment also provides insights that the multimodal aggregation models provide a better collective estimate than traditional computational social choice mechanisms (e.g., median, mean, Borda rule). Lastly, to improve from the equal-weighted multimodal aggregation approach, we investigate the effect on the collective estimates of assigning different priority weights to the cardinal and ordinal input modalities.[3]

[2] A modified version of Chapter 4 and 5 has been published in Yoo and Escobedo (2021).

[3] A shorter preliminary version of Chapter 6 has been published in Kemmer *et al.* (2020).

Chapter 2

OVERVIEW OF AGGREGATION FRAMEWORKS

This chapter aims to provide the notation conventions to be used throughout the proposal and an overview of aggregation frameworks.

2.1 Notation and Preliminary Conventions

V	A set of alternatives (i.e., $V = \{v_1, v_2, v_3, ..., v_n\}$), where v_i denotes an alternative i and n is the number of alternatives
V^k	The k-th subset of alternatives, where $V^k \subseteq V$, $k \geq 1$
$\mathcal{P}(V)$	A family of all possible partitions of V (e.g., if $\{V^1, V^2\} \in \mathcal{P}(V)$, then $V^1 \cup V^2 = V$ and $V^1 \cap V^2 = \emptyset$)
L	A set of judges
A	A set of input rankings (ordinal-valued evaluations)
$\boldsymbol{a}^\ell$	The ranking from judge ℓ ($\ell = 1, 2, 3, .., \lvert L \rvert$), where $\boldsymbol{a}^\ell \in A$
a_i^ℓ	Rank position of v_i in the evaluation from judge ℓ, where $\ell = 1, 2, 3, .., \lvert L \rvert$
$v_i \succ v_j$	v_i is preferred over v_j (i.e., $a_i < a_j$ for some ranking $\boldsymbol{a}$)
$v_i \approx v_j$	v_i is tied with v_j (i.e., $a_i = a_j$ for some ranking $\boldsymbol{a}$)
$v_i \succeq v_j$	v_i is preferred over or tied with v_j (i.e., $a_i \leq a_j$ for some ranking $\boldsymbol{a}$)
p_{ij}	The number of judges who prefer v_i over v_j (i.e., $\lvert \left\{ \boldsymbol{a}^\ell \in A : a_i^\ell < a_j^\ell \right\} \rvert$)
t_{ij}	The number of judges who tie v_i and v_j (i.e., $\lvert \left\{ \boldsymbol{a}^\ell \in A : a_i^\ell = a_j^\ell \right\} \rvert$)
$v_i \overset{m}{\succ} v_j$	A majority of judges prefers, rather than disprefers, v_i over v_j (i.e., $p_{ij} > p_{ji}$)
$v_i \overset{m}{\approx} v_j$	No majority of judges prefers or disprefers v_i over v_j (i.e., $p_{ij} = p_{ji}$)
$v_i \overset{M}{\succ} v_j$	A decisive majority of judges prefers v_i over v_j (i.e., $p_{ij} > p_{ji} + t_{ij}$)
$v_i \overset{M}{\approx} v_j$	No decisive majority of judges prefers v_i over v_j, or vice versa (i.e., $t_{ij} \geq \lvert p_{ij} - p_{ji} \rvert$)

Table 2.1: Symbols and Notations

Denoting $V = \{v_1, \ldots, v_n\}$ as a set of alternatives, a judge's *ranking* or ordinal evaluation of V is characterized by a vector $\boldsymbol{a}$ of dimension of n, whose i-th element denotes the ordinal position assigned to alternative v_i. If $a_i < a_j$, $\boldsymbol{a}$ is said to *prefer* v_i to v_j (or to *disprefer* v_j to v_i), and when $a_i = a_j$, $\boldsymbol{a}$ is said to *tie* v_i and v_j, where $1 \leq i, j \leq n$ and $i \neq j$. Additionally, when a_i is assigned the null value "$\bullet$", v_i is said to be unranked within $\boldsymbol{a}$; the alternatives explicitly ranked in $\boldsymbol{a}$ are denoted by the subset $V_{\boldsymbol{a}} \subseteq V$ (i.e., $a_i \neq \bullet$ for $v_i \in V_{\boldsymbol{a}}$). For example, in the 5-alternative ranking $\boldsymbol{a} = (1, 2, 2, \bullet, 4), v_1$ is preferred over v_2, v_3, and v_5; v_2 and v_3 are tied for the second position but both are preferred over v_5; v_4 is left unranked; and $V_{\boldsymbol{a}} = V \backslash \{v_4\}$. The following definitions highlight the primary *ranking spaces* by which they can be categorized.

Definition 1. *Let $\Omega = \{\bullet, 1, \ldots, n\}^n$ denote the broadest ranking space consisting of all (i) strict, (ii) non-strict, (iii) complete, and (iv) incomplete rankings—corresponding to rankings (i) without ties, (ii) with and without ties, (iii) full, and (iv) partial and full, respectively. Since non-strict and incomplete rankings also encompass strict and complete rankings, respectively, Ω is denoted henceforth as the space of non-strict incomplete rankings.*

Definition 2. *Let $\Omega_C = \{1, \ldots, n\}^n$ denote the space of complete rankings over n alternatives, which consists of all non-strict (and strict) rankings where every alternative is explicitly ranked (i.e., partial evaluations are disallowed).*

Definition 3. *Let $\Omega_S = \{\bullet, 1, \ldots, n\}^n$ denote the space of strict rankings over n alternatives, which consists of all incomplete (and complete) rankings where no alternatives are tied.*

From the above definitions, it is evident that $\Omega_C \subset \Omega$, $\Omega_S \subset \Omega$, and Ω_C and Ω_S are incomparable.

The principal focus of this work is on deterministic metric-based methods for comparing and aggregating rankings, which are regarded as the most robust methodologies within Operation Research and Social Choice (Brandt *et al.*, 2016). Among them, distance-based and coefficient-based frameworks are the methodologies mainly used in robust rank aggregation due to their mathematically rigorous (i.e., axiomatic) foundations. The distance-based framework seeks a solution that minimizes the *cumulative disagreement* with the input rankings, while the coefficient-based framework seeks a solution that maximizes the *cumulative agreement* with the input rankings. Accordingly, these methods are often referred to as *consensus* ranking aggregation methods. Using these two frameworks, we can describe the rank aggregation problem. Letting $d(\cdot) : \Omega^2 \rightarrow \mathbb{R}^1_{+\cup\{0\}}$ denote an arbitrary ranking distance function, the distance-based rank aggregation problem is stated formally as:

$$\arg\min_{\boldsymbol{r}\in\Omega_C} \sum_{\ell=1}^{|L|} d(\boldsymbol{r}, \boldsymbol{a}^\ell), \tag{2.1}$$

where $\boldsymbol{a}^\ell \in \Omega$ for $\ell = 1, \ldots, |L|$.

Alternatively, let $\tau(\cdot) : \Omega^2 \rightarrow [-1, 1]^1$ denote an arbitrary ranking correlation function. The correlation-based rank aggregation problem is stated formally as:

$$\arg\max_{\boldsymbol{r}\in\Omega_C} \sum_{\ell=1}^{|L|} \tau(\boldsymbol{r}, \boldsymbol{a}^\ell), \tag{2.2}$$

where $\boldsymbol{a}^\ell \in \Omega$ for $\ell = 1, \ldots, |L|$.

Expression (2.1) can be intuitively interpreted as the problem of finding a ranking $\boldsymbol{r}$ that minimizes disagreement—quantified according to d—collectively with non-strict incomplete rankings; Expression (2.2) can be intuitively interpreted as the problem of finding a ranking $\boldsymbol{r}$ that maximizes agreement—quantified according to τ—collectively with the same inputs. For certain distance and correlation-coefficient pairings (see Equation (2.9), (3.9), (3.13)), the two respective optimization problems are equivalent. It is imperative to point out that, although the input rankings are allowed to be incomplete to allow flexibility of preference expression, the consensus ranking is required to lie in the space of complete rankings—that is, $\boldsymbol{r} \in \Omega_C$ is a constraint of both problems.

2.2 Distance-based Models

A distance function is typically advocated as the most suitable for aggregating inputs through a set of mathematical axioms it uniquely satisfies. Kemeny and Snell (1962) introduced a first axiomatic distance for ranking in Ω_C, which measures the disagreement among people. The consensus ranking framework (i.e., finding the solution that has minimum disagreement with the group) based on the Kemeny-Snell distance has competitive advantages over other aggregation frameworks. Known widely as Kemeny aggregation, this framework is less vulnerable to manipulation than scoring methods (Feld and Grofman, 1988; Favardin *et al.*, 2002; Endriss *et al.*, 2016). The Kemeny-Snell distance, denoted as d_{KS}, is defined as follows (note that sign(x) returns 1 if $x > 0$, 0 if $x = 0$, and -1 if $x < 0$):

$$d_{KS}(\boldsymbol{a}, \boldsymbol{b}) = \frac{1}{\gamma} \sum_{i=1}^{n} \sum_{j=1}^{n} |\text{sign}(a_i - a_j) - \text{sign}(b_i - b_j)| \tag{2.3}$$

where $\boldsymbol{a}, \boldsymbol{b} \in \Omega_C$ and γ is a constant associated with a chosen minimum positive distance unit. In Kemeny and Snell (1962), $\gamma = 2$, corresponding to a minimum distance unit of 1 (since each alternative pair is counted twice in the above expression). Put simply, $d_{KS}(\boldsymbol{a}, \boldsymbol{b})$ measures the number of pairwise rank reversals required to turn $\boldsymbol{a}$ into $\boldsymbol{b}$. They also argued that the distance should follow a set of intuitive metric-based axioms.

KS-Axiom 1 (Nonnegativity)**.** *$d(\boldsymbol{a}, \boldsymbol{b}) \geq 0$; and $d(\boldsymbol{a}, \boldsymbol{b}) = 0$ if and only if $\boldsymbol{a}$ and $\boldsymbol{b}$ are the same ranking.*

KS-Axiom 2 (Commutativity)**.** *$d(\boldsymbol{a}, \boldsymbol{b}) = d(\boldsymbol{b}, \boldsymbol{a})$.*

KS-Axiom 3 (Triangular Inequality)**.** *$d(\boldsymbol{a}, \boldsymbol{b}) + d(\boldsymbol{b}, \boldsymbol{c}) \geq d(\boldsymbol{a}, \boldsymbol{c})$, and the equality holds if and only if $\boldsymbol{b}$ is between $\boldsymbol{a}$ and $\boldsymbol{c}$. Note that $\boldsymbol{b}$ is said to be between $\boldsymbol{a}$ and $\boldsymbol{c}$ if, for each (v_i, v_j), the preference judgment of $\boldsymbol{b}$ either (i) agrees with $\boldsymbol{a}$ or (ii) agrees with $\boldsymbol{c}$ or (iii) $\boldsymbol{a}$ prefers v_i, $\boldsymbol{c}$ prefers v_j, and $\boldsymbol{b}$ ties them.*

KS-Axiom 4 (Anonymity)**.** *If $\boldsymbol{a}'$ results from $\boldsymbol{a}$ by a permutation of the alternatives in V, and $\boldsymbol{b}'$ results from $\boldsymbol{b}$ by the same permutation, then $d(\boldsymbol{a}, \boldsymbol{b}) = d(\boldsymbol{a}', \boldsymbol{b}')$.*

KS-Axiom 5 (Extension)**.** *If two rankings $\boldsymbol{a}$ and $\boldsymbol{b}$ agree except for a set V' of k elements, which is a segment of both, then $d(\boldsymbol{a}, \boldsymbol{b})$ may be computed as if these k alternatives were the only alternatives being ranked.*

KS-Axiom 6 (Scaling)**.** *The minimum positive unit is 1.*

Distance d_{KS} was extended to handle incomplete rankings in (Cook *et al.*, 2007b; Dwork *et al.*, 2001). The corresponding distance function between $\boldsymbol{a}, \boldsymbol{b} \in \Omega$ is defined as:

$$d_{PKS}(\boldsymbol{a}, \boldsymbol{b}) = d_{KS}(\boldsymbol{a}|_{(V_a \bigcap V_b)}, \boldsymbol{b}|_{(V_a \bigcap V_b)}), \tag{2.4}$$

where $\boldsymbol{a}|_{(V_a \bigcap V_b)}$, $\boldsymbol{b}|_{(V_a \bigcap V_b)}$ denote the projections of each ranking onto the subset of alternatives *evaluated in both rankings*. In other words, d_{PKS} enforces the intuitive interpretation that ranking disagreements should be based only on the alternatives ranked in common by $\boldsymbol{a}$ and $\boldsymbol{b}$. However, Moreno-Centeno and Escobedo (2016) suggests that utilizing d_{PKS} may be undesirable for the group decision-making context due to an associated systematic bias. They show that despite the aligned preferences of a large majority, a few judges with opposing preferences can dominate the resulting consensus ranking by simply evaluating more alternatives.

The normalized projected Kemeny Snell distance, written here succinctly as d_{NPKS}, was developed in Moreno-Centeno and Escobedo (2016) to overcome the aforementioned drawback of d_{PKS}. The d_{NPKS} distance is equivalent to d_{KS} when the inputs are restricted to space Ω_C, but it uniquely satisfies an intuitive set of axioms desired of any distance defined in space Ω. The corresponding distance function between $\boldsymbol{a}, \boldsymbol{b} \in \Omega$ is defined as:

$$d_{NPKS}(\boldsymbol{a},\boldsymbol{b}) = \begin{cases} \frac{d_{KS}(\boldsymbol{a}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})}, \boldsymbol{b}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})})}{\bar{n}(\bar{n}-1)/2} & \text{if } \bar{n} \geq 2, \\ 0 & \text{otherwise,} \end{cases} \tag{2.5}$$

where $\bar{n} := |V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}}|$. Moreno-Centeno and Escobedo (2016) modified Kemeny and Snell's axioms to obtain a set of axioms appropriate for a distance between incomplete rankings, which is given as follows:

KS'-Axiom 1 (Relevance)**.** $d(\boldsymbol{a},\boldsymbol{b}) = d(\boldsymbol{a}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})}, \boldsymbol{b}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})})$.

KS'-Axiom 2 (Nonnegativity)**.** $d(\boldsymbol{a},\boldsymbol{b}) \geq 0$; *and* $d(\boldsymbol{a},\boldsymbol{b}) = 0$ *if and only if* $\boldsymbol{a}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})}$ *and* $\boldsymbol{b}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})}$ *are the same ranking.*

KS'-Axiom 3 (Commutativity)**.** $d(\boldsymbol{a},\boldsymbol{b}) = d(\boldsymbol{b},\boldsymbol{a})$.

KS'-Axiom 4 (Relaxed Triangular Inequality)**.** $d(\boldsymbol{a}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}} \bigcap V_{\boldsymbol{c}})}, \boldsymbol{b}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}} \bigcap V_{\boldsymbol{c}})}) + d(\boldsymbol{b}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}} \bigcap V_{\boldsymbol{c}})}, \boldsymbol{c}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}} \bigcap V_{\boldsymbol{c}})}) \geq d(\boldsymbol{a}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}} \bigcap V_{\boldsymbol{c}})}, \boldsymbol{c}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}} \bigcap V_{\boldsymbol{c}})})$ *and the equality holds if and only if* $\boldsymbol{b}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}} \bigcap V_{\boldsymbol{c}})}$ *is between* $\boldsymbol{a}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}} \bigcap V_{\boldsymbol{c}})}$ *and* $\boldsymbol{c}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}} \bigcap V_{\boldsymbol{c}})}$.

KS'-Axiom 5 (Anonymity)**.** *If* $\boldsymbol{a}'$ *results from* $\boldsymbol{a}$ *by a permutation of the alternatives in* V*, and* $\boldsymbol{b}'$ *results from* $\boldsymbol{b}$ *by the same permutation, then* $d(\boldsymbol{a},\boldsymbol{b}) = d(\boldsymbol{a}',\boldsymbol{b}')$.

KS'-Axiom 6 (Extension)**.** *If two rankings* $\boldsymbol{a}$ *and* $\boldsymbol{b}$ *agree except for a set* V' *of* k *elements, which is a segment of both, then* $d(\boldsymbol{a},\boldsymbol{b})$ *may be computed as if these* k *alternatives were the only alternatives being ranked.*

KS'-Axiom 7 (Normalization)**.** *$d(\boldsymbol{a},\boldsymbol{b}) \leq 1$; and $d(\boldsymbol{a},\boldsymbol{b}) = 1$ if and only if $\boldsymbol{b}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})}$ is the reverse ranking of $\boldsymbol{a}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})}$ (the latter must be a linear ordering).*

These axioms are uniquely satisfied by d_{NPKS}. Note that d_{PKS} satisfies all the axioms that d_{NPKS} satisfies except the normalization axiom (KS'-Axiom 6); specifically, d_{PKS} satisfies KS-Axiom 6, where as d_{NPKS} satisfies KS'-Axiom 6 (Moreno-Centeno and Escobedo, 2016).

Another distance metric referenced in this work is the Kendall distance, which is adapted from the Kendall-τ correlation coefficient (Kendall, 1938), defined as:

$$d_{\tau}(\boldsymbol{a},\boldsymbol{b}) = \sum_{1 \leq i < j \leq n} \mathbb{1}_{[(a_i - a_j)(b_i - b_j) < 0]}.$$

This distance counts the number of the pairwise inversions between $\boldsymbol{a}$ and $\boldsymbol{b}$ and is equivalent to d_{KS} when the rankings are strict, although the distances are scaled differently. Specifically, when one pair of items has (strict) opposing preferences, d_{KS} accrues a value of 2 (based on one of the Kemeny-Snell axioms), while d_{τ} accrues a distance a value of 1. Hence, the distances are related by the equation $d_{KS}(\boldsymbol{a},\boldsymbol{b}) = 2d_{\tau}(\boldsymbol{a},\boldsymbol{b})$. Since the original Kendall-τ distance is defined only for strict rankings, Brancotte *et al.* (2015) redesigned the Kendall-τ distance for non-strict rankings, which is defined as follows:

$$d_{\tau'}(\boldsymbol{a},\boldsymbol{b}) = \sum_{1 \leq i < j \leq n} \mathbb{1}_{((a_i < a_j) \cap (b_i > b_j)) \cup ((a_i > a_j) \cap (b_i < b_j)) \cup ((a_i = a_j) \cap (b_i \neq b_j)) \cup ((a_i \neq a_j) \cap (b_i = b_j))}.$$

The main difference between d_{KS} and the Kendall-τ distance for non-strict rankings is that when one ranking ties two specific alternatives and the other ranking does not, the Kendall-τ distance for non-strict rankings returns the same distance as when the two rankings have opposite strict preferences. Conversely, the d_{KS} distance returns half of the distance value in the former case relative to the latter case.

Moreover, there exists an aggregation framework for ratings (or intensity rankings). Cook and Kress (1985) quantifies the distance between complete ratings by considering the intensity of the preferences. Given complete ratings $\boldsymbol{a}$ and $\boldsymbol{b}$:

$$d_{CK}(\boldsymbol{a}, \boldsymbol{b}) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} |(a_i - a_j) - (b_i - b_j)|$$

where a_i and b_i are the rating scores of alternatives v_i and v_j in ratings $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively. Here are the set of axioms that d_{CK} satisfies:

CK-Axiom 1 (Nonnegativity)**.** $d(\boldsymbol{a}, \boldsymbol{b}) \geq 0$.

CK-Axiom 2 (Commutativity)**.** $d(\boldsymbol{a}, \boldsymbol{b}) = d(\boldsymbol{b}, \boldsymbol{a})$.

CK-Axiom 3 (Triangular Inequality)**.** $d(\boldsymbol{a}, \boldsymbol{b}) + d(\boldsymbol{b}, \boldsymbol{c}) \geq d(\boldsymbol{a}, \boldsymbol{c})$, *and the equality holds if and only if* $\boldsymbol{b}$ *is between* $\boldsymbol{a}$ *and* $\boldsymbol{c}$.

CK-Axiom 4 (Proportionality)**.** *The distance between any two adjacent ratings is proportional to the degree of adjacency.*

CK-Axiom 5 (Scaling)**.** *The minimum positive unit is 1.*

Extended from d_{CK}, the normalized projected Cook-Kress distance, denoted as d_{NPCK} is developed for incomplete rating aggregation (Fishbain and Moreno-Centeno, 2016). Given incomplete ratings $\boldsymbol{a}$ and $\boldsymbol{b}$, d_{NPCK} is defined as follows:

$$d_{NPCK}(\boldsymbol{a}, \boldsymbol{b}) = \begin{cases} \frac{d_{CK}(\boldsymbol{a}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})}, \boldsymbol{b}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})})}{4R \cdot \left\lceil \frac{|V_{\boldsymbol{a}} \cap V_{\boldsymbol{b}}|}{2} \right\rceil \cdot \left\lfloor \frac{|V_{\boldsymbol{a}} \cap V_{\boldsymbol{b}}|}{2} \right\rfloor} & \text{if } \bar{n} \geq 2, \\ 0 & \text{otherwise,} \end{cases} \tag{2.6}$$

where $\bar{n} := |V_{\boldsymbol{c}} \bigcap V_{\boldsymbol{d}}|$ and where $R := U - L$ is the range of the ratings. The following axioms are obtained by slightly modifying the axioms for a distance between complete ratings.

CK'-Axiom 1 (Relevance)**.** $d(\boldsymbol{a}, \boldsymbol{b}) = d(\boldsymbol{a}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})}, \boldsymbol{b}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})})$.

CK'-Axiom 2 (Nonnegativity)**.** $d(\boldsymbol{a}, \boldsymbol{b}) \geq 0$.

CK'-Axiom 3 (Commutativity)**.** *$d(\boldsymbol{a}, \boldsymbol{b}) = d(\boldsymbol{b}, \boldsymbol{a})$.*

CK'-Axiom 4 (Relaxed Triangular Inequality)**.** *$d(\boldsymbol{a}|_{(V_a \bigcap V_b \bigcap V_c)}, \boldsymbol{b}|_{(V_a \bigcap V_b \bigcap V_c)}) + d(\boldsymbol{b}|_{(V_a \bigcap V_b \bigcap V_c)}, \boldsymbol{c}|_{(V_a \bigcap V_b \bigcap V_c)}) \geq d(\boldsymbol{a}|_{(V_a \bigcap V_b \bigcap V_c)}, \boldsymbol{c}|_{(V_a \bigcap V_b \bigcap V_c)})$ and the equality holds if and only if $\boldsymbol{b}|_{(V_a \bigcap V_b \bigcap V_c)}$ is between for incomplete ratings $\boldsymbol{a}|_{(V_a \bigcap V_b \bigcap V_c)}$ and $\boldsymbol{c}|_{(V_a \bigcap V_b \bigcap V_c)}$.*

CK'-Axiom 5 (Proportionality)**.** *The distance between any two adjacent ratings is proportional to the degree of adjacency.*

CK'-Axiom 6 (Normalization)**.** *$d(\boldsymbol{a}, \boldsymbol{b}) \leq 1$; and $d(\boldsymbol{a}, \boldsymbol{b}) = 1$ if and only if $\boldsymbol{b}|_{(V_a \bigcap V_b)}$ and $\boldsymbol{a}|_{(V_a \bigcap V_b)}$ are opposite ratings.*

Similar to d_{CK} and d_{NPCK}, the separation-deviation model (SD) can be used where the input is given as pairwise comparison preferences of alternative and point-wise score evaluation (Hochbaum, 2010). The two major components of this model are: *separation* and *deviation*. The separation term takes into account both the difference between the pairwise comparison of two alternatives i and j in the aggregated outcome and each participant's evaluations (separation), which is equivalent to the difference of intensities in ratings as in d_{CK}, and the difference between the value of alternative i in the aggregated outcome and in each participant's evaluation (deviation) (Hochbaum, 2010). Note that r_i represents the rating value of alternative v_i in the aggregated outcome and a_i^ℓ represents the rating value of alternative i in the ℓ-th judge's evaluation.

$$\begin{aligned}
&\underset{\boldsymbol{r}}{\text{minimize}} && \sum_{\ell=1}^{|L|}\Big(\sum_{i,j=1}^{n} s_{ij}^{\ell}((r_i - r_j) - (b_i^\ell - b_j^\ell)) + \sum_{i=1}^{n} d_i^\ell(r_i - b_i^\ell)\Big) && \text{(2.7a)}\\
&\text{subject to} && L \leq r_i \leq U \quad i = 1, ..., n && \text{(2.7b)}\\
& && r_i \in \mathbb{Z}^+_{\cup\{0\}} \qquad i = 1, ..., n. && \text{(2.7c)}
\end{aligned}$$

The function s_{ij}^ℓ penalizes the difference between the separation gap of alternatives i and j in the aggregated outcome and in ℓ-th participant and the function d_i^ℓ penalizes the difference between the deviation gap in the aggregated outcome and in ℓ-th participant with respect to alternative i. Here, r_i is constrained to be integer and the upper and lower bounds of r_i are $\max(a_i^\ell)$ and $\min(a_i^\ell)$, respectively. In order to ensure that the model is solvable in polynomial time, the penalty functions s_{ij}^ℓ and d_i^ℓ must be convex. For linear s_{ij}^ℓ and d_i^ℓ, because the constraint coefficient matrix is totally unimodular, the resulting problem can be solved as a linear program and $\boldsymbol{r}$ is guaranteed to be integral (Hochbaum, 2010; Escobedo *et al.*, 2021).

2.3 Coefficient-based Models

Coefficient-based frameworks are another popular methodology in rank aggregation; while ranking distances measure disagreement, ranking correlation coefficients measure similarity (i.e., agreement). They have been investigated primarily in statistics literature—e.g., (Kendall, 1938; Ahlgren *et al.*, 2003; Yilmaz *et al.*, 2008). Kendall (1938) developed a coefficient-based framework, which is closely linked to the Kemeny-Snell distance. The original methodology called Kendall-τ is a non-parametric correlation coefficient that measures the agreement among *strict rankings*, i.e., rankings that do not allow ties; it was extended to handle *non-strict rankings*, i.e., rankings that allow ties in Kendall (1948). Emond and Mason (2002) provided another version of Kendall-τ correlation coefficient for non-strict rankings and demonstrated that the Kendall-τ extended correlation coefficient, τ_x, returns the same optimal solutions as the Kemeny aggregation framework, when the inputs are also complete. It is defined as follows:

$$\tau_x(\boldsymbol{a}, \boldsymbol{b}) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{ij}}{n(n-1)},$$

where a_{ij} (and b_{ij}) is the $(i, j)-$element of the *ranking-matrix* of complete rankings $\boldsymbol{a}$ (and $\boldsymbol{b}$), $[a_{ij}]$, given by:

$$a_{ij} = \begin{cases} 1 & \text{if } \ a_i \leq a_j, \\ -1 & \text{if } \ a_i > a_j, \\ 0 & \text{if } i = j. \end{cases} \tag{2.8}$$

While the original τ coefficient (and the corresponding representation for d_{KS}) treats a tie as an expression of indifference by assigning it a value of 0 (see Emond and Mason (2002)), τ_x treats a tie as an expression of positive agreement by assigning it a value of 1 in the ranking-matrix. There exists a formal connection between key axiomatic-distance and correlation-coefficient. Emond and Mason (2002) prove that τ_x is connected to d_{KS} via the following equation:

$$\tau_x(\boldsymbol{a}, \boldsymbol{b}) = 1 - \frac{\gamma \, d_{KS}(\boldsymbol{a}, \boldsymbol{b})}{n(n-1)}, \tag{2.9}$$

where $\gamma > 0$ is the minimum d_{KS} distance unit (see Equation (2.3)). The above equation illustrates the connection between the measure of agreement (the correlation coefficient) and the measure of disagreement (the distance). To recognize this, it is important to explain that τ_x achieves values of 1 and -1 when there is complete agreement and complete disagreement, respectively, between two rankings $\boldsymbol{a}$ and $\boldsymbol{b}$. Hence, it can be interpreted that the expression of $\tau_x(\boldsymbol{a}, \boldsymbol{b})$ starts from a default assumption of perfect agreement between $\boldsymbol{a}$ and $\boldsymbol{b}$ (i.e., a correlation value of 1), and then it subtracts any disagreements between $\boldsymbol{a}$ and $\boldsymbol{b}$ from this perfect agreement, as quantified by $d_{KS}(\boldsymbol{a}, \boldsymbol{b})$. In the case when i and j are tied, d_{KS} subtracts 0 (i.e., indifference), as expected; however, in doing so, τ_x keeps the default assumption of agreement between i and j.

From this connection, the respective problems give equivalent optimal solutions, that is,

$$\arg\min_{\boldsymbol{r}} \sum_{\ell=1}^{|L|} d_{KS}(\boldsymbol{a}^{\ell}, \boldsymbol{r}) = \arg\max_{\boldsymbol{r}} \sum_{\ell=1}^{|L|} \tau_x(\boldsymbol{a}^{\ell}, \boldsymbol{r}), \tag{2.10}$$

where $\boldsymbol{r}$ is a complete ranking and $\boldsymbol{a}^\ell$ is the evaluation from judge $\ell \in L$. This connection also renders τ_x with a similar axiomatic foundation as d_{KS}. At the same time, it suggests that the inadequacies of the latter to handle incomplete rankings (Moreno-Centeno and Escobedo, 2016) carry over to the former. Chapter 3 will explore this premise.

We remark that alternative correlation coefficients have been defined for the space of incomplete rankings using concepts from fuzzy set theory, which deals with the representation of incomplete or vague information. In this context, missing ranking values are expressed as an interval (Slowinski, 2012)—which serves to estimate the missing or incomparable information. This treatment is useful in various contexts and covered in various works (e.g., Grzegorzewski (2004, 2006, 2009); Grzegorzewski and Ziembinska (2011)), but it does not conform with the neutral treatment highlighted in the dissertation. Therefore, it is not considered for the remainder of the dissertation.

2.4 Other Models

This section introduces some relevant models for aggregation problem that are motivated from other popular voting mechanisms but are neither based on ranking distances nor ranking correlation coefficients.

Average

The aggregated rating $\boldsymbol{r}$ from the average method is:

$$\boldsymbol{r} = (\frac{\sum_{\ell=1}^{|L|} b_1^\ell}{|L(1)|}, \frac{\sum_{\ell=1}^{|L|} b_2^\ell}{|L(2)|}, ..., \frac{\sum_{\ell=1}^{|L|} b_n^\ell}{|L(n)|})$$

where $L(i)$ denotes the subset of participants who evaluated alternative (an image) i.

Median

The median method finds the halfway point of the cardinal estimates after arranging the estimates in order from least to greatest. Specifically, assuming $|L|$ is even, the aggregated rating $\boldsymbol{r}$ from the median method is:

$$\boldsymbol{r} = (\frac{\bar{b}_{1\frac{|L(1)|}{2}} + \bar{b}_{1\frac{|L(1)|}{2}+1}}{2}, \frac{\bar{b}_{2\frac{|L(2)|}{2}} + \bar{b}_{2\frac{|L(2)|}{2}+1}}{2}, ..., \frac{\bar{b}_{n\frac{|L(n)|}{2}} + \bar{b}_{n\frac{|L(n)|}{2}+1}}{2})$$

where $\bar{b}_{ij}$ is the jth value in the list of arranged estimates of alternative i sorted from least to greatest.

In the social choice theory, three rules mentioned below are generally called social choice functions, which map multiple preference rankings and/or ratings into a single winner from the set of candidates, because they select one single alternative rather than the ranking of the set of alternatives (note that, as the distance and correlation coefficient-based frameworks can return a full ranking of alternatives, they are classified as social welfare functions). To obtain a full ranking, each alternative is ordered according to the score given by the rules.

Plurality Rule

The plurality rule selects an alternative with the most first-place votes. The function for determining whether alternative i is in the first place in ranking evaluation from participant ℓ is given by (Brandt *et al.*, 2016):

$$f(a_i^\ell) = \begin{cases} 1 & \text{if } a_i^\ell = 1, \\ 0 & \text{else.} \end{cases}$$

The plurality rule assigns a score to each alternative, where a score of alternative i is defined as:

$$\text{plurality}(i) = \sum_{\ell=1}^{|L|} f(a_i^\ell).$$

The final ranking of the plurality rule can be obtained by ordering the alternatives in descending order based on the score.

However, the plurality rule disregards the rank positions of alternatives which are not in first place. In Table 2.2, v_1 is selected as a plurality winner because it has the most first-place votes. However, v_1 is selected as the most dispreferred alternative three times. Intuitively, v_2 should have been selected as a winner because v_2 retains the top-2 position in all the input rankings.

		Rankings						
		$\boldsymbol{a}^1$	$\boldsymbol{a}^2$	$\boldsymbol{a}^3$	$\boldsymbol{a}^4$	$\boldsymbol{a}^5$	$\boldsymbol{a}^6$	Plurality score
Alternatives	v_1	1	5	1	5	1	5	3
	v_2	2	2	1	2	1	2	2
	v_3	3	4	4	1	3	3	1
	v_4	4	3	1	2	5	4	1
	v_5	5	1	5	4	4	1	2

Table 2.2: An Example Showing the Disadvantage of the Plurality Rule

Borda Rule

The Borda rule is a well-known method that assigns a score to each alternative in a ballot according to how many alternatives it defeats and choose the alternative with the highest score as a winner (Brandt *et al.*, 2016). Mathematically, assuming that there exist n alternatives and the highest score is $n-1$ (because there can exist maximum $n-1$ alternatives which are ranked lower than the first-placed alternative), the Borda rule assigns a score to each alternative, where a score of alternative i is defined as:

$$\text{Borda}(i) = \sum_{\ell=1}^{|L|}(n - a_i^\ell).$$

To determine a final ranking of the Borda rule, the alternatives are ordered in descending order based on the score.

This method can yield inconsistent outcomes due to the vulnerability (Dummett, 1998; Favardin *et al.*, 2002), especially when the rankings are incomplete (Moreno-Centeno and Escobedo, 2016). In the example in Table 2.3, the score of 4 is assigned to the most preferred alternative and the Borda score of the least preferred alternative is assigned accordingly, based on the definition of the Borda score. This example shows that the most dispreferred alternative has the highest Borda score. Because it is the most frequently evaluated alternative, it just receives more scores than others regardless of the rank position.

		Rankings				
		$\boldsymbol{a}^1$	$\boldsymbol{a}^2$	$\boldsymbol{a}^3$	$\boldsymbol{a}^4$	Borda score
Alternatives	v_1	●	1	●	1	8
	v_2	1	2	●	●	7
	v_3	2	●	1	●	7
	v_4	3	3	2	2	10

Table 2.3: An Example Showing the Disadvantage of the Borda Rule

Copeland Rule

The Copeland rule chooses the alternative with the highest number of pairwise wins minus defeats as a winner, which is mathematically written as (Brandt *et al.*, 2016):

$$\text{Copeland}(i) = \sum_{j \in V, j \neq i} \sum_{\ell=1}^{|L|} (|\{a^\ell : a_i^\ell < a_j^\ell\}| - |\{a^\ell : a_i^\ell > a_j^\ell\}|).$$

To determine a final ranking of the Copeland rule, the alternatives are ordered in descending order based on the assigned score.

These models are more computationally efficient and frequently discussed in computational social choice research (Galton, 1907; Mannes, 2009; Brandt *et al.*, 2016; Müller-Trede *et al.*, 2018; Winkler *et al.*, 2019; Da and Huang, 2020), but cannot fulfill certain fundamental properties associated with voting fairness (e.g., the Condorcet criterion (Condorcet, 1785) and its extensions (Young and Levenglick, 1978; Young, 1988)).

Chapter 3

A NEW CORRELATION COEFFICIENT FOR COMPARING AND AGGREGATING NON-STRICT AND INCOMPLETE RANKINGS

To the best of our knowledge, there has not been a ranking correlation coefficient explicitly tailored to the space of non-strict incomplete rankings, Ω, under a neutral treatment of incompleteness. Therefore, Section 3.1 introduces the ranking correlation coefficient $\hat{\tau}_x$ for non-strict incomplete rankings. Section 3.2 provides the properties and axioms it satisfies. Finally, Section 3.3 establishes the equivalence of $\hat{\tau}_x$ with the axiomatic distance d_{NPKS} as well as the equivalence of τ_x with d_{PKS} when the input rankings lie in space Ω.

3.1 Derivation of the New Correlation Coefficient

To quantify the similarity between non-strict incomplete rankings via correlation coefficients, a fundamental requirement is that the correlation between any pair of rankings $\boldsymbol{a}, \boldsymbol{b} \in \Omega$ must lie within the interval $[-1, 1]$. The -1 and 1 values must be achieved whenever $\boldsymbol{a}$ and $\boldsymbol{b}$ completely agree and completely disagree, respectively; otherwise, a value from the interior of the interval should be commensurate with the level of similarity. As explained in Chapter 2, τ_x cannot fulfill these essential requirements. Hence, this subsection derives a new correlation coefficient that satisfies these properties as well as a set of metric-like axioms tailored to space Ω. As a first step, we define a new ranking-matrix $[a_{ij}]$ representation for $\boldsymbol{a} \in \Omega$ as:

$$a_{ij} = \begin{cases} 1 & \text{if } \ a_i \leq a_j, \\ -1 & \text{if } \ a_i > a_j, \\ 0 & \text{if } i = j, \text{or } a_i = \bullet, \text{or } a_j = \bullet \end{cases} \tag{3.1}$$

where $1 \leq i, j \leq n$. This ranking-matrix can be obtained by extending Equation (2.8) to also assign $a_{ij} = 0$ whenever alternative i or j (or both) is unranked in $\boldsymbol{a}$ and, thus, it is equivalent to the τ_x ranking-matrix when the input rankings are complete. This extension was cursorily proposed in Emond and Mason (2002), although it was neither explicitly defined nor implemented therein. It is chosen as the basis of the new correlation coefficient because its treatment of ties is equivalent to the Kemeny Snell "half-flip" metric, which assigns only half of a rank reversal between $\boldsymbol{a}$ and $\boldsymbol{b}$ whenever one ties (v_i, v_j) but the other professes a strict preference for v_i over v_j, or vice versa.

As a second step, consider ranking-matrices $[a_{ij}]$ and $[b_{ij}]$ respectively defined according to Equation (3.1) and their associated *matrix inner product*:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{ij}.$$

When $\boldsymbol{a}$ and $\boldsymbol{b}$ rank every alternative, the number of non-zeros in each ranking-matrix and the maximum matrix inner product are both equal to $n(n-1)$. The reasons are that the ranking-matrix diagonal elements are all 0 and that $a_{ij} b_{ij} = 1$ for all $i \neq j$ when $b_{ij} = a_{ij}$. It is also straightforward to discern that a minimum matrix inner product of $-n(n-1)$ can be achieved only if $\boldsymbol{a}$ does not contain ties and $b_{ij} = -a_{ij}$ for all $i \neq j$.

When $\boldsymbol{a}$ or $\boldsymbol{b}$ does not rank every alternative, for each v_i such that either $a_i = \bullet$ or $b_i = \bullet$, the ith ranking-matrix row and column are set to zero, thereby decreasing the maximum and increasing the minimum matrix inner products by $2(n-1)$. Put otherwise, such a matrix inner product may be calculated as if the ith row and column of both ranking-matrices do not exist. Hence, the maximum and minimum inner products of $[a_{ij}]$ and $[b_{ij}]$ are reduced to $\bar{n}(\bar{n}-1)$ and $-\bar{n}(\bar{n}-1)$, respectively, where $\bar{n} = |V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}}|$. Accordingly, a new correlation function can be derived to achieve the full expected correlation interval $[-1, 1]$. It is named the *scaled Kendall tau-extended correlation coefficient*, written succinctly as $\hat{\tau}_x$, and is defined as:

$$\hat{\tau}_x(\boldsymbol{a}, \boldsymbol{b}) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{ij}}{\bar{n}(\bar{n}-1)}, \tag{3.2}$$

which may be rewritten in terms of τ_x via the equation:

$$\hat{\tau}_x(\boldsymbol{a}, \boldsymbol{b}) = \frac{n(n-1)}{\bar{n}(\bar{n}-1)} \tau_x(\boldsymbol{a}, \boldsymbol{b}). \tag{3.3}$$

This alternative expression emphasizes that, by scaling $\tau_x(\boldsymbol{a}, \boldsymbol{b})$ by the factor $\frac{n(n-1)}{\bar{n}(\bar{n}-1)} \geq 1$, $\hat{\tau}_x$ removes the impact of irrelevant pairwise preference comparisons—the pairs of alternatives unranked by $\boldsymbol{a}$ or $\boldsymbol{b}$—from their correlation. As a result, the extrema correlation values -1 and 1 can be achieved when comparing two appropriate non-strict incomplete rankings. Clearly, when $V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}} = V$, scaling factor in Equation (3.3) equals 1, meaning $\hat{\tau}_x$ is equivalent to τ_x in space Ω_C.

3.2 Axiomatic Foundation of the New Correlation Coefficient

This section presents a set of intuitive metric-like axioms that $\hat{\tau}_x$ satisfies and the formal proofs.

KT-Axiom 1 (Relevance)**.** *The correlation discounts the unevaluated alternatives:*

$$\hat{\tau}_x(\boldsymbol{a}, \boldsymbol{b}) = \hat{\tau}_x(\boldsymbol{a}|_{(V_a \bigcap V_b)}, \boldsymbol{b}|_{(V_a \bigcap V_b)})$$

Proof. From the definition of $\hat{\tau}_x$, the ranking-matrix elements for unevaluated alternatives are assigned to be 0 and, thus, the corresponding numerator terms $a_{ij}b_{ij}$ become 0. Therefore, it is valid to calculate the correlation coefficient by focusing on the mutually evaluated alternatives. □

KT-Axiom 2 (Commutativity)**.** *The correlation value is independent of the order in which* $\boldsymbol{a}$ *and* $\boldsymbol{b}$ *are compared:*

$$\hat{\tau}_x(\boldsymbol{a}, \boldsymbol{b}) = \hat{\tau}_x(\boldsymbol{b}, \boldsymbol{a}).$$

Proof. From the definition of $\hat{\tau}_x$, we can write the following equations:

$$\begin{aligned}
\hat{\tau}_x(\boldsymbol{a}, \boldsymbol{b}) &= \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij}}{\bar{n}(\bar{n}-1)} \\
&= \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}a_{ij}}{\bar{n}(\bar{n}-1)} \\
&= \hat{\tau}_x(\boldsymbol{b}, \boldsymbol{a}) \qquad \square
\end{aligned}$$

KT-Axiom 3 (Neutrality)**.** *The correlation value is independent of the particular labeling of the alternatives:*

If $\boldsymbol{a}' = \pi(\boldsymbol{a})$ *and* $\boldsymbol{b}' = \pi(\boldsymbol{b})$*, then* $\hat{\tau}_x(\boldsymbol{a}, \boldsymbol{b}) = \hat{\tau}_x(\boldsymbol{a}', \boldsymbol{b}')$*, where* $\pi := \{1, 2, ..., n\} \to \{1, 2, ..., n\}$ *is a permutation function.*

Proof. Without loss of generality, assume that only two alternatives are permuted at a time, namely the k-th and l-th alternatives, with $k < l$. Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be the pre-permutation ranking-matrices corresponding to ranking $\boldsymbol{a}$ and $\boldsymbol{b}$, which are illustrated as:

$$A = \begin{pmatrix}
a_{11} & \cdots & a_{1k} & \cdots & a_{1l} & \cdots & a_{1n} \\
\vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\
a_{k1} & \cdots & a_{kk} & \cdots & a_{kl} & \cdots & a_{kn} \\
\vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\
a_{l1} & \cdots & a_{lk} & \cdots & a_{ll} & \cdots & a_{ln} \\
\vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\
a_{n1} & \cdots & a_{nk} & \cdots & a_{nl} & \cdots & a_{nn}
\end{pmatrix} \qquad B = \begin{pmatrix}
b_{11} & \cdots & b_{1k} & \cdots & b_{1l} & \cdots & b_{1n} \\
\vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\
b_{k1} & \cdots & b_{kk} & \cdots & b_{kl} & \cdots & b_{kn} \\
\vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\
b_{l1} & \cdots & b_{lk} & \cdots & b_{ll} & \cdots & b_{ln} \\
\vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\
b_{n1} & \cdots & b_{nk} & \cdots & b_{nl} & \cdots & b_{nn}
\end{pmatrix}.$$

Let $\boldsymbol{a}'$ and $\boldsymbol{b}'$ be the post-permutation rankings with corresponding ranking-matrices

$A' = [a'_{ij}]$ and $B' = [b'_{ij}]$, which are illustrated as:

$$A' = \begin{pmatrix} a_{11} & \cdots & a_{1l} & \cdots & a_{1k} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{l1} & \cdots & a_{ll} & \cdots & a_{lk} & \cdots & a_{ln} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{k1} & \cdots & a_{kl} & \cdots & a_{kk} & \cdots & a_{kn} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{n1} & \cdots & a_{nl} & \cdots & a_{nk} & \cdots & a_{nn} \end{pmatrix} \qquad B' = \begin{pmatrix} b_{11} & \cdots & b_{1l} & \cdots & b_{1k} & \cdots & b_{1n} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ b_{l1} & \cdots & b_{ll} & \cdots & b_{lk} & \cdots & b_{ln} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ b_{k1} & \cdots & b_{kl} & \cdots & b_{kk} & \cdots & b_{kn} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ b_{n1} & \cdots & b_{nl} & \cdots & b_{nk} & \cdots & b_{nn} \end{pmatrix}.$$

Note that unshaded elements in A' and B' remain the same after permutation (i.e., $a_{ij} = a'_{ij}$ for $i, j \neq k, l$). Since the k-th row (column) and l-th row (column) of A and B are exchanged, the remaining entries are given by:

$$a'_{kk} = a_{ll}, \quad a'_{ll} = a_{kk}, \quad a'_{ik} = a_{il}, \quad a'_{il} = a_{ik}, \quad a'_{ki} = a_{li}, \quad a'_{li} = a_{ki} \tag{3.4}$$

for every $i \neq k, l$ (the new entries for $\boldsymbol{b}$ are defined similarly). The permutation will affect only the numerator of $\hat{\tau}_x(\boldsymbol{a}, \boldsymbol{b})$. In particular, by using Expression (3.4), the following equations can be derived:

$$\begin{aligned} \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij} &= \sum_{i \neq k,l}^{n}\sum_{j \neq k,l}^{n} a_{ij}b_{ij} + \sum_{j \neq k,l}(a_{kj}b_{kj} + a_{lj}b_{lj}) + \sum_{i \neq k,l}(a_{ik}b_{ik} + a_{il}b_{il}) + \sum_{i=k,l}\sum_{j=k,l} a_{ij}b_{ij} \\ &= \sum_{i \neq k,l}^{n}\sum_{j \neq k,l}^{n} a'_{ij}b'_{ij} + \sum_{j \neq k,l}(a'_{lj}b'_{lj} + a'_{kj}b'_{kj}) + \sum_{i \neq k,l}(a'_{il}b'_{il} + a'_{ik}b'_{ik}) + \sum_{i=k,l}\sum_{j=k,l} a'_{ji}b'_{ji} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a'_{ij}b'_{ij}. \end{aligned}$$

Therefore, $\hat{\tau}_x(\boldsymbol{a}, \boldsymbol{b}) = \hat{\tau}_x(\boldsymbol{a}', \boldsymbol{b}')$. Since any permutation can be described as a sequence of permutations of two alternatives at a time, $\hat{\tau}_x(\boldsymbol{a}, \boldsymbol{b}) = \hat{\tau}_x(\boldsymbol{a}', \boldsymbol{b}')$ holds for any permutation π. □

KT-Axiom 4 (Reduction)**.** *If $\boldsymbol{a}$ and $\boldsymbol{b}$ agree except for a set $V' \subseteq V$, then $\hat{\tau}_x(\boldsymbol{a}, \boldsymbol{b})$ may be computed by focusing only on the alternatives in V':*

$$\hat{\tau}_x(\boldsymbol{a}, \boldsymbol{b}) = 1 + 2\hat{\tau}_x(\boldsymbol{a}|_{V'}, \boldsymbol{b}|_{V'}).$$

Proof. By definition of the ranking-matrix representation of $\hat{\tau}_x$, if rankings $\boldsymbol{a}$ and $\boldsymbol{b}$ have positive agreement for alternatives v_i, v_j (i.e., one prefers v_i over v_j and the other also prefers v_i over v_j, or both tie v_i and v_j), the corresponding numerator $a_{ij}b_{ij}$ becomes 1. Otherwise, $a_{ij}b_{ij}$ becomes -1. Let p_c be the number of pairs in concordance and p_{dc} be the number of pairs in discordance. Then, $\bar{n}(\bar{n}-1) = 2p_c+2p_{dc}$ because there are two elements of ranking-matrix for v_i and v_j, a_{ij} and a_{ji} (b_{ij} and b_{ji}, respectively). The calculation of $\hat{\tau}_x$ can be decomposed as follows:

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij}}{\bar{n}(\bar{n}-1)} = \frac{1}{\bar{n}(\bar{n}-1)} \times 2p_c + \frac{-1}{\bar{n}(\bar{n}-1)} \times 2p_{dc}.$$

By replacing $2p_c$ with $\bar{n}(\bar{n}-1) - 2p_{dc}$,

$$\begin{aligned}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij}}{\bar{n}(\bar{n}-1)} &= \frac{1}{\bar{n}(\bar{n}-1)} \times (\bar{n}(\bar{n}-1) - 2p_{dc}) + \frac{-1}{\bar{n}(\bar{n}-1)} \times 2p_{dc} \\ &= 1 - \frac{2}{\bar{n}(\bar{n}-1)} \times 2p_{dc} \\ &= 1 + \frac{2\sum_{i=1}^{n}\sum_{j=1}^{n} a'_{ij}b'_{ij}}{\bar{n}(\bar{n}-1)}\end{aligned}$$

Hence, if $\boldsymbol{a}$ and $\boldsymbol{b}$ agree except for a set $V' \subseteq V$, then $\hat{\tau}_x$ can be calculated by focusing on the alternatives where $\boldsymbol{a}$ and $\boldsymbol{b}$ disagree. That is,

$$\hat{\tau}_x(\boldsymbol{a}, \boldsymbol{b}) = 1 + 2\hat{\tau}_x(\boldsymbol{a}', \boldsymbol{b}') \qquad \square$$

KT-Axiom 5 (Relaxed Triangle Inequality)**.** *Relationship among the three possible paired comparisons from three incomplete rankings:*

$$\begin{aligned}\hat{\tau}_x(\boldsymbol{a}|_{(V_{\boldsymbol{a}}\bigcap V_{\boldsymbol{b}}\bigcap V_{\boldsymbol{c}})}, \boldsymbol{b}|_{(V_{\boldsymbol{a}}\bigcap V_{\boldsymbol{b}}\bigcap V_{\boldsymbol{c}})}) + \hat{\tau}_x(\boldsymbol{b}|_{(V_{\boldsymbol{a}}\bigcap V_{\boldsymbol{b}}\bigcap V_{\boldsymbol{c}})}, \boldsymbol{c}|_{(V_{\boldsymbol{a}}\bigcap V_{\boldsymbol{b}}\bigcap V_{\boldsymbol{c}})}) \\ \leq \hat{\tau}_x(\boldsymbol{a}|_{(V_{\boldsymbol{a}}\bigcap V_{\boldsymbol{b}}\bigcap V_{\boldsymbol{c}})}, \boldsymbol{c}|_{(V_{\boldsymbol{a}}\bigcap V_{\boldsymbol{b}}\bigcap V_{\boldsymbol{c}})}) + 1;\end{aligned}$$

and equality holds if and only if $\boldsymbol{b}|_{(V_{\boldsymbol{a}}\bigcap V_{\boldsymbol{b}}\bigcap V_{\boldsymbol{c}})}$ is between the other two projected rankings; here, $V_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}} := V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}} \bigcap V_{\boldsymbol{c}}$ for concise representation.

Proof. Let $\bar{n} = |V_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}}|$. To investigate the relationship between $\hat{\tau}_x(\boldsymbol{a},\boldsymbol{b})$, $\hat{\tau}_x(\boldsymbol{b},\boldsymbol{c})$, and $\hat{\tau}_x(\boldsymbol{a},\boldsymbol{c})$, begin by writing their corresponding definitions:

$$\hat{\tau}_x(\boldsymbol{a},\boldsymbol{b}) = \frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} a_{ij}b_{ij}}{\bar{n}(\bar{n}-1)}, \hat{\tau}_x(\boldsymbol{b},\boldsymbol{c}) = \frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} b_{ij}c_{ij}}{\bar{n}(\bar{n}-1)}, \hat{\tau}_x(\boldsymbol{a},\boldsymbol{c}) = \frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} a_{ij}c_{ij}}{\bar{n}(\bar{n}-1)}.$$

From these definitions, we can form the expression:

$$\hat{\tau}_x(\boldsymbol{a},\boldsymbol{b}) + \hat{\tau}_x(\boldsymbol{b},\boldsymbol{c}) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}(a_{ij}+c_{ij})}{\bar{n}(\bar{n}-1)}.$$

There are three possibilities for the sum and product of a_{ij} and c_{ij}:

$$\begin{array}{lll} 1)\ a_{ij} = c_{ij} = 1 & \Longrightarrow & a_{ij}+c_{ij} = 2, \quad\ \ a_{ij}c_{ij} = 1 \\ 2)\ a_{ij} = 1, c_{ij} = -1, \text{ or } a_{ij} = -1, c_{ij} = 1 & \Longrightarrow & a_{ij}+c_{ij} = 0, \quad\ \ a_{ij}c_{ij} = -1 \\ 3)\ a_{ij} = c_{ij} = -1 & \Longrightarrow & a_{ij}+c_{ij} = -2, \ \ a_{ij}c_{ij} = 1 \end{array}$$

Now, when $\boldsymbol{b}|_{(V_{\boldsymbol{a}}\bigcap V_{\boldsymbol{b}}\bigcap V_{\boldsymbol{c}})}$ is between the other two projected rankings (i.e., b_{ij} is equal to either a_{ij} or c_{ij} or both, or b_{ij} may also equal 1 when a_{ij} and b_{ij} disagree), b_{ij} can be determined from the values of a_{ij} and c_{ij} as follows:

$$\begin{array}{lll} 1)\ a_{ij} = c_{ij} = 1 & \Longrightarrow & b_{ij} = 1 \\ 2)\ a_{ij} = 1, c_{ij} = -1, \text{ or } a_{ij} = -1, c_{ij} = 1 & \Longrightarrow & b_{ij} = 1 \text{ or } -1 \\ 3)\ a_{ij} = c_{ij} = -1 & \Longrightarrow & b_{ij} = -1 \end{array}$$

Referencing the above cases, if $\boldsymbol{b}|_{(V_{\boldsymbol{a}}\bigcap V_{\boldsymbol{b}}\bigcap V_{\boldsymbol{c}})}$ is between the other two projected rankings, the following equality always holds for each i, j:

$$b_{ij}(a_{ij}+c_{ij}) = a_{ij}c_{ij} + 1. \tag{3.5}$$

Therefore, summing over all i, j yields the following inequality:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}(a_{ij}+c_{ij}) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}c_{ij} + \bar{n}(\bar{n}-1). \tag{3.6}$$

By a similar analysis, if $\boldsymbol{b}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}} \bigcap V_c)}$ is not between the other two projected rankings,

$$b_{ij}(a_{ij} + c_{ij}) < a_{ij}c_{ij} + 1, \tag{3.7}$$

and summing over all i, j yields the following inequality:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}(a_{ij} + c_{ij}) < \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}c_{ij} + \bar{n}(\bar{n} - 1). \tag{3.8}$$

Combining equations (3.6) and (3.8) yields the inequality:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}(a_{ij} + c_{ij}) \leq \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}c_{ij} + \bar{n}(\bar{n} - 1).$$

Therefore, dividing by $\bar{n}(\bar{n} - 1)$, we obtain the desired expression. □

KT-Axiom 6 (Scaling)**.** *The correlation range is between -1 and 1, inclusively:*

$$-1 \leq \hat{\tau}_x(\boldsymbol{a}, \boldsymbol{b}) \leq 1;$$

with $\hat{\tau}_x(\boldsymbol{a}, \boldsymbol{b}) = 1$ if and only if $\boldsymbol{a}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})} = \boldsymbol{b}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})}$ and $\hat{\tau}_x(\boldsymbol{a}, \boldsymbol{b}) = -1$ if and only if $\boldsymbol{b}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})}$ is the reverse ranking of $\boldsymbol{a}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})}$ (the latter must be a linear ordering).

Proof. ($\Leftarrow$) Assume $\boldsymbol{a}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})}$ and $\boldsymbol{b}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})}$ are the same ranking. Then,

$$\hat{\tau}_x(\boldsymbol{a}, \boldsymbol{b}) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij}}{\bar{n}(\bar{n} - 1)} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}a_{ij}}{\bar{n}(\bar{n} - 1)} = \frac{\bar{n}(\bar{n} - 1)}{\bar{n}(\bar{n} - 1)} = 1.$$

If $\boldsymbol{b}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})}$ is the reverse ranking of $\boldsymbol{a}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})}$ and $\boldsymbol{b}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})}$ and $\boldsymbol{a}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})}$ are linear orderings, then b_{ij} always has the opposite value of a_{ij}. That is, $b_{ij} = -a_{ij}$, which leads to the following inequality:

$$\hat{\tau}_x(\boldsymbol{a}, \boldsymbol{b}) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij}}{\bar{n}(\bar{n} - 1)} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(-a_{ij})}{\bar{n}(\bar{n} - 1)} = \frac{-\bar{n}(\bar{n} - 1)}{\bar{n}(\bar{n} - 1)} = -1.$$

($\Rightarrow$) Let $\hat{\tau}_x(\boldsymbol{a}, \boldsymbol{b}) = 1$. Then, $\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij}$ should be $\bar{n}(\bar{n}-1)$, which means that $a_{ij}b_{ij} = 1$ for every $v_i, v_j \in V_{\boldsymbol{a}} \cap V_{\boldsymbol{b}}$. That is, $\boldsymbol{a}$ and $\boldsymbol{b}$ agree on all their preferences. On the other hand, to achieve $\hat{\tau}_x(\boldsymbol{a}, \boldsymbol{b}) = -1$, $\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij}$ should equal $-\bar{n}(\bar{n} -$

1), which implies that $a_{ij}b_{ij} = -1$ for every $v_i, v_j \in V_{\boldsymbol{a}} \cap V_{\boldsymbol{b}}$ and that $\boldsymbol{a}|_{(V_a \bigcap V_b)}$ and $\boldsymbol{b}|_{(V_a \bigcap V_b)}$ are linear orderings. That is, $\boldsymbol{a}|_{(V_a \bigcap V_b)}$ and $\boldsymbol{b}|_{(V_a \bigcap V_b)}$ express opposing strict preferences over all alternative pairs. Therefore, if $\hat{\tau}_x(\boldsymbol{a}|_{(V_a \bigcap V_b)}, \boldsymbol{b}|_{(V_a \bigcap V_b)}) = 1$, $\boldsymbol{a}|_{(V_a \bigcap V_b)}$ and $\boldsymbol{b}|_{(V_a \bigcap V_b)}$ are the same ranking, and if $\hat{\tau}_x(\boldsymbol{a}|_{(V_a \bigcap V_b)}, \boldsymbol{b}|_{(V_a \bigcap V_b)}) = -1$, $\boldsymbol{b}|_{(V_a \bigcap V_b)}$ is the reverse ranking of $\boldsymbol{a}|_{(V_a \bigcap V_b)}$. □

As suggested by Axiom 1, the $\hat{\tau}_x$ similarity between two incomplete rankings can be equivalently calculated by simply dropping the alternatives unranked by either ranking (i.e., by projecting them to the subset of alternatives evaluatded by both). It brings the pragmatic benefit of eliminating the unenforceable/unrealistic requirement of having to allocate an equal number of alternatives for each judge to evaluate, which may be difficult to enforce due to differing expertise, disagreeing schedules, unplanned exemptions, etc. (Hochbaum and Levin, 2010). Indeed, it is advisable to avoid assigning fewer subjective evaluation tasks to mitigate cognitive errors (Basili and Vannucci, 2015; Saaty and Ozdemir, 2003). While this may seem to remove the incomplete data from the researcher's view when comparing two incomplete rankings, we emphasize that the consensus ranking problem (see Equation (2.1) or Equation (2.2)) involves accruing the comparisons between the candidate solution (always a complete ranking) and each input ranking (which may be incomplete or complete). In this context, Axiom 1 ensures that each input incomplete ranking influences only the consensus ranking elements corresponding to its ranked alternatives.

3.3 Key Pairings between Distances and Correlation Coefficients

This section establishes a formal connection between $\hat{\tau}_x$ and d_{NPKS} as well as between another key axiomatic-distance and correlation-coefficient pairing in space Ω. Together these results fill a significant gap in the literature because although Emond and Mason (2002) made a connection between distance and correlation-based methods for complete rankings (see Equation (2.9)), they conjectured that a parallel connection could not be established for incomplete rankings. The statements and proofs of the theorems and corollaries are in the following paragraphs.

Theorem 1 (Linear transformation between $\hat{\tau}_x$ and d_{NPKS})**.** *Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be arbitrary rankings over $n = |V|$ alternatives drawn from the space of non-strict incomplete rankings, Ω. Then, the $\hat{\tau}_x$ correlation coefficient and the d_{NPKS} distance are connected through the equation:*

$$d_{NPKS}(\boldsymbol{a}, \boldsymbol{b}) = \frac{1}{2} - \frac{1}{2}\hat{\tau}_x(\boldsymbol{a}, \boldsymbol{b}). \tag{3.9}$$

Proof. For succinctness, denote $\bar{\boldsymbol{a}} = \boldsymbol{a}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})}$ and $\bar{\boldsymbol{b}} = \boldsymbol{b}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})}$ as the rankings over $\bar{n} \leq n$ alternatives obtained by projecting $\boldsymbol{a}$ and $\boldsymbol{b}$ onto the subset of alternatives $\bar{V} = V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}}$ ranked in common. Notice that $\bar{\boldsymbol{a}}$ and $\bar{\boldsymbol{b}}$ are complete rankings over the same reduced universe of $\bar{n}$ alternatives (i.e., they lie in space Ω_C relative to $\bar{V}$). As such, using 1/2 as the minimum d_{KS} distance unit, the corresponding τ_x and d_{KS} values for $\bar{\boldsymbol{a}}$ and $\bar{\boldsymbol{b}}$ are equated as follows (Emond and Mason, 2002):

$$\tau_x(\bar{\boldsymbol{a}}, \bar{\boldsymbol{b}}) = 1 - \frac{4\, d_{KS}(\bar{\boldsymbol{a}}, \bar{\boldsymbol{b}})}{\bar{n}(\bar{n}-1)},$$

which expressed in terms of d_{KS} yields the equivalent relationship:

$$\begin{align} d_{KS}(\bar{\boldsymbol{a}}, \bar{\boldsymbol{b}}) &= \frac{\bar{n}(\bar{n}-1)}{4} - \frac{\bar{n}(\bar{n}-1)\tau_x(\bar{\boldsymbol{a}}, \bar{\boldsymbol{b}})}{4} \tag{3.10}\\ &= \frac{\bar{n}(\bar{n}-1)}{4} - \frac{\bar{n}(\bar{n}-1)\sum_{i=1}^{\bar{n}}\sum_{j=1}^{\bar{n}} \bar{a}_{ij}\bar{b}_{ij}}{4\bar{n}(\bar{n}-1)} \tag{3.11}\\ &= \frac{\bar{n}(\bar{n}-1)}{4} - \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij}}{4}, \tag{3.12} \end{align}$$

where Equation (3.12) cancels a common factor in the second term and utilizes the fact that unranked items in either ranking vector contribute nothing to the sum—that is the matrix inner products are identical in the original and projected spaces. Now, multiplying both sides of Equation (3.12) by $[\bar{n}(\bar{n}-1)/2]^{-1}$ gives:

$$\begin{aligned} \frac{d_{KS}(\bar{\boldsymbol{a}},\bar{\boldsymbol{b}})}{\bar{n}(\bar{n}-1)/2} &= \frac{1}{2} - \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij}}{2\bar{n}(\bar{n}-1)} \\ \Rightarrow \qquad d_{NPKS}(\boldsymbol{a},\boldsymbol{b}) &= \frac{1}{2} - \frac{1}{2}\hat{\tau}_x(\boldsymbol{a},\boldsymbol{b}) \qquad \square \end{aligned}$$

Theorem 2 (Linear transformation between τ_x and d_{PKS})**.** *Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be arbitrary rankings of $n = |V|$ alternatives from space Ω. Then, the τ_x correlation coefficient and the d_{PKS} distance are connected through the equation:*

$$d_{PKS}(\boldsymbol{a},\boldsymbol{b}) = \frac{\bar{n}(\bar{n}-1)}{4} - \frac{n(n-1)}{4}\tau_x(\boldsymbol{a},\boldsymbol{b}), \tag{3.13}$$

where $\bar{n} = |\bar{V}| = |V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}}|$ (i.e., the number of alternatives explicitly ranked by both $\boldsymbol{a}$ and $\boldsymbol{b}$).

Proof. From Theorem 1, we have that:

$$d_{NPKS}(\boldsymbol{a},\boldsymbol{b}) = \frac{1}{2} - \frac{1}{2}\hat{\tau}_x(\boldsymbol{a},\boldsymbol{b}),$$

which can be expanded via Equations (2.5) and (3.2) as:

$$\frac{d_{KS}(\boldsymbol{a}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})}, \boldsymbol{b}|_{(V_{\boldsymbol{a}} \bigcap V_{\boldsymbol{b}})})}{\bar{n}(\bar{n}-1)/2} = \frac{1}{2} - \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij}}{2\bar{n}(\bar{n}-1)}.$$

Thus, multiplying both sides by $\bar{n}(\bar{n}-1)/2$ yields:

$$\begin{aligned} d_{PKS}(\boldsymbol{a},\boldsymbol{b}) &= \frac{\bar{n}(\bar{n}-1)}{4} - \frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij} \\ &= \frac{\bar{n}(\bar{n}-1)}{4} - \frac{n(n-1)}{4}\left[\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij}}{n(n-1)}\right] \end{aligned}$$

which completes the proof since the bracketed expression matches the definition of $\tau_x(\boldsymbol{a},\boldsymbol{b})$. □

The following two corollaries are a direct result of these connections.

Corollary 1. *The rank aggregation optimization problems typified by $\hat{\tau}_x$ and d_{NPKS} are equivalent and, thus, provide identical consensus rankings. Similarly, the rank aggregation optimization problems typified by τ_x and d_{PKS} are equivalent.*

Proof. The first part of corollary is established through the following series of equations:

$$\begin{aligned}
\arg\min_{r\in\Omega_C} \sum_{\ell=1}^{|L|} d_{NPKS}(\boldsymbol{r}, \boldsymbol{a}^{\ell}) &= \arg\max_{r\in\Omega_C} \sum_{\ell=1}^{|L|} -d_{NPKS}(\boldsymbol{r}, \boldsymbol{a}^{\ell}) && (3.14)\\
&= \arg\max_{r\in\Omega_C} \sum_{\ell=1}^{|L|} -\left[\frac{1}{2} - \frac{1}{2}\hat{\tau}_x(\boldsymbol{r}, \boldsymbol{a}^{\ell})\right] && (3.15)\\
&= \arg\max_{r\in\Omega_C} \sum_{\ell=1}^{|L|} \hat{\tau}_x(\boldsymbol{r}, \boldsymbol{a}^{\ell}), && (3.16)
\end{aligned}$$

where the last equation results from the fact that scalars common to every term in the sum and constant terms do not impact the optimal solution.

Similarly, the second part of the corollary can be proved via the following series of equations:

$$\begin{aligned}
&\arg\min_{r\in\Omega_C} \sum_{\ell=1}^{|L|} d_{PKS}(\boldsymbol{r}, \boldsymbol{a}^{\ell}) && (3.17)\\
&= \arg\max_{r\in\Omega_C} \sum_{\ell=1}^{|L|} -d_{PKS}(\boldsymbol{r}, \boldsymbol{a}^{\ell}) && (3.18)\\
&= \arg\max_{r\in\Omega_C} \sum_{\ell=1}^{|L|} -\left[\frac{(|V_{\boldsymbol{r}} \cap V_{\boldsymbol{a}^{\ell}}|)(|V_{\boldsymbol{r}} \cap V_{\boldsymbol{a}^{\ell}}| - 1)}{4} - \frac{n(n-1)}{4}\tau_x(\boldsymbol{r}, \boldsymbol{a}^{\ell})\right] && (3.19)\\
&= \arg\max_{r\in\Omega_C} \sum_{\ell=1}^{|L|} \frac{n(n-1)}{4}\tau_x(\boldsymbol{r}, \boldsymbol{a}^{\ell}) - \frac{(|V_{\boldsymbol{a}^{\ell}}|)(|V_{\boldsymbol{a}^{\ell}}| - 1)}{4} && (3.20)\\
&= \arg\max_{r\in\Omega_C} \sum_{\ell=1}^{|L|} \tau_x(\boldsymbol{r}, \boldsymbol{a}^{\ell}) && (3.21)
\end{aligned}$$

where Equation (3.19) ensues from Theorem 2; where Equation (3.20) results from the fact that, since $\boldsymbol{r}$ must be a complete ranking, $|V_{\boldsymbol{r}} \cap V_{\boldsymbol{a}^{\ell}}| = |V_{\boldsymbol{a}^{\ell}}|$ for every ℓ; and, where Equation (3.21) results from the fact that scalars common to every term in the sum as well as constant terms (i.e, the second term in Equation (3.20) is independent of any candidate solution) have no bearing on the optimal solution. □

Corollary 2. *The correlation-based non-strict incomplete rank aggregation problem is $\mathcal{NP}$-hard.*

Proof. The distance-based non-strict incomplete rank aggregation problem was proven to be $\mathcal{NP}$-hard in Moreno-Centeno and Escobedo (2016). Since solving the correlation-based rank aggregation problem is equivalent to solving the distance-based rank aggregation problem by Corollary 1, the former problem is also $\mathcal{NP}$-hard. □

The distance-correlation connection provides mutual support for the usefulness of the respective measures. In particular, τ, τ_x, and $\hat{\tau}_x$ are strengthened by the robust properties and social choice foundations of the Kemeny aggregation framework (Brandt *et al.*, 2016; Kemeny and Snell, 1962; Young and Levenglick, 1978; Young, 1988). Meanwhile, d_{KS}, d_{PKS}, and d_{NPKS} benefit from the computational advantages engendered by the linear transformation from correlation-based framework, which allows sidestepping nonlinear terms in the definition of distance metrics. These advantages are bolstered by the optimization methodologies developed in Chapter 4.

3.4 Concluding Remarks

In this chapter, we make several contributions to the area of robust rank aggregation. Principally, it develops the $\hat{\tau}_x$ ranking correlation coefficient, which fulfills the standard definitions of statistical correlation in the space of non-strict incomplete rankings. This ranking measure is applicable to situations where no assumptions should be made regarding individual preferences over unranked objects. Its formal derivation and axiomatic foundation ensure that $\hat{\tau}_x$ assigns equitable voting power to each input ranking in the aggregation process, irrespective of the number of objectives it ranks. By also connecting $\hat{\tau}_x$ with d_{NPKS}, this work enhances distance and correlation-based robust methodologies for rank aggregation including the development of expedient optimization methodologies, which appears in the next ensuing chapters.

Chapter 4

MATHEMATICAL PROGRAMMING IN COMPUTATIONAL SOCIAL CHOICE

This chapter is organized as follows. In Section 4.1, the existing mathematical programming formulations for rank aggregation are introduced. Section 4.2 and Section 4.3 introduce an integer programming formulation and a binary programming formulation for the Kemeny rank aggregation with non-strict complete and incomplete rankings. In Section 4.4, the performance of the binary programming formulation is tested through a set of computational experiments.

4.1 Existing Mathematical Programming Formulations for Rank Aggregation

Cook *et al.* (2007a) developed a binary programming formulation for rank aggregation problems related to the Kemeny-Snell distance, and Conitzer *et al.* (2006) developed an integer programming formulation related to the Kendall-τ distance when the input rankings are strict. Because these formulations do not deal with non-strict rankings, Brancotte *et al.* (2015) provided a revised integer programming formulation for the Kendall-τ distance, given as follows:

$$\begin{aligned}
\underset{\boldsymbol{x}}{\text{minimize}} \quad & \sum_{\{v_i,v_j\}\subseteq V} (w_{j\leq i} * x_{i<j} + w_{i\leq j} * x_{j<i} + (w_{i<j} + w_{j<i}) * x_{i=j}) && \text{(4.1a)}\\
\text{subject to} \quad & x_{i<j} + x_{j<i} + x_{i=j} = 1, \quad \forall v_i, v_j \in V && \text{(4.1b)}\\
& x_{i<k} - x_{i<j} - x_{j<k} \geq -1, \quad \forall v_i, v_j, v_k \in V && \text{(4.1c)}\\
& 2x_{i<j} + 2x_{j<i} + 2x_{j<k} + 2x_{k<j} - x_{i<k} - x_{k<i} \geq 0, \forall v_i, v_j, v_k \in V && \text{(4.1d)}\\
& x_{i<j}, x_{j<i}, x_{i=j} \in \mathbb{B}, \quad \forall v_i, v_j \in V. && \text{(4.1e)}
\end{aligned}$$

where $w_{i\leq j}$ denotes the number of rankings with $v_i \succeq v_j$ and $w_{i<j}$ denotes the number

of rankings with $v_i \succ v_j$. Constraint (4.1b) ensures that all pairs of alternatives are assigned exactly one of the three possible relative ordinal positions: preferred, dispreferred, or tied (the first two are strict and the third is non-strict). Constraints (4.1c) and (4.1d) enforce transitivity of strict and non-strict ordinal relationships. We note that this formulation's objective function does not align with the definition of d_{KS} (see Equation (2.3)), based on the different treatment of ties of $d_{\tau'}$. Using d_{KS}, when there are two rankings, where one ties v_i and v_j and the other one does not, this should return half the distance compared to when the ordinal relationships strictly oppose each other. For example, with $\boldsymbol{a}^1 = (1, 2), \boldsymbol{a}^2 = (1, 1), \boldsymbol{a}^3 = (2, 1)$, this yields $d_{KS}(\boldsymbol{a}^1, \boldsymbol{a}^2) = 1$, while $d_{KS}(\boldsymbol{a}^1, \boldsymbol{a}^3) = 2$. To solve the Kemeny rank aggregation problem using Brancotte *et al.* (2015)'s formulation, it is necessary to modify its objective function to match the treatment of ties of d_{KS}. This is done in the following proposition.

Proposition 1. *To adopt the treatment of ties of d_{KS}, Brancotte* et al. *(2015)'s formulation is modified as follows:*

$$\underset{\boldsymbol{x}}{\textit{minimize}} \sum_{\{v_i, v_j\} \subseteq V} (w_{j<i} + \frac{1}{2} w_{j=i}) x_{i<j} + (w_{i<j} + \frac{1}{2} w_{i=j}) x_{j<i} + \frac{1}{2}(w_{i<j} + w_{j<i}) x_{i=j}$$

$$\textit{subject to constraints } (4.1\text{b}) - (4.1\text{e}).$$

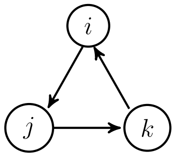
This modified version is used throughout the experiments to allow for fair comparison.

4.2 Generalized Integer Programming Formulation for Rank Aggregation

In this section, we leverage the correlation coefficient interpretation of Kemeny rank aggregation to derive a new integer programming formulation that is applicable to non-strict complete and incomplete rankings. The key to this formulation relies on devising a constraint set that ensures that the values of matrix induce a complete and consistent set of preferences, i.e., a complete and non-strict ranking. To this end, we develop a graph-based representation of the ranking-matrix (see Equation (2.8)) of a non-strict complete ranking.

Definition 4. *Let $G = (V, E)$ be unweighted directed graph for representing a non-strict complete ranking $\boldsymbol{r}$ as follows: V is the set of nodes (alternatives) and an each pair of nodes is connected by one or two directed edges $E \subseteq V \times V$ according to the preference relationship between each pair of alternatives. It includes the directed edge (i, j) if $r_i < r_j$ (i.e., $v_i \succ v_j$) and it includes the directed edges (i, j) and (j, i) if $r_i = r_j$ (i.e., $v_i \approx v_j$).*

From Definition 4, given a non-strict complete ranking, it is straightforward to construct its digraph (or matrix) representation. However, not every unweighted digraph will correspond to a complete and consistent set of preferences, since certain ones can induce preference cycles. For example, Figure (4.1a) can be represented via the matrix in Figure (4.1b), but these representations do not yield a non-strict complete ranking due to the preferential cycle ($v_i \succ v_j$, $v_j \succ v_k$, and $v_k \succ v_i$).



(a) Example of an unweighted directed graph

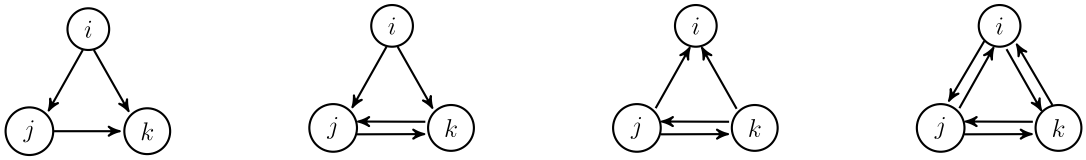
$$\begin{bmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}$$

(b) Matrix representation of the digraph

Figure 4.1: Not Every Unweighted Digraph (or Its Matrix Representation) Yields a Complete and Consistent Set of Preferences

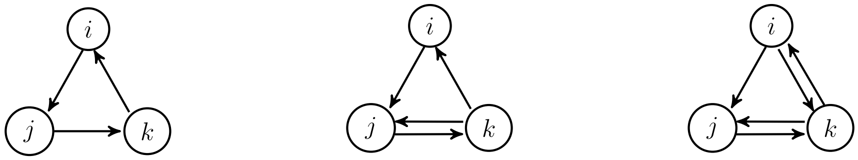
Hereafter, we define a *ranking-matrix graph* as an unweighted directed graph that induces a non-strict complete ranking (i.e., it does not create any preferential cycles). To identify a ranking-matrix graph structure, certain conditions are needed.

For starters, since the solution must be a complete ranking, each pair of nodes in G must be connected by at least one directed edge. The following theorem specifies the remaining conditions for an arbitrary unweighted digraph to be a ranking-matrix graph. To this end, *a uni-cycle* is defined as a simple path that starts and ends on the same vertex in *one* direction *but not in the reverse direction. A bi-cycle* is defined as a simple path that starts and ends on the same vertex and can be traversed in *both* directions. According to these definitions, a bi-cycle and a uni-cycle are mutually exclusive. Additionally, it is not possible to have a uni-cycle of size 2 because, if there exists a directed edge from i to j and a directed edge from j to i, this creates a bi-cycle. The focus of the theorem and proof is to prevent graphs with uni-cycles since such structures can be associated with inconsistent sets of preferences (i.e., non-transitivity). To be more succinct and precise, we denote a graph without uni-cycles as a *uni-cycle-free graph* and a graph with at least one uni-cycle as a *unicyclic graph.* Figures 4.2 and 4.3 show the possible uni-cycle-free graphs and unicyclic graphs, respectively, over three alternatives.



(a) G'^1 (b) G'^2 (c) G'^3 (d) G'^4

Figure 4.2: Unicycle-free Graphs



(a) G'^5 (b) G'^6 (c) G'^7

Figure 4.3: Unicyclic Graphs

Theorem 3. *Let $G = (V, E)$ be an unweighted directed graph for representing a non-*

strict complete ranking $\boldsymbol{r}$ as follows: V is the set of nodes (alternatives) and an each pair of nodes is connected by one directed edge, that is, either $(i, j) \in E$ or $(j, i) \in E$ if $r_i < r_j$ (i.e., $v_i \succ v_j$) or two directed edges (i, j) and (j, i) if $r_i = r_j$ (i.e., $v_i \approx v_j$). Graph G is a ranking-matrix graph if and only if it does not contain uni-cycles (i.e., it is a unicycle-free graph).

Proof. Let $G' = (V', E')$ be a subgraph of G, that is, $V' \subseteq V$, $E' = (V' \times V') \cap E$. A bi-cycle exists whenever, for every adjacent pair of nodes i_k, i_{k+1} in a path $i_1, i_2, i_3, ..., i_p$, we have $(i_k, i_{k+1}) \in E'$ and $(i_{k+1}, i_k) \in E'$, for $1 \leq k \leq p-1 < |V'|$. Clearly, when there exists a bi-cycle in G', it means every alternative included in the bi-cycle is tied. That is, when there exist directed cycles $i_1, i_2, ..., i_{p-1}, i_p, i_1$ and $i_1, i_p, i_{p-1}, ..., i_2, i_1$ with distinct nodes $i_1, i_2, ..., i_{p-1}, i_p \in V'$, then $i_1 \approx i_2 \approx ... \approx i_p$. Hence, whenever a graph of size 3 has a bi-cycle, it leads to a valid setting for the corresponding ranking-matrix entries. Specifically, this gives that $a_{i_\ell, i_{\ell'}} = 1$, $a_{i_{\ell'}, i_\ell} = 1$ for all i_ℓ, $i_{\ell'} \in \{i_1, i_2, ..., i_p\}$, where $i_\ell \neq i_{\ell'}$.

Recall that we can focus on a graph of length 3 or greater since a graph of size 2 cannot contain uni-cycle. Moreover, to check if these cycles exist in a digraph having at least one directed edge between every pair of nodes, it is sufficient to concentrate on finding unicyclic triads (i.e., uni-cycles of size 3), as a preference graph without unicyclic triads cannot have any higher-order uni-cycles (Gass, 1998). Hence, $|V'| = 3$ from this point, without loss of generality. To continue, Figure 4.2 lists distinct isomorphic classes of unicycle-free digraphs of size 3, while Figure 4.3 lists distinct isomorphic classes of unicyclic digraphs of size 3. Even though more unicyclic and unicycle-free graphs of size 3 are possible, it is only necessary to consider three respective isomorphic classes given in Figures 4.2 and 4.3; all other graphs can be represented by permuting the labels of the given graphs.

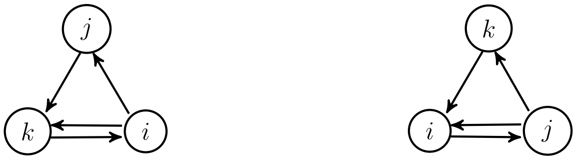


Figure 4.4: Isomorphically Equivalent Digraphs of G'^6

For Figure (4.3a), because $v_i \succ v_j$, $v_j \succ v_k$, and $v_k \succ v_i$, the preference relation includes the cycle $v_i \succ v_j \succ v_k \succ v_i$, which does not yield a proper ranking of three alternatives). For Figure (4.3b), because $v_i \succ v_j$, $v_j \approx v_k$, and $v_k \succ v_i$, the preference relation includes the cycle $v_i \succ v_j \approx v_k \succ v_i$, which also does not yield a complete and consistent set of preferences. For Figure (4.3c), because $v_i \succ v_j$, $v_j \approx v_k$, and $v_k \approx v_i$, the preference relation $v_i \succ v_j \approx v_k \approx v_i$ is not a consistent set of preferences. Therefore, the graphs with a uni-cycle do not yield a complete and consistent set of preferences, meaning they cannot correspond to a ranking. Using proof by exhaustion, we have demonstrated that graph G is a ranking-matrix graph if and only if it is a unicycle-free graph. □

The results of this theorem and the following corollary will be used to derive a new integer programming formulation for Kemeny aggregation.

Corollary 3. *The ranking-matrix* $\mathbf{S} \in \mathbb{Z}^{n\times n}$*, along with corresponding auxiliary binary variables* $\mathbf{Y} \in \mathbb{B}^{n\times n}$*, induces a complete and consistent set of preferences of* n *alternatives (i.e., a non-strict complete ranking) if the following constraints are satisfied for some setting of* $\mathbf{S}$ *and* $\mathbf{Y}$*:*

$$s_{ij} - s_{kj} - s_{ik} \geq -1 \quad i,j,k = 1,...,n; \quad i \neq j \neq k \neq i \tag{4.2a}$$

$$s_{ij} + s_{ji} \geq 0 \quad i,j = 1,...,n; \quad i < j \tag{4.2b}$$

$$s_{ii} = 0 \quad i = 1,...,n; \tag{4.2c}$$

$$s_{ij} - 2y_{ij} = -1 \quad i,j = 1,...,n; \quad i = j \tag{4.2d}$$

$$s_{ij} \in \{-1,0,1\}, y_{ij} \in \{0,1\} \quad i,j = 1,...,n. \tag{4.2e}$$

Proof. For ranking-matrix $\mathbf{S} = [s_{ij}]$ to represent a complete and consistent set of preferences, the following conditions must be met—note that the definition of s_{ij} is exactly the same as the definition of a_{ij} given by Equation (2.8). First, the diagonal elements must be set to 0, that is, $s_{ii} = 0$, which is represented by constraint (4.2c). The off-diagonal elements must be non-zero values, specifically, they must be equal to 1 or -1. This is enforced via auxiliary binary variables y_{ij} in constraint (4.2d). Moreover, s_{ij} and s_{ji} cannot both be negative. Hence, constraint (4.2b) restricts at least one of s_{ij} and s_{ji} to be positive, when $i \neq j$. Constraint (4.2e) explicitly states the respective domains of s_{ij} and y_{ij}.

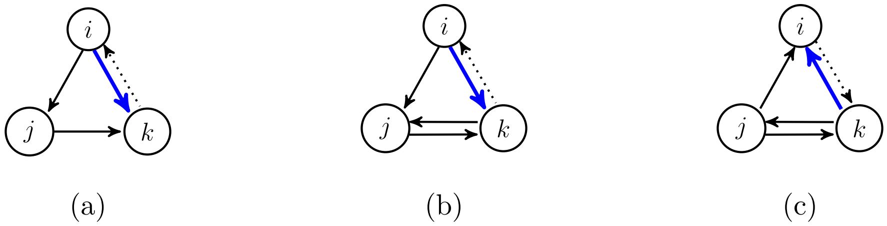


Figure 4.5: The Dotted Directed Edge Creates a Uni-cycle

The proof of Theorem 3 explains that to check whether a uni-cycle of any length exists within a ranking-matrix digraph, it is sufficient to verify that no uni-cycles of length 3 exist. The possible unicycle-free and unicyclic and digraphs over three alternatives are shown in Figures 4.2 and 4.3. Each unicyclic graph of size 3 can be obtained by adding a particular directed edge of one of the unicycle-free graphs shown in Figure 4.2 or by replacing a particular directed edge with its reverse directed edge. This is depicted Figure 4.5. Specifically, adding the dotted directed edge or replacing the thick (blue) edge with the dotted directed edge creates a uni-cycle. For example, replacing the thick (blue) edge with the dotted directed edge in Figure (4.5a) and Figure (4.5b) yields Figure (4.3a) and Figure (4.3b), respectively. Also, adding the dotted directed edge in Figure (4.5b) yields Figure (4.3c), which is also a unicyclic graph.

This implies that the unicyclic graphs can be avoided by eliminating these additional or replacement edges from occurring. More specifically, as in Figures (4.5a) and (4.5b), whenever there is a directed edge from i to j, but not one from j to i, which gives that $s_{ij} = 1$, $s_{ji} = -1$, and a directed edge from j to k (with or without one from k to j), which gives that $s_{jk} = 1$ (here, the value of s_{kj} does not matter), the edge between i and k should be directed from i to k (and not in the opposite direction), which gives that $s_{ki} = -1$. This condition can be written as:

$$s_{ji} = -1, s_{jk} = 1 \Longrightarrow s_{ki} = -1. \tag{4.3}$$

Moreover, as in Figure (4.5c), whenever there is a directed edge from j to i, but not one from i to j, which gives that $s_{ji} = 1, s_{ij} = -1$, and a directed edge from k to j (with or without one from j to k), which gives that $s_{kj} = 1$ (here, the value of s_{jk} does not matter), the edge between k and i should be directed from k to i (and not in the opposite direction), which gives that $s_{ik} = -1$. This condition can be written as:

$$s_{ij} = -1, s_{kj} = 1 \Longrightarrow s_{ik} = -1. \tag{4.4}$$

Conditions (4.3) and (4.4) can be equivalently satisfied via the following linear constraints:

$$s_{ji} - s_{jk} \geq s_{ki} - 1 \tag{4.5}$$

$$s_{ij} - s_{kj} \geq s_{ik} - 1. \tag{4.6}$$

In fact, the formulation can be further reduced. By swapping labels i and j in constraint (4.6) (since it holds for any permutation of the labels), we can derive constraint (4.5) (i.e., it is redundant). Therefore, constraints (4.2a)-(4.2e) provide the full set of constraints. □

4.3 Generalized Binary Programming Formulation for Rank Aggregation

We develop the first exact integer programming formulation for generalized Kemeny aggregation, that is, for non-strict complete and incomplete rankings, which is given by:

$$\begin{aligned} \underset{\mathbf{S}}{\text{maximize}} \quad & \sum_{i}\sum_{j} c_{ij} s_{ij} \\ \text{subject to} \quad & \text{constraints (4.2a)-(4.2e)} \end{aligned}$$

where $[c_{ij}] \in \mathbb{Z}^{n \times n}$ is the cumulative ranking-matrix of the input rankings, defined as $c_{ij} = \sum_{\ell=1}^{|L|} a_{ij}^{\ell}$, for $i, j = 1, ..., n$ (see Equation (2.8)), when the rankings are complete. When they are incomplete, the cumulative ranking-matrices defined by Yoo *et al.* (2020) can be utilized, which correspond to the incomplete-ranking distances introduced in Dwork *et al.* (2001) and Moreno-Centeno and Escobedo (2016). For the former it is defined as $c_{ij} = \sum_{l=1}^{|L|} = \frac{\alpha_{ij}^{l}}{n(n-1)}$, where n is the total number of alternatives, and for the latter it is defined as $c_{ij} = \sum_{\ell=1}^{|L|} \frac{\alpha_{ij}^{\ell}}{n^{\ell}(n^{\ell}-1)}$, where n^{ℓ} is the number of each alternative evaluated by judge ℓ. In effect, these expressions normalize the ranking-matrix values of each judge according to the total number of alternatives or to the number of alternatives evaluated by each judge. The formulation can be adapted for other incomplete-ranking measures (distances or correlation coefficients) that can be summarized via a respective ranking-matrix.

The Generalized Kemeny-aggregation Binary Programming formulation (GKBP) is obtained by substituting for s_{ij} with $(2y_{ij} - 1)$ in the integer programming formulation, which gives:

$$
\begin{aligned}
&\underset{\mathbf{y}}{\text{maximize}} && \sum_i \sum_j c_{ij}(2y_{ij} - 1) && && \text{(4.7a)}\\
&\text{subject to} && y_{ij} - y_{kj} - y_{ik} \geq -1 && i,j,k = 1,...,n; \quad i \neq j \neq k \neq i && \text{(4.7b)}\\
& && y_{ij} + y_{ji} \geq 1 && i,j = 1,...,n; \quad i < j && \text{(4.7c)}\\
& && y_{ii} = 0 && i = 1,...,n; && \text{(4.7d)}\\
& && y_{ij} \in \{0,1\} && i,j = 1,...,n; \quad i \neq j. && \text{(4.7e)}
\end{aligned}
$$

Combining Equation (4.2d) with the definition of ranking-matrix $[s_{ij}]$ gives the implicit definition of y_{ij}, which represents the ordinal relationship between alternatives v_i, v_j. Upon inspection, GKBP has n^2 variables while Brancotte *et al.* (2015)'s formulation has $\frac{3}{2}n^2 - 2n$ variables; additionally, the new formulation has $n(n-1)(n-2)/2$ fewer constraints. Hence, GKBP has approximately $O(n^2)$-fewer variables and $O(n^3)$-fewer constraints than Brancotte *et al.* (2015)'s formulation. Taking advantage of the above binary programming formulation, Escobedo *et al.* (2021) derives an equivalent mathematical programming formulation of Equation (6.2), which is further described in Section 6.3.

On a more fundamental level, GKBP can be connected to the theory of order polytopes. An order polytope $\mathbf{P^n_O}$ is the convex hull of vertices that represent the possible members of a specific type of binary relation on n alternatives. Notable examples are the linear order polytope $\mathbf{P^n_{LO}}$—the convex hull of binary relations that are total, irreflexive, and transitive—and the weak order polytope $\mathbf{P^n_{WO}}$—the convex hull of binary relations that are total, reflexive, and transitive—since their vertices correspond to strict complete rankings and non-strict complete rankings, respectively. Previous works formulated Kemeny aggregation for strict complete rankings as a special case of the formulation of the linear ordering problem, whose aim is to find a linear ordering that maximizes the sum of weights c_{ij} in a weighted directed graph (Newman and Vempala, 2001; Martí and Reinelt, 2011). The ensuing theorem makes an analogous connection between GKBP and the weak order polytope. To the best of our knowledge, this is the first work to establish such a connection.

Theorem 4. *The GKBP constraints provide a formulation for the weak-order polytope.*

Proof. Fiorini and Fishburn (2004) provides a binary programming formulation for $\mathbf{P^n_{WO}}$. For $i, j \in \{1...n\}$, the decision variable x_{ij} is defined as:

$$x_{ij} = \begin{cases} 1 & \text{if } j \succeq i \\ 0 & \text{otherwise,} \end{cases}$$

which is equivalent to the definition of y_{ji} in GKBP. The constraints for $\mathbf{P^n_{WO}}$ are given as:

$$x_{ij} \leq 1 \tag{4.8a}$$

$$x_{ij} + x_{ji} \geq 1 \tag{4.8b}$$

$$x_{ik} - x_{ij} - x_{jk} \geq -1. \tag{4.8c}$$

By substituting x_{ij} with y_{ji}, the constraints become:

$$y_{ji} \leq 1 \tag{4.9a}$$

$$y_{ji} + y_{ij} \geq 1 \tag{4.9b}$$

$$y_{ki} - y_{ji} - y_{kj} \geq -1. \tag{4.9c}$$

Notice that the combination of Equation (4.9a) and (4.9b) gives the domain of y_{ij}. Equation (4.9b) matches with the constraint (4.7c) in GKBP. Furthermore, the above constraints hold for any permutation of the labels; therefore, changing i to j, j to k, and k to i yields:

$$y_{ij} - y_{kj} - y_{ik} \geq -1. \tag{4.10}$$

Equation (4.10) is the same as constraint (4.7b) in GKBP. Therefore, the constraints of GKBP provide a logically equivalent formulation of the weak-order polytope. □

From this theorem, the underpinnings of the GKBP formulation are strengthened through their connection with the theory of order polytopes. In fact, the GKBP constraints are equivalent to the basic family of facet-defining inequalities (see Fiorini and Fishburn (2004)). This connection gives the formulation inherent computational advantages since the facet-defining inequalities of $\mathbf{P}^{\mathbf{n}}_{\mathbf{WO}}$ could help obtain tighter lower bounds for the Kemeny aggregation problem within the branch and bound algorithm, thereby expediting solution times (Nemhauser and Wolsey, 1988).

4.4 Computational Experiments

The computational studies compare the performance of two formulations for rank aggregation: GKBP (formulation (4.7)) and the modified version of Brancotte *et al.* (2015) stated in Proposition 1. The test datasets consist of probabilistic instances constructed based on the concept of Mallows distribution (see Section 4.4.1) and benchmark instances from PrefLib, a library of preference data (Mattei and Walsh, 2013). Prior to describing the experiments, recall that GKBP finds a ranking that maximizes agreement quantified according to the Kendall τ-extended correlation coefficient and the modified Brancotte *et al.* (2015) model finds a ranking that minimizes disagreement quantified according to d_{KS}. Due to the connection of this distance-correlation coefficient pairing (see Equation (2.10)), the two respective problems are equivalent, which allows for a fair comparison of their performance.

The experiments were performed on machines equipped with 36GB of RAM memory shared by two Intel Xeon E5-2680 processors running at 2.40 GHz; code was written in Python and the formulations were solved using CPLEX solver version 12.8.0 (IBM Knowledge Center, 2017).

4.4.1 Instances from probabilistic distribution

The formulations are first tested on randomized instances constructed from rankings sampled from a probabilistic distribution with an underlying ground truth and an adjustable level of noise/error. This choice allows for the generation of instances with differing levels of difficulty, thereby enabling a systematic comparison of the formulations. Among the existing options for generating randomized rankings, the Mallows-ϕ model (Mallows, 1957; Diaconis, 1988; Marden, 1996; Critchlow, 2012) is the most popular and has been used similarly in other works—e.g., (Lu and Boutilier, 2014; Betzler *et al.*, 2014; Asfaw *et al.*, 2017; Crispino *et al.*, 2019; Yoo *et al.*, 2020).

The Mallows-ϕ model is a Kendall-τ distance-based model (i.e., the Kemeny-Snell distance-based model when the rankings are strict and complete), which is parameterized by a "ground truth" (or reference) ranking $\underline{\boldsymbol{a}}$ and "dispersion" $\phi \in (0, 1]$. These parameters are used to quantify the probability of obtaining a complete ranking $\boldsymbol{a}$ as:

$$P(\boldsymbol{a}) = P(\boldsymbol{a}|\underline{\boldsymbol{a}}, \phi) = \frac{\phi^{d_{KS}(\boldsymbol{a},\underline{\boldsymbol{a}})}}{Z},$$

where Ω_C is the space of complete rankings. When sampling from this distribution, as ϕ gets closer to 0, the generated ranking converges to $\underline{\boldsymbol{a}}$; as ϕ gets closer to 1, any complete ranking has equal probability of occurring (i.e., this becomes the uniform distribution). Note that the Mallows' model can be used to generate a set of noisy rankings from a ground truth and a given dispersion parameter that is shared by all of the rankings. On the other hand, Kemeny aggregation returns the maximum likelihood estimator of a model in which each judge provides a noisy estimate of one ground truth ranking with each judge possessing the same dispersion or noise parameter ϕ. Hence, each process can be interpreted as being the inverse of the other. Prior works have developed efficient algorithms for sampling rankings from the Mallows-ϕ model

(Doignon *et al.*, 2004; Ceberio *et al.*, 2015; Irurozki *et al.*, 2016). However, it is inefficient to sample rankings directly from the Mallows-ϕ model. Instead, we use a slightly modified version of the repeated insertion model of Doignon *et al.* (2004), which was originally designed for strict complete rankings. The next paragraph describes how ties and incompleteness are added to the rankings generated by the repeated insertion model so as to provide suitable instances for testing the featured formulations.

Algorithm 1 and 2 describe how non-strict complete and incomplete ranking instances are constructed and guided based on the concept of Mallows distribution. Before introducing the algorithms, we provide needed definitions. Let $\boldsymbol{a}^{-1}$ be an *alternative-ordering* induced from rankings by sorting the alternatives from best to worst, according to their ranks. For example, for $\boldsymbol{a} = (1, 5, 2, 4, 3)$, $\boldsymbol{a}^{-1} = (v_1, v_3, v_5, v_4, v_2)$. Extending this notation, $\boldsymbol{a}^{-1}(i)$ specifies the ith-highest ranked alternative in $\boldsymbol{a}$ (Doignon *et al.*, 2004); in the aforementioned example, $\boldsymbol{a}^{-1}(3) = v_5$. When $\boldsymbol{a}$ is a non-strict ranking, the alternative-ordering is obtained by putting alternatives with the same rank position into preference equivalence classes; for example, for $\boldsymbol{a} = (1, 3, 3, 1, 5)$, $\boldsymbol{a}^{-1} = (\langle v_1, v_4\rangle, \langle v_2, v_3\rangle, v_5)$. Additionally, $\underline{\boldsymbol{a}}^{-1}|_{V_{\boldsymbol{a}}}$ is an alternative ordering that is projected to the alternatives in $V_{\boldsymbol{a}}$, where $V_{\boldsymbol{a}}$ is an alternative set which is evaluated by $\boldsymbol{a}$. Finally, UniDist(L, U) denotes the discrete uniform distribution, where L and U are the minimum and maximum values of the distribution.

Algorithm 1 Generating non-strict complete rankings

```
**Input**: Dispersion: $\phi$, reference alternative ordering: $\underline{\boldsymbol{a}}^{-1}$
**Output**: A set of non-strict complete rankings
for $i = 1, 2, ..., |V|$ do
    for $j = 1, 2, ..., i$ do
        $a^{-1}(j) \leftarrow \underline{a}^{-1}(i)$ with probability: $p_{ij} = \phi^{i-j}/(1 + \phi + \cdots + \phi^{i-1})$
while (the number of alternatives involved in ties) $\leq 0.5n$ do
    $u \leftarrow$ UniDist$(1, h - 1)$, where $h$ is the worst (highest-ranked) position in $\boldsymbol{a}$
    $v \leftarrow a^{-1}(u + 1)$, and then $a_v \leftarrow u$
```

To generate non-strict rankings, a random number u is repeatedly drawn from a discrete uniform distribution $U(1, h-1)$, where h is the highest-valued (worst) rank position in the current ranking. The alternative in rank position u is tied with the alternative in the next rank position higher than u. Ties are repeatedly inserted until the number of tied alternatives reaches a specified threshold, which is set to $0.5n$. For example, let $\boldsymbol{a} = (1, 2, 3, 3, 5)$ and $u = 3$. The next rank position higher than u is 5, and v_5 is the alternative with this rank. Therefore, $\boldsymbol{a}$ becomes $(1, 2, 3, 3, 3)$, and the process stops because the number of tied alternatives reaches the threshold (i.e., $3 > 0.5 \cdot 5 = 2.5$).

Algorithm 2 Generating non-strict incomplete rankings

```
**Input**: Dispersion: $\phi$, alternative set for $\boldsymbol{a}$: $V_{\boldsymbol{a}}$, projected reference alternative ordering: $\underline{\boldsymbol{a}}^{-1}|_{V_a}$
**Output**: A set of non-strict incomplete rankings
for $i = 1, 2, ..., |V_a|$ do
    for $j = 1, 2, ..., i$ do
        $a^{-1}(j) \leftarrow \underline{a}^{-1}(i)|_{V_a}$ with probability: $p_{ij} = \phi^{i-j}/(1 + \phi + \cdots + \phi^{i-1})$
for $i = 1, 2, ..., |V|$ do
    if $v_i \in V_{\boldsymbol{a}}$ then
        $a_i \leftarrow$ rank position of $v_i$ in $\boldsymbol{a}^{-1}$
    else
        $a_i \leftarrow \bullet$
while (the number of alternatives involved in ties) $\leq 0.5n$ do
    $u \leftarrow \text{UniDist}(1, h-1)$, where $h$ is the worst (highest-ranked) position in $\boldsymbol{a}$
    $v \leftarrow a^{-1}(u+1)$, and then $a_v \leftarrow u$
```

To generate incomplete rankings, the extended repeated insertion model is applied on a subset $V' \subset V$ and then marks the alternatives $V \backslash V'$ as unranked. Ties are inserted to incomplete rankings using the same procedure as with complete rankings but restricted to the alternatives in V'.

Configurations of probabilistic distribution experiments

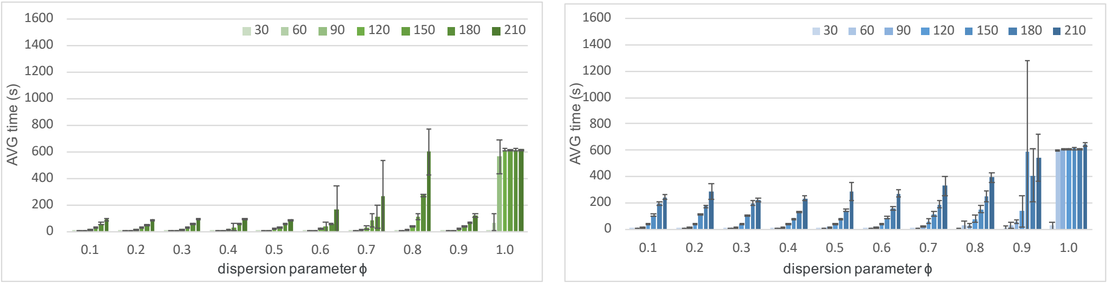
The first set of instances is constructed and guided based on the concept of Mallows distribution; specifically, instances are obtained by sampling complete rankings from the Mallows-ϕ distribution and inserting ties and/or incompleteness, as described in the preceding subsection. We first investigate the effect of varying the dispersion parameter, $\phi \in \{0.1, 0.2, ..., 0.9, 1.0\}$, and the number of alternatives, $n \in \{30, 60, 90, ..., 210\}$; the number of judges is fixed to 50.

For each of the parameter configurations detailed above, in all upcoming experiments, the computing times of each formulation are individually recorded for 10 corresponding instances, which are summarized via average (AVG) and standard deviation (SD) values (represented via error bars). When a formulation cannot return an optimal solution within a 600-second (10-minute) time limit, the relative optimality gap is recorded (a solution is considered as optimal when the relative optimality gap is less than equal to 0.0001). Note that the relative optimality gap for a maximization problem is defined as:

$$\begin{aligned}&\text{Relative optimality gap}\\&\quad=\frac{\text{best relaxation bound} - \text{objective function value for best integer solution}}{\text{objective function value for best integer solution}+\text{1e-10}}\end{aligned}$$

For example, for an instance solved with one of the featured formulations, the objective function value of the current best integer solution was 0.312 and the relative optimality gap was 0.842; this indicates that the objective function value of the optimal solution could be as high as 0.575. As shown in Figure 4.6 and 4.7, the computing times for some instances exceed 600 seconds; this occurs because CPLEX can be slow to terminate when a new incumbent solution is found close to the time limit (IBM Support, 2019).

As shown in Figure 4.6, for non-strict complete ranking instances, GKBP finds the optimal solution in less time than Brancotte *et al.* (2015)'s formulation for most values of n and ϕ. In general, for both formulations, computing times increase with the value of ϕ and n. GKBP returns the optimal solution for all instances within the

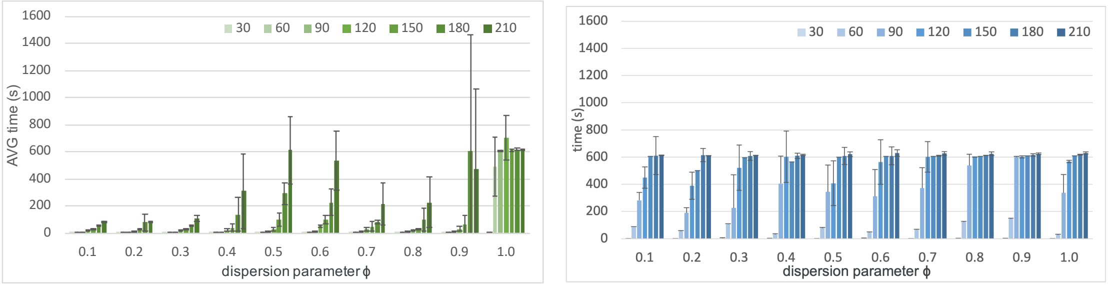


(a) Computing times via GKBP

(b) Computing times via Brancotte et al. (2015)

Figure 4.6: The Average Computing Time for Non-strict Complete Rankings (Note That the Color of These Charts Denotes the Number of Alternatives in the Instances)

time limit, except for $\phi = 1.0$ with $n \geq 90$, while Brancotte *et al.* (2015) is not able to solve some instances with $\phi = 0.7$ with $n = 150$, $\phi = 0.8$ with $n = 210$, and most instances with $\phi \geq 0.9$. Despite the fact that GKBP found an optimal solution faster than Brancotte's model for most of the tested instances, when both reached the time limit without an optimal solution, the optimality gaps of GKBP were at times larger. For example, the average relative optimality gaps over the 10 instances with $\phi = 1.0$ and $n = 90$ were 2.99 for GKBP and 1.00 for Brancotte *et al.* (2015)'s model.



(a) Computing times via GKBP

(b) Computing times via Brancotte et al. (2015)

Figure 4.7: The Average Computing Time for Non-strict Incomplete Rankings (Note That the Color of These Charts Denotes the Number of Alternatives in the Instances)

Figure 4.7 displays the computing times for non-strict incomplete ranking instances. Before explaining the results, we note that Brancotte *et al.* (2015)'s model is not originally designed to handle incomplete rankings, while GKBP can handle incomplete rankings using the cumulative ranking-matrices defined by Yoo *et al.* (2020). To compare the models under the same treatment of incomplete rankings, we normalize $d_{\tau'}$ with the same normalization factor as d_{NPKS} (see Section 4.3). Compared to complete rankings, it takes longer to reach optimality for most values of n and ϕ. Similar to the prior results, GKBP reaches optimality in a shorter amount of time than Brancotte *et al.* (2015). GKBP attains the optimal solution except four instances with $\phi = 0.9$ and $n = 180, 210$ and most instances with $\phi = 1.0$ and $n \geq 60$, while Brancotte *et al.* (2015) cannot solve most instances, except those for all ϕ with $n = 30, 60, 90$, and 120. For example, for an instance with $\phi = 0.9$ and $n = 210$, the relative optimality gap of GKBP is 0.0001, which is considered as optimal, while that of Brancotte *et al.* (2015) is 1.00. We remark that the performance of Brancotte *et al.* (2015)'s model over these instances is worse without the inclusion of the normalization factor.

4.4.2 Instances from Preflib benchmark dataset

The second set of instances is selected from the library of preference data Preflib (Mattei and Walsh, 2013), specifically the "Order with Ties - Complete List (TOC)" dataset. This benchmark dataset consists of 378 instances with differing numbers of alternatives and rankings (i.e., judges) obtained from various domains, and they include real-world data (e.g., figure skating competitions, cross-country skiing and ski jump championship results). The instances include results from Formula One racing and human computation activities, which tend to be relatively less subjective and possess a higher level of collective similarity, as well as data from elections and pure preferences (e.g., the Sushi data set), which tend to be more subjective and possess a lower level of collective similarity. For instance, Milosz and Hamel (2018) estimated the Mallows dispersion parameter ϕ of the "Websearch" instance to be 0.0265, which is a relatively low value. Although we do not have information on the specific dispersion values apart from this single instance, Preflib instances encompass a wide range of subjectivity, meaning the inputs are expected to have varying degrees of collective similarity, as suggested by this discussion.

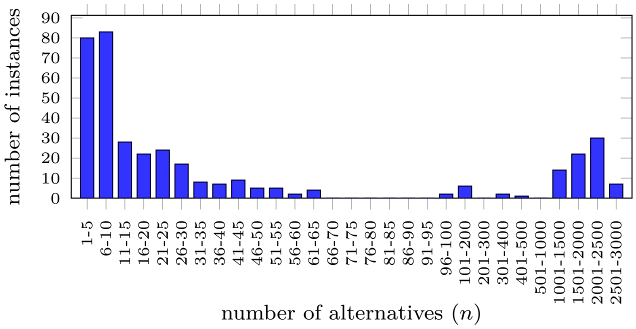


Figure 4.8: The Number of Alternatives (n) of the Instances in the Preflib Dataset

Figure 4.8 summarizes the distribution of the instances according to ranges of n (number of alternatives). As shown in the figure, most instances have $n \leq 65$, but there are a few instances with $n \geq 1000$.

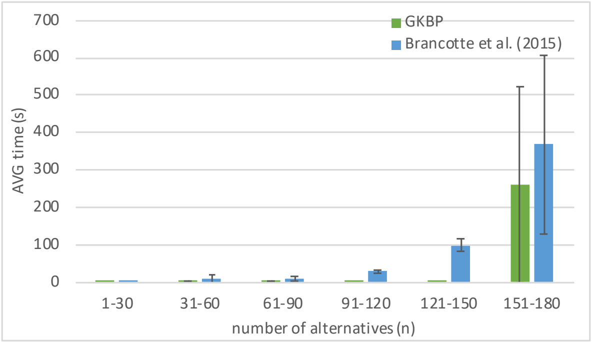


Figure 4.9: The Computing Time of GKBP and Brancotte *et al.* (2015)'s Formulation

For this experiment, only the instances with $n < 300$ are considered; many instances with $n > 300$ resulted in termination likely due to insufficient memory. In all, there are 302 instances with $3 \leq n \leq 170$; for clarity, the instances are grouped by intervals of 30 over the range of n in Figure 4.9. GKBP is faster, on average, than Brancotte *et al.* (2015)'s formulation over all instances with $n \leq 300$. The longest computing time for GKBP is 338.94 seconds when there are 170 alternatives, while the time limit is reached with a relative optimality gap of 0.46, on average, for Brancotte *et al.* (2015). From this analysis, it is evident that GKBP can also solve benchmark problems noticeably faster.

4.5 Concluding Remarks

In this chapter, we developed a new binary programming formulation for robust rank aggregation. The major benefit of the generalized Kemeny-aggregation binary programming formulation is that it is applicable to a wide variety of ordinal preferences including complete and incomplete rankings, with and without ties. Additionally, it has fewer variables and constraints compared to a closely related integer programming formulation for the generalized Kendall-τ distance, leading to computational savings, as demonstrated by the featured set of experiments. The additional benefit of the formulation is that it was leveraged to develop a joint ranking and rating aggregation model by Escobedo *et al.* (2021). Moreover, the connection between the facet-defining inequalities of the weak order polytope and the binary programming formulation demonstrates the theoretical rigor of the proposed exact approach. Lastly, the results of the featured experiments on randomized instances and benchmark data show their substantial computational advantages.

Chapter 5

A STRUCTURAL SOCIAL CHOICE PROPERTY

To further expedite the solution process of Kemeny rank aggregation, this chapter will devise a structural social choice property, which allows a decomposition of a large size problem into a partition of smaller subproblems. Section 5.1 introduces the Condorcet criterion, which is the basis of our new social choice property that is discussed in Section 5.2. Section 5.3 shows the decomposition algorithm for the new property. In Section 5.4, we investigate the effectiveness of the social choice property. Section 5.5 presents the final remarks and future work.

5.1 The Condorcet Criterion

Condorcet (1785) proposed a social choice property, named thereafter as the Condorcet Criterion (CC), stating that if a majority of voters prefers one alternative ahead of all other alternatives, that alternative should alone obtain the best position in the voting outcome. Formally, recalling that p_{ij} is the number of judges who prefer alternative v_i over alternative v_j, this property can be written as:

$$\begin{aligned} &\text{If } \exists v_i \in V, \text{ s.t. } p_{ij} > p_{ji} \text{ (i.e., } v_i \overset{m}{\succ} v_j) \ \forall v_j \in V \backslash \{v_i\} \\ &\qquad \implies v_i^* \succ v_j^*, \text{ or equivalently, } \boldsymbol{r}_i^* < \boldsymbol{r}_j^*, \end{aligned}$$

where $\boldsymbol{r}^*$ is the final aggregate ranking and $v_i^* \succ v_j^*$ indicates that v_i is ranked strictly better than v_j in $\boldsymbol{r}^*$. At the preliminary screening stage of decision-making, keeping a diverse and large set of candidates, rather than selecting few candidates, provides decision-makers with broader options. For this reason, CC provides limited usefulness for decision-making since it can only identify one winning alternative (i.e., the

Condorcet winner) or one losing alternative (i.e., the Condorcet loser), when it is satisfied. An extended version of the Condorcet winner is the Smith set; the winning (losing) Smith set is the smallest non-empty set of alternatives that defeats (is defeated by, resp.) every alternative outside the set in a pairwise election (Smith, 1973). Truchon *et al.* (1998) provided another natural extension of CC, called the Extended Condorcet Criterion (XCC). This property requires that if V can be organized into a partition $\mathcal{V} := \{V^1, ..., V^K\} \in \mathcal{P}(V)$, such that all alternatives in subset $V^k \in \mathcal{V}$ are pairwise preferred over all alternatives in subset $V^{k'} \in \mathcal{V}$ by a majority (i.e., $V^k \overset{m}{\succ} V^{k'}$), where $k < k'$, then the alternatives in V^k must be ranked strictly better than all alternatives in $V^{k'}$ in the optimal ranking. Formally, this property can be written as:

$$\text{If } \exists \mathcal{V} := \{V^1, V^2, ..., V^K\} \in \mathcal{P}(V), \text{ s.t. } V^k \overset{m}{\succ} V^{k'}, \text{ for } 1 \leq k < k' \leq K$$
$$\implies v_i^* \succ v_j^*, \text{ or equivalently, } \boldsymbol{r}_i^* < \boldsymbol{r}_j^*, \ \forall v_i \in V^k, \ \forall v_j \in V^{k'}.$$

Table 5.1 illustrates how XCC can be applied to the Kemeny aggregation problem. In the example, since $\mathcal{V}^{\text{XCC}} := \{\{v_1, v_2\}, \{v_3\}, \{v_4\}\}$ is a partition satisfying XCC, the Kemeny optimal ranking is expected to place v_1 and v_2 ahead of v_3 and v_4, and to place v_3 ahead of v_4 (the optimal ordering between v_1 and v_2 cannot be determined from the application of this property alone). As shown in the table, the Kemeny optimal rankings (three in this case) are all consistent with XCC. Note that Kemeny rank aggregation (as well as other distance-based methods) may yield more than one optimal solution (Young and Levenglick, 1978; Dwork *et al.*, 2001). Indeed, Muravyov (2014) explained that the number of optimal solutions in Kemeny aggregation can at times be greater than the number of input rankings and that these solution rankings may rank the alternatives in significantly different ways, which can lead to ambiguity—this is called *Paradox of Kemeny.* This is explained to a great extent by the fact that the Kemeny rank aggregation problem can be characterized as finding the median ranking among the given set of rankings. Medians do not need to be unique and, therefore, Kemeny optimal rankings are not guaranteed to be unique as well (Kemeny and Snell, 1962).

	$\boldsymbol{a}^1$	$\boldsymbol{a}^2$	$\boldsymbol{a}^3$	$\boldsymbol{a}^4$	$\boldsymbol{a}^5$	$\boldsymbol{a}^6$	$\boldsymbol{a}^7$	$\boldsymbol{a}^8$	Kemeny optimal rankings		
v_1	1	2	1	2	1	2	1	2	1	1	2
v_2	2	1	2	1	2	1	2	1	1	2	1
v_3	3	3	3	4	4	4	3	3	3	3	3
v_4	4	4	4	3	3	3	4	4	4	4	4

Table 5.1: The Kemeny Optimal Solutions Are Consistent With XCC

Although both CC and XCC have been implemented to refine the complexity of Kemeny aggregation (i.e., providing parameterized complexity with respect to the number of subsets and the size of subsets in $\mathcal{V}^{\text{XCC}}$), they are not appropriate for non-strict rankings. In the example showcased in Table 5.2, we have that $p_{12} = 5$, $p_{21} = 3$, $t_{12} = 1$, $p_{23} = 5$, $p_{32} = 3$, $t_{23} = 1$, $p_{13} = 3$, $p_{31} = 0$ and $t_{13} = 6$. According to the definition of XCC, the final optimal solution is expected to be $(1, 2, 3)$, since $\mathcal{V}^{\text{XCC}} := \{\{v_1\}, \{v_2\}, \{v_3\}\}$ is a partition satisfying XCC—that is, $p_{ij} > p_{ji}$ for $1 \leq i < j \leq 3$. However, v_1, v_2, and v_3 are tied in the aggregate ranking when optimizing with the Kemeny-Snell distance and allowing ties. Effectively, this implies that Kemeny aggregation for non-strict rankings is not consistent with XCC. In order to overcome this inadequacy, we define a new social choice property in the ensuing subsection.

	$\boldsymbol{a}^1$	$\boldsymbol{a}^2$	$\boldsymbol{a}^3$	$\boldsymbol{a}^4$	$\boldsymbol{a}^5$	$\boldsymbol{a}^6$	$\boldsymbol{a}^7$	$\boldsymbol{a}^8$	$\boldsymbol{a}^9$	Kemeny optimal ranking	XCC solution
v_1	1	1	1	1	1	1	2	2	2	1	1
v_2	1	3	3	3	2	2	1	1	1	1	2
v_3	3	1	1	1	3	2	2	2	2	1	3

Table 5.2: The Optimal Solution Is Not Consistent With CC and XCC

5.2 An Extended Condorcet Criterion with Ties

To introduce a substitute to XCC that is suitable for non-strict rankings, we first define an important concept. We say that *a decisive majority* prefers an alternative v_i over an alternative v_j, written as $v_i \overset{M}{\succ} v_j$, if $p_{ij} > p_{ji} + t_{ij}$, that is, the number of people who prefer v_i over v_j is greater than the number of people who prefer v_j over v_i plus those who tie them. If neither $v_i \overset{M}{\succ} v_j$ nor $v_j \overset{M}{\succ} v_i$, there is no decisive majority that prefers v_i over v_j, and vice versa, written as $v_i \overset{M}{\approx} v_j$. Similarly, we say that a decisive majority prefers all alternatives in the partition V^k over all alternatives in the partition $V^{k'}$, written as $V^k \overset{M}{\succ} V^{k'}$, if $p_{ij} > p_{ji} + t_{ij}, \forall v_i \in V^k, \forall v_j \in V^{k'}$. If neither $V^k \overset{M}{\succ} V^{k'}$ nor $V^{k'} \overset{M}{\succ} V^k$, there is no decisive majority that prefers V^k over $V^{k'}$, and vice versa, written as $V^k \overset{M}{\approx} V^{k'}$.

Definition 5. *Let $\mathcal{V} := \{V^1, V^2, ...V^K\}$ s.t. $V^k \overset{M}{\succ} V^{k'}$ for $1 \leq k < k' \leq K$. The Non-strict Extended Condorcet Criterion (NXCC) requires that all $v_i \in V^k$ must precede all $v_j \in V^{k'}$ in the final ranking. That is,*

$$\begin{gathered} \textit{If } \exists \mathcal{V} := \{V^1, V^2, ..., V^K\} \in \mathcal{P}(V), \ \textit{s.t. } V^k \overset{M}{\succ} V^{k'}, \ \textit{for } 1 \leq k < k' \leq K \\ \implies v_i^* \succ v_j^*, \textit{ or equivalently, } \boldsymbol{r}_i^* < \boldsymbol{r}_j^*, \forall v_i \in V^k, \ \forall v_j \in V^{k'}. \end{gathered}$$

A basic implication of this property is that, the more subsets the partition has, the smaller the sizes of the subproblems that need to be solved (since each subset will tend to have fewer alternatives). Note that when all rankings are strict and complete, NXCC is exactly XCC because $t_{ij} = 0$.

5.3 Decomposition Algorithm

To apply NXCC, it is necessary to determine the ordered partition of subsets of alternatives—in which lower-indexed subsets are each preferred over higher-indexed subsets by a decisive majority—from the data. This can be done via Algorithm 3, which has a worst-case complexity of $O(n^2)$, where n is the number of alternatives. At iteration $i-1$, the algorithm inserts v_i before, within, or after the existing subsets in the working partition $\mathcal{V} = \{V^1, V^2, ..., V^{\kappa(i)}\}$, where $\kappa(i)$ is the number of subsets prior to the iteration. To determine its precise point of insertion, v_i is compared at most to the alternatives in all subsets from V^1 to $V^{\kappa(i)}$, which takes at most $(i-1)$ comparisons, each of which is assumed to take constant time. This has to be done for $i = 2, \ldots, n$. Therefore, the algorithm requires at most $n(n-1)/2$ such comparisons, resulting in a worst-case complexity of $O(n^2)$.

Algorithm 3 Decomposition Algorithm

```
    Input: $\{p_{ij}\}$,$\{p_{ji}\}$,$\{t_{ij}\}$
    Output: An ordered partition of subsets $\mathcal{V} = \{V^1, V^2, ..., V^K\}$
$\mathcal{V} = \{\{v_1\}\}$
for $i = 2, 3, ..., |V|$ do
    if $(\exists v_j \in V^1$ s.t., $t_{ij} \geq |p_{ij} - p_{ji}|)$,  or
        $(\exists v_j \in V^1$ s.t., $p_{ij} > p_{ji} + t_{ij}$ and $\exists v_{j'} \in V^1 \backslash \{v_j\}$ s.t., $p_{j'i} > p_{ij'} + t_{ij'})$  then
        Put $v_i$ in $V^1$ and $k \leftarrow 2$
    else if $\forall v_j \in V^1$ s.t., $p_{ij} > p_{ji} + t_{ij}$ then
        Insert $v_i$ before $V^1$, increment the index of subsets after $V^{\sigma(i)}$ by 1, and $k \leftarrow 3$
    else if $\forall v_j \in V^1$ s.t., $p_{ji} > p_{ij} + t_{ij}$ then
        Insert $v_i$ after $V^1$, increment the index of subsets after $V^{\sigma(i)}$ by 1, and $k \leftarrow 3$
    while $k \leq |\mathcal{V}|$ do
        if $(\exists v_j \in V^k$ s.t., $t_{ij} \geq |p_{ij} - p_{ji}|)$,  or
            $(\exists v_j \in V^k$ s.t., $p_{ij} > p_{ji} + t_{ij}$ and $\exists v_{j'} \in V^k \backslash \{v_j\}$ s.t., $p_{j'i} > p_{ij'} + t_{ij'})$ then
            Merge subsets from $V^{\sigma(i)}$ to $V^k$
            Decrease the index of subsets after $V^k$ by $k - \sigma(i)$ and $k \leftarrow \sigma(i) + 1$
        else if $\forall v_j \in V^k$ s.t., $p_{ij} > p_{ji} + t_{ij}$ then
            $k \leftarrow k + 1$
        else if $\forall v_j \in V^k$ s.t., $p_{ji} > p_{ij} + t_{ij}$ then
            if $|\sigma(i) - k| = 1$ and $|V^{\sigma(i)}| = 1$ then
                Move $V^{\sigma(i)}$ after $V^k$ and increment the index of subsets after $V^{\sigma(i)}$ by 1
            else
                Merge subsets from $V^{\sigma(i)}$ to $V^k$
                Decrease the index of subsets after $V^k$ by $k - \sigma(i)$ and $k \leftarrow \sigma(i) + 1$
```

* $\sigma(i)$ is the index of the subset containing v_i.

The main difference between NXCC and XCC is that ties are or are not considered, respectively, to determine the majority's strict pairwise preferences. Specifically, XCC does not consider ties to be relevant to the conclusion that $v_i \in V^k$ should be strictly preferred over all $v_j \in V^k$. On the other hand, NXCC requires that to arrive at this conclusion, the number of judges who strictly prefer v_i over v_j should be greater than those who do not—which includes those who tie them or who strictly prefer v_j over v_i. Table 5.2 illustrates that the outcome of XCC decomposition is not consistent with Kemeny aggregation for non-strict rankings. Therein, since $t_{13} > p_{13} - p_{31}$ (the number of judges who tie v_1 and v_3 is greater than the net difference between the number of judges who have a strict preference), it cannot be concluded that v_1 should be ahead of v_3 in the final optimal ranking. Hence, these two alternatives and every other alternative between them cannot be ordered a priori into separate subsets, which is the outcome obtained when NXCC is applied to the example. The ensuing paragraphs formally prove that the Kemeny optimal solution satisfies NXCC when the rankings are non-strict and complete. This is done through Lemma 1 and Theorem 5. Beforehand, it is useful to introduce some additional notation.

Notation 1. *The reduced instance associated with two subsets of alternatives V^k, $V^{k'} \subset V$, written as $A_{[k \cup k']} = A_{[V^k \cup V^{k'}]}$ is the submatrix induced by rows $V^k \cup V^{k'}$ of A. Similarly, $\boldsymbol{a}^{\ell}_{[k \cup k']} = \boldsymbol{a}^{\ell}_{[V^k \cup V^{k'}]}$ and $\boldsymbol{r}^*_{[k \cup k']} = \boldsymbol{r}^*_{[V^k \cup V^{k'}]}$ are the reduced evaluation from judge ℓ and the optimal reduced ranking with respect to $V^k \cup V^{k'}$, respectively.*

When exactly two alternatives are considered, this notation is modified as follows.

Notation 2. *The reduced instance $A_{\{i,j\}} = A_{\{v_i,v_j\}}$ is the submatrix induced by alternatives v_i and v_j. Similarly, $\boldsymbol{a}^{\ell}_{\{i,j\}} = \boldsymbol{a}^{\ell}_{\{v_i,v_j\}}$ and $\boldsymbol{r}^*_{\{i,j\}} = \boldsymbol{r}^*_{\{v_i,v_j\}}$ are the reduced evaluation from judge ℓ and the optimal reduced ranking with respect to v_i and v_j, respectively.*

Furthermore, the above notation is combined to specify the ranking position of single alternatives within a reduced problem space.

Notation 3. *The ordinal position of v_i in the reduced evaluation from judge $\boldsymbol{a}^\ell$ with respect to $V^k \cup V^{k'}$ is denoted as $a^\ell_{i|[k\cup k']}$.*

Using this notation, the cumulative distance between $\{\boldsymbol{a}^\ell\}_{\ell=1}^{|L|}$ and $\boldsymbol{r}$, accrued by only alternatives v_i and v_j, can be written as $\sum_{\ell\in L} d(\boldsymbol{a}^\ell_{\{i,j\}}, \boldsymbol{r}_{\{i,j\}})$. Similarly, the distance between $\{\boldsymbol{a}^\ell\}_{\ell=1}^{|L|}$ and $\boldsymbol{r}$, accrued by all alternatives in V^k and all alternatives in $V^{k'}$, can be written as $\sum_{\ell\in L} d(\boldsymbol{a}^\ell_{[k\cup k']}, \boldsymbol{r}_{[k\cup k']})$.

Lemma 1. *Let $V^1, V^2 \subset V$ with $V^1 \cap V^2 = \emptyset$. Consider the reduced aggregation problem consisting of input rankings $A_{[1\cup 2]}$, that is, the part of the evaluations involving only $V^1 \cup V^2$. If $V^1 \overset{M}{\succ} V^2$, every alternative $v_i \in V^1$ should obtain a better position than every alternative $v_j \in V^2$ in the optimal solution to the reduced problem, that is, $r^*_{i|[1\cup 2]} < r^*_{j|[1\cup 2]}$, where $\boldsymbol{r}^*_{[1\cup 2]} := \underset{\boldsymbol{r}_{[1\cup 2]}}{\arg\min} \sum_{\ell\in L} d(\boldsymbol{a}^\ell_{[1\cup 2]}, \boldsymbol{r}_{[1\cup 2]})$.*

Proof. We prove this by contradiction. Let $\boldsymbol{r}^*_{[1\cup 2]}$ be a Kemeny optimal ranking to the reduced problem involving only V^1 and V^2 and assume that $r^*_{i|[1\cup 2]} \geq r^*_{j|[1\cup 2]}$, for at least one alternative pair v_i, v_j, where $v_i \in V^1, v_j \in V^2$—i.e., there exists at least one alternative in V^1 which is tied or dispreferred over at least one alternative in V^2. Additionally, denote the ranking where all alternatives in V^1 are preferred over all alternatives in V^2 by $\bar{\boldsymbol{r}}^*_{[1\cup 2]}$.

Initially, choose an arbitrary alternative $v_i \in V^1$. Then, the Kemeny-Snell distance between v_i and every alternative $v_j \in V^2$ is calculated as:

$$\sum_{v_j\in V^2}\sum_{\ell\in L} d(\boldsymbol{a}^\ell_{\{i,j\}|[1\cup 2]}, \boldsymbol{r}^*_{\{i,j\}|[1\cup 2]}) = \frac{1}{\gamma}\sum_{v_j\in V^2}\sum_{\ell\in L} |\text{sign}(a^\ell_{i|[1\cup 2]} - a^\ell_{j|[1\cup 2]}) - \text{sign}(r^*_{i|[1\cup 2]} - r^*_{j|[1\cup 2]})|$$

where γ is a positive constant associated with the minimum distance unit (see Equation (2.3)). Without loss of generality, γ can be ignored in the remainder of the proof because it is a constant term which only affects the objective-value scaling. Moreover, since there are three possible ordinal relationships between v_i and v_j in $\boldsymbol{r}^*$, we have that,

$$\sum_{\ell \in L} d(\boldsymbol{a}^\ell_{\{i,j\}}, \boldsymbol{r}^*_{\{i,j\}}) = \begin{cases} \sum_{\ell \in L} |\text{sign}(a^\ell_i - a^\ell_j) - (-1)| = 2p_{ji} + t_{ij} & \text{if } r^*_i < r^*_j \\ \sum_{\ell \in L} |\text{sign}(a^\ell_i - a^\ell_j) - 0| = p_{ij} + p_{ji} & \text{if } r^*_i = r^*_j \\ \sum_{\ell \in L} |\text{sign}(a^\ell_i - a^\ell_j) - 1| = 2p_{ij} + t_{ij} & \text{if } r^*_i > r^*_j \\ \qquad\qquad 0 & \text{otherwise.} \end{cases}$$

Therefore, the distance between v_i and all $v_j \in V^2$ can be factored as,

$$\sum_{v_j \in V^2} \sum_{\ell \in L} d(\boldsymbol{a}^\ell_{\{i,j\}|[1 \cup 2]}, \boldsymbol{r}^*_{\{i,j\}|[1 \cup 2]}) = \sum_{\substack{v_j \in V^{k'} \\ \text{s.t. } r^*_{i|[1 \cup 2]} > r^*_{j|[1 \cup 2]}}} (2p_{ij} + t_{ij}) + \sum_{\substack{v_j \in V^{k'} \\ \text{s.t. } r^*_{i|[1 \cup 2]} = r^*_{j|[1 \cup 2]}}} (p_{ij} + p_{ji}) + \sum_{\substack{v_j \in V^{k'} \\ \text{s.t. } r^*_{i|[1 \cup 2]} < r^*_{j|[1 \cup 2]}}} (2p_{ji} + t_{ij}).$$

From the assumption that $V^1 \overset{M}{\succ} V^2$ for all $v_i \in V^1$ and $v_j \in V^2$, because $p_{ij} > p_{ji} + t_{ij}$, we can derive the following inequalities involving the second ($r^*_i = r^*_j$) and third ($r^*_i > r^*_j$) cases above:

$$\begin{aligned} 2p_{ji} + t_{ij} &< \quad p_{ij} + p_{ji} \\ 2p_{ji} + t_{ij} &< \ 2p_{ij} + t_{ij}, \end{aligned}$$

which results in the following relationship:

$$\sum_{\substack{v_j \in V^{k'} \\ \text{s.t. } r^*_{i|[1 \cup 2]} > r^*_{j|[1 \cup 2]}}} (2p_{ij} + t_{ij}) + \sum_{\substack{v_j \in V^{k'} \\ \text{s.t. } r^*_{i|[1 \cup 2]} = r^*_{j|[1 \cup 2]}}} (p_{ij} + p_{ji}) + \sum_{\substack{v_j \in V^{k'} \\ \text{s.t. } r^*_{i|[1 \cup 2]} < r^*_{j|[1 \cup 2]}}} (2p_{ji} + t_{ij}) > \sum_{v_j \in V^2} (2p_{ji} + t_{ij}).$$

That is,

$$\sum_{v_j \in V^2} \sum_{\ell \in L} d(\boldsymbol{a}^\ell_{\{i,j\}|[1 \cup 2]}, \boldsymbol{r}^*_{\{i,j\}|[1 \cup 2]}) > \sum_{v_j \in V^2} \sum_{\ell \in L} d(\boldsymbol{a}^\ell, \bar{\boldsymbol{r}}^*_{\{i,j\}|[1 \cup 2]})$$

which means the optimal ranking $\boldsymbol{r}^*$ where v_i is tied or dispreferred over some alternatives in V^2 returns a longer cumulative distance than the ranking $\bar{\boldsymbol{r}}^*$ where v_i

is preferred over all $v_j \in V^2$. This contradicts the assumption that $\boldsymbol{r}^*$ is the optimal ranking, since it does not return the shortest cumulative distance. Therefore, if $V^1 \overset{M}{\succ} V^2$, assigning v_i with a better ranking position than all alternatives $v_j \in V^2$ returns a strictly shorter distance than a ranking where alternative v_i is tied with or dispreferred over an arbitrary alternative v_j, that is $r^*_{i|[1\cup 2]} < r^*_{j|[1\cup 2]}$, $\forall v_j \in V^2$. Since v_i was chosen arbitrarily, this holds for all alternatives in V^1. Hence, if $V^1 \overset{M}{\succ} V^2$, every alternative $v_i \in V^1$ should obtain a better position than every alternative $v_j \in V^2$ in the optimal solution. □

Theorem 5. *Kemeny aggregation satisfies NXCC when the inputs are non-strict and complete.*

Proof. Define the partition $\mathcal{V} = \{V^1, V^2, ..., V^K\}$, where $V^k = \{v^k_1, v^k_2, ..., v^k_{|V^k|}\}$ and assume that $V^k \overset{M}{\succ} V^{k'}$ for every k, k', where $1 \leq k < k' \leq K$. Let $\boldsymbol{r}^*$ be a Kemeny optimal ranking. In order to prove that Kemeny aggregation satisfies NXCC when rankings are non-strict and complete—which means that all elements in V^k should precede all elements in $V^{k'}$ in $\boldsymbol{r}^*$—the Kemeny-Snell distance between V^k and $V^{k'}$ can be calculated for all $k < k'$. To this end, the cumulative Kemeny-Snell distance is expanded as follows:

$$
\begin{aligned}
\sum_{\ell \in L} d(\boldsymbol{a}^\ell, \boldsymbol{r}^*) &= \sum_{i=1}^{|V|-1} \sum_{j=i+1}^{|V|} \sum_{\ell \in L} d(\boldsymbol{a}^\ell_{\{i,j\}}, \boldsymbol{r}^*_{\{i,j\}}) \\
&= \sum_{k=1}^{K} \sum_{k'=1}^{K} \sum_{i=1}^{|V^k|} \sum_{j=1}^{|V^{k'}|} \sum_{\ell \in L} d(\boldsymbol{a}^\ell_{\{i,j\}}, \boldsymbol{r}^*_{\{i,j\}}) \\
&= \underbrace{\sum_{k=1}^{K} \sum_{i=1}^{|V^k|-1} \sum_{j=i+1}^{|V^k|} \sum_{\ell \in L} d(\boldsymbol{a}^\ell_{\{i,j\}}, \boldsymbol{r}^*_{\{i,j\}})}_{\text{within subset } V^k \text{ (intrasubset distances)}} + \underbrace{\sum_{k=1}^{K-1} \sum_{k'=k+1}^{K} \sum_{i=1}^{|V^k|} \sum_{j=i}^{|V^{k'}|} \sum_{\ell \in L} d(\boldsymbol{a}^\ell_{\{i,j\}}, \boldsymbol{r}^*_{\{i,j\}})}_{\text{between subsets } V^k \text{ and } V^{k'} \text{ (intersubset distances)}} .
\end{aligned}
$$

If all $v_i \in V^k$ can be proved to be strictly preferred or strictly dispreferred over all $v_j \in V^{k'}$ in the optimal solution, for all $k < k'$, this guarantees an optimal ordered partition $\mathcal{V} = \{V^1, V^2, ..., V^K\}$, with $V^1 \succ V^2 \succ ... \succ V^K$; the ordering of the alternatives within each subset (i.e., between $v_i, v_{i'} \in V^k$) can be performed in a subsequent step and is ignored for the rest of the proof. We can derive the following bound on the optimal cumulative intersubset distances:

$$\min_{\boldsymbol{r}^*} \sum_{k=1}^{K-1} \sum_{k'=k+1}^{K} \sum_{\ell \in L} d(\boldsymbol{a}^{\ell}_{[k \cup k']}, \boldsymbol{r}^*) \geq \sum_{k=1}^{K-1} \sum_{k'=k+1}^{K} \min_{\boldsymbol{r}^*_{[k \cup k']}} \sum_{\ell \in L} d(\boldsymbol{a}^{\ell}_{[k \cup k']}, \boldsymbol{r}^*_{[k \cup k']}). \tag{5.1}$$

The optimal solutions from the right-hand side of inequality (5.1) correspond to the $K(K-1)/2$ optimal reduced orderings of all subset pairs from $\{V^1, V^2, ..., V^K\} \in \mathcal{V}$, which can produce preference cycles (i.e., contradictions) when combined. For example, assuming $v_i \in V^1, v_j \in V^2, v_k \in V^3$, the optimal solutions to the reduced problems can be $r^*_{i|[1 \cup 2]} < r^*_{j|[1 \cup 2]}$, $r^*_{j|[2 \cup 3]} < r^*_{k|[2 \cup 3]}$ and $r^*_{k|[1 \cup 3]} < r^*_{i|[1 \cup 3]}$. On the other hand, the optimal solution from the left-hand side of inequality (5.1) gives a Kemeny optimal ordering of the K subsets. In other words, the collection of right-hand side two-subset subproblems can be interpreted as a relaxed version of the left-hand side rank aggregation problem. The relaxed problem does not enforce the preference transitivity between all subsets orderings. However, when the optimal solutions to the two-subset optimal reduced orderings yield a combined solution that is feasible to the left-hand side problem, an optimal solution for the original rank aggregation problem has been found, since the objective values will also be equal.

By Lemma 1, when $V^k \overset{M}{\succ} V^{k'}$, the optimal solution to the reduced problem induced by each pair V^k and $V^{k'}$ places every alternative $v_i \in V^k$ ahead of every alternative $v_j \in V^{k'}$, for all $k < k'$. Hence, the two-subset orderings can be combined into a K-subset partial ordering $\mathcal{V}^* = \{V^1, V^2, ..., V^K\}$ with $V^1 \succ V^2 \succ ... \succ V^K$, without contradictions or preferences cycles. In particular, the combined solution to the reduced subproblems is feasible to the original problem and $r^*_i < r^*_j$, where $v_i \in V^k$, $v_j \in V^{k'}$, for all $k < k'$, giving an optimal partial ranking of all alternatives in V. Hence, Kemeny aggregation satisfies NXCC when the rankings are non-strict and complete. □

Besides offering significant potential computational advantages, NXCC brings various practical benefits for decision-making. This includes the ability to focus on the most relevant alternatives. It is usually not known a priori which alternatives will be the ones to occupy the top, middle, or bottom positions in the consensus ranking; the exact positions are ultimately revealed through the aggregation process. One exception by which it may be possible to know such information ahead of time is through the application of the NXCC property developed in this work. In particular, NXCC takes advantage of the pairwise comparison information to determine whether there exist subsets of alternatives that will always be preferred over other subsets of alternatives. By determining the NXCC partition, decision-makers can focus on the exact ordering of the alternatives contained in just the top and/or bottom subsets. That is, alternatives that belong to subsets in the middle of the partition can be dropped from consideration while formally guaranteeing that the relative ordering of the remaining alternatives will not be affected. An additional related benefit is the ability to certifiably rule out certain outcomes even when the consensus ranking is not unique. That is, even if an instance may have multiple alternative optimal solutions but only one is obtained by the exact solution method, the NXCC decomposition would guarantee that alternatives in higher-indexed subsets will never be ranked ahead of alternatives in lower-indexed subsets. Put otherwise, it could help determine that many alternatives will never occupy the top positions of any optimal ranking.

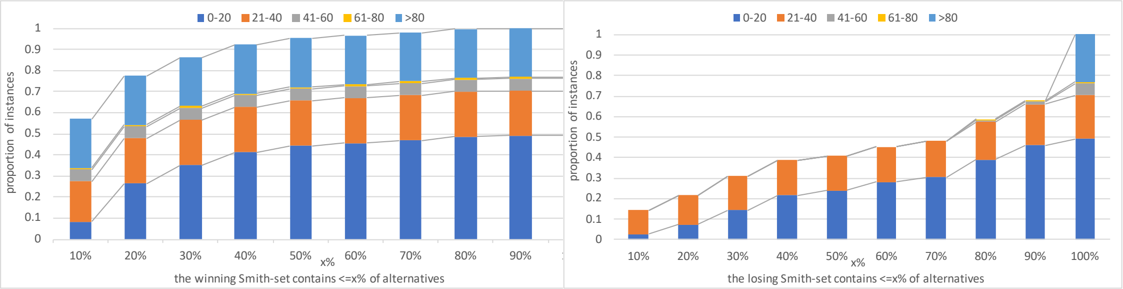
5.4 Computational Performance of NXCC

The purpose of this experiment is to provide an estimate of how NXCC could improve the computing time of Kemeny aggregation via partitioning and bring about other practical benefits. To highlight the practicality of the property, this work experiments on the Preflib benchmark dataset consisting of 378 instances described in Section 4.4.2. During the experiment, the following information is recorded: (1) the existence of a Condorcet winner and loser, (2) whether the partition is non-trivial after the decomposition (i.e., $|\mathcal{V}| \geq 2$, where $|\mathcal{V}|$ is the number of subsets in the partition), and (3) the size of the winning Smith set and the losing Smith set (see Section 5.1); Table 5.3 and Figure 5.1 display this information. Checking the existence of a Condorcet winner and the size of the winning and losing Smith set are two indicators of the effectiveness of NXCC. In particular, a smaller winning Smith set helps to narrow down the winners or most relevant candidates. Likewise, a larger losing Smith set helps decision-makers rule out many irrelevant alternatives, since the complement of the losing Smith set becomes smaller. In addition, practitioners are often most interested in the alternatives that obtain the top positions. For instance, in recommendation system, it is more important to suggest a set of items that is most likely to be preferred (i.e., top alternative sets), rather than suggesting middle or the least preferred items (Davidson *et al.*, 2010). The following analysis helps better understand and quantify the practical benefits of NXCC decomposition.

Key Instance Characteristics	Proportion of Instances
Condorcet winner	191 out of 378 (50.53%)
Condorcet loser	62 out of 378 (16.40%)
Non-trivial partition ($|\mathcal{V}| \geq 2$)	230 out of 378 (60.85%)

Table 5.3: NXCC Helps Identify the Most Relevant Candidates in the Instances from the Preflib Dataset

More detailed information about the Smith sets is visualized in Figure 5.1, which shows the cumulative distribution of the proportion of instances in the dataset with respect to the percentage of alternatives in the winning Smith set and the losing Smith set (note that the color legend of these charts serves to group the tested instances according to the range of number of alternatives they consider). Specifically, each bar graph represents the number of instances whose Smith set contains at most $x\%$ of the total number of alternatives. For example, Figure (5.1a) shows that approximately 58% of instances have a small winning Smith set (i.e., at most $.1n$ alternatives in the winning Smith set); in particular, at most 10% of the alternatives (i.e., $.1n$ alternatives) are contained in the winning (top) Smith set for the majority of the instances with more than 80 alternatives (colored in light blue in the graph). Moreover, Figure (5.1b) shows that approximately 60% of the instances have a large losing Smith set; in particular, more than 90% of the alternatives are contained in the losing (bottom) Smith set for instances with more than 80 alternatives (colored in light blue in the graph). From these observations, we conclude that NXCC yields small winning Smith sets and large losing Smith sets for more than 50% of the instances, which means that NXCC decomposition can significantly simplify the identification of relevant candidates from these benchmark instances.



(a) Cumulative distribution - Winning Smith Set Size

(b) Cumulative distribution - Losing Smith Set Size

Figure 5.1: NXCC Yields Small Winning Smith Sets and Large Losing Smith Sets for More than 50% of the Instances

Next, we test the computational benefits of the decomposition. To do so, this study compares the computing times of solving the full (non-decomposed) instances (see the computing time of solving these instances in Figure 4.9) and solving the decomposed instances. The latter include the partitioning time and the time of solving the subproblem for each subset in sequential manner. The experiment was conducted on the PrefLib instances with a non-trivial partition (i.e., $|\mathcal{V}| \geq 2$) and $n \leq 300$ (because CPLEX could not solve non-decomposed instances with more alternatives); the number of instances that meet these conditions is 177. Table 5.4 shows the *relative improvement* in computing time (i.e., reduction in computing time), which is defined as follows:

$$\frac{\text{(time to solve non-decomposed problem)} - \sum \text{(time to solve decomposed subproblems)}}{\text{(time to solve full (non-decomposed) problems)}} \times 100\%$$

$\lvert\mathcal{V}\rvert$	2	3	4	5-10	11-20	21-30
number of instances	72	33	20	21	12	19
relative improvement	25%	44%	72%	65%	67%	96%

Table 5.4: Applying GKBP for Each Subset after Partitioning Reduces the Computing Time by at Least 25%

As shown in Table 5.4, the computing time is reduced when a higher number of subsets is obtained after the decomposition. For example, when $|\mathcal{V}| = 4$, the computing time is reduced by 72%, on average, whereas the computing time is reduced by 25% on average when $|\mathcal{V}| = 2$. If each subset was solved by multiple processors at the same time, the computing time could be further reduced. Hence, using distributed computing resources, the more finely decomposed $\mathcal{V}$ is, the faster that large instances could be solved with the combination of GKBP and NXCC. It is pertinent to point out, however, that most of the benchmark instances with less than 150 alternatives can be solved within a minute using GKBP (without decomposition). To further assess the potential computational improvements of NXCC, we also apply the decomposition algorithm to non-strict complete ranking instances generated using the procedure described in Section 4.4.1. Here, we restrict n to larger values, specifically $n \in \{90, 120, 150, 180, 210\}$, and generate 10 instances for each value. Moreover, we fix $\phi = 0.8$, since this is the setting after which certain instances cannot be solved to optimality by GKBP within ten minutes (see Figure (4.6a)). In other words, such instances are relatively difficult but they can still be solved by GKBP within a reasonable amount of time.

Table 5.5 shows the (absolute) improvement in computing time, which is calculated as the difference between the time to solve full (non-decomposed) problem and the cumulative time to solve the decomposed subproblems (i.e., the numerator in the relative improvement calculation).

n	90	120	150	180	210
(absolute) improvement (s)	12.12	47.31	125.84	269.35	560.42+
standard deviation (s)	0.50	3.60	29.60	72.41	207.23

Table 5.5: The Effectiveness of NXCC Is More Prominent on Larger, More Difficult Instances

As shown in Table 5.5, after applying NXCC, computing times improve significantly for every n; most strikingly, decomposed instances with 210 alternatives are solved within few seconds after decomposition, but the original non-decomposed instances were not solved within 600 seconds (note that non-decomposed instances with 210 alternatives are not solved within the time limit, meaning the absolute improvement is at least what is shown on the table). We surmise that NXCC is more effective in these instances because they are moderately difficult to solve.

5.5 Concluding Remarks

This chapter introduced a computationally expedient social choice property, which unlike the original Condorcet criterion and the extended Condorcet criterion, aligns with the Kemeny-Snell distance when dealing with rankings with ties. Moreover, the structural decomposition enabled by the novel social choice property has polynomial-time computational complexity, which decomposes large-size problems into smaller subproblems, while guaranteeing that the optimal solutions to the subproblems can be joined to provide the overall optimal solution. Through the combination of the binary programming formulation and the social choice property, certain instances that could only be solved approximately can be solved exactly in a reasonable amount of time, even when the input evaluations are relative non-cohesive and/or contain hundreds of alternatives.

Chapter 6

OVERCOMING ANCHORING EFFECTS IN MULTIMODAL INPUT ELICITATION TO EXTRACT MORE ACCURATE CROWD ESTIMATES

"For the many, of whom each individual is but an ordinary person, when they meet together may very likely be better than the few good, if regarded not individually but collectively, just as a feast to which many contribute is better than a dinner provided out of a single purse. Hence, the many are better judges than a single man of music and poetry; for some understand one part, and some another, and among them they understand the whole." — Aristotle, Politics

6.1 Crowdsourcing

"Who wants to be a millionaire" is an internationally popular television quiz show. A contestant wins a top prize of $1,000,000 by answering fifteen multiple-choice questions, each of which is worth a specific amount of money and is of increasing difficulty. Contestants can request different types of assistance when they get stuck on a question, including 'Ask the Audience' and 'Ask the Expert'. For the 'Ask the Audience' option, the audience is asked the question and a quick poll is elicited; its historical accuracy is 91%. On the other hand, the 'Ask the Expert' option has a historical accuracy of 65% (Economist, 2004). The show has evinced that it is better to trust the opinion of crowds rather than a single expert, demonstrating the benefits of crowdsourcing. Crowdsourcing is formally defined as an outsourcing of work to a large group of people that were traditionally assigned to a single person (Quinn and Bederson, 2011). This concept has been applied in various settings; for example, prediction markets are popular crowd-based approaches for predicting future events which are unknown, such as presidential election outcomes and future product sales (Rothschild, 2009; Atanasov *et al.*, 2017). Specifically, market participants are buying or selling contracts (or shares of stocks) based on their beliefs or predictions, and then these collective predictions (i.e., market prices) naturally approach a static point (a market equilibrium), which represents the probability of a future outcome. An online user-generated review site is one of the crowd-based approaches where multiple reviewers leave ratings and they are aggregated to provide people with useful information (Lee *et al.*, 2015). In transportation information systems, crowdsourcing has been utilized to collect information, such as bike routes and traffic congestion, and to solicit feedback on the quality of transit service (Misra *et al.*, 2014). Crowdsourced human inputs can be also used to collect data and to obtain labels for data samples in data analytics (Xintong *et al.*, 2014; Najafabadi *et al.*, 2015; Zheng *et al.*, 2018) and to complement the computer's ability within a human-computer system (Demartini *et al.*, 2017). As suggested above, crowdsourcing is primarily utilized in two ways: (1) to collect diverse information and (2) to uncover an unknown ground truth by eliciting and aggregating individual collective estimates. The second main purpose is the focus of this chapter. As described in the aforementioned practical examples, eliciting opinions from multiple people is beneficial for obtaining reliable judgments. Among the options of eliciting quantitative evaluations over multiple items (i.e., objects, alternatives) from crowds, cardinal inputs (ratings) and ordinal inputs (rankings) are the most common forms. They are used in various applications such as online reviews (Aral, 2014), academic paper competitions (Hochbaum and Levin, 2006b), information retrieval (Farah and Vanderpooten, 2007; Yilmaz *et al.*, 2008), and similarity search (Fagin *et al.*, 2003; Ye *et al.*, 2016; Gao and Xu, 2019).

As mentioned earlier, there is a longstanding debate as to whether rankings or ratings should be adopted for opinion elicitation. In this chapter, both ordinal and cardinal estimates are elicited, and they are aggregated via optimization-based aggregation models and traditional aggregation/voting rules (more detailed explanations are included in Section 6.3). Reflecting on the aforementioned discussion, this chapter examines the following research questions:

1. What could be the best way to elicit individual opinions? Is ranking better than rating, or vice versa?

2. Does multimodal information help achieve better collective estimates?

3. Can multimodal input elicitation cause anchoring effects?

4. Does prioritizing ranking over rating information (or vice versa) improve the collective estimates from multimodal aggregation models?

To answer these research questions, we design a crowdsourcing experiment based on an extended version of the dot estimation task, which is considered a benchmark task in crowdsourced computation since it allows for an objective comparison of the collective estimates to a known ground truth. The standard dot estimation task asks participants to estimate the number of dots in different images (Horton, 2010); the estimates are then aggregated to provide the collective estimate(s). Our work explores the idea of combining cardinal and ordinal inputs to improve collective cardinal and ordinal estimates. We note that the main differences from Kemmer *et al.* (2020) are that (1) we use two user interfaces to test whether one interface is more convenient than the other and whether one leads to more accurate collective estimates, and (2) we prioritize the multimodal information by assigning priority weights on each estimate in order to test the importance of each modality of information in multimodal aggregation, which is discussed in Section 6.6.

We offer three main contributions. First, we provide empirical evidence on the effect of utilizing different elicitation methods and user interfaces on the quality of elicited opinions. Second, we show that the anchoring effects encountered in the elicitation of multimodal estimates can counteract the benefit of the wisdom of crowds. Third, we empirically justify that assigning asymmetric priority weights to the different information elements can improve the quality of collective estimates.

The organization of the rest of the chapter is as follows. Section 6.2 reviews the related literature and describes a set of hypotheses that will be tested. Section 6.3 introduces the featured aggregation methods. Section 6.4 includes a description of the crowdsourcing experiment. Section 6.5 evaluates the effects of the different elicitation methods and user interfaces tested in the experiment, and Section 6.6 evaluates the impact of assigning asymmetric priority weights to multimodal information. Lastly, Section 6.7 concludes with a discussion of the contributions and the practical implications of this work.

6.2 Hypothesis Development

6.2.1 Multimodality in decision-making

Multimodality is broadly defined and used differently across various applications—e.g., image-text representation (Kruk *et al.*, 2019), visual-acoustical features (Sun *et al.*, 2020). However, we mainly focus on different types of individual input predictions in decision-making throughout this chapter, specifically ranking and rating estimates.

As mentioned in Section 1.2, rankings and ratings are often regarded as competing alternatives for eliciting preference data. Ratings enable the expression of preference intensity, however, rating scales are subjective and different from one person to another (Ammar and Shah, 2012). Rankings circumvent the issue of inconsistent subjective scales by focusing on pairwise comparisons between items, but it is not straightforward how to determine the quality of the evaluated items in the absolute sense. As mentioned earlier, most previous works employ only one modality of decisions. Based on a thorough and systematic literature review, this is the only work, to the best of our knowledge, which discusses multimodality in the context of crowdsourcing.

We investigate not only the usefulness of ranking and rating measures respectively, but also the effectiveness of utilizing multimodal information jointly (Section 6.5). For this purpose, we define the following hypothesis:

Hypothesis 1. *Aggregating multimodal estimates can improve crowd wisdom relative to what is achieved from aggregating inputs from unimodal estimates.*

Moreover, we hypothesize that assigning asymmetric priority weights in multimodal aggregation affects collective estimation (Section 6.6).

Hypothesis 2. *Assigning asymmetric priority weights to different modalities can improve crowd wisdom relative to what is achieved from assigning symmetric priority weights.*

6.2.2 Anchoring

As countless experiments have demonstrated, when people make a subsequent numerical estimation, they tend to start with an implicit or explicit reference point and adjust their estimates upwards or downwards based on their reference points—a cognitive heuristic formally termed as an *anchoring-and-adjustment* (Tversky and Kahneman, 1974). This heuristic generally creates an anchoring bias, which negatively affects the quality of estimates especially when the initial reference point largely deviates from the true value and adjustments are insufficient. In crowdsourcing, Simoiu *et al.* (2019) examines the effect of showing the cumulative crowd response, the most recent, and the most confident responses to participants, and it discovers that showing the cumulative crowd response degrades the performance because participants heavily rely on it. However, for the other two responses shown, participants appropriately ignore initial inaccuracies of the provided responses and, consequently, the crowd performance is relatively unaffected.

Anchoring is also discussed in the field of managerial decision-making and behavioral operations management. For example, within an inventory management setting, a manager needs to make a decision on the price and production quantity of their products. Many studies show that decision-makers anchor on the mean demand and/or the prior order quantity for the production quantity decision of the next period (Schweitzer and Cachon, 2000; Benzion *et al.*, 2008; Bolton and Katok, 2008; Becker-Peth *et al.*, 2013), which leads to the underordering/overordering behavior (Long and Nasiry, 2015; Ramachandran *et al.*, 2018). The anchoring effect is also observed in the connection with consumer's preferences and recommendation systems (Adomavicius *et al.*, 2013; Xiao and Benbasat, 2018). Specifically, a rating provided by a recommendation system may serve as an anchor, biasing the consumers' own preference ratings. Because consumers' preferences are used as an input of these systems, the biased preference ratings may be harmful to maintaining a good quality of recommendations. As discussed in a wide array of research, the anchoring effect plays a significant role in making a decision. In our research, we investigate whether anchoring on the ordinal estimates influences the quality of the cardinal estimates. The third hypothesis is as follows:

Hypothesis 3. *Anchoring effects caused by eliciting cardinal estimates from self-provided ordinal estimates can negatively affect the quality of collective estimates.*

6.3 Aggregation Methods

The collected individual estimates are integrated using a number of aggregation methods described in this section. Beforehand, some notation conventions are described. Let $\boldsymbol{a}^\ell$ and $\boldsymbol{b}^\ell$ denote the ordinal and cardinal estimate vectors, respectively, gathered from participant ℓ. The subset of alternatives (i.e., images) evaluated in $\boldsymbol{a}$ (resp., $\boldsymbol{b}$) is denoted as $V_{\boldsymbol{a}}$ (resp., $V_{\boldsymbol{b}}$)—this additional notation is needed because participants are asked to evaluate only a subset of all images. Also, let a_i^ℓ denote the rank position of alternative i in the ordinal estimate from participant ℓ. Ordinal estimates are assumed to be strict (i.e., ties are not allowed).

We have used five traditional voting rule-based methods and four optimization-based models. Specifically, the five traditional voting rule-based methods include average, median, plurality rule, Copeland rule, and Borda rule (see Section 2.1) and the following optimization-based models are used.

Ordinal aggregation (ranking aggregation)

The Ordinal Aggregation (OA) model is a ranking-based aggregation model, which minimizes d_{NPKS} for incomplete rankings. The mathematical formulation for ordinal aggregation is the Generalized Kemeny-aggregation Binary Programming formulation (See Section 4.3).

Cardinal aggregation (rating aggregation)

The Cardinal Aggregation (CA) model is a rating-based aggregation model, which minimizes d_{NPCK} for incomplete ratings, and is mathematically written as follows:

$$\min_{\boldsymbol{r}} \sum_{\ell=1}^{|L|} d_{NPCK}(\boldsymbol{b}^\ell, \boldsymbol{r}).$$

where $\boldsymbol{r}$ is the consensus rating. Escobedo *et al.* (2021) introduces the integer programming formulation used to solve CA herein as follows:

$$\underset{\mathbf{r},\mathbf{t}}{\text{maximize}} \quad \sum_{\ell=1}^{|L|} -4C^{\ell} \sum_{(i,j)\in E^{\ell}} t_{ij}^{\ell} \tag{6.1a}$$

$$\text{subject to} \quad t_{ij}^{\ell} - \mu(r_i - r_j) \geq -p_{ij}^{\ell} \qquad (i,j) \in E^{\ell}, \ell = 1, ..., |L| \tag{6.1b}$$

$$t_{ij}^{\ell} + \mu(r_i - r_j) \geq p_{ij}^{\ell} \qquad (i,j) \in E^{\ell}, \ell = 1, ..., |L| \tag{6.1c}$$

$$r_i \leq \frac{U - L}{\mu} \qquad i = 1, ..., n \tag{6.1d}$$

$$r_i \in \mathbb{Z}^{+}_{\cup\{0\}} \qquad i = 1, ..., n. \tag{6.1e}$$

Parameters C^{ℓ} inside the objective function denote the normalization, which is a denominator in Equation (2.6) and defined as:

$$C^{\ell} = (4R \cdot \left\lceil \frac{n^{\ell}}{2} \right\rceil \cdot \left\lfloor \frac{n^{\ell}}{2} \right\rfloor)^{-1}$$

(note that only the number of alternatives evaluated by $\boldsymbol{a}^{\ell}$ is considered in this expression because $\boldsymbol{r}$ is always a complete rating). Auxiliary variables t_{ij}^{ℓ} are introduced to linearize the objective function and are defined as $t_{ij}^{\ell} = \left|\mu(r_i - r_j) - p_{ij}^{k}\right|$, where $p_{ij}^{\ell} = b_i^{\ell} - b_j^{\ell}$, and they require additional constraints (Constraints (6.1b) and (6.1c)). Recalling that the definition of edge set E in Section 6.4.3, parameter E^{ℓ} represents an edge set comprised of the pairwise comparison of nodes provided by a participant ℓ. Additionally, parameter μ specifies the minimum separation gap in rating values in the solution, which is 1 in our experiment. Lastly, r_i represents the rating value of alternative i in the aggregated outcome, which must be calibrated accordingly to obtain its original scale (i.e., $r_i \leftarrow L + \mu r_i$).

Cardinal and ordinal aggregation

The Cardinal and Ordinal Aggregation (COA) model jointly aggregates a set of rating and ranking information by utilizing d_{NPKS} and d_{NPCK}. This optimization model returns the optimal rating $\boldsymbol{r}$, and the optimal ranking is induced by ordering the values of $\boldsymbol{r}$ in nonincreasing order, written as rank($\boldsymbol{r}$). In other words, the rating-ranking solution is perfectly correlated. Define λ_a and λ_b as the weights assigned to the rating distance values and to the ranking distance values, respectively; these parameters allow changing the relative importance of the two input modalities. The COA optimization model is defined as follows:

$$\min_{\boldsymbol{r}} \sum_{\ell=1}^{|L|} \lambda_a \; d_{NPCK}(\boldsymbol{b}^{\ell}, \boldsymbol{r}) + \sum_{\ell=1}^{|L|} \lambda_b \; d_{NPKS}(\boldsymbol{a}^{\ell}, \text{rank}(\boldsymbol{r})). \tag{6.2}$$

Escobedo *et al.* (2021) derives the full mixed integer programming formulation used to solve COA, which is written as follows (assume that λ_a and λ_b are equal to 1):

$$\begin{aligned}
\underset{\boldsymbol{r},\mathbf{t},\mathbf{y}}{\text{maximize}} \quad & \sum_{\ell=1}^{|L|} -4C^{\ell} \sum_{(i,j)\in\mathcal{E}^{\ell}} t_{ij}^{\ell} + \sum_{\ell=1}^{|L|} D^{\ell} \sum_{(i,j)\in\mathcal{E}^{\ell}} y_{ij} && && (6.3\text{a})\\
\text{subject to} \quad & t_{ij}^{\ell} - \mu(r_i - r_j) \geq -p_{ij}^{\ell} && (i,j) \in \mathcal{E}^{\ell}, \ell = 1, ..., |L| && (6.3\text{b})\\
& t_{ij}^{\ell} + \mu(r_i - r_j) \geq p_{ij}^{\ell} && (i,j) \in \mathcal{E}^{\ell}, \ell = 1, ..., |L| && (6.3\text{c})\\
& r_i - r_j \geq M_1 y_{ij} - 1 && i,j = 1, ..., n; i \neq j && (6.3\text{d})\\
& -r_i + r_j \geq M_2(1 - y_{ij}) && i,j = 1, ..., n; i \neq j && (6.3\text{e})\\
& r_i \leq \frac{U - L}{\mu} && i = 1, ..., n && (6.3\text{f})\\
& r_i \in \mathbb{Z}^{+}_{\cup\{0\}} && i = 1, ..., n && (6.3\text{g})\\
& y_{ij} \in \{0,1\} && i,j = 1, ..., n. && (6.3\text{h})
\end{aligned}$$

Parameters D^{ℓ} inside the objective function denote the normalization constant, which is a denominator in Equation (2.5) and is defined as:

$$D^{\ell} = \frac{1}{2n^{\ell}(n^{\ell} - 1)}.$$

Note that M_1 and M_2 are constants large enough so that constraint (6.3d) is always satisfied when $y_{ij} = 1$ and constraint (6.3e) is always satisfied when $y_{ij} = 0$, for any feasible setting of $\boldsymbol{r}$.

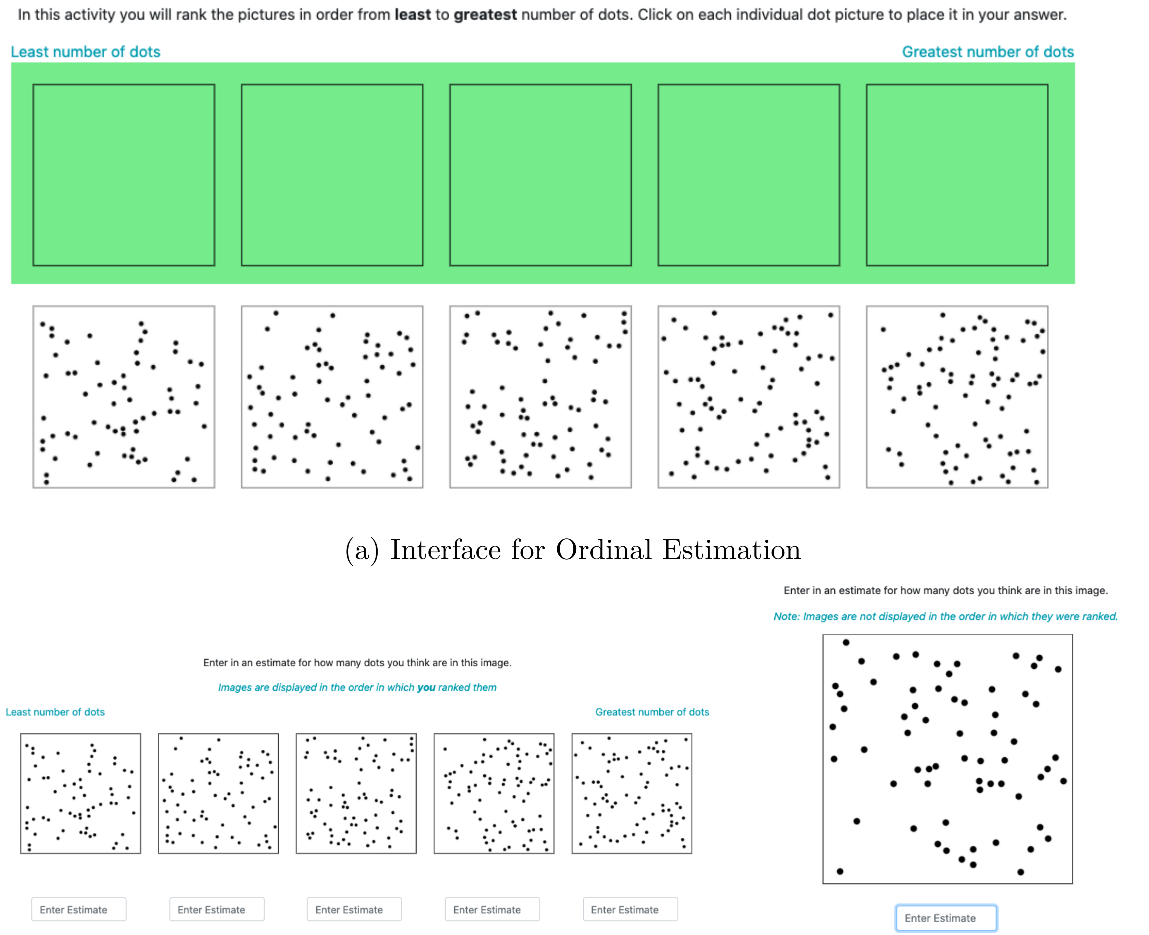
Separation-deviation model

The Separation-Deviation (SD) model is another multimodal model, which takes into account both the difference between the pairwise comparison of two alternatives i and j in the aggregated outcome and in each participant's evaluations (separation) and the difference between the value of alternative i in the aggregated outcome and in each participant's evaluation (deviation) (Hochbaum, 2010) (for the mathematical formulation, see Section 2.2).

6.4 Experimental Design and Data Collection

6.4.1 Experimental settings

The experiment consists of an expanded dot estimation task. The task prompts participants to first order a subset of images (i.e., ordinal estimates) based on the number of dots each contains (order from the least number of dots to the greatest number of dots), as shown in Figure (6.1a). Second, it prompts them to estimate the number of dots (i.e., cardinal estimates) contained in each of the same images. There are two possible ways to estimate the number of dots. One way is that the images are shown together in the order that the participants provide the ordinal estimates, as shown in Figure (6.1b)—in this input elicitation option, the ordinal estimates and the cardinal estimates are associated, called *Setting A* throughout the paper. The other way is that the images are shown individually in randomized order, as shown in Figure (6.1c), which helps disassociate the cardinal estimates from the ordinal estimates, called *Setting B* throughout the paper. Comparing two different user interfaces is intended to test whether one of the two interfaces is more convenient for participants, and whether one leads to more accurate collective estimates than the other.



In this activity you will rank the pictures in order from **least** to **greatest** number of dots. Click on each individual dot picture to place it in your answer.

(a) Interface for Ordinal Estimation

Enter in an estimate for how many dots you think are in this image.

*Images are displayed in the order in which **you** ranked them*

(b) Interface for Cardinal Estimation in Setting A

Enter in an estimate for how many dots you think are in this image.

Note: Images are not displayed in the order in which they were ranked.

(c) Interface for Cardinal Estimation in Setting B

Figure 6.1: User Interface for Estimation Tasks

There are four problems of varying sizes, specifically 2-image, 3-image, 5-image, and 6-image problems (the number indicates the size of the subset of images seen by each participant in the ordinal estimation task), each with its own data set of 30 images ranging from 50 to 79 dots. Each participant only evaluates a subset of the 30 images. If a participant does not complete the task correctly, he or she is prompted to try the same question again. Participants are able to modify their responses before submitting them in case they make a mistake or change their minds. The experiment was deployed on Amazon Mechanical Turk (Amazon MTurk), which is a crowdsourcing platform that provides a vast and diverse pool of workers. Hence, it is often considered a representative of the population at large (Paolacci *et al.*, 2010; Berinsky *et al.*, 2012; Stewart *et al.*, 2015; Chandler and Shapiro, 2016; Mortensen and Hughes, 2018); importantly, this characteristic helps to satisfy the second required condition (diversity) to achieve crowd wisdom.

6.4.2 Participants

A total of 600 participants (300 for each setting) were recruited through Amazon MTurk. To obtain a diverse, yet reliable estimate, the experiment set a qualification for the participants. Specifically, the participants should be located in the United States and have a previous task approval Human Intelligence Tasks (HIT) rate of at least 90%. Participants who completed the experiment were paid $1.00 for approximately five minutes of work. At the end of the study, participants were asked to fill out a demographic survey. The demographic survey was completed by 273 of the 300 participants for setting A and by 286 of the 300 participants for setting B; the detailed information for each setting is provided in Appendix A.

6.4.3 Data collection and processing

In each of the four problems of the experiment, there are 30 images to be evaluated. However, each participant only evaluates a subset of the 30 images, according to the problem size—2, 3, 5, and 6 images respectively. To ensure that all 30 images of each problem are seen by the same number of people and that each individual evaluates a different random subset of the images, we developed a special task allocation scheme, described as follows. First, the 30 images are randomly permuted and partitioned into subsets according to problem size. For example, in the 6-image problem, the 30 images are partitioned into 5 subsets, each containing six images (i.e., 30/5=6).

Similarly, a permutation of the 30 images in the 2, 3, and 5-image problems result in partitions with 15, 10 and 6 subsets, respectively. Each subset of images is seen by one person. In other words, it takes 15, 10, 6, and 5 participants for all 30 images to be seen exactly once in the 2, 3, 5 and 6-image problems, respectively. After all 30 images are allocated among the respective number of participants, a new permutation of the images is generated and the process is repeated. Applying this allocation mechanism over 300 participants yields that each image is seen 20 (=300/15), 30 (=300/10), 50 (=300/6), and 60 (=300/5) times in the 2, 3, 5, and 6-image problems, respectively. The permutation and allocation can be also mathematically written as follows. Let $\boldsymbol{\Pi} = \{\boldsymbol{\pi^1}, \boldsymbol{\pi^2}, ..., \boldsymbol{\pi^{300/(30/p)}}\}$ be the set of permutations of 30 images and $\boldsymbol{\pi^i} = \{\pi_1^i, \pi_2^i, ...\pi_{30}^i\}$ be the ith permutation, for $i = 1, .., 300/(30/x)$, where p is the problem size. In this case, the set of permutations is $\boldsymbol{\Pi} = \{\boldsymbol{\pi^1}, \boldsymbol{\pi^2}, ..., \boldsymbol{\pi^{50}}\}$ and its last permutation is $\boldsymbol{\pi^{50}} = \{\pi_1^{50}, \pi_2^{50}, ...\pi_{30}^{50}\}$ for 5-image problem. To give a better understanding of the task allocation scheme, Figure 6.2 shows a visual description for task allocation scheme in the 5-image problem.

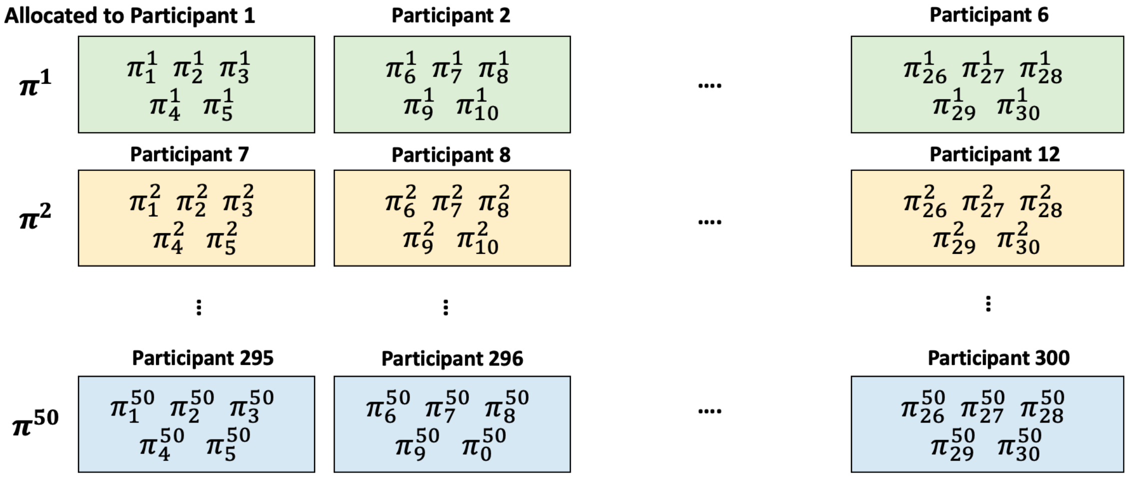

Allocated to Participant 1
Participant 2
Participant 6
$\boldsymbol{\pi^1}$
$\pi_1^1\ \pi_2^1\ \pi_3^1\ \pi_4^1\ \pi_5^1$
$\pi_6^1\ \pi_7^1\ \pi_8^1\ \pi_9^1\ \pi_{10}^1$
....
$\pi_{26}^1\ \pi_{27}^1\ \pi_{28}^1\ \pi_{29}^1\ \pi_{30}^1$
Participant 7
Participant 8
Participant 12
$\boldsymbol{\pi^2}$
$\pi_1^2\ \pi_2^2\ \pi_3^2\ \pi_4^2\ \pi_5^2$
$\pi_6^2\ \pi_7^2\ \pi_8^2\ \pi_9^2\ \pi_{10}^2$
....
$\pi_{26}^2\ \pi_{27}^2\ \pi_{28}^2\ \pi_{29}^2\ \pi_{30}^2$
Participant 295
Participant 296
Participant 300
$\boldsymbol{\pi^{50}}$
$\pi_1^{50}\ \pi_2^{50}\ \pi_3^{50}\ \pi_4^{50}\ \pi_5^{50}$
$\pi_6^{50}\ \pi_7^{50}\ \pi_8^{50}\ \pi_9^{50}\ \pi_0^{50}$
....
$\pi_{26}^{50}\ \pi_{27}^{50}\ \pi_{28}^{50}\ \pi_{29}^{50}\ \pi_{30}^{50}$

Figure 6.2: Depiction of Task Allocation among the 300 Participants for the 5-image Problem

We characterize the goodness of each participant-to-image allocation based on the maximum *number of hops* (Hochbaum and Levin, 2006a) in the pairwise comparison graph. The pairwise comparison graph $G = (V, E)$ is undirected, with each node $i \in V$ representing one of the alternatives (i.e., images) and each edge $(i, j) \in E$ representing that the pair of alternatives was evaluated by at least one participant. The number of hops between $i, j \in V$ is the length of the shortest path between the two nodes, and the maximum number of hops in G is the longest among all of the shortest paths. As an example of these concepts, consider four images i, j, k, and ℓ, and assume that images i and j are evaluated by one participant, j and k are evaluated by a second participant, k and ℓ are evaluated by a third participant, and no one evaluates pairs i

			Number of times each image was seen											
Problem size	Statistics	Setting	5	10	15	20	25	30	35	40	45	50	55	60
2-image	avg	A	2.31	1.74	1.60	1.50								
		B	2.29	1.75	1.60	1.51								
	max	A	4.00	3.00	2.25	2.00								
		B	4.00	3.00	2.75	2.00								
3-image	avg	A	1.74	1.49	1.35	1.24	1.17	1.12						
		B	1.74	1.49	1.33	1.23	1.16	1.11						
	max	A	3.00	2.00	2.00	2.00	2.00	2.00						
		B	3.00	2.00	2.00	2.00	2.00	2.00						
5-image	avg	A	1.49	1.24	1.12	1.06	1.03	1.01	1.01	1.00	1.00	1.00		
		B	1.48	1.23	1.12	1.06	1.03	1.01	1.01	1.00	1.00	1.00		
	max	A	2.00	2.00	2.00	2.00	2.00	2.00	1.92	1.64	1.30	1.00		
		B	2.00	2.00	2.00	2.00	2.00	2.00	1.93	1.71	1.40	1.00		
6-image	avg	A	1.41	1.16	1.07	1.03	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		B	1.39	1.15	1.06	1.02	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	max	A	2.00	2.00	2.00	2.00	1.99	1.88	1.62	1.40	1.26	1.17	1.08	1.00
		B	2.00	2.00	2.00	2.00	1.97	1.71	1.34	1.14	1.05	1.02	1.00	1.00

Table 6.1: Average Number of Hops in Pairwise Comparison Graphs Calculated for Different Subsets of the Data

and k, j and ℓ, and i and ℓ. Then, the number of hops between i and j, j and k, and k and ℓ is 1. Additionally, the number of hops between i and k and j and ℓ is 2, and the number of hops between i and ℓ is 3. To maintain the robustness and reliability of the relative comparisons, it is recommended to have a maximum of 2 hops and no more than 3 hops (Hochbaum and Levin, 2006a).

Table 6.1 shows the average and maximum number of hops for different subsets of the datasets in the experiments (note that the minimum number of hops is always 1); the hops statistics for setting A and setting B are virtually identical. As expected, the higher problem sizes have lower hops statistics (since more images are assigned to participants).

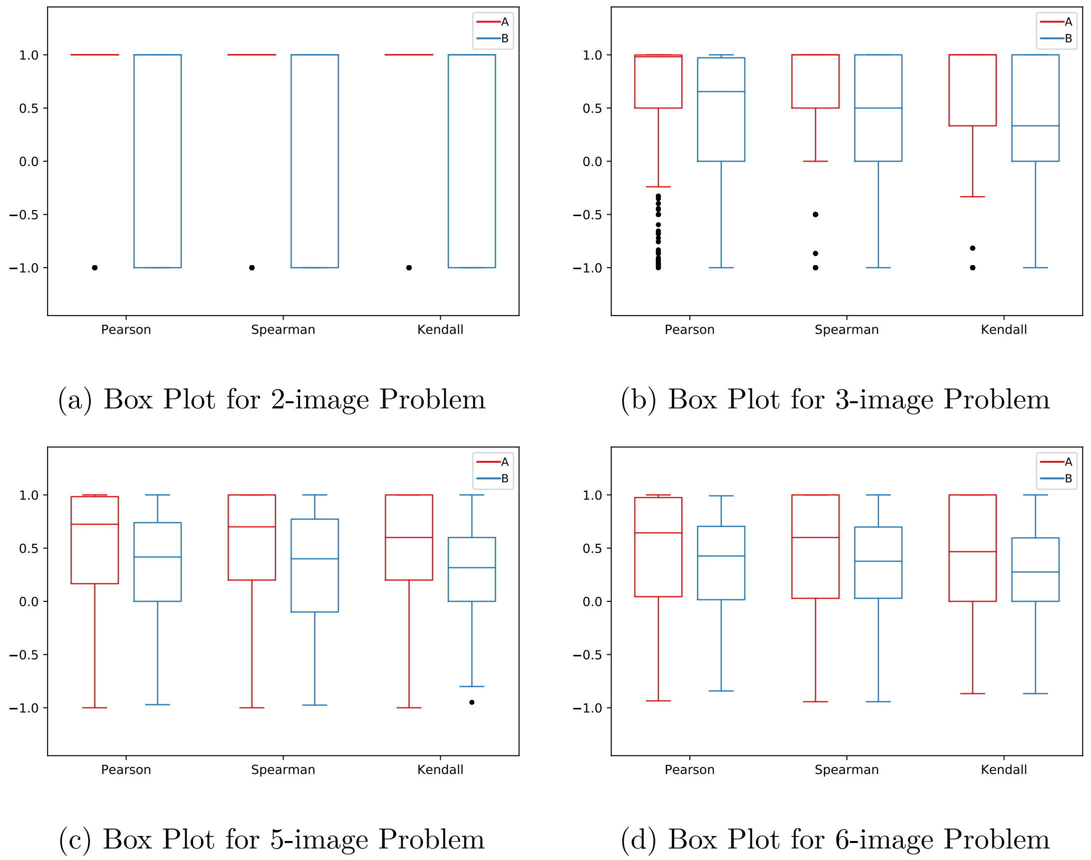
6.5 Study 1. A/B Testing for Multimodal Information Elicitation

As is described in Section 6.4, the experiment asks participants to provide ordinal estimates (i.e., the order of the subset of images) and cardinal estimates (i.e., the number of dots in images). All participants provide ordinal estimates in the same manner; however, participants provide cardinal estimates in one of two possible ways: half of the participants provide estimates over one image at a time, shown in a randomized order (not according to the self-provided ordinal estimate); and the other half of the participants provides estimates over all images shown at the same time, based on self-provided estimates in the ordinal estimation task (from the least to the greatest number of dots). In effect, this experimental setting serves as an *A/B test* (Kohavi and Longbotham, 2017) regarding the impact of two participant input elicitation choices. A/B testing is popularly used for e-commerce, mobile application, and website optimization (Kaufmann *et al.*, 2014; Huang *et al.*, 2019). It seeks to compare two user interfaces to determine via a two-sample hypothesis testing which is more effective in terms of user engagement, satisfaction, or another metric of interest. In this study, we seek to determine if eliciting subsequent estimates based on previous self-provided estimates is convenient for decision-makers and if it helps achieving better collective estimates. To that end, we compare how the cardinal-input elicitation choice affects estimation accuracy and how long it takes for participants to complete each task.

6.5.1 Correlation between ordinal and cardinal estimates

Before evaluating the accuracy of collective estimates, we discuss how the individual cardinal estimates (specifically, ordinal estimates induced from the cardinal estimates) are discrepant from their corresponding ordinal estimates. To do so, the association between the cardinal and ordinal estimates is calculated using the three different correlation coefficients: original Kendall-τ, Spearmans' ρ, and Pearson. These three correlation coefficients have been frequently used in the literature (e.g., see Russell and Gray (1994); Bonett and Wright (2000); Bolboaca and Jäntschi (2006)), and their domains are between -1 and 1.

As shown in Figure 6.3, the cardinal and ordinal estimates are more strongly correlated (although not perfectly) in Setting A and weakly correlated in Setting B. The weak correlation observed in Setting B can be explained by the higher propensity for self-contradictions between the two types of information provided by each participant. Even though the instructions indicated that images are displayed in the order in which the participants ranked them from those perceived to have the fewest to the most number of dots (see Figure (6.1c)), contradictions in Setting A were also observed, but they were significantly less frequent (we conjecture the majority of these may be due to user error). Statistical analysis revealed that the mean of correlation between the input ordinal and cardinal estimates of the two settings was significantly different, with a p-value < 0.001. An additional insight from this analysis is that, as



(a) Box Plot for 2-image Problem

(b) Box Plot for 3-image Problem

(c) Box Plot for 5-image Problem

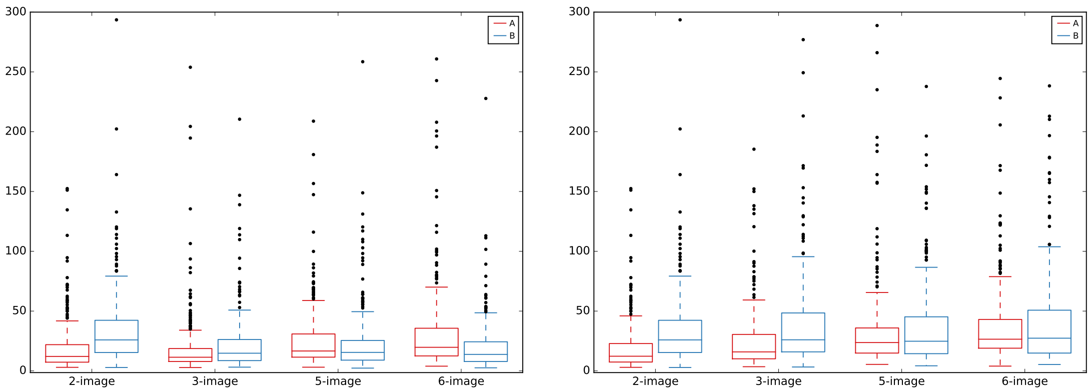
(d) Box Plot for 6-image Problem

Figure 6.3: Correlation Between Ordinal and Cardinal Estimates for Each Setting

problem size increases, cardinal estimates and the ordinal estimates become less correlated. This could be due to an increase in the cognitive load and a higher possibility for self-contradiction when more images are evaluated.

6.5.2 Task completion time

In addition to checking if different settings create a discrepancy between the individual cardinal and ordinal estimates, we evaluate if one setting is more convenient than the other for the participants. To test this, we compare the task completion time for two different settings, under the assumption that the shorter the task completion time is, the more convenient the setting is.



(a) Ordinal Estimation

(b) Cardinal Estimation

Figure 6.4: Task Completion Time for Each Setting

The two-sample hypothesis testing revealed that the means of the cardinal estimation task completion time for each setting are not statistically significantly different with p-values of 0.24 and 0.39 for the 3-image and 6-image problems, respectively (p-value for the 2-image and 5-image problem are less than 0.001 and 0.03, respectively). Therefore, we cannot conclude that eliciting cardinal estimates on top of self-provided ordinal estimates is relatively more convenient for participants. Similarly, we cannot determine that one of the settings is more convenient than the other in the ordinal estimation task because the computing time for one setting is not consistently better than the other based on the prior observation.

Moreover, the completion times of cardinal estimation and ordinal estimation are statistically different for Setting A and B, except for the 2-image problem with p-values less than 0.05 for the 3-, 5-, and 6-image problems). Specifically, cardinal estimation generally takes longer than ordinal estimation. This result implies that providing the order of images is more convenient than providing the estimated number of dots in images for participants, which is possibly because the ordinal estimation only requires relative comparison between the images, as opposed to the absolute and exact estimates.

6.5.3 Accuracy of collective estimates

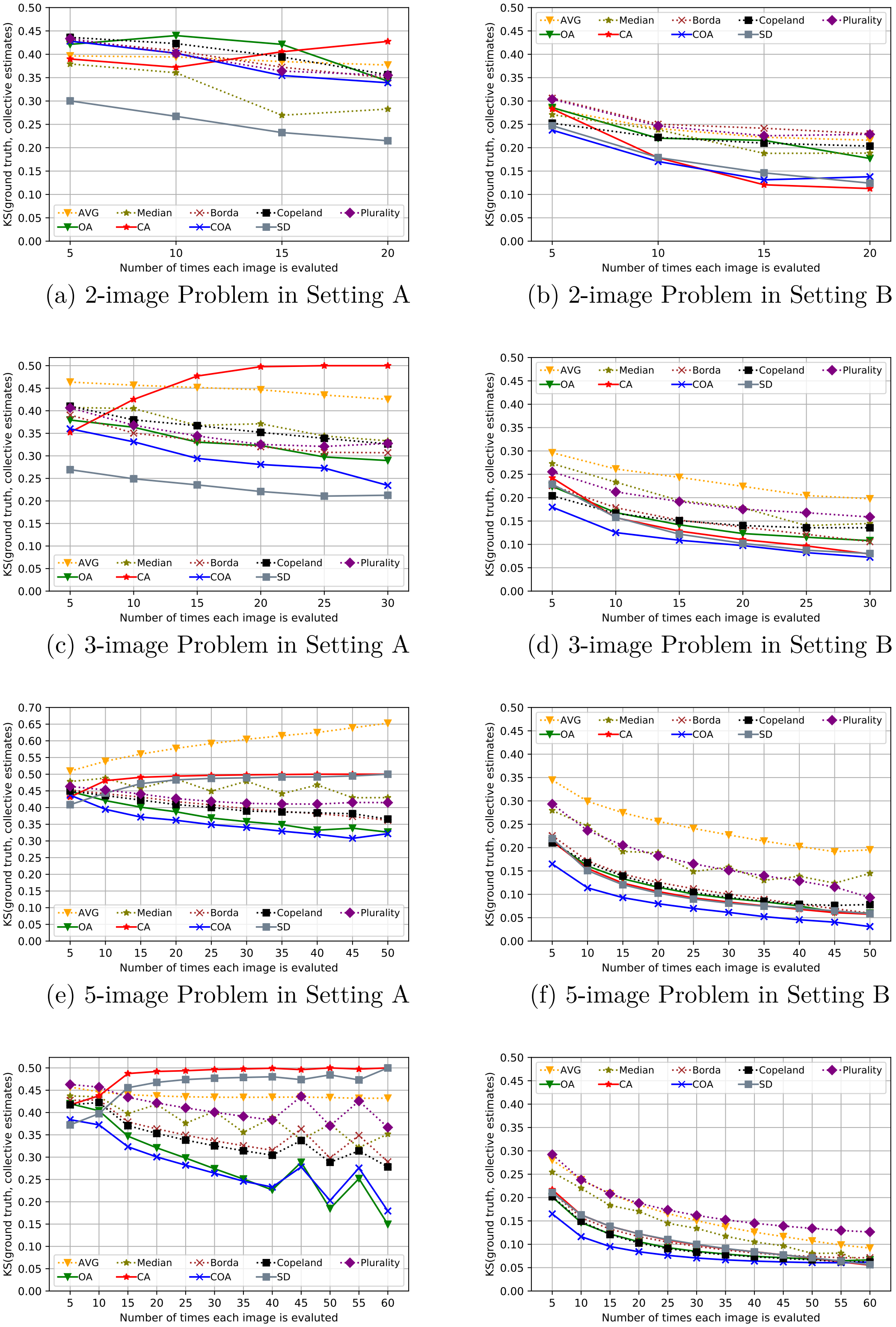
To evaluate the accuracy of collective estimates, the distance from the collective estimates to the ground truth is calculated; specifically, the distance between the collective ordinal estimate and the ground truth is quantified via d_{KS} (see Equation (2.3)) and the distance between the collective cardinal estimate and the ground truth is quantified via the Euclidean distance normalized by the range of the cardinal estimates (i.e., maximum possible cardinal estimate - minimum possible cardinal estimate).

Moreover, in the COA model, the weights from the cumulative ranking distance (i.e., the sum involving d_{NPKS} in Equation (2.5)) and from the cumulative rating distance (i.e., sum involving d_{NPCK} in Equation (2.6)) are set equal to each other (i.e., symmetrically). Similarly, in the SD model, the weights from the separation and from the deviation are set equal to each other (varying the priority weights will be discussed in Section 6.6). We remark that the rating-based aggregation models and the average and median methods only return a collective cardinal estimate $\boldsymbol{r}$. The collective ordinal estimate is induced by ordering the values of $\boldsymbol{r}$ in non-increasing order.

The experiments were performed on machines equipped with 36GB of RAM memory shared by two Intel Xeon E5-2680 processors running at 2.40 GHz; code was written in Python and the optimization-based models were solved using CPLEX solver version 12.8.0. A time limit was set to 10 minutes for solving the optimization-based aggregation models to return the collective estimates, and the optimality gap was recorded (see Table B.1 and B.2 in Appendix B).

Ordinal Estimation

By comparing the left column and right column of Figure 6.5, it is evident that more accurate collective ordinal estimates are attained in Setting B, i.e., when a possible anchoring effect is attenuated by eliciting ordinal and cardinal estimates independently. Additionally, optimization-based aggregation models generally outperform traditional aggregation/voting rules in this context. As shown in the left column of Figure 6.5, COA provides a more accurate collective ordinal estimate than the other models when individual ordinal estimates and cardinal estimates are independently elicited. As shown in the right column of Figure 6.5, SD provides a more accurate collective ordinal estimate than the other models when the individual cardinal estimates are elicited dependent on the individual ordinal estimates. Both results indicate that multimodal aggregation outperforms unimodal aggregation in deriving better collective estimates, especially when fewer people are available to perform the crowdsourcing task. This observation highlights the practical benefits of multimodal aggregation, as recruiting enough participants to attain the benefits of crowd wisdom with traditional methods can come at a high cost. Furthermore, as shown in the left column of Figure 6.5, the distance between the ground truth and the collective ordinal estimates decreases as the number of participants increases; that is, the accuracy of the collective ordinal estimates improves as more people are included, which reflects the concept of the wisdom of crowds. However, when the cardinal estimates are collected on top of the self-provided ordinal estimates, some models return collective estimates that do not align with the concept of the wisdom of crowds. These include model CA, the averaging method, and model SD for the 5-image and 6-image problems.



(a) 2-image Problem in Setting A

(b) 2-image Problem in Setting B

(c) 3-image Problem in Setting A

(d) 3-image Problem in Setting B

(e) 5-image Problem in Setting A

(f) 5-image Problem in Setting B

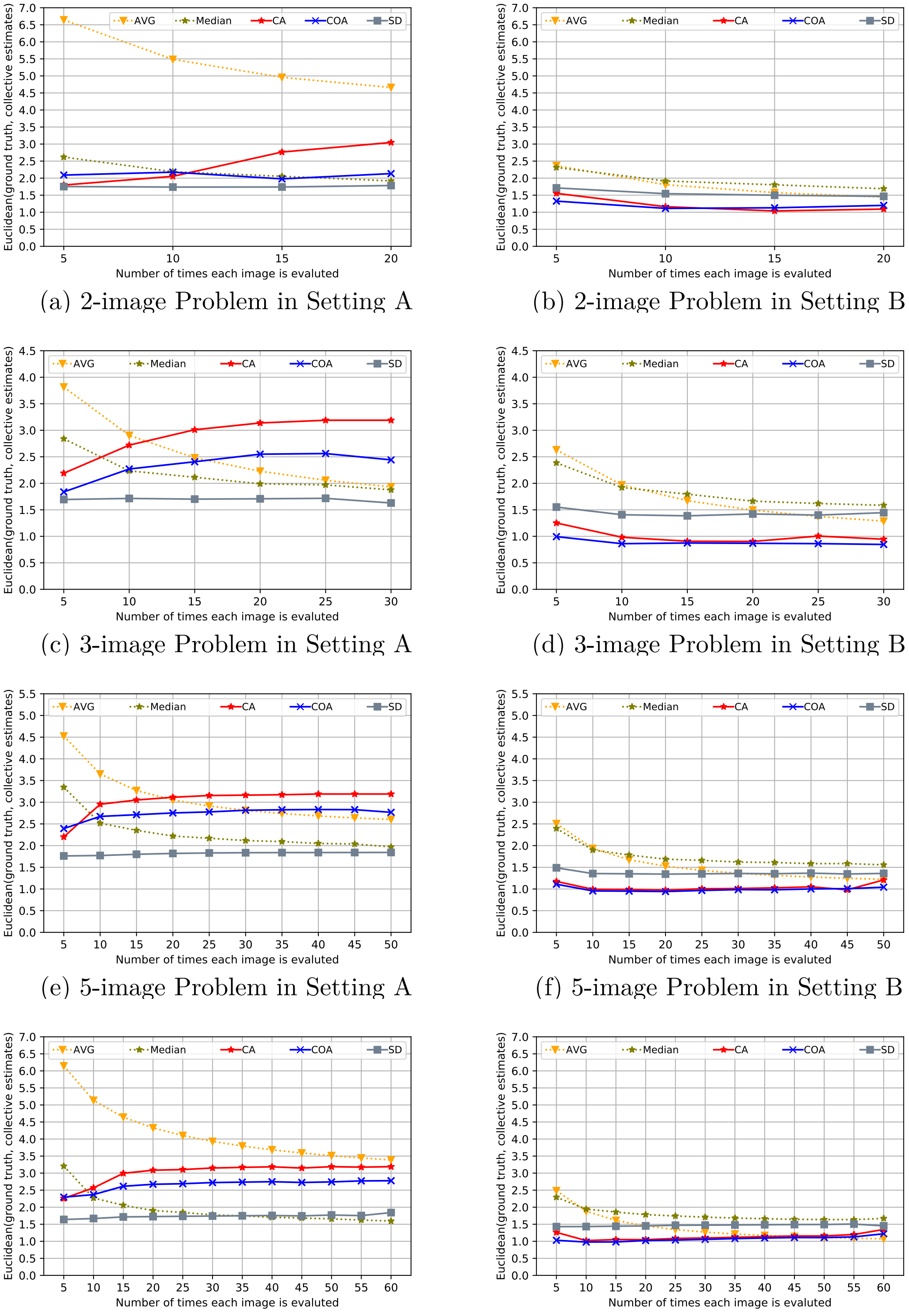
(g) 6-image Problem in Setting A

(h) 6-image Problem in Setting B

Figure 6.5: Accuracy of Collective Ordinal Estimates

Cardinal Estimation

Similar to the ordinal estimation task, more accurate collective cardinal estimates are obtained in Setting B, where both multimodal aggregation models COA and SD outperform all other aggregation techniques. However, a key difference from the ordinal estimation task in this setting is that the size of the problem does not appear to affect the accuracy of cardinal estimation. This seems to support that the elicitation of cardinal estimates on one image at a time in Setting B helps participants to devote similar efforts on each of these tasks. When cardinal estimates are elicited on top of the self-provided ordinal estimates in Setting A so as to reduce cognitive load, the collective estimates seem to improve less markedly as problem size increases. In fact, unlike the average and median method results, the optimization model estimates often degrade when more participants are added in this setting (e.g., see Figures (6.6a), (6.6c), (6.6e), and (6.6g)). One likely reason for this counterintuitive result is a compounding effect of ordinal estimation errors on cardinal estimation, which effectively makes the multimodal aggregation models vulnerable to the anchoring effect. This will be further evaluated in the ensuing section when the ordinal and cardinal inputs are given different priority weights in these models. We note that all aggregation models and traditional/voting rules were solved within the time limit, except for COA. Although the COA model often reached the time limit before the optimal solution could be found, its suboptimal collective estimates were often closer to the ground truth than the optimal collective estimates returned by the other methods.



(a) 2-image Problem in Setting A

(b) 2-image Problem in Setting B

(c) 3-image Problem in Setting A

(d) 3-image Problem in Setting B

(e) 5-image Problem in Setting A

(f) 5-image Problem in Setting B

(g) 6-image Problem in Setting A

(h) 6-image Problem in Setting B

Figure 6.6: Accuracy of Collective Cardinal Estimates

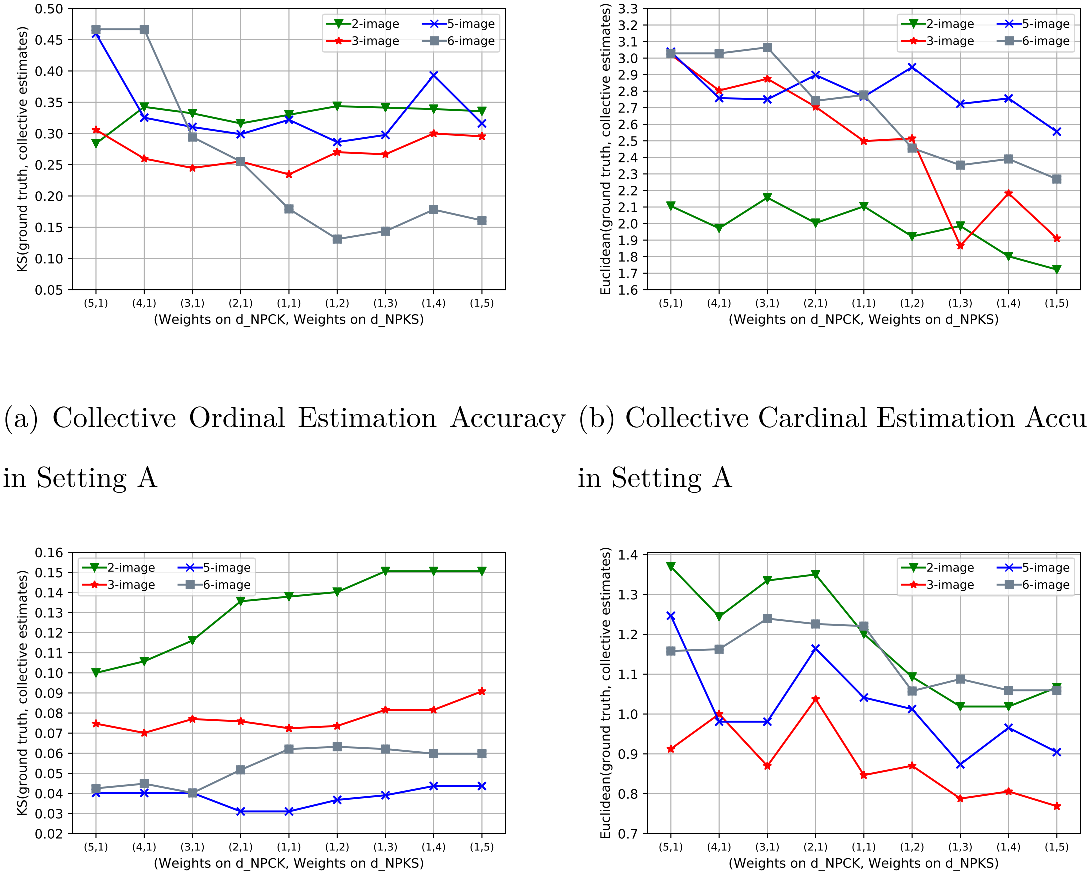
In conclusion, the experiment offers more insight about how multimodal information can enhance the wisdom of crowd effect, but also how it may cause anchoring effects and negatively impact the collective estimate quality, which confirms Hypothesis 1 and 3.

6.6 Study 2. Weighted Multimodal Aggregation

As shown in the previous section, having both cardinal and ordinal estimates helps attain better collective estimates compared to using only one input modality. This is because ordinal and cardinal information can be used in complementary fashion. Similarly, considering both separation and deviation components can improve collective estimation accuracy. Extending from this outcome, we test whether assigning different priority weights on multimodal information affects the collective estimation. To do so, first, we assign priority to the cardinal and ordinal estimates in COA model by varying the weights on d_{NPCK} and d_{NPKS}, denoted as λ_a and λ_b, respectively (see Equation (6.2)). The left and right coordinates in the x-axis represent the weight of d_{NPCK} and d_{NPKS}, respectively (e.g., (5,1) indicates λ_a=5 and λ_b=1 in COA).

The first two figures in Figure 6.7 demonstrate that, when cardinal and ordinal estimates are associated, having more weight on ordinal estimates leads to more accurate collective estimates. This result matches our elicitation approach in Setting A. When cardinal estimates are elicited on top of the self-provided ordinal estimates, ordinal estimates play an important role. This outcome ultimately implies that the initial decision largely affects the subsequent decision (i.e., anchoring), which seems to confirm Hypothesis 3.

Contrary to the first two figures, the last two figures in Figure 6.7 demonstrate that the collective cardinal estimates tend to improve when a higher relative weight is assigned to the ordinal inputs; for example, the collective estimates from the 3-image



(a) Collective Ordinal Estimation Accuracy in Setting A

(b) Collective Cardinal Estimation Accuracy in Setting A

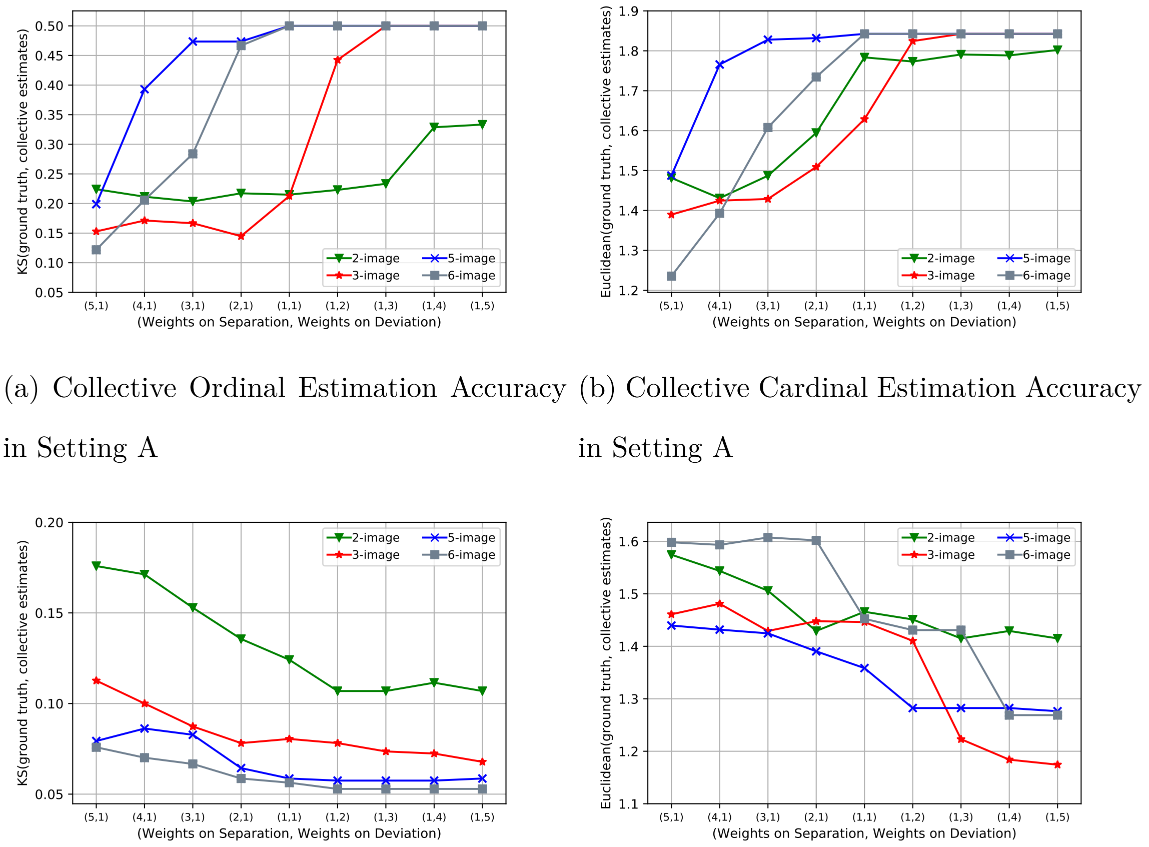
(c) Collective Ordinal Estimation Accuracy in Setting B

(d) Collective Cardinal Estimation Accuracy in Setting B

Figure 6.7: Accuracy of Collective Estimates from the Cardinal and Ordinal Aggregation Model Obtained from Changing Priority Weights of the Input Modalities

cardinal estimation problem with weights (1,2) are closer to the ground truth than the collective estimates from the problems with weights (2,1). Similarly, the collective estimates from the 5-image cardinal estimation problem with weights (1,5) are closer to the ground truth than the collective estimates from the problems with weights (5,1). These results justify that multimodal information can be more intelligently used to achieve even better collective estimates. However, there is no common set of weights that returned the best collective estimates for all 2, 3, 5, and 6-image problems.

Moreover, SD is another multimodal aggregation model; it considers not only the difference of intensities between two alternatives in the aggregated outcome and in each participant's evaluation, but also the difference between the point estimate of alternative in the aggregated outcome and in each participant's evaluation. We herein assign asymmetric priority weights to the separation and deviation elements.



(a) Collective Ordinal Estimation Accuracy in Setting A

(b) Collective Cardinal Estimation Accuracy in Setting A

(c) Collective Ordinal Estimation Accuracy in Setting B

(d) Collective Cardinal Estimation Accuracy in Setting B

Figure 6.8: Accuracy of Collective Estimates from the Separation-deviation Model According to the Priority Weights on Separation and Deviation

The first two figures show that providing a higher priority weight on separation information returns better collective estimates when the cardinal estimates are elicited on top of the self-provided ordinal estimates. However, the bottom two figures in Figure 6.8 demonstrate that, when the cardinal and ordinal estimates are disassociated in elicitation, providing a higher priority weight on deviation information returns better collective estimates.

6.7 General Discussion

6.7.1 Main contributions

This research makes various contributions to the broad understanding of multimodal information aggregation in crowdsourcing. First, it empirically shows that the choices of elicitation methods and their implementation within the participant interface affect the quality of elicited opinions. When participants are asked to provide their cardinal estimates disassociated from their ordinal estimates, it was found that individual cardinal estimates (more specifically, the induced ranking from them) and ordinal estimates have a higher tendency to be contradictory than when the participants are asked to provide cardinal estimates on top of their self-provided ordinal estimates. This was quantified by comparing the two respective correlations. This indicates that elicitation methods can affect individual decision-making. Additionally, although cardinal and ordinal estimates had a higher tendency to be contradictory when they were disassociated, the collective estimates were more accurate than the collective estimates from the other elicitation interface, which had lower self-contradictions. Our research explains this via the anchoring effect; to the best of our knowledge, this is the first research that connects multimodal aggregation in crowdsourcing with this cognitive bias. Once the anchor (ordinal estimate) has poor quality, the subsequent decision (cardinal estimate) is likely to be poor as well. Because anchoring caused error to propagate from one elicited modality to another, this yielded a poor collective estimation, and it ultimately counteracted the benefit of the wisdom of crowds. The diminished crowd wisdom could also connect to a violation of the required conditions for the wisdom of crowds. There are four conditions to have the wisdom of crowds: independence, diversity, decentralization, and aggregation. When the cardinal estimates are disassociated from the ordinal estimates, the wisdom of crowds prevails. However, when cardinal estimates are associated with the ordinal estimates, the cardinal estimates may no longer be considered independent, which diminishes the wisdom of crowds.

Furthermore, we empirically justify that assigning asymmetric priority weights to the different information elements (e.g., cardinal-ordinal, separation-deviation) can improve the quality of collective estimates. While Kemmer *et al.* (2020) assign equal weights to ordinal and cardinal estimates in the COA model and equal weights on separation and deviation information to the SD model, the new experiment tries to impose asymmetric weights to each modality of estimates to test whether this improves or deteriorates the accuracy of collective estimates. We found that having more weight on ordinal estimates returns a better collective cardinal estimate in the cardinal and ordinal aggregation model and having more weight on the deviation component returns a better collective estimate in the separation-deviation aggregation model. These results support the importance and the effectiveness of mitigating potential anchoring effects from multimodal estimate elicitation; specifically, the weighted cardinal and ordinal aggregation shows that the two types of modalities can be used in complementary fashion to obtain a better collective estimate. Lastly, the optimal priority weights for each input modality depends on the type of estimation activities (e.g., ordinal estimation vs. cardinal estimation).

6.7.2 Practical implications

Our findings have practical and essential implications for decision-makers who handle ranking and rating information. First, we demonstrate that multimodal input elicitation can improve the quality of recommendation systems and review systems. Specifically, recommendation systems are designed to incorporate user-reported ratings and provide a list of suggested items (e.g., movies, restaurants). However, user-reported ratings can be biased or inconsistent due to the subjective scales or cognitive bias (e.g., anchoring bias), which may provide a distorted view of user preferences and ultimately contaminate inputs of recommendation systems, leading to decreased quality of future recommendations and negatively influencing consumers' decision (Adomavicius *et al.*, 2013). Therefore, eliciting both ranking and rating inputs enable the collection of richer opinions and their aggregation could return more robust and increased quality of recommendations.

Moreover, our research evinces that having multimodal information allows efficient collective decision-making in terms of time and budget. Specifically, the results of our experiment show that aggregating multimodal information yields a better collective decision when recruiting people is not feasible. Consumers recently participate in product development and marketing, which is considered a type of crowdsourcing (Djelassi and Decoopman, 2013; Huang *et al.*, 2014). However, collecting information from consumers requires time and cost; therefore, obtaining a better decision from fewer people is often considered an important practical benefit. Therefore, eliciting both cardinal and ordinal inputs could allow for more effective decisions with fewer users recruited, thereby providing benefits in terms of time and cost.

Lastly, our experiment provides practical guidance for decision-makers to determine which aggregation methods should be used. Specifically, optimization-based aggregation models are highly recommended when the quality of decisions is prioritized over computing time. Although simple average and median can return a collective decision in a shorter amount of time, the quality of the collective decision is inferior to that of collective decision from multimodal optimization-based aggregation models. Speed and accuracy are often in opposition in collective decision-making and their trade-offs are considered important (Franks *et al.*, 2003). Some applications prioritize accuracy over speed, such as image annotation for disease diagnosis (Irshad *et al.*, 2014). However, other applications prioritize speed over accuracy, where real-time decision-making is required, such as real-time transportation availability. Based on the context of decision-making, aggregation models can be selected appropriately.

Chapter 7

GENERAL DISCUSSION AND CONCLUSIONS

Rank aggregation is widely used in group decision-making and many other applications where it is of interest to consolidate heterogeneous ordered lists. As society becomes more connected and technologically advanced, individual opinions take different forms of expressions and people need to make decisions based on various factors. This brings extra challenges in computational social choice and limits the applicability of the existing aggregation frameworks. This dissertation introduces robust mathematical frameworks that resolve the existing challenges. In particular, it develops the correlation coefficient that can be applied to various ranking formats, including non-strict and incomplete rankings. The correlation coefficient is designed to enforce a neutral treatment of incompleteness whereby no assumptions are made about individual preferences involving unranked objects. Also, it satisfies key social choice properties that have been shown to engender improved decisions.

Moreover, this work also introduces a binary programming formulation for aggregating various types of rankings, including non-strict and incomplete rankings. The binary programming formulation leverages the equivalence of two ranking aggregation problems, namely that of minimizing the Kemeny-Snell distance and of maximizing the Kendall-tau correlation, to compare the newly introduced binary programming formulation to a modified version of an existing integer programming formulation associated with the Kendall-tau distance. The new formulation has fewer variables and constraints, which leads to faster solution times, and it also has a special connection with the weak-order polytope.

Additionally, to further expedite the solution process, this work develops a new social choice property, the Non-strict Extended Condorcet Criterion, which can be regarded as a natural extension of the well-known Condorcet criterion and the Extended Condorcet criterion. Unlike its parent properties, the new property is adequate for handling complete rankings with ties. This property allows us to develop a structural decomposition algorithm that can solve large instances of the NP-hard Kemeny rank aggregation problem exactly within a practical amount of time. Its practicality is formally tested using the instances constructed from a probabilistic distribution and benchmark instances from a library of preference data.

Branching out from the theoretical computational social choice research, this work applies the principled aggregation frameworks to the context of crowdsourcing. Specifically, the crowdsourcing experiment demonstrates how the quality and efficiency of crowdsourced collective estimates can be improved by aggregating multiple modalities of input and how multimodal aggregation models can mitigate anchoring effects.

This dissertation opens a great possibility for future research. Future studies will develop the social choice property that can be applied to non-strict incomplete rankings. Specifically, the non-strict extended Condorcet criterion is only applicable to complete rankings with ties. For example, when rankings include too much incomplete information (i.e, many pairs of alternatives are not compared), it is not reasonable to use the current definition of a decisive majority since, for it to be useful, it would require one same preference relation to be made by more than half of judges for each pair of items. If, for example, only two of ten judges evaluate two specific alternatives and their preferences agree on these two alternatives, it is hard to say that there exists a decisive majority because more than three-fourths of judges do not evaluate those two alternatives. Thus, the extension of the non-strict extended Condorcet criterion can yield a practical contribution to the field of computational social choice.

Moreover, future work will involve more computational experiments to enhance the computational speed of Kemeny aggregation, parallel programming and/or valid inequalities (e.g., Doignon and Fiorini (2001); Escobedo and Yasmin (2021)) will be further explored.

Furthermore, future crowdsourcing research will use a variety of methods to cope with unreliable crowdsourcing workers. Although we have used the ad-hoc elimination criteria to remove some outlying crowdsourcing workers according to its responses, it is more analytical to filter out spammers in a more systematic way. Honeypot traps are a popular method to rule out spammers. They work by placing an undemanding question randomly (workers do not know which question is a honeypot trap) and eliminating those workers who did not answer the question correctly. Using such a method allows mitigating the occurring errors from insincere workers and ultimately obtaining a better crowdsourcing outcome (Lee *et al.*, 2010; Chittilappilly *et al.*, 2016; Mortensen *et al.*, 2017).

Finally, future crowdsourcing tasks involve user-generated subjective data, which does not have a ground truth and is more subjective than a typical wisdom of crowds task such as estimating the number of dots. Because business applications often deal with subjective data (e.g., consumer preferences), expanding the scope of the crowdsourcing task into these and other similar contexts could unveil additional practical applications of our research.

REFERENCES

Adomavicius, G., J. C. Bockstedt, S. P. Curley and J. Zhang, “Do recommender systems manipulate consumer preferences? a study of anchoring effects”, Information Systems Research **24**, 4, 956–975 (2013). 6.2.2, 6.7.2

Ahlgren, P., B. Jarneving and R. Rousseau, “Requirements for a cocitation similarity measure, with special reference to pearson’s correlation coefficient”, Journal of the Association for Information Science and Technology **54**, 6, 550–560 (2003). 2.3

Ailon, N., “Aggregation of partial rankings, p-ratings and top-m lists”, Algorithmica **57**, 2, 284–300 (2010). 1.1

Ailon, N., M. Charikar and A. Newman, “Aggregating inconsistent information: ranking and clustering”, Journal of the ACM (JACM) **55**, 5, 23 (2008). 1.1

Ali, A. and M. Meilă, “Experiments with kemeny ranking: What works when?”, Mathematical Social Sciences **64**, 1, 28–40 (2012). 1.1

Ammar, A. and D. Shah, “Ranking: Compare, don’t score”, in “2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)”, pp. 776–783 (IEEE, 2011). 1.2

Ammar, A. and D. Shah, “Efficient rank aggregation using partial data”, ACM SIGMETRICS Performance Evaluation Review **40**, 1, 355–366 (2012). 1.2, 6.2.1

Amodio, S., A. DAmbrosio and R. Siciliano, “Accurate algorithms for identifying the median ranking when dealing with weak and partial rankings under the kemeny axiomatic approach”, European Journal of Operational Research **249**, 2, 667–676 (2016). 1.1

Aral, S., “The problem with online ratings”, MIT Sloan Management Review **55**, 2, 47 (2014). 6.1

Arrow, K. J. *et al.*, “Social choice and individual values”, (1951). 1

Asfaw, D., V. Vitelli, Ø. Sørensen, E. Arjas and A. Frigessi, “Time-varying rankings with the bayesian mallows model”, Stat **6**, 1, 14–30 (2017). 4.4.1

Atanasov, P., P. Rescober, E. Stone, S. A. Swift, E. Servan-Schreiber, P. Tetlock, L. Ungar and B. Mellers, “Distilling the wisdom of crowds: Prediction markets vs. prediction polls”, Management science **63**, 3, 691–706 (2017). 6.1

Au, R. K. and K. Watanabe, “Numerosity underestimation with item similarity in dynamic visual display”, Journal of vision **13**, 8, 5–5 (2013). 1.3

Bartholdi, J., C. A. Tovey and M. A. Trick, “Voting schemes for which it can be difficult to tell who won the election”, Social Choice and Welfare **6**, 157–165 (1989). 1.1

Basili, M. and S. Vannucci, “Choice overload and height ranking of menus in partially-ordered sets”, Entropy **17**, 11, 7584–7595 (2015). 3.2

Becker-Peth, M., E. Katok and U. W. Thonemann, “Designing buyback contracts for irrational but predictable newsvendors”, Management Science **59**, 8, 1800–1816 (2013). 6.2.2

Benzion, U., Y. Cohen, R. Peled and T. Shavit, “Decision-making and the newsvendor problem: an experimental study”, Journal of the Operational Research Society **59**, 9, 1281–1287 (2008). 6.2.2

Berinsky, A. J., G. A. Huber and G. S. Lenz, “Evaluating online labor markets for experimental research: Amazon. com’s mechanical turk”, Political analysis **20**, 3, 351–368 (2012). 6.4.1

Betzler, N., R. Bredereck and R. Niedermeier, “Theoretical and empirical evaluation of data reduction for exact kemeny rank aggregation”, Autonomous Agents and Multi-Agent Systems **28**, 5, 721–748 (2014). 1.1, 4.4.1

Bolboaca, S.-D. and L. Jäntschi, “Pearson versus spearman, kendall’s tau correlation analysis on structure-activity relationships of biologic active compounds”, Leonardo Journal of Sciences **5**, 9, 179–200 (2006). 6.5.1

Bolton, G. E. and E. Katok, “Learning by doing in the newsvendor problem: A laboratory investigation of the role of experience and feedback”, Manufacturing & Service Operations Management **10**, 3, 519–538 (2008). 6.2.2

Bonett, D. G. and T. A. Wright, “Sample size requirements for estimating pearson, kendall and spearman correlations”, Psychometrika **65**, 1, 23–28 (2000). 6.5.1

Brancotte, B., B. Yang, G. Blin, S. Cohen-Boulakia, A. Denise and S. Hamel, “Rank aggregation with ties: Experiments and analysis”, Proceedings of the VLDB Endowment **8**, 11, 1202–1213 (2015). (document), 2.2, 4.1, 4.1, 1, 4.3, 4.4, 4.4.1, 4.4.1, 4.9, 4.4.2

Brandt, F., V. Conitzer, U. Endriss, J. Lang and A. D. Procaccia, *Handbook of computational social choice* (Cambridge University Press, 2016). 2.1, 2.4, 2.4, 2.4, 3.3

Budescu, D. V. and E. Chen, “Identifying expertise to extract the wisdom of crowds”, Management Science **61**, 2, 267–280 (2015). 1.3

Ceberio, J., E. Irurozki, A. Mendiburu and J. A. Lozano, “A review of distances for the mallows and generalized mallows estimation of distribution algorithms”, Computational Optimization and Applications **62**, 2, 545–564 (2015). 4.4.1

Chandler, J. and D. Shapiro, “Conducting clinical research using crowdsourced convenience samples”, Annual review of clinical psychology **12** (2016). 6.4.1

Charras, P., G. Brod and J. Lupiáñez, “Is 26+ 26 smaller than 24+ 28? estimating the approximate magnitude of repeated versus different numbers”, Attention, Perception, & Psychophysics **74**, 1, 163–173 (2012). 1.3

Chittilappilly, A. I., L. Chen and S. Amer-Yahia, "A survey of general-purpose crowdsourcing techniques", IEEE Transactions on Knowledge and Data Engineering **28**, 9, 2246–2266 (2016). 7

Cohen, D., P. Jeavons, C. Jefferson, K. E. Petrie and B. M. Smith, "Symmetry definitions for constraint satisfaction problems", in "International Conference on Principles and Practice of Constraint Programming", pp. 17–31 (Springer, 2005). 1.1

Condorcet, M., "Marquis de (1785)", Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix (1785). 2.4, 5.1

Conitzer, V., A. Davenport and J. Kalagnanam, "Improved bounds for computing kemeny rankings", in "AAAI", vol. 6, pp. 620–626 (2006). 4.1

Cook, W. D., "Distance-based and ad hoc consensus models in ordinal preference ranking", European Journal of Operational Research **172**, 2, 369–385 (2006). 1

Cook, W. D., B. Golany, M. Penn and T. Raviv, "Creating a consensus ranking of proposals from reviewers' partial ordinal rankings", Computers & Operations Research **34**, 4, 954–965 (2007a). 1, 4.1

Cook, W. D., B. Golany, M. Penn and T. Raviv, "Creating a consensus ranking of proposals from reviewers? partial ordinal rankings", Computers & Operations Research **34**, 4, 954–965 (2007b). 2.2

Cook, W. D. and M. Kress, "Ordinal ranking with intensity of preference", Management science **31**, 1, 26–32 (1985). 2.2

Crispino, M., E. Arjas, V. Vitelli, N. Barrett, A. Frigessi *et al.*, "A bayesian mallows approach to nontransitive pair comparison data: How human are sounds?", The Annals of Applied Statistics **13**, 1, 492–519 (2019). 4.4.1

Critchlow, D. E., *Metric methods for analyzing partially ranked data*, vol. 34 (Springer Science & Business Media, 2012). 4.4.1

Cushman, P., J. T. Hoeksema, C. Kouveliotou, J. Lowenthal, B. Peterson, K. G. Stassun and T. von Hippel, "Impact of declining proposal success rates on scientific productivity", arXiv preprint arXiv:1510.01647 (2015). 1.1

Da, Z. and X. Huang, "Harnessing the wisdom of crowds", Management Science **66**, 5, 1847–1867 (2020). 1.3, 2.4

Davenport, A. and J. Kalagnanam, "A computational study of the kemeny rule for preference aggregation", in "AAAI", vol. 4, pp. 697–702 (2004). 1.1

Davidson, J., B. Liebald, J. Liu, P. Nandy, T. Van Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingston *et al.*, "The youtube video recommendation system", in "Proceedings of the fourth ACM conference on Recommender systems", pp. 293–296 (2010). 5.4

de Borda, J.-C., “Mémoire sur les élections au scrutin, histoire de l?académie royale des sciences”, Paris, France (1781). 1

Demartini, G., D. E. Difallah, U. Gadiraju and M. Catasta, “An introduction to hybrid human-machine information systems”, Foundations and Trends in Web Science **7**, 1, 1–87 (2017). 6.1

Desarkar, M. S., S. Sarkar and P. Mitra, “Preference relations based unsupervised rank aggregation for metasearch”, Expert Systems with Applications **49**, 86–98 (2016). 1.1

Diaconis, P., “Group representations in probability and statistics”, Lecture notes-monograph series **11**, i–192 (1988). 4.4.1

Djelassi, S. and I. Decoopman, “Customers’ participation in product development through crowdsourcing: Issues and implications”, Industrial Marketing Management **42**, 5, 683–692 (2013). 6.7.2

Doignon, J.-P. and S. Fiorini, “Facets of the weak order polytope derived from the induced partition projection”, SIAM journal on discrete mathematics **15**, 1, 112–121 (2001). 7

Doignon, J.-P., A. Pekeč and M. Regenwetter, “The repeated insertion model for rankings: Missing link between two subset choice models”, Psychometrika **69**, 1, 33–54 (2004). 4.4.1

Dummett, M., “The borda count and agenda manipulation”, Social Choice and Welfare **15**, 2, 289–296 (1998). 1, 2.4

Dwork, C., R. Kumar, M. Naor and D. Sivakumar, “Rank aggregation methods for the web”, in “Proceedings of the 10th international conference on the World Wide Web”, pp. 613–622 (ACM, New York, NY, USA, 2001). 1, 1.1, 2.2, 4.3, 5.1

Economist, T., “When the many know best”, URL `https://www.economist.com/books-and-arts/2004/05/27/when-the-many-know-best` (2004). 6.1

Emond, E. J. and D. W. Mason, *A new technique for high level decision support* (Department of National Defence Canada, Operational Research Division, Directorate of Operational Research (Corporate, Air & Maritime), 2000). 1.1

Emond, E. J. and D. W. Mason, “A new rank correlation coefficient with application to the consensus ranking problem”, Journal of Multi-Criteria Decision Analysis **11**, 1, 17–28 (2002). 2.3, 2.3, 3.1, 3.3, 3.3

Endriss, U., S. Obraztsova, M. Polukarov and J. S. Rosenschein, “Strategic voting with incomplete information”, (2016). 1, 2.2

Escobedo, A. R., E. Moreno-Centeno and R. Yasmin, “An axiomatic distance methodology for aggregating multimodal evaluations”, Optimization-Online preprint 8223 (2021). 1, 2.2, 4.3, 4.5, 6.3, 6.3

Escobedo, A. R. and R. Yasmin, “Derivations of large classes of facet-defining inequalities of the weak order polytope using ranking structures”, arXiv preprint arXiv:2008.03799 (2021). 7

Fagin, R., R. Kumar, M. Mahdian, D. Sivakumar and E. Vee, “Comparing and aggregating rankings with ties”, in “Proceedings of the twenty-third ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems”, pp. 47–58 (2004). 1.1

Fagin, R., R. Kumar and D. Sivakumar, “Efficient similarity search and classification via rank aggregation”, in “Proceedings of the 2003 ACM SIGMOD international conference on Management of data”, pp. 301–312 (ACM, 2003). 1, 1.1, 6.1

Farah, M. and D. Vanderpooten, “An outranking approach for rank aggregation in information retrieval”, in “Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval”, pp. 591–598 (ACM, 2007). 1, 6.1

Favardin, P., D. Lepelley and J. Serais, “Borda rule, copeland method and strategic manipulation”, Review of Economic Design **7**, 2, 213–228 (2002). 1, 2.2, 2.4

Feld, S. L. and B. Grofman, “The borda count in n-dimensional issue space”, Public Choice **59**, 2, 167–176 (1988). 1, 2.2

Fields, E. B., G. E. Okudan and O. M. Ashour, “Rank aggregation methods comparison: A case for triage prioritization”, Expert Systems with Applications **40**, 4, 1305–1311 (2013). 1

Fiorini, S. and P. C. Fishburn, “Weak order polytopes”, Discrete mathematics **275**, 1-3, 111–127 (2004). 1, 4.3, 4.3

Fishbain, B. and E. Moreno-Centeno, “Self calibrated wireless distributed environmental sensory networks”, Scientific reports **6**, 24382 (2016). 2.2

Franks, N. R., A. Dornhaus, J. P. Fitzsimmons and M. Stevens, “Speed versus accuracy in collective decision making”, Proceedings of the Royal Society of London. Series B: Biological Sciences **270**, 1532, 2457–2463 (2003). 6.7.2

Galton, F., “Vox populi”, (1907). 1.3, 2.4

Gao, Y. and K. Xu, “prankaggreg: A fast clustering based partial rank aggregation”, Information Sciences **478**, 408–421 (2019). 1, 6.1

Gass, S., “Tournaments, transitivity and pairwise comparison matrices”, Journal of the Operational Research Society **49**, 6, 616–624 (1998). 4.2

Gebuis, T. and B. Reynvoet, “The role of visual information in numerosity estimation”, PloS one **7**, 5, e37426 (2012). 1.3

Goldstein, D. G., R. P. McAfee and S. Suri, “The wisdom of smaller, smarter crowds”, in “Proceedings of the fifteenth ACM conference on Economics and computation”, pp. 471–488 (2014). 1.3

Goldstone, R. L., “Feature distribution and biased estimation of visual displays.”, Journal of Experimental Psychology: Human Perception and Performance **19**, 3, 564 (1993). 1.3

Good, I., “The number of orderings of n candidates when ties are permitted”, Fib. Quart **13**, 11–18 (1975). 1.1

Gross, O. A., “Preferential arrangements”, The American Mathematical Monthly **69**, 1, 4–8 (1962). 1.1

Grzegorzewski, P., “On measuring association between preference systems”, in “Fuzzy Systems, 2004. Proceedings. 2004 IEEE International Conference on”, vol. 1, pp. 133–137 (IEEE, 2004). 2.3

Grzegorzewski, P., “The coefficient of concordance for vague data”, Computational Statistics & Data Analysis **51**, 1, 314–322 (2006). 2.3

Grzegorzewski, P., “Kendall’s correlation coefficient for vague preferences”, Soft Computing **13**, 11, 1055–1061 (2009). 2.3

Grzegorzewski, P. and P. Ziembinska, “Spearman’s rank correlation coefficient for vague preferences”, in “International Conference on Flexible Query Answering Systems”, pp. 342–353 (Springer, 2011). 2.3

Heiser, W. J., “Geometric representation of association between categories”, Psychometrika **69**, 4, 513–545 (2004). 1

Helson, H., “Adaptation-level theory: an experimental and systematic approach to behavior.”, (1964). 1.3

Hochbaum, D. S., “The separation, and separation-deviation methodology for group decision making and aggregate ranking”, in “Risk and Optimization in an Uncertain World”, pp. 116–141 (INFORMS, 2010). 2.2, 2.2, 6.3

Hochbaum, D. S. and A. Levin, “The k-allocation problem and its variants”, in “International Workshop on Approximation and Online Algorithms”, pp. 253–264 (Springer, 2006a). 6.4.3

Hochbaum, D. S. and A. Levin, “Methodologies and algorithms for group-rankings decision”, Management Science **52**, 9, 1394–1408 (2006b). 1, 6.1

Hochbaum, D. S. and A. Levin, “How to allocate review tasks for robust ranking”, Acta informatica **47**, 5-6, 325–345 (2010). 3.2

Hollingsworth, W. H., J. P. Simmons, T. R. Coates and H. A. Cross, “Perceived numerosity as a function of array number, speed of array development, and density of array items”, Bulletin of the Psychonomic Society **29**, 5, 448–450 (1991). 1.3

Horton, J. J., “The dot-guessing game: A ‘fruit fly’for human computation research”, Available at SSRN 1600372 (2010). 6.1

Huang, N., G. Burtch, B. Gu, Y. Hong, C. Liang, K. Wang, D. Fu and B. Yang, "Motivating user-generated content with performance feedback: Evidence from randomized field experiments", Management Science **65**, 1, 327–345 (2019). 6.5

Huang, Y., P. Vir Singh and K. Srinivasan, "Crowdsourcing new product ideas under consumer learning", Management science **60**, 9, 2138–2159 (2014). 6.7.2

IBM Knowledge Center, "Ibm ilog cplex optimization studio v12.8.0 documentation", (2017). 4.4

IBM Support, URL `https://www.ibm.com/support/pages/apar/RS03137` (2019). 4.4.1

Irshad, H., L. Montaser-Kouhsari, G. Waltz, O. Bucur, J. Nowak, F. Dong, N. W. Knoblauch and A. H. Beck, "Crowdsourcing image annotation for nucleus detection and segmentation in computational pathology: evaluating experts, automated methods, and the crowd", in "Pacific symposium on biocomputing Co-chairs", pp. 294–305 (World Scientific, 2014). 6.7.2

Irurozki, E., B. Calvo, J. A. Lozano *et al.*, "Permallows: An r package for mallows and generalized mallows models", Journal of Statistical Software **71**, 12, 1–30 (2016). 4.4.1

Jayles, B. and R. H. Kurvers, "Debiasing the crowd: selectively exchanging social information improves collective decision making", arXiv preprint arXiv:2003.06863 (2020). 1.3

Kahneman, D., S. P. Slovic, P. Slovic and A. Tversky, *Judgment under uncertainty: Heuristics and biases* (Cambridge university press, 1982). 1.3

Kaufmann, E., O. Cappé and A. Garivier, "On the complexity of a/b testing", in "Conference on Learning Theory", pp. 461–481 (PMLR, 2014). 6.5

Keisler, J., "Value of information in portfolio decision analysis", Decision analysis **1**, 3, 177–189 (2004). 1.1

Kemeny, J. G. and L. J. Snell, "Preference ranking: An axiomatic approach", in "Mathematical Models in Social Science", pp. 9–23 (Ginn, Boston, 1962). 1, 2.2, 2.2, 3.3, 5.1

Kemmer, R., Y. Yoo, A. Escobedo and R. Maciejewski, "Enhancing collective estimates by aggregating cardinal and ordinal inputs", in "Proceedings of the AAAI Conference on Human Computation and Crowdsourcing", vol. 8, pp. 73–82 (2020). 3, 6.1, 6.7.1

Kendall, M. G., "A new measure of rank correlation", Biometrika **30**, 1/2, 81–93 (1938). 2.2, 2.3

Kendall, M. G., "The treatment of ties in ranking problems", Biometrika pp. 239–251 (1945). 1.1

Kendall, M. G., “Rank correlation methods”, (1948). 2.3

Kenyon-Mathieu, C. and W. Schudy, “How to rank with few errors”, in “Proceedings of the thirty-ninth annual ACM symposium on Theory of computing”, pp. 95–103 (ACM, 2007). 1.1

Kim, M., F. Farnoud and O. Milenkovic, “Hydra: gene prioritization via hybrid distance-score rank aggregation”, Bioinformatics **31**, 7, 1034–1043 (2015). 1.2

Kohavi, R. and R. Longbotham, “Online controlled experiments and a/b testing.”, Encyclopedia of machine learning and data mining **7**, 8, 922–929 (2017). 6.5

Kolde, R., S. Laur, P. Adler and J. Vilo, “Robust rank aggregation for gene list integration and meta-analysis”, Bioinformatics **28**, 4, 573–580 (2012). 1.1

Kruger, J. and D. Dunning, “Unskilled and unaware of it: how difficulties in recognizing one’s own incompetence lead to inflated self-assessments.”, Journal of personality and social psychology **77**, 6, 1121 (1999). 1.3

Kruk, J., J. Lubin, K. Sikka, X. Lin, D. Jurafsky and A. Divakaran, “Integrating text and image: Determining multimodal document intent in instagram posts”, arXiv preprint arXiv:1904.09073 (2019). 1.2, 6.2.1

Lee, K., J. Caverlee and S. Webb, “The social honeypot project: protecting online communities from spammers”, in “Proceedings of the 19th international conference on World wide web”, pp. 1139–1140 (2010). 7

Lee, Y.-J., K. Hosanagar and Y. Tan, “Do i follow my friends or the crowd? information cascades in online movie ratings”, Management Science **61**, 9, 2241–2258 (2015). 6.1

Li, K., X. Zhang and G. Li, “A rating-ranking method for crowdsourced top-k computation”, in “Proceedings of the 2018 International Conference on Management of Data”, pp. 975–990 (2018). 1.2

Liberti, L., “Automatic generation of symmetry-breaking constraints”, in “International Conference on Combinatorial Optimization and Applications”, pp. 328–338 (Springer, 2008). 1.1

Lin, S., “Rank aggregation methods”, Wiley Interdisciplinary Reviews: Computational Statistics **2**, 5, 555–570 (2010a). 1, 1.1

Lin, S., “Space oriented rank-based data integration”, Statistical Applications in Genetics and Molecular Biology **9**, 1 (2010b). 1, 1.1

Long, X. and J. Nasiry, “Prospect theory explains newsvendor behavior: The role of reference points”, Management Science **61**, 12, 3009–3012 (2015). 6.2.2

Lu, T. and C. Boutilier, “Effective sampling and learning for mallows models with pairwise-preference data”, The Journal of Machine Learning Research **15**, 1, 3783–3829 (2014). 4.4.1

Mallows, C. L., “Non-null ranking models. i”, Biometrika **44**, 1/2, 114–130 (1957). 4.4.1

Mandal, M. and A. Mukhopadhyay, “Multiobjective pso-based rank aggregation: Application in gene ranking from microarray data”, Information Sciences **385**, 55–75 (2017). 1, 1.1

Mannes, A. E., “Are we wise about the wisdom of crowds? the use of group judgments in belief revision”, Management Science **55**, 8, 1267–1279 (2009). 1.3, 2.4

Mao, A., A. D. Procaccia and Y. Chen, “Social choice for human computation”, in “Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence”, (Citeseer, 2012). 1.3

Mao, A., A. D. Procaccia and Y. Chen, “Better human computation through principled voting”, in “Twenty-Seventh AAAI Conference on Artificial Intelligence”, (2013). 1.3

Marbach, D., J. C. Costello, R. Küffner, N. M. Vega, R. J. Prill, D. M. Camacho, K. R. Allison, A. Aderhold, R. Bonneau, Y. Chen *et al.*, “Wisdom of crowds for robust gene network inference”, Nature methods **9**, 8, 796 (2012). 1, 1.1

Marden, J. I., *Analyzing and modeling rank data* (CRC Press, 1996). 4.4.1

Martí, R. and G. Reinelt, in “The Linear Ordering Problem”, (Springer, 2011). 1, 4.3

Mattei, N. and T. Walsh, “Preflib: A library for preferences http://www. preflib. org”, in “International Conference on Algorithmic DecisionTheory”, pp. 259–270 (Springer, 2013). 4.4, 4.4.2

Miller, G. A., “The magical number seven, plus or minus two: Some limits on our capacity for processing information”, Psychol. Rev. **63**, 2, 81–97, URL `http://www.musanim.com/miller1956/` (1956). 1.1

Milosz, R. and S. Hamel, “Exploring the median of permutations problem”, Journal of Discrete Algorithms **52**, 92–111 (2018). 4.4.2

Misra, A., A. Gooze, K. Watkins, M. Asad and C. A. Le Dantec, “Crowdsourcing and its application to transportation data collection and management”, Transportation Research Record **2414**, 1, 1–8 (2014). 6.1

Moreno-Centeno, E. and A. R. Escobedo, “Axiomatic aggregation of incomplete rankings”, IIE Transactions **48**, 6, 475–488 (2016). 1, 1.1, 2.2, 2.2, 2.2, 2.3, 2.4, 3.3, 4.3

Moreno-Centeno, E. and R. M. Karp, “The implicit hitting set approach to solve combinatorial optimization problems with an application to multigenome alignment”, Operations Research **61**, 2, 453–468 (2013). 1.1

Mortensen, K. and T. L. Hughes, “Comparing amazon’s mechanical turk platform to conventional data collection methods in the health and medical research literature”, Journal of General Internal Medicine **33**, 4, 533–538 (2018). 6.4.1

Mortensen, M. L., G. P. Adam, T. A. Trikalinos, T. Kraska and B. C. Wallace, "An exploration of crowdsourcing citation screening for systematic reviews", Research synthesis methods **8**, 3, 366–386 (2017). 7

Müller-Trede, J., S. Choshen-Hillel, M. Barneron and I. Yaniv, "The wisdom of crowds in matters of taste", Management Science **64**, 4, 1779–1803 (2018). 1.3, 2.4

Muravyov, S. V., "Dealing with chaotic results of kemeny ranking determination", Measurement **51**, 328–334 (2014). 5.1

Najafabadi, M. M., F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald and E. Muharemagic, "Deep learning applications and challenges in big data analytics", Journal of Big Data **2**, 1, 1 (2015). 6.1

Navajas, J., T. Niella, G. Garbulsky, B. Bahrami and M. Sigman, "Aggregated knowledge from a small number of debates outperforms the wisdom of large crowds", Nature Human Behaviour **2**, 2, 126–132 (2018). 1.3

Nemhauser, G. L. and L. A. Wolsey, "Integer programming and combinatorial optimization", Wiley, Chichester. GL Nemhauser, MWP Savelsbergh, GS Sigismondi (1992). Constraint Classification for Mixed Integer Programming Formulations. COAL Bulletin **20**, 8–12 (1988). 4.3

Newman, A. and S. Vempala, "Fences are futile: On relaxations for the linear ordering problem", in "International Conference on Integer Programming and Combinatorial Optimization", pp. 333–347 (Springer, 2001). 4.3

Paolacci, G., J. Chandler and P. G. Ipeirotis, "Running experiments on amazon mechanical turk", Judgment and Decision making **5**, 5, 411–419 (2010). 6.4.1

Quinn, A. J. and B. B. Bederson, "Human computation: a survey and taxonomy of a growing field", in "Proceedings of the SIGCHI conference on human factors in computing systems", pp. 1403–1412 (2011). 6.1

Ramachandran, K., N. Tereyağoğlu and Y. Xia, "Multidimensional decision making in operations: An experimental investigation of joint pricing and quantity decisions", Management Science **64**, 12, 5544–5558 (2018). 6.2.2

Rothschild, D., "Forecasting elections: Comparing prediction markets, polls, and their biases", Public Opinion Quarterly **73**, 5, 895–916 (2009). 6.1

Russell, P. A. and C. D. Gray, "Ranking or rating? some data and their implications for the measurement of evaluative response", British journal of Psychology **85**, 1, 79–92 (1994). 6.5.1

Saaty, T. L. and M. S. Ozdemir, "Why the magic number seven plus or minus two", Mathematical and Computer Modelling **38**, 3-4, 233–244 (2003). 3.2

Schilling, M. S., N. Oeser and C. Schaub, "How effective are decision analyses? assessing decision process and group alignment effects", Decision Analysis **4**, 4, 227–242 (2007). 1.1

Schweitzer, M. E. and G. P. Cachon, “Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence”, Management Science **46**, 3, 404–420 (2000). 6.2.2

Sherali, H. D. and J. C. Smith, “Improving discrete model representations via symmetry considerations”, Management Science **47**, 10, 1396–1407 (2001). 1.1

Siddharthan, A., C. Lambin, A.-M. Robinson, N. Sharma, R. Comont, E. O’mahony, C. Mellish and R. V. D. Wal, “Crowdsourcing without a crowd: Reliable online species identification using bayesian models to minimize crowd size”, ACM Transactions on Intelligent Systems and Technology (TIST) **7**, 4, 1–20 (2016). 1.3

Simoiu, C., C. Sumanth, A. Mysore and S. Goel, “Studying the “wisdom of crowds” at scale”, in “Proceedings of the AAAI Conference on Human Computation and Crowdsourcing”, vol. 7, pp. 171–179 (2019). 6.2.2

Slowinski, R., *Fuzzy sets in decision analysis, operations research and statistics*, vol. 1 (Springer Science & Business Media, 2012). 2.3

Smith, J. H., “Aggregation of preferences with variable electorate”, Econometrica: Journal of the Econometric Society pp. 1027–1041 (1973). 5.1

Stewart, N., C. Ungemach, A. J. Harris, D. M. Bartels, B. R. Newell, G. Paolacci, J. Chandler *et al.*, “The average laboratory samples a population of 7,300 amazon mechanical turk workers”, Judgment and Decision making **10**, 5, 479–491 (2015). 6.4.1

Steyvers, M., B. Miller, P. Hemmer and M. D. Lee, “The wisdom of crowds in the recollection of order information”, in “Advances in neural information processing systems”, pp. 1785–1793 (2009). 1.3

Streib, N., S. J. Young and J. Sokol, “A major league baseball team uses operations research to improve draft preparation”, Interfaces **42**, 2, 119–130 (2012). 1

Sun, Z., P. Sarma, W. Sethares and Y. Liang, “Learning relationships between text, audio, and video via deep canonical correlation for multimodal language analysis”, in “Proceedings of the AAAI Conference on Artificial Intelligence”, vol. 34, pp. 8992–8999 (2020). 1.2, 6.2.1

Surowiecki, J., *The wisdom of crowds* (Anchor, 2005). 1.3

Truchon, M. *et al.*, “An extension of the condorcet criterion and kemeny orders”, Cahier **9813** (1998). 5.1

Tversky, A. and D. Kahneman, “Judgment under uncertainty: Heuristics and biases”, science **185**, 4157, 1124–1131 (1974). 1.3, 6.2.2

Wald, R., T. M. Khoshgoftaar, D. Dittman, W. Awada and A. Napolitano, “An extensive comparison of feature ranking aggregation techniques in bioinformatics”, in “2012 IEEE 13th International Conference on Information Reuse & Integration (IRI)”, pp. 377–384 (IEEE, 2012). 1.1

Wang, J. and N. B. Shah, "Your 2 is my 1, your 3 is my 9: Handling arbitrary miscalibrations in ratings", arXiv preprint arXiv:1806.05085 (2018). 1.2

Wang, J. and N. B. Shah, "Ranking and rating rankings and ratings", in "Proceedings of the AAAI Conference on Artificial Intelligence", vol. 34, pp. 13704–13707 (2020). 1.2

Winkler, R. L., Y. Grushka-Cockayne, K. C. Lichtendahl Jr and V. R. R. Jose, "Probability forecasts and their combination: A research perspective", Decision Analysis **16**, 4, 239–260 (2019). 1.3, 2.4

Xiao, B. and I. Benbasat, "An empirical examination of the influence of biased personalized product recommendations on consumers' decision making outcomes", Decision Support Systems **110**, 46–57 (2018). 6.2.2

Xintong, G., W. Hongzhi, Y. Song and G. Hong, "Brief survey of crowdsourcing for data mining", Expert Systems with Applications **41**, 17, 7987–7994 (2014). 6.1

Ye, M., C. Liang, Y. Yu, Z. Wang, Q. Leng, C. Xiao, J. Chen and R. Hu, "Person reidentification via ranking aggregation of similarity pulling and dissimilarity pushing", IEEE Transactions on Multimedia **18**, 12, 2553–2566 (2016). 1, 6.1

Yi, S. K., M. Steyvers, M. Lee and M. Dry, "Wisdom of the crowds in minimum spanning tree problems", in "Proceedings of the Annual Meeting of the Cognitive Science Society", vol. 32 (2010). 1.3

Yilmaz, E., J. A. Aslam and S. Robertson, "A new rank correlation coefficient for information retrieval", in "Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval", pp. 587–594 (ACM, 2008). 1, 2.3, 6.1

Yoo, Y. and A. R. Escobedo, "A new binary programming formulation and social choice propertyfor expediting the solution to Kemeny rank aggregation", Decision Analysis [Available at `optimization-online.org/DB_HTML/2020/08/7958.html`] (2021). 2

Yoo, Y., A. R. Escobedo and J. K. Skolfield, "A new correlation coefficient for comparing and aggregating non-strict and incomplete rankings", European Journal of Operational Research **285**, 3, 1025–1041 (2020). 1.1, 1, 4.3, 4.4.1, 4.4.1

Young, H. P., "Condorcet's theory of voting", American Political science review **82**, 04, 1231–1244 (1988). 2.4, 3.3

Young, H. P. and A. Levenglick, "A consistent extension of condorcets election principle", SIAM Journal on applied Mathematics **35**, 2, 285–300 (1978). 2.4, 3.3, 5.1

Zheng, F., R. Tao, H. R. Maier, L. See, D. Savic, T. Zhang, Q. Chen, T. H. Assumpção, P. Yang, B. Heidari *et al.*, "Crowdsourcing methods for data collection in geophysics: State of the art, issues, and future directions", Reviews of Geophysics **56**, 4, 698–740 (2018). 6.1

APPENDIX A

DEMOGRAPHIC SURVEY RESULTS

	A		B	
Total	273		286	
Gender				
Female	78	28.57%	116	40.56%
Male	195	71.43%	168	58.74%
Other	0	0%	2	0.70%
Age				
10-19	0	0.00%	2	0.70%
20-29	81	29.67%	95	33.22%
30-39	122	44.69%	89	31.12%
40-49	35	12.82%	54	18.88%
50-59	24	8.79%	35	12.24%
60-69	11	4.03%	9	3.15%
70-79	0	0.00%	2	0.70%
Education				
2-year degree	12	4.40%	29	10.14%
4-year degree	177	64.84%	124	43.36%
College	22	8.06%	51	17.83%
Master’s	45	16.48%	32	11.19%
Professional (MD, KJD, etc)	2	0.73%	5	1.75%
Doctoral	0	0%	1	0.35%
High-school/GED	15	5.49%	43	15.03%
Less than high-school	0	0%	1	0.35%
Employment status				
Employed	253	92.67%	234	81.82%
Unemployed	20	7.33%	52	18.18%
Native English Speaker				
Native	271	99.27%	284	99.30%
Non-native	2	0.73%	2	0.70%

Table A.1: Demographic Survey Results

APPENDIX B

OPTIMALITY GAP FOR THE CARDINAL AND ORDINAL AGGREGATION MODEL

	Number of times each image was seen											
Problem size	5	10	15	20	25	30	35	40	45	50	55	60
2-image	0.30	0.27	0.23	0.21								
3-image	0.27	0.25	0.24	0.22	0.21	0.21						
5-image	0.41	0.40	0.37	0.36	0.35	0.34	0.32	0.32	0.31	0.32		
6-image	0.37	0.37	0.32	0.30	0.28	0.26	0.25	0.23	0.28	0.18	0.25	0.15

Table B.1: Optimality Gaps for Setting A

	Number of times each image was seen											
Problem size	5	10	15	20	25	30	35	40	45	50	55	60
2-image	0.00	0.00	0.00	0.00								
3-image	0.00	0.00	0.00	0.00	0.01	0.00						
5-image	0.04	0.00	0.02	0.02	0.00	0.01	0.00	0.00	0.00	0.00		
6-image	0.00	0.03	0.02	0.02	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00

Table B.2: Optimality Gaps for Setting B