Computational Perspectives on Democracy

Anson Kahng

CMU-CS-21-126

August 2021

Computer Science Department
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Ariel Procaccia (Chair)
Chinmay Kulkarni
Nihar Shah
Vincent Conitzer (Duke University)
David Pennock (Rutgers University)

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

This research was sponsored by the National Science Foundation under grant numbers IIS-1350598, CCF-1525932, and IIS-1714140, the Department of Defense under grant number W911NF1320045, the Office of Naval Research under grant number N000141712428, and the JP Morgan Research grant.

Keywords: Computational Social Choice, Theoretical Computer Science, Artificial Intelligence

For Grandpa and Harabeoji.

Abstract

Democracy is a natural approach to large-scale decision-making that allows people affected by a potential decision to provide input about the outcome. However, modern implementations of democracy are based on outdated information technology and must adapt to the changing technological landscape. This thesis explores the relationship between computer science and democracy, which is, crucially, a two-way street—just as principles from computer science can be used to analyze and design democratic paradigms, ideas from democracy can be used to solve hard problems in computer science.

Question 1: What can computer science do for democracy?

To explore this first question, we examine the theoretical foundations of three democratic paradigms: liquid democracy, participatory budgeting, and multiwinner elections. Each of these paradigms broadly redistributes power from the few to the many: For instance, liquid democracy allows people to choose delegates more flexibly and participatory budgeting enables citizens to directly influence government spending toward public projects. However, because these paradigms are relatively new, their theoretical properties are relatively unexplored. We analyze each of these three settings from the point of view of computational social choice, which is a mathematical framework for collective decision-making. In particular, we focus on a combination of robustness, fairness, and efficiency with the end goal of providing actionable advice for future iterations of these paradigms.

Question 2: What can democracy do for computer science?

Toward this end, we explore two settings in which democratic principles can be used to augment approaches to making difficult decisions—in our case, automating ethical decision-making and hiring in online labor markets. Both of these problems are difficult in the sense that there is no universally agreed-upon function to optimize, making them a poor fit for traditional approaches in computer science. Instead, we try to emulate a world in which we can get input from people in order to arrive at a "societal" decision. In each of these settings, we first propose and analyze a theoretical approach that leads a single decision, and then, in collaboration with HCI researchers, run experiments in the real world to test the efficacy and practicability of our approaches in the real world.

Acknowledgments

First and foremost, I must thank Ariel Procaccia, my advisor, for being the best advisor anyone could hope for. You have taught me so much, not only technically, but also with respect to writing effective papers, giving better talks, mentoring junior students, and providing me with a sustainable example of healthy (and productive!) work-life balance. I will strive to learn from and emulate your unrivaled blend of brilliance, eloquence, and patience. Thank you for everything.

To my thesis committee, Vince Conitzer, Chinmay Kulkarni, David Pennock, and Nihar Shah, thank you for your insightful comments and wonderful questions that have helped me structure and present my work. I truly appreciate all the comments and questions, even if I may not have all the answers yet!

To the professors at Harvard that led me down the grad school path: David and Yiling. Thank you for opening my eyes to the wonders of Econ-CS (or CS-Econ here at CMU), for supervising my first attempt at independent research (David), and for your continued support and guidance over the years.

To my collaborators: Gerdus Benadè, Allissa Chan, Rupert Freeman, Paul Gölz, Bernhard Haeupler, Ellis Hershkowitz, Gregory Kehne, Ji Tae Kim, Yasmine Kotturi, Chinmay Kulkarni, David Kurokawa, Daniel Kusbit, Min Kyung Lee, Siheon Lee, Simon Mackenzie, Ritesh Noothigattu, David Pennock, Dominik Peters, Ariel Procaccia, Alex Psomas, Daniel See, and Xinran Yuan. I truly enjoyed collaborating with each and every one of you, and thank you for bringing your expertise and hard work to the projects we worked on together.

To my friends: Ellis, Alex, Greg, Roie, Mark, Kevin, Paul, Bailey, Jalani, Ellen, Ryan, Costin, Sidu, Ved, Kevin, Carl, Lisa, Amna, Michelle, Ramya, Chara, Robi, Andrew, Justin, and all others. Thank you for your continued support and companionship through this journey. Grad school would have been an altogether different and darker time without you.

To my family: Dad, Mom, Alex, Aidan, Aunt Lyn Sue, Grandma, Halmoni, and everyone else. Thank you for loving and supporting me since the very beginning, and for instilling in me an innate curiosity and love of learning. I owe you everything, and hope to make you proud.

To Soyeun. Last but certainly not least, thank you for your unwavering support through the highs and lows of grad school. I truly could not have done it without your invariably sage advice, spot-on presentation notes, and emotional and moral support.

Contents

List of Figures

List of Tables

Democracy is the worst form of government except for all those other forms that have been tried from time to time.

Winston Churchill

0 Introduction

Group decision-making is a fundamental challenge in human society. How should groups of people (e.g., nations, states, cities, or neighborhoods) make collective decisions based on heterogeneous opinions?

Democracy offers a compelling answer to this question: Let the people themselves decide. Societies over the years have ascribed to this philosophy to varying degrees. The ancient Greeks were early champions of democracy, and Athenians notably used democratic processes like sortition (random selection from the public) and voting (soliciting structured input from the public) at all levels of society. Since then, democracy has ebbed and flowed, but mostly flowed. Global powers up until the 17th century were largely monarchies or oligarchies, but John Locke's liberal democratic framework espoused in his seminal work, *Two Treatises of Government* [162], set in motion a democratic wave that has continued into the 21st century.

However, although many countries have ostensibly embraced democratic principles, the world is currently in a state of democratic retrograde, otherwise known as democratic backsliding. Implementations of modern democracy largely adhere to the model of representative democracy, where the general public may directly elect representatives in legislative bodies, but these representatives may make decisions that do not accurately reflect the will of the people. This, among other factors, has led to a marked decrease in trust in democratic systems in recent years. Notably, advances in technology have not been blameless in this erosion of trust. In fact, social networks have led to increased division and greater discord online by facilitating the creation of echo chambers, an increase in polarization, and the proliferation of fake news.

All this is to say, humanity still has significant work to do in figuring out how to make democracy "work" as a stable, harmonious decision-making paradigm in which people affected by large-scale decisions get a say in what these decisions are. Luckily, despite the rather negative picture painted by the foregoing paragraphs, not all hope is lost for democratic ideals. In particular, as new technologies develop, academics and practitioners alike have turned to the new frontiers of digital democracy, or e-democracy, where technology is explicitly used to promote and strengthen democratic practices. Throughout this thesis, we hope to contribute to work in this vein, and the first part of this thesis focuses on the theoretical analysis of democratic paradigms that redistribute power into the hands of the people by allowing them to directly influence decisions.

However, the relationship between computer science and democracy is not a one-way street: Just as we can ask how tools from computer science can help evaluate and design new democratic paradigms, we can also ask how principles of democracy can help address difficult problems in computer science. This line of research is related to participatory artificial intelligence, wherein human values are taken into consideration in order to design more democratic and ethical machine learning systems.

Overall, we hope to address the following broad questions.

> **Question 1**: How can we formally analyze existing democratic paradigms with an eye toward creating better iterations of these paradigms in the future?
>
> **Question 2**: How can we apply democratic principles to augment traditional computer science-based approaches in order to solve difficult real-world problems?

Background: Computational Social Choice

Computational social choice is an interdisciplinary discipline that combines tools from social choice theory and theoretical computer science. Below, we provide a short overview of relevant topics; please see [47; 211; 197; 94; 114; 199] for a significantly more detailed introduction.

Social choice theory formalizes the problem of aggregating individual preferences to make a collective decision. In general, most problems in social choice can be thought of as preference aggregation, in which a mechanism receives a collection of opinions (e.g., ordinal rankings of alternatives) from agents and must return a single outcome, which may take the form of a single winner, a subset of winners, or a complete ranking, among other formats. Within this framework, notable branches of social choice include voting theory, which concerns the design and analysis of voting rules that take opinions as input and apply a well-specified mechanism (which can equivalently be thought of as a function) to make a final decision; resource allocation and fair division, where a common resource must be divided among a set of agents with different utility functions; and ranking systems, in which the set of agents and alternatives coincide, i.e., agents provide ranking information among themselves. Each of these topics will be touched upon in this thesis: Chapters 1, 3 and 4 primarily concern voting theory; Chapter 2 considers resource allocation; and Chapter 5 involves ranking systems.

Computational social choice introduces tools from theoretical science like complexity theory, approximation algorithms, and worst-case analysis to traditional social choice theory. Early work in computational social choice focused on the computational complexity of evaluating and manipulating voting rules; however, in this thesis, we focus on a combination of the following topics, each of which is often a desideratum of mechanisms in practice.

- *Robustness*: For our purposes, a voting rule is robust if it is resistant to noise. In other words, we desire rules with guarantees that hold with high probability even in the presence of perturbations. Chapters 1, 4 and 5 analyze mechanisms from the perspective of formal robustness guarantees.
- *Fairness and representation*: Fairness has many definitions, depending on the context. In the context of this thesis, we primarily consider proportionality, which maintains that groups of people deserve some amount of representation in the outcome of a decision in accordance with their size and/or cohesiveness. Chapters 2 and 3 explore notions of fairness and proportionality in participatory budgeting and multiwinner elections, respectively.
- *Strategyproofness*: A voting rule is strategyproof if each agent is (weakly) best-off reporting her true beliefs to the mechanism, i.e., dishonest manipulation has no benefits. We often desire strategyproofness in order to incentivize good behavior among agents. The work in Chapter 5 concerns the design of strategyproof peer ranking mechanisms, both in theory and practice.
- *Computability*: Some problems we wish to solve are computationally intractable, and voting rules must be computationally tractable in order to be useful. Therefore, in some cases, we must turn our attention to polynomial-time approximation algorithms. The work in Chapters 1 and 2 involves analysis of this flavor.

As an introductory note on voting theory, we note that many seminal results in the field are negative. In particular, there are three fundamental impossibility results in the field of voting theory: Condorcet's paradox, Arrow's Impossibility Theorem, and the Gibbard-Satterthwaite Theorem. Condorcet's paradox [88] states that, on a population level, majority judgments between pairs of alternatives may not be transitive; i.e., that there can exist settings where alternative a is preferred to alternative b by at least half of the voters, b is preferred to c by at least half of the voters, but over half of the voters prefer c to a, causing a cycle. Arrow's Impossibility Theorem [11] states that, under mild assumptions (i.e., the rule is not a dictatorship and satisfies unanimity) and in settings with at least three alternatives, every deterministic voting rule must violate a property known as independence of irrelevant alternatives (IIA). Finally, the Gibbard-Satterthwaite Theorem [116; 203] maintains that any deterministic, non-dictatorship voting rule for three or more candidates is not strategyproof; i.e., that agents can benefit by misreporting their preferences. These fundamental impossibility results mean that, in our work, we must restrict either the domain of (strategic) voter behavior or weaken the theoretical results we hope to prove.

Finally, we note that computational social choice has limitations inherent in its model of decision-making. For instance, it is generally assumed that voters' utilities do not change throughout the process, i.e., that they possess the same values before, during, and after voting. This is not necessarily the case in practice, and work on *deliberation* explicitly takes this into account. Additionally, computational social choice does not provide guidance about how to choose the set of alternatives over which voters opine, and the voting format—e.g., approval, veto, knapsack, or ranked votes—necessarily restricts the amount of information that voters can provide to any voting mechanism.

0.1 Structure

Part I: How computer science can help democracy

The first part of this thesis asks how computer science can help democracy. In particular, we explore the theoretical foundations of three new democratic paradigms: liquid democracy, participatory budgeting, and proportional multiwinner elections. On a high level, both liquid democracy and participatory budgeting allow people to more directly influence decision-making processes, and multiwinner elections allow people to choose committees of winners instead of only a single winner.

For each of the three paradigms in this part, we focus primarily on one or more of three theoretical desiderata of voting rules: *robustness*, *fairness*, and *efficiency*. In the context of this thesis, the robustness of a voting rule is its resilience against worst-case noise, fairness broadly refers to notions of proportionality and representation, and efficiency will be measured in terms of utilitarian social welfare, or the sum of all agents' utilities.

Liquid Democracy (Chapter 1) *Liquid democracy* is a collective decision making paradigm that allows voters to transitively delegate their votes. Our first work on this subject studied liquid democracy through an algorithmic lens [137]. In our model, there are two alternatives, one correct and one incorrect, and we are interested in the probability that the majority opinion is correct. Our main question is whether there exist delegation mechanisms that are *guaranteed* to outperform direct voting, in the sense of being always at least as likely, and sometimes more likely, to make a correct decision. Even though we assume that voters can only delegate their votes to more informed voters, we show that *local* delegation mechanisms, which only take the local neighborhood of each voter as input (and, arguably, capture the spirit of liquid democracy), cannot provide the foregoing guarantee. By contrast, we design a non-local delegation mechanism that does provably outperform direct voting under mild assumptions about voters.

The above result corroborates a common critique of liquid democracy: Often, a small subset of agents may gain massive influence. To address this, we propose to change the current practice by allowing agents to specify multiple delegation options instead of just one. Then, we seek to control the flow of votes in a way that balances influence as much as possible. Specifically, we analyze the problem of choosing delegations to approximately minimize the maximum number of votes entrusted to any agent, by drawing connections to the literature on confluent flow. We also introduce a random graph model for liquid democracy, and draw on prior work on the *power of choice*, which allows us to establish a doubly exponential separation between the maximum weight of a voter in the case where each voter provides a single delegation and the maximum weight of a voter in the case where each voter provides even two possible delegations [119].

District-Fair Participatory Budgeting (Chapter 2) *Participatory budgeting* is a democratic paradigm in which local governments solicit input from their constituents to make budget decisions about public funds. In practice, it often takes the form of cities asking their residents to vote over a list of public projects to fund. Furthermore, cities often use participatory budgeting on a district level, where each district in a city holds a separate election to spend their portion of the budget (generally allocated proportionally to population). However, district-level elections may yield poor social welfare because no single district has enough money to fund large, widely beneficial projects. On the other hand, making decisions solely based on global social welfare may be unfair to some districts: A social-welfare-maximizing solution may not fund any projects preferred by a particular district.

Thus, we study how to fairly maximize social welfare in participatory budgeting subject to a *district fairness*, which is a fairness constraint that promises each district at least as much utility as it would have received under a district-level participatory budgeting process. We show that, although optimizing social welfare subject to district fairness is NP-hard, we can efficiently construct a lottery over welfare-optimal outcomes that is district-fair in expectation. Moreover, we show that, when we are allowed to slightly relax fairness, we can efficiently compute a deterministic, almost-district-fair solution that is welfare-maximizing, but which may overspend the budget by a small constant factor [125].

Proportionality in Multiwinner Elections (Chapter 3) Finally, we analyze mechanisms for multiwinner elections, where a group of agents, or voters, selects a committee from a set of candidates based on the agents' preferences. In our setting, each agent expresses her preferences through an approval vote, where she designates a subset of candidates she approves for the committee, and all votes are then aggregated to select a winning committee from the pool of candidates.

The property we would like to satisfy is that of proportionality, which intuitively says that a $c \leq 1$ fraction of voters who agree on a c fraction of the alternatives should be able to guarantee themselves control over a c fraction of the committee. We propose a measure of proportionality for elections where the size of the committee is not fixed beforehand and hope to establish a clear separation between the theoretical guarantees of deterministic and randomized rules under this notion of proportionality [108].

Part II: How democracy can help computer science

In the second part of this thesis, we ask how traditional democratic principles can help solve hard problems in computer science. In particular, we explore virtual democracy, which provides a principled approach to automating ethical decision-making, and impartial peer ranking, in which we design mechanisms to leverage individual expertise in settings with conflicts of interest.

In both of these settings, we try to solve problems that do not have clear objective functions to optimize. For instance, when trying to automate ethical decision-making, the very concept of morality differs from person to person depending on their worldview; the same phenomenon holds when evaluating individual expertise and fit for a job. Therefore, standard optimization techniques are not a good fit for these types of problems because we often cannot even identify a widely agreed-upon metric over which to optimize. Instead, we will design mechanisms that directly ask participants for their opinions and then use these opinions in a principled way to make a final decision.

Virtual Democracy (Chapter 4) *Virtual democracy* is an approach to automate decisions by learning models of the preferences of individual people, and, at runtime, aggregating the *predicted* preferences of those people on the dilemma at hand. One of the key questions is which aggregation method — or *voting rule* — to use; we offer a novel statistical viewpoint that provides guidance. Specifically, we seek voting rules that are *robust* to prediction errors, in that their output on people's true preferences is likely to coincide with their output on noisy estimates thereof. We prove that the classic Borda count rule is robust in this sense, whereas any voting rule belonging to the wide family of pairwise-majority consistent (PMC) rules is not [138].

In concert with our theoretical results, we have worked closely with a local nonprofit organization in Pittsburgh, 412 Food Rescue, to build a framework for virtual democracy that enables people to build algorithmic policy for their communities. Through this framework, we study how to design algorithmic policy in order to balance varying interests in a moral, legitimate way. Our findings suggest that the framework successfully enabled participants to build models that they felt confident represented their own beliefs. Participatory algorithm design also improved both procedural fairness and the distributive outcomes of the algorithm, raised participants' algorithmic awareness, and helped identify inconsistencies in human decision-making in the governing organization [155].

Impartial Aggregation (Chapter 5) We study rank aggregation algorithms that take as input the opinions of players over their peers, represented as rankings, and output a social ordering of the players (which reflects, e.g., relative contribution to a project or fit for a job). To prevent strategic behavior, these algorithms must be *impartial*; that is, players should not be able to influence their own position in the output ranking. We design several randomized algorithms that are impartial and closely emulate given (non-impartial) rank aggregation rules in a rigorous sense [136].

We complement these theoretical results with a study of impartial algorithms applied to online labor markets through *HirePeer*, a novel alternative approach to hiring at scale we created that leverages peer assessment to elicit honest assessments of fellow workers' job application materials, which it then aggregates using an impartial ranking algorithm [148]. Surprisingly, we find that applying peer assessment to online hiring—even without impartial ranking algorithms to remove conflicts of interest—was an accurate and pedagogically beneficial practice.

If Internet is the new printing press, then what is democracy for the Internet era? ... How can we get our representatives, our elected representatives, to represent us?

Pia Mancini.

1
Liquid Democracy

In this chapter, we examine theoretical properties of liquid democracy, a democratic paradigm that allows for transitive vote delegation. First, in a model with a ground truth, we demonstrate that a large class of decentralized delegation mechanisms for liquid democracy is susceptible to the concentration of voting power in relatively few voters. We also demonstrate that a simple centralized delegation mechanism for liquid democracy can avoid this problem. Second, informed by the findings above, we turn our sights to centralized delegation mechanisms that effectively reduce the maximum weight of any voter. In particular, we show that allowing voters to give multiple possible delegations leads to a significant reduction in the maximum weight of any voter.

1.1 An Algorithmic Perspective on Liquid Democracy

Liquid democracy is a modern approach to voting in which voters can either vote directly or delegate their vote to other voters. In contrast to the classic *proxy voting* paradigm [173], the key innovation underlying liquid democracy is that proxies—who were selected by voters to vote on their behalf—may delegate their own vote to a proxy, and, in doing so, further delegate all the votes entrusted to them. Put another way (to justify the liquid metaphor), votes may freely flow through the directed delegation graph until they reach a sink, that is, a vertex with outdegree 0. When the election takes place, each voter who did not delegate his vote, but rather voted, is weighted by the total number of votes delegated to him, including his own. In recent years, this approach has been implemented and used on a large scale, notably by eclectic political parties such as the German Pirate Party (Piratenpartei) and Sweden's Demoex (short for Democracy Experiment).

One reason for the success of liquid democracy is that it is seen as a practical com-promise between *direct democracy* (voters vote directly on every issue) and *representative democracy*, and, in a sense, is the best of both worlds. Direct democracy is particularly problematic, as nicely articulated by Green-Armytage [121]:

> "Even if it were possible for every citizen to learn everything they could possibly know about every political issue, people who did this would be able to do little else, and massive amounts of time would be wasted in duplicated effort. Or, if every citizen voted but most people did not take the time to learn about the issues, the results would be highly random and/or highly sensitive to overly simplistic public relations campaigns."

Another example is polling the audience in *Who Wants to Be a Millionaire*, in which the audience would like to help but sometimes gets the question wrong because people who don't know the correct answer systematically favor a specific incorrect answer.

By contrast, under liquid democracy, voters who did not invest an effort to learn about the issue at hand (presumably, most voters) would ideally delegate their votes to well-informed voters. This should intuitively lead to collective decisions that are less random, and more likely to be correct, than those that would be made under direct democracy.

Our goal is to rigorously investigate the intuition that liquid democracy "outperforms" direct democracy from an algorithmic viewpoint. Indeed, we are interested in *delegation mechanisms*, which decide how votes should be delegated based on how relatively informed voters are, and possibly even based on the structure of an underlying social network. Our main research question is:

> *Are there delegation mechanisms that are* guaranteed *to yield more accurate decisions than direct voting?*

1.1.1 Overview of the Model and Results

We focus on a (common) setting where a decision is to be made on a binary issue, i.e., one of two alternatives must be selected. To model the idea of accuracy, we assume that one alternative is correct, and the other is incorrect. Each voter i has a competence level p_i, which is the probability he would vote correctly if he cast a ballot himself.

Voters may delegate their votes to neighbors in a social network, represented as a directed graph. At the heart of our model is the assumption that voters may only delegate their votes to strictly more competent neighbors (and, therefore, there can be no delegation cycles). Specifically, we say that voter i *approves* voter j if $p_j > p_i + \alpha$, for a parameter $\alpha \geq 0$; voters may only delegate to approved neighbors. In defense of this strong assumption, we note that the first of our two theorems—arguably the more interesting of the two—is an *impossibility* result, so assuming that delegation necessarily boosts accuracy only *strengthens* it.

As mentioned above, we are interested in studying *delegation mechanisms*, which decide how votes are delegated (possibly randomly), based on the underlying graph and the approval relation between voters. We pay special attention to *local* delegation mechanisms, which make delegation decisions based only on the neighborhood of each voter. Local mechanisms capture the spirit of liquid democracy in that voters make independent delegation decisions based solely on their own viewpoint, without guidance from a central authority. By contrast, non-local mechanisms intuitively require a centralized algorithm that coordinates delegations.

Recall that our goal is to design delegation mechanisms that are *guaranteed* to be more accurate than direct voting. To this end, we define the *gain* of a mechanism with respect to a given instance as the difference between the probability that it makes a correct decision (when votes are delegated and weighted majority voting is applied) and the probability that direct voting makes a correct decision on the same instance. The desired guarantee can be formalized via two properties of mechanisms: *positive gain (PG)*, which means that there are some sufficiently large instances in which the mechanism has positive gain that is bounded away from 0; and *do no harm (DNH)*, which requires that the loss (negative gain) of the mechanism goes to 0 as the number of voters grows. These properties are both weak; in particular, PG is a truly minimal requirement which, in a sense, mainly rules out direct voting itself as a delegation mechanism.

In Section 1.1.4, we study local delegation mechanisms and establish an impossibility result: such mechanisms cannot satisfy both PG and DNH. In a nutshell, the idea is that for any local delegation mechanism that satisfies PG we can construct an instance where few voters amass a large number of delegated votes, that is, delegation introduces significant *correlation* between the votes. The instance is such that, when the high-weight voters are incorrect, the weighted majority vote is incorrect; yet direct voting is very likely to lead to a correct decision.

In Section 1.1.5, we show that non-local mechanisms can circumvent the foregoing impossibility. Specifically, we design a delegation mechanism, GreedyCap, that satisfies the PG and DNH properties under mild assumptions about voter competencies. It does so by imposing a cap on the number of votes that can be delegated to any particular voter, thereby avoiding excessive correlation.

In conclusion, our work highlights the significance, and potential dangers, of delegating many votes to few voters. Importantly, there is evidence that this can happen in practice. For example, Der Spiegel reported[1] that one member of the German Pirate Party, a linguistics professor at the University of Bamberg, amassed so much weight that his "vote was like a decree." Although recent work by Kling et al. [146] highlights the fact that, in practice, high-weight voters vote reasonably and do not abuse their power, our results corroborate the intuition that this situation should ideally be avoided.

1.1.2 Related Work

There is a significant body of work on delegative democracy and proxy voting [173; 221; 5]. In particular, Cohensius et al. [74] study a model where voters' positions on an issue are points in a metric space. In their version of direct democracy, a small subset of active voters report their positions, and an aggregation method (such as the median or mean when the metric space is the real line) outputs a single position. Under proxy voting,

[1] http://www.spiegel.de/international/germany/liquid-democracy-web-platform-makes-professor-most-powerful-pirate-a-818683.html

each inactive voter delegates his vote to the closest active voter. Cohensius et al. identify conditions under which proxy voting gives a more accurate outcome than direct voting, where the measure is proximity of the outcome to the aggregation method applied to all voters' positions.

To the best of our knowledge, there are only two papers prior to the initial publication of our work that provide theoretical analyses of liquid democracy. The first is the aforementioned paper by Green-Armytage [121]. He considers a setting where, similarly to Cohensius et al. [74], voters are identified with points on the real line, but in his model votes are noisy estimates of those positions. Green-Armytage defines the *expressive loss* of a voter as the squared distance between his vote and his position and proves that delegation (even transitive delegation) can only decrease the expressive loss in his model. He also defines *systematic loss* as the squared distance between the median vote and the median position, but discusses this type of loss only informally (interestingly, he does explicitly mention that correlation can lead to systematic loss in his model).

The second paper is by Christoff and Grossi [72]. They introduce a model of liquid democracy based on the theory of binary aggregation (i.e., their model has a mathematical logic flavor). Their results focus on two problems: the possibility of delegation cycles, and logical inconsistencies that can arise when opinions on interdependent propositions are expressed through proxies. Both are nonissues in our model (although the need to avoid cycles is certainly a concern in practice).

Further afield, there is a rich body of work in computational social choice [47] on the aggregation of objective opinions [76; 78; 96; 236; 235; 163; 192; 14; 15; 169; 64; 65; 193]. As in our work, the high-level goal is to pinpoint the correct outcome based on noisy votes. However, previous work in this area does not encompass any notion of vote delegation.

One seminal result in the aggregation of objective opinions—in particular, when deciding between two options, one of which is correct and the other of which is incorrect—is the Condorcet Jury Theorem [123], which states that if voters are independent and each have probability greater than 1/2 of choosing the correct outcome, then the probability of choosing correctly approaches one as the size of the electorate increases. Note that the Condorcet Jury Theorem is directly applicable to the setting of direct democracy, but not immediately to the (weighted) setting of liquid democracy. Researchers have also studied voting rules in a networked setting, but without delegation, from the perspective of maximum likelihood estimation [81; 80].

There are also several papers that have explored the theoretical foundations of liquid democracy. Notably, a paper by Brill and Talmon [51] considers liquid democracy in the setting of ordinal elections in which the electorate wishes to construct a complete ordering over alternatives, as opposed to deciding a binary issue as in this work. In this framework, each voter may specify a partial ordering over the alternatives and delegate to others in order to construct a complete ranking. However, decisions made by delegates may violate transitivity with respect to each voter's partial ordering, and even checking whether delegated votes satisfy transitivity is NP-hard. In order to circumvent issues of transitivity, they introduce a novel class of voting rules for liquid democracy based on distance rationalization, which take as input (perhaps intransitive) delegation graphs and output the "closest" consensus profile.

Bloembergen et al. [36] consider a game-theoretic version of liquid democracy in which voters must determine whether or not it is rational to delegate their votes to others. They introduce a *delegation game* in which each voter has a hidden true "type" that she knows imperfectly, and the goal of each voter is to communicate her true type to the mechanism either directly (by voting) or indirectly (by delegating). While this setting distills the problem of finding delegates that represent one's own opinion, it focuses on proving the existence of Nash equilibria under certain assumptions and provides only weak performance bounds in the setting we consider.

Finally, Abramowitz and Mattei [2] propose a variant of representative democracy that incorporates ideas of liquid democracy by allowing voters to alter the voting weights of their representatives depending on the issue at hand. Although this circumvents some issues of liquid democracy—for instance, because delegations are no longer transitive, delegation cycles cannot occur—the proposed system is considerably more constrained than general liquid democracy.

1.1.3 The Model

We represent an instance of our problem using a directed, labeled graph $G = (V, E, \vec{p})$. $V = \{1, \ldots, n\}$ is a set of n *voters*, also referred to as *vertices* (we use the two terms interchangeably). E represents a (directed) social network in which the existence of an edge (i, j) means that voter i knows (of) voter j. We denote the neighborhood of voter i to be the set of neighbors that i knows of, or $N_G(i) = \{j \in V : i \text{ knows of } j\}$.

We assume that the voters vote on a binary issue; there is a correct alternative and an incorrect alternative. Each voter $i \in V$ is labeled by his *competence level* p_i. This is the probability that i has the correct opinion about the issue at hand, i.e., the probability that i will vote correctly.

Our setting is also parameterized by $\alpha \in (0, 1)$. Given this parameter and a labeled graph $G = (V, E, \vec{p})$, we define an approval relation between voters: $i \in V$ *approves* $j \in V$ if $(i, j) \in E$ and $p_j \geq p_i + \alpha$. In words, i approves his neighbor j if the difference in their competence levels is at least than α. Note that the approval relation is acyclic because $\alpha > 0$. Denote

$$A_G(i) = \{j \in V : i \text{ approves } j\}.$$

Delegation Mechanisms

The liquid democracy paradigm is implemented through a *delegation mechanism* M, which takes as input a labeled graph G, and outputs, for each voter i, a delegation probability distribution over $A_G(i) \cup \{i\}$ that represents the probability that i will delegate his vote to each of his approved neighbors, or to himself (which means he does not delegate his vote).

To determine whether a delegation mechanism M makes a correct decision on a labeled graph $G = (V, E, \vec{p})$, we use the following 4-step process (which is described in words to avoid introducing notation that will not be used again):

1. Apply M to G to output a delegation probability distribution for each voter i.

2. Sample the probability distribution for each vertex to obtain an acyclic *delegation graph*. Each sink i of the delegation graph (i.e., vertex with no outgoing edges) has weight equal to the number of vertices with directed paths to i, including i itself.
3. Each sink i votes for the correct alternative with probability p_i, and for the incorrect alternative with probability $1 - p_i$.
4. A decision is made based on the weighted majority vote.[2]

We denote the probability that the mechanism M makes a correct decision on graph G via this 4-step process by $P_M(G)$.

Local Mechanisms

We are particularly interested in a special class of delegation mechanisms that we call *local* mechanisms. Intuitively, local mechanisms capture the natural setting where each voter makes an independent delegation decision without central coordination or knowledge of global properties about the delegation graph. Formally, a *local delegation mechanism* is a delegation mechanism such that the probability distribution of each vertex i depends only on (a) the subset $A_G(i)$ of neighbors that i approves, (b) an arbitrary ranking π_i over $A_G(i)$, and (c) $N_G(i)$, or i's neighborhood. Note that the ranking π_i does not have any inherent meaning; it is simply a way to distinguish specific neighbors. In particular, local mechanisms assume that each voter has knowledge of the identities of his approved and non-approved neighbors; a local delegation mechanism is applied to π_i and $N_G(i)$ in order to output a delegation probability distribution for voter i.

For instance, say that in a setting with $\alpha = 0.15$, Alice ($p_{Alice} = 0.6$) has four neighbors: Bob ($p_{Bob} = 0.8$), Carla ($p_{Carla} = 0.9$), Dean ($p_{Dean} = 0.5$), and Evelyn ($p_{Evelyn} = 0.7$). Alice approves of Bob and Carla, and let π_{Alice} = Carla $\succ$ Bob. Then, the local delegation mechanism takes π_{Alice} and the set of Alice's neighbors, and returns a probability distribution over delegating to Bob, delegating to Carla, and voting directly.

Let us give some examples of local delegation mechanisms:

- Voters do not delegate their votes. This *direct voting* mechanism plays a special role in our model, and we denote it by D.
- Each voter delegates his vote to a random approved neighbor, if he has any.
- Each voter delegates his vote to a random approved neighbor, if he has approved neighbors but has even more non-approved neighbors.
- Each voter delegates his vote deterministically to a single approved neighbor (e.g., the first in his local ordering π_i), if he has any. The ranking π_i is needed only in order to enable this type of mechanism.

By contrast, the following delegation mechanisms are not local:

- Each voter delegates his vote to his most competent approved neighbor. (Voters cannot distinguish between their approved neighbors, except through the information given by the "arbitrary" ranking π_i.)

[2] Ties can be broken arbitrarily.

- Let there exist a distinguished voter with global identifier V_1. If V_1 appears in the approval set of any voter, that voter delegates to V_1 with probability 1.
- Each voter delegates his vote only if all agents in his approval set have global identifiers that are even integers.

Desiderata

Recall that we are interested in comparing the likelihood of making correct decisions via delegative voting with that of direct voting. To this end, define the *gain* of delegation mechanism M on labeled graph G as

$$\text{gain}(M, G) = P_M(G) - P_D(G).$$

We would like to design delegation mechanisms that have positive gain (bounded away from zero) in some situations, and which never lose significantly to direct voting. Formally, we are interested in the following two desirable axioms:

- A mechanism M satisfies the *positive gain (PG)* property if there exist $\gamma > 0, n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ there exists a graph G_n on n vertices such that $\text{gain}(M, G_n) \geq \gamma$.
- A mechanism M satisfies the *do no harm (DNH)* property if for all $\varepsilon > 0$, there exists $n_1 \in \mathbb{N}$ such that for all graphs G_n on $n \geq n_1$ vertices, $\text{gain}(M, G_n) \geq -\varepsilon$.

The choice of quantifiers here is of great significance. PG asks for the *existence* of (large enough) instances where the gain is at least γ, for a constant γ. By contrast, DNH essentially requires that any loss would go to 0 as the size of the graph goes to infinity. That is, there may certainly be small instances where delegative voting loses out to direct voting, but that should not be the case in the large.

We note that PG and DNH are defined over worst-case instances. Another natural question to ask is about the expected gain of delegation mechanisms: For a random graph and choice of competence levels, is a given mechanism expected to outperform direct voting? However, we leave this to future work.

1.1.4 Impossibility for Local Mechanisms

In our model, we make the strong assumption that voters can only delegate their vote to other voters who are more competent than they are, and, in particular, delegation chains can significantly boost the competence of any particular vote. Under this assumption, it seems natural to expect that delegative voting will always do at least as well as direct voting in every situation, and strictly better in some situations. This should intuitively be true under local mechanisms, say, when each voter delegates his vote to an arbitrary approved neighbor (if he has any). The following example helps build intuition for what can go wrong.

Example 1. *Consider the labeled graph $G_n = (V, E, \vec{p})$ over n vertices, where $E = \{(i, 1) : i \in V \setminus \{1\}\}$, i.e., G is a star with 1 at the center. Moreover, $p_1 = 4/5$, $p_i = 2/3$ for all*

$i \in V \setminus \{1\}$, and $\alpha = 1/10$. Then, as n grows larger, $P_D(G_n)$ goes to 1 by the Law of Large Numbers, or, equivalently, by the Condorcet Jury Theorem [123]. By contrast, all leaves approve the center, and a naïve local delegation mechanism M would delegate all their votes. In that case, the decision depends only on the vote of the center, so $P_M(G_n) = 4/5$ for all $n \in \mathbb{N}$, and gain(M, G_n) converges to $-1/5$. We conclude that M violates the DNH property.

One might hope that there are "smarter" local delegation mechanisms, that, say, recognize that when a voter only has one approved neighbor, his vote should not be delegated. However, our first result shows that this is not the case: local delegation mechanisms cannot even satisfy the two minimal requirements of PG and DNH.

Theorem 1.1. *For any $\alpha_0 \in (0,1)$ such that $i \in V$* approves *$j \in V$ if $(i,j) \in E$ and $p_j > p_i + \alpha_0$, there is no local mechanism that satisfies the PG and DNH properties.*

The first step in the proof is to better understand the way in which local mechanisms are constrained. This is captured by the following lemma.

Lemma 1.2. *Let M be a local mechanism. Then M satisfies the PG property only if there exist $k, m, \rho > 0$ such that, if a voter approves k out of his m total neighbors, then the total probability of delegation to any of these approved neighbors is exactly ρ.*

Proof. Suppose that PG holds. Let $\gamma > 0$ and fix a labeled graph G such that gain$(M, G) \geq \gamma > 0$. In order for this to be the case, there must exist some vertex i that delegates with positive probability ρ. Let k be the number of neighbors in G that i approves, and let m be his total number of neighbors in G; this yields the desired tuple (k, m, ρ).[3] □

The crux of the theorem's proof is the construction of a graph that, from the local viewpoint of many of the vertices, looks like the neighborhood prescribed by Lemma 1.2. Specifically, a *k-center m-uniform star* consists of vertices called *leaves* that are each connected to k central vertices (the centers) as well as $m-k$ other leaves. Each leaf vertex has competence level p_ℓ, and each center vertex has competence level p_c, such that $p_c > p_\ell + \alpha$. We set the value of k and m to be the values whose existence is guaranteed by Lemma 1.2, which means that the construction of a k-center m-uniform star satisfies the property that each leaf delegates to some center vertex with probability ρ. Throughout the proof, we will let $n_c = k$ be the number of centers, and n_ℓ will denote the number of leaves.

At a high level, we show that the loss of any local mechanism can approach $(1 - p_c)^k$, which is constant given k. We do this by constructing a graph that consists of a k-center m-uniform star with an independent disconnected component consisting of n_d vertices of competence level p_d. We set the parameters so that the direct voting mechanism D decides correctly with high probability. By contrast, under the local delegation mechanism M, enough leaves delegate their votes to the centers so that if all centers were to vote incorrectly, which happens with probability $(1 - p_c)^k$, then M would decide incorrectly. While the basic idea is simple enough, the formal construction is quite delicate, as many different parameters must be carefully balanced.

[3] Note that the conclusion is invariant to the ranking π_i.

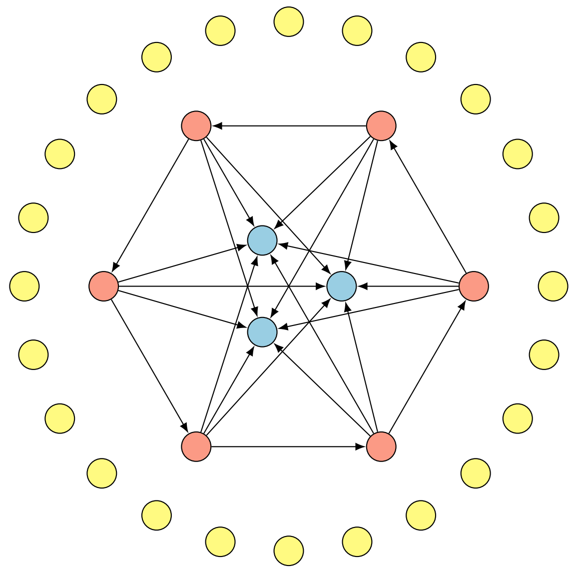


Figure 1.1: Graph G for $n_\ell = 6$ leaves (shown in red), $n_c = 3$ centers (shown in blue), $n_d = 24$ disconnected vertices (shown in yellow), and $m = 4$.

Proof of Theorem 1.1. Let M be a local mechanism that satisfies PG. By Lemma 1.2, there must exist at least one (k, m, ρ) tuple for M that satisfies the lemma's conclusion. For any n_1 prescribed by DNH and any α_0, we can construct a graph G_n with $n \geq n_1$ such that DNH does not hold.

Let G be a graph of size $n = n_c + n_\ell + n_d$ that consists of a k-center m-uniform star and a disconnected component containing n_d disconnected points (see Figure 1.1). Each center has competence level p_c, each leaf in the star has competence level p_ℓ, and each point in the disconnected component has competence level p_d. Given (k, m, ρ), n_1, and α_0, note that the following constraints must hold.

$$n_\ell \geq m - n_c \tag{1.1}$$

$$n = n_\ell + n_c + n_d \geq n_1 \tag{1.2}$$

$$p_c > p_\ell + \alpha_0 \tag{1.3}$$

We will prove that the construction above instantiated with the following parameter values violates DNH for any input of (k, m, ρ), n_1, and $\alpha = \alpha_0 + \varepsilon'$ for $\varepsilon' = \frac{1-\alpha_0}{2} > 0$, for sufficiently small δ (i.e., as $\delta \to 0$).

We begin by defining the sizes of each component: n_c, n_ℓ, and n_d.

$$n_c = k \tag{1.4}$$

$$n_\ell = \frac{n_1 m}{\alpha\delta} \tag{1.5}$$

$$n_d = C_1 \frac{n_1 m}{\alpha\delta} \tag{1.6}$$

Note that n_d depends on a constant C_1, which, along with another constant σ, is defined next.

$$C_1 = \frac{\left(\frac{\left(\frac{p_\ell \rho}{\sigma} n_\ell - p_\ell \sqrt{n_\ell}\right)}{2}\right)^2 - n_c}{n_\ell} - 1 \tag{1.7}$$

$$\sigma = \sqrt{-\frac{\ln\left(\frac{\delta}{2}\right)}{2}} \tag{1.8}$$

Now, we define the competency values for each component, p_c, p_ℓ, and p_d. Note that there is a range of acceptable compentency values for p_d.

$$p_c = \frac{1+\alpha}{2} \tag{1.9}$$

$$p_\ell = \frac{1-\alpha}{2} \tag{1.10}$$

$$p_d \in \left[\left(\frac{n/2 - n_\ell p_\ell}{n_d}\right) + \frac{\sigma\sqrt{n}}{n_d}, \left(\frac{n/2 - n_\ell p_\ell}{n_d}\right) + \frac{(n_\ell \rho - \tau)p_\ell - \sigma\sqrt{n}}{n_d}\right] \tag{1.11}$$

Finally, we define τ, another constant that will be useful for establishing concentration guarantees in the proof.

$$\tau = \sqrt{-\frac{\left(\ln\frac{\delta}{2}\right)n_\ell}{2}} \tag{1.12}$$

The following claim asserts that the construction is feasible.

Claim 1. *$C_1 > 0$ and the range of values for p_d in (1.11) is nonempty.*

Proof. From above, we have

$$C_1 = \frac{\left(\frac{\left(\frac{p_\ell \rho}{\sigma} n_\ell - p_\ell \sqrt{n_\ell}\right)}{2}\right)^2 - k}{n_\ell} - 1$$

and rearranging terms yields

$$2\sqrt{(C_1+1)n_\ell + k} = \frac{p_\ell \rho}{\sigma} n_\ell - p_\ell \sqrt{n_\ell}.$$

Now, note that $n_d = C_1 n_\ell$ and therefore $(C_1+1)n_\ell + k = n_d + n_\ell + k = n$. Additionally, note that $\sqrt{n_\ell} = \frac{\tau}{\sigma}$. Substituting this in, we have

$$2\sqrt{n} = \frac{p_\ell \rho n_\ell - p_\ell \tau}{\sigma}$$

and therefore

$$\sigma\sqrt{n} = p_\ell(\rho n_\ell - \tau) - \sigma\sqrt{n}. \tag{1.13}$$

Now, we note that $\sigma\sqrt{n} - kp_c < \sigma\sqrt{n}$ and

$$\sqrt{n} > \sqrt{n - k - (\rho n_\ell - \tau)}$$

because both k and $\rho n_\ell - \tau = 2\sigma\sqrt{n}/p_\ell$ are greater than 0, and $n - k - (\rho n_\ell - \tau) > n - k - n_\ell \geq 0$. Now, from (1.13), we can conclude that

$$\begin{aligned}\sigma\sqrt{n} - kp_c < \sigma\sqrt{n} &= p_\ell(\rho n_\ell - \tau) - \sigma\sqrt{n}\\ &< (\rho n_\ell - \tau)p_\ell - \sigma\sqrt{n - k - (\rho n_\ell - \tau)},\end{aligned}$$

which means

$$p_d \in \left[\left(\frac{n/2 - n_\ell p_\ell}{n_d}\right) + \frac{\sigma\sqrt{n} - kp_c}{n_d}, \left(\frac{n/2 - n_\ell p_\ell}{n_d}\right) + \frac{(n_\ell\rho - \tau)p_\ell - \sigma\sqrt{n - k - (n_\ell\rho - \tau)}}{n_d}\right]$$

is non-empty, and our value for p_d is admissible.

Lastly, we have to show that C_1 is itself admissible; i.e., that the following holds:

$$\frac{\left(\frac{\left(\frac{p_\ell\rho}{\sigma}n_\ell - p_\ell\sqrt{n_\ell}\right)}{2}\right)^2 - k}{n_\ell} - 1 > 0.$$

Rearranging and expanding, we obtain

$$\frac{p_\ell n_\ell \rho}{\sigma} - p_\ell\sqrt{n_\ell} \geq 2\sqrt{n_\ell + k}.$$

Now, note that both sides are positive as $\delta \to 0$. Indeed, the right hand side consists of positive terms and the left hand side simplifies to $p_\ell\sqrt{n_\ell}(\rho\sqrt{n_\ell}/\sigma - 1)$, which is positive iff $\rho\sqrt{n_\ell} > \sigma$, which is true as $\delta \to 0$ because $1/\delta$ grows more quickly than $\ln(2/\delta)$. Therefore, squaring both sides yields

$$\left(\frac{p_\ell n_\ell \rho}{\sigma}\right)^2 + (p_\ell)^2 n_\ell - 2\frac{(p_\ell)^2\rho(n_\ell)^{3/2}}{\sigma} \geq 4(n_\ell + k).$$

Now, substituting in our value for n_ℓ, we obtain

$$\left[\left(\frac{p_\ell\rho}{\sigma}\right)\left(\frac{n_1 m}{\alpha\delta}\right)\right]^2 + (p_\ell)^2\left(\frac{n_1 m}{\alpha\delta}\right) - \frac{2(p_\ell)^2\rho}{\sigma}\left(\frac{n_1 m}{\alpha\delta}\right)^{3/2} - 4\left(\frac{n_1 m}{\alpha\delta}\right) - 4k. \tag{1.14}$$

As $\delta \to 0$, (1.14) becomes dominated by the highest-order $1/\delta$ term, and therefore is always positive for any assignment to the other variables because the rest of them are constrained to be strictly positive. □

Because $\alpha, \delta \in (0, 1)$, the value of n_ℓ in (1.5) is greater than both n_1 and m, hence constraints (1.1) and (1.2) are immediately satisfied. Moreover, constraint (1.3) is satisfied by (1.9) and (1.10).

Turning to the proof that DNH is violated, let cor_D, del_M, and $nondel_M$ be the random variables corresponding to the number of correct votes under D, the number of delegated correct votes under M, and the number of non-delegated correct votes under M. Additionally, let ε, τ (defined again below), and ξ be as follows.

$$\varepsilon = \sqrt{-\frac{\left(\ln \frac{\delta}{2}\right) n}{2}},$$
$$\tau = \sqrt{-\frac{\left(\ln \frac{\delta}{2}\right) n_\ell}{2}}, \text{ and}$$
$$\xi = \sqrt{-\frac{\left(\ln \frac{\delta}{2}\right) (n - n_c - (\rho n_\ell - \tau))}{2}}.$$

Our goal is to bound the expectations of cor_D, del_M, and $nondel_M$. First, we examine $\mathbb{E}[cor_D]$. We would like to show that

$$\mathbb{E}[cor_D] \geq n/2 + \varepsilon. \tag{1.15}$$

Expanding out the expected value, this is equivalent to

$$p_c n_c + p_\ell n_\ell + p_d n_d \geq n/2 + \varepsilon.$$

From (1.11), we have

$$p_d \geq \frac{n/2 - p_\ell n_\ell + \varepsilon}{n_d},$$

so it is sufficient to show that

$$p_c n_c + p_\ell n_\ell + n_d \left(\frac{n/2 - p_\ell n_\ell + \varepsilon}{n_d}\right) \geq n/2 + \varepsilon,$$

and simplifying yields $p_c n_c \geq 0$. This is true by Equation (1.9), because α and k are both constrained to be strictly positive.

Next, we examine $\mathbb{E}[del_M]$. We would like to show that

$$\mathbb{E}[del_M] = n_\ell \rho. \tag{1.16}$$

This is trivial to see, as del_M is a sum of n_ℓ Bernoulli random variables with "success" probability ρ.

Finally, we examine the "typical case" over $nondel_M$, or $\mathbb{E}[nondel_M | del_M = v]$ for all integers $v \in [n_\ell \rho - \tau, n_\ell \rho + \tau]$. Intuitively, this case considers the number of correct votes cast by still-independent vertices after "enough" leaf vertices have delegated their votes. If these votes do not make up a majority, then all centers voting incorrectly will cause the entire graph to vote incorrectly. We would like to show that

$$\mathbb{E}[nondel_M | del_M = v] \leq n/2 - \xi. \tag{1.17}$$

for all integers $v \in [n_\ell \rho - \tau, n_\ell \rho + \tau]$. Conditionally on del_M being in the prescribed range above, we see that in the worst case, $del_M = n_\ell \rho - \tau$, meaning the fewest possible voters delegate under this assumption. Given this, we would like to show that

$$p_d n_d + p_\ell(n_\ell - (\rho n_\ell - \tau)) \leq n/2 - \xi.$$

From Equation (1.11) we have

$$p_d \leq \frac{n/2 - p_\ell n_\ell + (n_\ell \rho - \tau)p_\ell - \xi}{n_d},$$

which yields

$$\left(\frac{n/2 - p_\ell n_\ell + (n_\ell \rho - \tau)p_\ell - \xi}{n_d}\right) n_d + p_\ell(n_\ell - (\rho n_\ell - \tau)) \\ \leq n/2 - \xi.$$

Simplifying yields $0 \leq 0$—a tautology. This establishes Equation (1.17).

We now wish to bound the probability of cor_D, del_M, and $nondel_M$ deviating by too much. We use Hoeffding's inequality [127], which states that given n independent Bernoulli random variables $X_i \in [0,1]$ and $X = \sum_i X_i$, the following concentration bound holds:

$$\Pr\left[|X - \mathbb{E}[X]| \geq \varepsilon\right] \leq 2\exp\left(\frac{-2\varepsilon^2}{n}\right). \tag{1.18}$$

First, we examine cor_D. From (1.18) and a straightforward substitution for ε, we obtain

$$\begin{aligned}\Pr\left(|cor_D - \mathbb{E}[cor_D]| \geq \varepsilon\right) &\leq 2\exp\left(\frac{-2\varepsilon^2}{n}\right) \\ &= 2\exp\left(-\frac{2\left[\sqrt{-\frac{\left(\ln\frac{\delta}{2}\right)n}{2}}\right]^2}{n}\right) \\ &= \delta.\end{aligned} \tag{1.19}$$

Likewise, for del_M, from (1.18) and a straightforward substitution for τ, we obtain

$$\begin{aligned}\Pr\left[|del_M - \mathbb{E}[del_M]| \geq \tau\right] &\leq 2\exp\left(\frac{-2\tau^2}{n_\ell}\right) \\ &= 2\exp\left(-\frac{2\left[\sqrt{-\frac{\left(\ln\frac{\delta}{2}\right)n_\ell}{2}}\right]^2}{n_\ell}\right) \\ &= \delta.\end{aligned} \tag{1.20}$$

Finally, for $nondel_M$, we are interested in upper-bounding

$$\Pr[|nondel_M - \mathbb{E}[nondel_M | del_M = v]| \geq \xi \mid del_M = v],$$

for every integer $v \in [n_\ell \rho - \tau, n_\ell \rho + \tau]$. As before, we apply Equation (1.18), and, as it turns out, we can derive an upper bound when $del_M = n_\ell \rho - \tau$. Therefore, we obtain that for every $v \in [n_\ell \rho - \tau, n_\ell \rho + \tau]$,

$$\begin{aligned}
&\Pr\left[|nondel_M - \mathbb{E}[nondel_M | del_M = v]| \geq \xi \mid del_M = v\right] \\
&\quad \leq 2 \exp\left(\frac{-2\xi^2}{n - n_c - (\rho n_\ell - \tau)}\right) \\
&\quad = 2 \exp\left(-\frac{2\left[\sqrt{-\frac{\left(\ln \frac{\delta}{2}\right)(n - n_c - (\rho n_\ell - \tau))}{2}}\right]^2}{n - n_c - (\rho n_\ell - \tau)}\right) \\
&\quad = \delta,
\end{aligned} \tag{1.21}$$

where the denominator comes from the (worst-case) total number of non-delegated votes under M.

From the above, we see that

$$\begin{aligned}
\Pr[cor_D > n/2] &\geq 1 - \delta, && \text{(by (1.15) and (1.19))} \\
\Pr[del_M \in (n_\ell \rho - \tau, n_\ell \rho + \tau)] &\geq 1 - \delta, && \text{(by (1.16) and (1.20))} \\
\Pr[nondel_M < n/2 \mid del_M = v] &\geq 1 - \delta, && \text{(by (1.17) and (1.21))}
\end{aligned}$$

where the last inequality holds for all integers $v \in [n_\ell \rho - \tau, n_\ell \rho + \tau]$.

Therefore, the lower bound on the probability of D deciding correctly is $p_d(G) \geq 1 - \delta$. We can lower-bound the probability of M deciding incorrectly in order to upper-bound $P_M(G)$. We slightly overload notation and let M be the event that M decides correctly, and $\neg M$ be the event that M decides incorrectly. Moreover, denote by V the event that $del_M \in [n_\ell \rho - \tau, n_\ell \rho + \tau]$. By definition, we have

$$\Pr[\neg M] = \Pr[\neg M | V] \Pr[V] + \Pr[\neg M | \neg V] \Pr[\neg V],$$

and because probabilities cannot be negative,

$$\Pr[\neg M] \geq \Pr[\neg M | V] \Pr[V].$$

Now, because $\Pr[V] \geq 1 - \delta$,

$$\Pr[\neg M] \geq \Pr[\neg M | V](1 - \delta).$$

Furthermore, we know that $\Pr[\neg M | V]$ is also lower-bounded by $(1 - p_c)^{n_c}(1 - \delta)$ because one setting under which M decides incorrectly is exactly when all centers vote incorrectly and $nondel_M < n/2$. It follows that

$$\Pr[\neg M] \geq (1 - p_c)^{n_c}(1 - \delta)(1 - \delta).$$

Therefore, taking the complement, we have an upper bound on the probability of M voting correctly of

$$\Pr[M] \leq 1 - (1 - p_c)^{n_c}(1 - \delta)^2,$$

and the total loss can be lower-bounded by

$$(1 - \delta) - (1 - (1 - p_c)^{n_c}(1 - \delta)^2) = (1 - p_c)^{n_c}(1 - \delta)^2 - \delta.$$

As $\delta \to 0$, this tends to $(1 - p_c)^{n_c} = (1 - p_c)^k$, which is constant and bounded away from 0. We conclude that M violates the DNH property. $\square$

We note that even if each voter had access to a ranking of his approved neighbors by competence, this impossibility still holds because the construction is such that all approved vertices have equal competence.

1.1.5 Possibility for Non-Local Mechanisms

The main idea underlying Theorem 1.1 is that liquid democracy can correlate the votes to the point where the mistakes of a few popular voters tip the scales in the wrong direction. As we show in the theorem's proof, this is unavoidable under local delegation mechanisms, which, intuitively, cannot identify situations in which certain voters amass a large number of votes. However, non-local delegation mechanisms can circumvent this issue. Indeed, consider the following delegation mechanism.

input: labeled graph G with n vertices, cap $C : \mathbb{N} \to \mathbb{N}$

$V' \leftarrow V$
while $V' \neq \emptyset$ **do**
  let $i \in \mathrm{argmax}_{j \in V'} |A_G^{-1}(j) \cap V'|$
  $J \leftarrow A_G^{-1}(i) \cap V'$
  if $|J| \leq C(n) - 1$ **then**
    $J' \leftarrow J$
  else
    let $J' \subseteq J$ such that $|J'| = C(n) - 1$
  end if
  vertices in J' delegate to i
  $V' \leftarrow V' \setminus (\{i\} \cup \{J'\})$
end while

Algorithm 1: GreedyCap

In words, the mechanism GreedyCap, given as Algorithm 1, receives as input a labeled graph G, and a *cap* C which is a function of n. It iteratively selects a voter with maximum approvals, and delegates votes to him, so that no more than $C(n) - 1$ votes are delegated to a single voter (that is, no voter can have weight more than $C(n)$). All voters involved in the current iteration are then eliminated from further consideration, which is why delegations under this mechanism are only 1-hop.

It is obvious that GreedyCap satisfies the PG property. Intuitively, for any value of α, it is always possible to construct large instances of graphs where a few voters delegate to more competent voters in a way that increases the probability of making the correct decision overall. However, although it seems at first glance that it should satisfy DNH as well (as it solves the excessive correlation problem), the following example shows that, without further assumptions, it does not.

Example 2. *Assume for ease of exposition that $\alpha < 1/3$. For any odd $n = 2k+1$, consider the labeled graph $G_n = (V, E, \vec{p})$ on n vertices, defined as follows: $E = \{(1,2)\}$ (i.e., the only edge in the graph is from 1 to 2), $p_1 = 1/3$, $p_2 = 2/3$, there are k vertices with $p_i = 1$, and $k-1$ vertices with $p_i = 0$. Even if $C(n) \equiv 2$, GreedyCap would delegate the vote of voter 1 to voter 2. Therefore, the mechanism decides correctly if and only if voter 2 votes correctly, which happens with probability 2/3. By contrast, under direct voting, it is enough for either voter 1 or voter 2 to vote correctly, which happens with probability 7/9. It follows that the loss of GreedyCap is 1/9—a constant. We conclude that GreedyCap violates DNH.*

The reason the example works is that the outcome completely depends on voters 1 and 2, as the others vote deterministically (competence level 0 or 1). To avoid this problem, we make the natural assumption that competence levels are bounded away from 0 and 1, i.e., voters are never horribly misinformed or perfectly informed. It turns out that this additional assumption is sufficient to guarantee that GreedyCap satisfies the DNH property.

Theorem 1.3. *Assume that there exists $\beta \in (0, 1/2)$ such that all competence levels are in $[\beta, 1-\beta]$. Then for any difference in competencies $\alpha \in (0, 1-2\beta)$, GreedyCap with cap $C : \mathbb{N} \to \mathbb{N}$ such that $C(n) \in \omega(1)$ and $C(n) \in o(\sqrt{\log n})$ satisfies the PG and DNH properties.*

We begin with a proof sketch, focusing on the DNH property (as PG is rather simple). Given n voters, we denote the number of correct votes under direct voting and GreedyCap by X_D and X_M, respectively, and consider two cases.

1. $|\mathbb{E}[X_D] - \frac{n}{2}| > \frac{n}{\log n}$.
2. $|\mathbb{E}[X_D] - \frac{n}{2}| \leq \frac{n}{\log n}$.

In Case 1, the direct voting mechanism has mean far away from $n/2$. When $\mathbb{E}[X_D] < n/2 - n/\log n$, we can show that P_D goes to 0 as n goes to infinity. This means that DNH is satisfied for any value of P_M. In the case where $\mathbb{E}[X_D] > n/2 + n/\log n$, we can show that P_M goes to 1 as n goes to infinity, which means that DNH is satisfied for any value of P_D.

In Case 2, the direct voting setting has mean close to $n/2$. From here, we consider two subcases.

1. The number of voters who delegate is greater than $n/g(n)$, where $g(n) \in o(\log n)$ and $g(n) \in \omega(C(n)^2)$. Note that this yields—hence the upper bound on $C(n)$ in the statement of Theorem 1.3.
2. The number of voters who delegate is at most $n/g(n)$.

In Subcase 1, because a relatively large fraction of voters delegate their votes to more competent neighbors, $\mathbb{E}[X_M] - \mathbb{E}[X_D]$ is large enough to offset the simultaneous increase in the variance of X_M, and, in the limit, P_M goes to 1. In Subcase 2, we again have $\mathbb{E}[X_M] \geq \mathbb{E}[X_D]$ due to delegation. Additionally, because so few voters delegate, the ratio of the variance of X_M and that of X_D converges to 1 as n approaches infinity, which means that (in the worst case) the difference between P_D and P_M converges to 0.

Before presenting the theorem's detailed proof, we establish three useful lemmas that establish uniform convergence as n grows large.

The first lemma is the Lindeberg Central Limit Theorem [161], reproduced below.

Lemma 1.4 (Lindeberg Central Limit Theorem [161])**.** *Let $\{X_{ni} : n \geq 1; i = 1, \ldots, k_n\}$ be a triangular array of random variables with $X_{n1}, \ldots, X_{nk_n}$ independent for each n, $\mathbb{E}[X_{ni}] = 0$, and $\sum_{i=1}^{k_n} \mathbb{E}[X_{ni}^2] = 1$. If for each fixed $\epsilon > 0$,*

$$\lim_{n\to\infty} \sum_{i=1}^{k_n} \mathbb{E}\left[X_{ni}^2 \mathbb{1}\{|X_{ni}| > \epsilon\}\right] = 0,$$

where $\mathbb{1}$ is an indicator random variable, then

$$Z_n = \sum_{i=1}^{k_n} X_{ni} \to N(0, 1).$$

In our proof, we will use the Lindeberg Central Limit Theorem to prove the following key lemma, which states that we can treat arbitrary instances of liquid democracy like normal distributions as the number of voters increases.

Lemma 1.5. *For all $\beta \in (0, 1/2)$ and $C(n) \in o(\sqrt{n})$, for all $\epsilon > 0$, there is a constant $n_2 \in \mathbb{N}$ such that for all $n \geq n_2$, given arbitrary competencies $p_1, \ldots, p_n$ where each $p_i \in (\beta, 1-\beta)$ for $\beta \in (0, 1/2)$ and weights $b_1, \ldots, b_n$ such that each $b_i \in [0, C(n)]$ and $\sum_{i=1}^{n} b_i = n$, we have*

$$\left| \Pr[X > n/2] - \Phi\left(\frac{\mathbb{E}[X] - n/2}{\sqrt{\mathrm{Var}[X]}}\right) \right| < \epsilon,$$

where X represents the number of correct votes in liquid democracy with these competencies and weights, and $\mathbb{E}[X] = \sum_{i=1}^{n} b_i \cdot p_i$ and $\mathrm{Var}[X] = \sum_{i=1}^{n} b_i^2 p_i(1 - p_i)$.

Proof. Let $\{Y_{nk} : 1 \leq k \leq n\}$ be a triangular array of independent Bernoulli random variables where Y_{ni} has success probability $p_{nk} \in [\beta, 1 - \beta]$ for $\beta \in (0, 1/2)$. Furthermore, define $\{b_{nk} : 1 \leq k \leq n\}$ be a triangular array of nonnegative integers such that $0 \leq b_{nk} \leq C(n)$ for all $1 \leq k \leq n$ and $\sum_{k=1}^{n} b_{nk} = n$. Lastly, define the triangular array $X_{nk} \coloneqq b_{nk} \cdot Y_{nk}$.

We will first show that the sum of all the X_{nk} random variables is approximately normal, or

$$Z_n = \sum_{k=1}^{n} X_{nk} \to Z \sim N\left(\sum_{k=1}^{n} b_{nk}\, \mathbb{E}[Y_k], \sum_{k=1}^{n} b_{nk}^2\, \mathrm{Var}[Y_k]\right). \tag{1.22}$$

In order to show Equation (1.22), define $s_n^2 = \sum_{k=1}^n \operatorname{Var}[X_{nk}] = \sum_{k=1}^n b_{nk}^2 \operatorname{Var}[Y_{nk}]$. Let

$$W_{nk} := \frac{X_{nk} - \mathbb{E}[X_{nk}]}{s_n} = \frac{X_{nk} - \mathbb{E}[X_{nk}]}{\sqrt{\sum_{k=1}^n b_{nk}^2 \operatorname{Var}[Y_{nk}]}}.$$

We must first show that the triangular array W_{nk} satisfies the preconditions of Lemma 1.4. Note that $X_{n1}, \ldots, X_{nn}$ are independent because the $\{Y_{nk}\}$ are independent. Also, by definition, $\mathbb{E}[W_{nk}] = 0$. Furthermore, we have

$$\begin{aligned}
\sum_{k=1}^n \mathbb{E}[W_{nk}^2] &= \sum_{k=1}^n \mathbb{E}\left[\frac{(X_{nk} - \mathbb{E}[X_{nk}])^2}{s_n^2}\right] \\
&= \frac{1}{s_n^2} \sum_{k=1}^n \mathbb{E}[(X_{nk} - \mathbb{E}[X_{nk}])^2] \\
&= \frac{1}{s_n^2} \sum_{k=1}^n b_{nk}^2 \, \mathbb{E}[(Y_{nk} - \mathbb{E}[Y_{nk}])^2] \\
&= \frac{1}{\sum_{k=1}^n b_{nk}^2 \operatorname{Var}[Y_{nk}]} \sum_{k=1}^n b_{nk}^2 \operatorname{Var}[Y_{nk}] \\
&= 1.
\end{aligned}$$

Now, we must check whether the last precondition holds, i.e., whether

$$\lim_{n\to\infty} \sum_{k=1}^n \mathbb{E}\left[W_{nk}^2 \mathbb{1}\{|W_{nk}| > \epsilon\}\right] = 0.$$

We have

$$\begin{aligned}
\lim_{n\to\infty} \sum_{k=1}^n \mathbb{E}\left[W_{nk}^2 \mathbb{1}\{|W_{nk}| > \epsilon\}\right] &= \lim_{n\to\infty} \sum_{k=1}^n \mathbb{E}\left[\frac{(X_{nk} - \mathbb{E}[X_{nk}])^2}{s_n^2} \mathbb{1}\left\{|b_{nk}(Y_{nk} - \mathbb{E}[Y_{nk}])| > \epsilon \cdot s_n\right\}\right] \\
&= \lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^n b_{nk}^2 \, \mathbb{E}\left[(Y_{nk} - \mathbb{E}[Y_{nk}])^2 \mathbb{1}\left\{|b_{nk}(Y_{nk} - \mathbb{E}[Y_{nk}])| > \epsilon \cdot s_n\right\}\right].
\end{aligned}$$

Let us now examine the $\mathbb{1}\left\{|b_{nk}(Y_{nk} - \mathbb{E}[Y_{nk}])| > \epsilon \cdot s_n\right\}$ term. On the left side of the inequality, we have that

$$|b_{nk}(Y_{nk} - \mathbb{E}[Y_{nk}])| \in o(\sqrt{n})$$

because $b_{nk} \in o(\sqrt{n})$, and $|Y_{nk} - \mathbb{E}[Y_{nk}]| \leq \max(\beta, 1-\beta) \leq 1 - \beta$ because we consider $\beta \in (0, 1/2)$. Furthermore, on the right side, we have

$$\begin{aligned}
s_n &= \sqrt{\sum_{k=1}^n b_{nk}^2 p_{nk}(1 - p_{nk})} \\
&\geq \sqrt{\sum_{k=1}^n b_{nk}^2 \beta(1 - \beta)} && (p_{nk} \in (\beta, 1-\beta)) \\
&\geq \sqrt{n\beta(1-\beta)} && (\textstyle\sum_{k=1}^n b_{nk} = n) \\
&\in \Omega(\sqrt{n}).
\end{aligned}$$

Therefore, as n approaches infinity, the probability that the left hand side exceeds the right hand side goes to 0, satisfying the final precondition. We may now apply the Lindeberg Central Limit Theorem to see that

$$Z'_n = \sum_{k=1}^{n} W_{nk} \to N(0,1) \text{ as } n \to \infty.$$

By straightforward scaling and shifting arguments, as above, we see that this implies that

$$Z_n = \sum_{k=1}^{n} X_{nk} \to N\left(\sum_{k=1}^{n} b_{nk}\,\mathbb{E}[Y_k], \sum_{k=1}^{n} b_{nk}^2\,\mathrm{Var}[Y_k]\right) \text{ as } n \to \infty,$$

as desired.

This means that all instances with arbitrary competence and weight values approach a normal distribution as the size of the instance grows. We now must show uniform convergence. In order to do so, first assume toward a contradiction that there exists an ϵ such that for infinitely many n, there exist b_n and p_n such that the resulting instance is not within ϵ of normal. Then, we may construct a sequence including all of these failing instances, which would violate Equation (1.22). $\square$

We also require the following simple lemma.

Lemma 1.6. *Given a normally distributed variable $X \sim \mathcal{N}(\mathbb{E}[X], \mathrm{Var}[X])$ with $\mathbb{E}[X] \in [\mu_{min}, \mu_{max}]$ and $\mathrm{Var}[X] \in [\sigma^2_{min}, \sigma^2_{max}]$, then the following is true.*

$$\begin{aligned}
&\textit{Case 1: if } \mu_{max} > k: \\
&\qquad \Pr[X > k] \le \Pr[Y \sim \mathcal{N}(\mu_{max}, \sigma^2_{min}) > k] \\
&\qquad \Pr[X > k] \ge \Pr[Y \sim \mathcal{N}(\mu_{min}, \sigma^2_{max}) > k] \\
&\textit{Case 2: if } \mu_{max} < k: \\
&\qquad \Pr[X > k] \le \Pr[Y \sim \mathcal{N}(\mu_{max}, \sigma^2_{max}) > k] \\
&\qquad \Pr[X > k] \ge \Pr[Y \sim \mathcal{N}(\mu_{min}, \sigma^2_{min}) > k]
\end{aligned}$$

Proof. For both upper bounds, we want to minimize the value of $\Phi\left(\frac{k-\mathbb{E}[X]}{\mathrm{Var}[X]}\right)$. Because Φ is monotonically increasing, this is equivalent to minimizing the value of $\frac{k-\mathbb{E}[X]}{\mathrm{Var}[X]}$. It is clear that $k - \mu_{max} < k - \mu_{min}$. Now, if $k - \mu_{max} < 0$, then

$$\frac{k-\mu_{max}}{\sigma_{min}} < \frac{k-\mu_{max}}{\sigma_{max}}.$$

However, if $k - \mu_{max} > 0$, then

$$\frac{k-\mu_{max}}{\sigma_{max}} < \frac{k-\mu_{max}}{\sigma_{min}}.$$

For both lower bounds, we want to maximize the value of $\Phi\left(\frac{k-\mathbb{E}[X]}{\mathrm{Var}[X]}\right)$. Because Φ is monotonically increasing, this is equivalent to maximizing the value of $\frac{k-\mathbb{E}[X]}{\mathrm{Var}[X]}$. As in the above case, it is clear that $k-\mu_{min} > k-\mu_{max}$. Now, if $k-\mu_{min} < 0$, then

$$\frac{k-\mu_{min}}{\sigma_{max}} > \frac{k-\mu_{min}}{\sigma_{min}}.$$

However, if $k-\mu_{min} > 0$, then

$$\frac{k-\mu_{min}}{\sigma_{min}} > \frac{k-\mu_{min}}{\sigma_{max}}.$$

□

Finally, by definition, $\mathrm{Erf}\,(\infty) = 1$ and $\mathrm{Erf}\,(-\infty) = -1$, where $\mathrm{Erf}\,(\cdot)$ denotes the (Gauss) error function,

$$\mathrm{Erf}\,(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}dt.$$

We will use this fact repeatedly throughout the proof of the theorem, which we now turn to.

Proof of Theorem 1.3. Given a total number of voters n, let us define two random variables, X_D and X_M, where X_D denotes the number of correct votes under the direct voting mechanism D, and X_M represents the (weighted) number of correct votes under GreedyCap. We are interested in comparing $P_D = \Pr[X_D > n/2]$ and $P_M = \Pr[X_M > n/2]$.

Let $V = \{V_1, \ldots, V_n, \ldots\}$ be a sequence of independent Bernoulli random variables in which V_i represents the vote of voter i; i.e., each V_i has success probability $p_i \in [\beta, 1-\beta]$ for $\beta \in (0, 1/2)$. Using V, we define a sequence of instances indexed by n, where each instance consists of the first n voters in V. Let $\{b^D_{ni} : 1 \leq i \leq n\}$ and $\{b^M_{ni} : 1 \leq i \leq n\}$ be triangular arrays of nonnegative integers that denote the weight of each voter under direct voting and GreedyCap, respectively. Under direct voting, $b^D_{ni} = 1$ for all $1 \leq i \leq n$. Note that, in this case, $0 \leq b^D_{ni} \leq C(n)$ for all voters i, and $\sum_{i=1}^n b^D_{ni} = n$. Now, in the delegative case, let $b^M_{ni} = w_{ni}$ for all $1 \leq i \leq n$, where $w_{ni} \in \mathbb{Z}_{\geq 0}$ is the total weight accumulated by voter i in instance n (note that voters who choose to delegate have weight zero). Because each voter cannot accumulate weight greater than $C(n)$, we have that $0 \leq w_{ni} \leq C(n)$ for all voters i, and $\sum_{i=1}^n w_{ni} = n$.

Note that, given a population of voters of size n, $X_D = \sum_{i=1}^n b^D_{ni} V_i = \sum_{i=1}^n V_i$ and $X_M = \sum_{i=1}^n b^M_{ni} V_i = \sum_{i=1}^n w_{ni} V_i$. Now, because X_D and X_M both satisfy the conditions under which Lemma 1.1.5 holds, we may apply Lemma 1.1.5 to establish that X_D and X_M are approximately normally distributed as n goes to infinity; i.e.,

$$X_D = \sum_{i=1}^n V_i \to N\left(\sum_{i=1}^n \mathbb{E}[V_i], \sum_{i=1}^n \mathrm{Var}[V_i]\right) \text{ as } n \to \infty$$

and

$$X_M = \sum_{i=1}^n w_{ni} V_i \to N\left(\sum_{i=1}^n w_{ni}\,\mathbb{E}[V_i], \sum_{i=1}^n w_{ni}^2\,\mathrm{Var}[V_i]\right) \text{ as } n \to \infty.$$

Therefore, we can use the following formulas.

$$P_D \approx \int_{n/2}^{n} \frac{1}{\sqrt{2\pi \operatorname{Var}[X_D]}} \exp\left(\frac{-(x-\mathbb{E}[X_D])^2}{2\operatorname{Var}[X_D]}\right) dx \tag{1.23}$$

$$P_M \approx \int_{n/2}^{n} \frac{1}{\sqrt{2\pi \operatorname{Var}[X_M]}} \exp\left(\frac{-(x-\mathbb{E}[X_M])^2}{2\operatorname{Var}[X_M]}\right) dx \tag{1.24}$$

Indeed, throughout this proof, we will assume that P_D and P_M are exactly equal to these quantities; this is because Lemma 1.1.5 says that as n goes to infinity, this approximation becomes arbitrarily accurate.

Note that, from above, the PG property means that there exists ε such that $P_M - P_D > \varepsilon$ for at least one graph G_n on n vertices for all suitably large n. Similarly, the DNH property corresponds to $P_D - P_M < \varepsilon$ for all graphs G_n on n vertices for suitably large n and all values of ε. We show that these two properties hold.

For the PG property, we construct a simple family of examples where the property is satisfied. Let the social graph G be composed of pairs of nodes with one competent voter and one incompetent voter with an edge pointing to the competent voter. The competent voters have competence $1-\beta$ and the incompetent voters have competence β. If the voters vote independently, the symmetry between the competent and incompetent voters makes it clear that $P_D = 1/2$. Under Algorithm 1, the incompetent voters all delegate to the competent voters. We now have $\frac{n}{2}$ independent voters who each have one vote of weight two and competence $1-\beta$. By the Condorcet Jury Theorem [123], it follows that P_M approaches 1.

In the remainder of the proof, therefore, we focus on establishing the DNH property. We first show that

$$\operatorname{Var}[X_D] \in [\beta(1-\beta)n, n/4]. \tag{1.25}$$

Indeed, $X_D = \sum_{i=1}^{n} V_i$, where V_i is the Bernoulli random variable representing the vote of voter i. In particular, $V_i \sim \text{Bernoulli}(p_i)$, where $p_i \in [\beta, 1-\beta]$ is the competence level of voter i. Because all voters vote independently, $\operatorname{Var}[X_D] = \sum_{i=1}^{n} \operatorname{Var}[V_i]$, and

$$\operatorname{Var}[V_i] = p_i(1-p_i) \in [\beta(1-\beta), (1/2)^2].$$

This establishes Equation (1.25).

Now, let us separate the instances into two cases:

1. $|\,\mathbb{E}[X_D] - \frac{n}{2}| > \frac{n}{\log n}$.
2. $|\,\mathbb{E}[X_D] - \frac{n}{2}| \le \frac{n}{\log n}$.

Case 1. In this case, we can give strong lower bounds on both P_D and P_M.

Subcase 1: $\mathbb{E}[X_D] < n/2 - n/\log n$. By Equation (1.25), $\operatorname{Var}[X_D] \le n/4 < n$. Because $\mathbb{E}[X_D] < n/2$, by Lemma 1.6 we have

$$P_D < \int_{\frac{n}{2}}^{n} \frac{1}{\sqrt{2\pi n^2}} e^{\frac{-\left(x-\frac{n}{2}+\frac{n}{\log n}\right)^2}{2n}}\, dx. \tag{1.26}$$

This is equivalent to

$$P_D < \frac{1}{2}\left(\operatorname{Erf}\left(\frac{\sqrt{n}(2+\log n)}{2\sqrt{2}\log n}\right) - \operatorname{Erf}\left(\frac{\sqrt{n}}{\sqrt{2}\log n}\right)\right).$$

As n approaches infinity, both arguments go to infinity, and therefore (as $\operatorname{Erf}(\infty) = 1$) P_D approaches 0. This means that, no matter the value of P_M, DNH is satisfied.

Subcase 2: $\mathbb{E}[X_D] > n/2 + n/\log n$. We now examine the maximum possible value of $\operatorname{Var}[X_M] = \sum_{i=1}^n w_{ni}^2 \operatorname{Var}[V_i]$, where w_{ni} is the total weight accumulated by voter i and, again, V_i is the Bernoulli random variable representing the vote of voter i. Additionally, $\operatorname{Var}[V_i] \in [\beta(1-\beta), 1/4]$, and applying this yields

$$\operatorname{Var}[X_M] \leq \frac{1}{4} \cdot \sum_{i=1}^n w_{ni}^2.$$

Because each voter can accumulate at most weight $C(n)$, by the convexity of x^2, we can see that this is maximized when the maximum number of voters have weight exactly $C(n)$. Therefore, we have

$$\operatorname{Var}[X_M] \leq \frac{1}{4} \cdot \sum_{i=1}^{\lceil n/C(n)\rceil} C(n)^2 < nC(n).$$

Because $\mathbb{E}[X_D] > n/2$, by Lemma 1.6 we have

$$P_M > \int_{\frac{n}{2}}^{n} \frac{1}{\sqrt{2\pi n C(n)}} e^{\frac{-\left(x-\frac{n}{2}+\frac{n}{\log n}\right)^2}{2nC(n)}}\, dx. \tag{1.27}$$

This simplifies to

$$P_M > \frac{1}{2}\left(\operatorname{Erf}\left(\frac{\sqrt{n}(\log n - 2)}{2\sqrt{2C(n)}\log n}\right) + \operatorname{Erf}\left(\frac{\sqrt{n}}{\sqrt{2C(n)}\log n}\right)\right).$$

As n approaches infinity, both arguments go to infinity, and P_M approaches 1. Therefore, no matter what the value of P_D, DNH is satisfied.

Case 2. In this case, we split the argument into two further subcases:

1. The number of voters who delegate is greater than $n/g(n)$, where $g(n)$ is $o(\log n)$ and $\omega(C(n)^2)$.
2. The number of voters who delegate is less than or equal to $n/g(n)$.

Subcase 1: Due to delegation, we have $\mathbb{E}[X_M] - \mathbb{E}[X_D] \geq n\alpha/g(n)$. We can now bound the mean by

$$\mathbb{E}[X_M] \geq \frac{n}{2} - \frac{n}{\log n} + \frac{n\alpha}{g(n)}.$$

Therefore, because $g(n) = o(\log n)$, $\mathbb{E}[X_M] > n/2$ as n increases. As before, we also know that $\mathrm{Var}[X_M]$ is bounded from above by $nC(n)$, and therefore, by Lemma 1.6,

$$P_M \geq \int_{\frac{n}{2}}^{n} \frac{1}{\sqrt{2\pi nC(n)}} e^{\frac{-\left(x - \frac{n}{2} - \frac{n}{\log n} - \frac{n\alpha}{g(n)}\right)^2}{2nC(n)}}\, dx. \tag{1.28}$$

We would like to show that this integral goes to 1 as n goes to infinity.

This is equivalent to

$$\frac{1}{2}\left(\mathrm{Erf}\left(\frac{\frac{n}{2} - \frac{n\alpha}{g(n)} + \frac{n}{\log n}}{\sqrt{2nC(n)}}\right) - \mathrm{Erf}\left(\frac{\sqrt{n}\left(-\frac{\alpha}{g(n)} + \frac{1}{\log n}\right)}{\sqrt{2C(n)}}\right)\right).$$

Note that as n goes to infinity, the first argument goes to infinity and the second argument goes to negative infinity when $g(n) = o(\log n)$. Therefore, P_M goes to 1, satisfying DNH.

Subcase 2: In this case, most voters remain independent. We will argue that although the delegation does impact the variance, this impact will get arbitrarily small as n grows larger, implying that the loss will get arbitrarily small.

Let us index the voters according to what happens in the delegation scheme. Let the first n_1 indexed voters represent those who remain independent and do not get delegated a vote. Let the next n_2 indexed voters be those who got delegated at least one vote. Finally, the last $n - n_1 - n_2$ indexed voters are those who delegated their vote to another voter. Based on our assumption above, we know that $\lim_{n\to\infty} \frac{n_1}{n} = 1$; most voters remain independent and unaffected by the delegation scheme.

Additionally, note that the mean will be slightly different in the two schemes, but this to our advantage because the mean will improve in the delegation scheme due to "uphill" delegation.

Therefore, given

$$P_D = \int_{\frac{n}{2}}^{n} \frac{1}{\sqrt{2\pi \,\mathrm{Var}[X_D]}} e^{\frac{-(x - \mathbb{E}[X_D])^2}{2\,\mathrm{Var}[X_D]}}\, dx$$

and

$$P_M = \int_{\frac{n}{2}}^{n} \frac{1}{\sqrt{2\pi \,\mathrm{Var}[X_M]}} e^{\frac{-(x - \mathbb{E}[X_M])^2}{2\,\mathrm{Var}[X_M]}}\, dx,$$

because $\mathbb{E}[X_M] \geq \mathbb{E}[X_D]$, we can say that

$$P_M \geq \int_{\frac{n}{2}}^{n} \frac{1}{\sqrt{2\pi \,\mathrm{Var}[X_M]}} e^{\frac{-(x - \mathbb{E}[X_D])^2}{2\,\mathrm{Var}[X_M]}}\, dx.$$

Now, we have to relate $\mathrm{Var}[X_M]$ and $\mathrm{Var}[X_D]$. Ideally, we want to show that they are multiplicatively close to each other.

We can decompose the variance of X_D.

$$\operatorname{Var}[X_D] = \sum_{i=1}^{n_1} p_i(1-p_i) + \sum_{i=n_1+1}^{n} p_i(1-p_i).$$

Likewise, we can decompose the variance of X_M.

$$\operatorname{Var}[X_M] = \sum_{i=1}^{n_1} p_i(1-p_i) + \sum_{i=n_1+1}^{n_1+n_2} w_{ni}^2 p_i(1-p_i) + \sum_{i=n_1+n_2+1}^{n} 0.$$

Therefore, we have

$$\begin{aligned}
\operatorname{Var}[X_M] - \operatorname{Var}[X_D] &= \sum_{i=n_1+1}^{n_1+n_2} (w_{ni}^2 - 1) p_i(1-p_i) \\
&\quad - \sum_{i=n_1+n_2+1}^{n} p_i(1-p_i) \\
&\le \sum_{i=n_1+1}^{n_1+n_2} (w_{ni}^2 - 1) p_i(1-p_i) \\
&\le \frac{n_2}{4}(\max w_{ni}{}^2 - 1) \\
&\le \frac{1}{4} \cdot \frac{n}{g(n)}(C(n)^2 - 1),
\end{aligned}$$

where the last inequality holds because $w_{ni} \le C(n)$, and n_2, the number of voters who are delegated to, is at most the number of voters who delegate, which is at most $n/g(n)$ by assumption.

This means that

$$\operatorname{Var}[X_M] \le \operatorname{Var}[X_D] + \frac{1}{4} \cdot \frac{n}{g(n)}(C(n)^2 - 1)$$

and therefore

$$\begin{aligned}
\frac{\operatorname{Var}[X_M]}{\operatorname{Var}[X_D]} &\le \frac{\operatorname{Var}[X_D] + \frac{1}{4} \cdot \frac{n}{g(n)}(C(n)^2 - 1)}{\operatorname{Var}[X_D]} \\
&= 1 + \frac{\frac{n}{g(n)}(C(n)^2 - 1)}{4 \operatorname{Var}[X_D]}.
\end{aligned}$$

Now, note that by Equation (1.25),

$$\operatorname{Var}[X_D] \ge n\beta(1-\beta)$$

and therefore

$$\begin{aligned}
\operatorname{Var}[X_M] &\le \operatorname{Var}[X_D] \left(1 + \frac{\frac{n}{g(n)}(C(n)^2 - 1)}{4n\beta(1-\beta)}\right) \\
&= \operatorname{Var}[X_D] \left(1 + \frac{1}{g(n)} \cdot \frac{C(n)^2 - 1}{4\beta(1-\beta)}\right).
\end{aligned}$$

Let

$$\eta = \frac{1}{g(n)} \cdot \frac{C(n)^2 - 1}{4\beta(1-\beta)}$$

and note that as n goes to infinity, η goes to 0 because we chose $g(n)$ to grow asymptotically more quickly than $C(n)^2$.

Therefore, revisiting the original integrals, we have

$$P_D = \int_{\frac{n}{2}}^{n} \frac{1}{\sqrt{2\pi \operatorname{Var}[X_D]}} e^{\frac{-(x-\mathbb{E}[X_D])^2}{2\operatorname{Var}[X_D]}}\, dx$$

and

$$P_M \geq \int_{\frac{n}{2}}^{n} \frac{1}{\sqrt{2\pi \operatorname{Var}[X_D](1+\eta)}} e^{\frac{-(x-\mathbb{E}[X_D])^2}{2\operatorname{Var}[X_D](1+\eta)}}\, dx.$$

Simplifying the above yields

$$P_D = \frac{1}{2}\left(\operatorname{Erf}\left(\frac{n - \mathbb{E}[X_D]}{\sqrt{2\operatorname{Var}[X_D]}}\right) - \operatorname{Erf}\left(\frac{n - 2\,\mathbb{E}[X_D]}{2\sqrt{2\operatorname{Var}[X_D]}}\right)\right) \tag{1.29}$$

and

$$\begin{aligned} P_M \geq \frac{1}{2}\Bigg(&\operatorname{Erf}\left(\frac{n - \mathbb{E}[X_D]}{\sqrt{2\operatorname{Var}[X_D](1+\eta)}}\right) \\ &-\operatorname{Erf}\left(\frac{n - 2\,\mathbb{E}[X_D]}{2\sqrt{2\operatorname{Var}[X_D](1+\eta)}}\right)\Bigg) \end{aligned} \tag{1.30}$$

Furthermore, again by Equation (1.25), we know that $\operatorname{Var}[X_D] \in [\beta(1-\beta)n, n/4]$ and therefore $\sqrt{\operatorname{Var}[X_D]} = \sqrt{cn}$, where $c \in [\beta(1-\beta), 1/4]$. From this, note that as n goes to infinity, the argument to the first error function in each expression goes to infinity.

Let

$$h_1(n) = \frac{n - 2\,\mathbb{E}[X_D]}{2\sqrt{2\operatorname{Var}[X_D]}} \tag{1.31}$$

be the argument to the second error function in (1.29), and let

$$h_2(n) = \frac{n - 2\,\mathbb{E}[X_D]}{2\sqrt{2\operatorname{Var}[X_D](1+\eta)}} \tag{1.32}$$

be the argument to the second error function in (1.30). As n goes to infinity, note that the argument to (1.31) must go to one of four states: infinity, negative infinity, zero, or a constant. In the case that it goes to infinity, negative infinity, or zero, the presence of the extra $\frac{1}{\sqrt{1+\eta}}$ term in (1.32) does nothing to change the sign of the arguments, and therefore they each converge to the same state (infinity, negative infinity, or zero) as n approaches infinity. When the argument to (1.31) goes to a constant, note that as n goes to infinity, η goes to 0, and therefore the two converge once again.

We conclude that (an upper bound on) the difference between P_D and P_M converges to 0, and hence DNH is satisfied. $\square$

1.2 Minimizing the Maximum Weight of Voters in Liquid Democracy

The foregoing section indicates that, even if delegations go only to more competent agents, a high concentration of power might still be harmful for social welfare, by neutralizing benefits corresponding to the Condorcet Jury Theorem. Concentration of power is also a concern in real-world implementations of liquid democracy: Certain individuals, the so-called super-voters, seem to amass enormous weight, whereas most agents do not receive any delegations. As Kling et al. [147] describe, super-voters in the Pirate Party were so controversial that “the democratic nature of the system was questioned, and many users became inactive.” Besides the negative impact of super-voters on perceived legitimacy, super-voters might also be more exposed to bribing. Although delegators can retract their delegations as soon as they become aware of suspicious voting behavior, serious damage might be done in the meantime. Furthermore, if super-voters jointly have sufficient power, they might find it more efficient to organize majorities through deals between super-voters behind closed doors, rather than to try to win a broad majority through public discourse.

While all these concerns suggest that the weight of super-voters should be limited, the exact metric to optimize for varies between them and is often not even clearly defined. For the purposes of this chapter, we choose to minimize the weight of the heaviest voter. As is evident in the Spiegel article, the weight of individual voters plays a direct role in the perception of super-voters. But even beyond that, we are confident that minimizing this measure will lead to substantial improvements across all presented concerns.

Just how can the maximum weight be reduced? One approach might be to restrict the power of delegation by imposing caps on the weight. However, as argued by Behrens et al. [26], delegation is always possible by coordinating outside of the system and copying the desired delegate’s ballot. Pushing delegations outside of the system would not alleviate the problem of super-voters, just reduce transparency. Therefore, we instead adopt a voluntary approach: If agents are considering multiple potential delegates, all of whom they trust, they are encouraged to leave the decision for one of them to a centralized mechanism. With the goal of avoiding high-weight agents in mind, our research challenge is twofold:

> *First, investigate the algorithmic problem of selecting delegations to minimize the maximum weight of any agent, and, second, show that allowing multiple delegation options does indeed provide a significant reduction in the maximum weight compared to the status quo.*

1.2.1 Our Approach and Results

We formally define our problem in Section 1.2.3. In addition to minimizing the maximum weight of any voter, we specify how to deal with delegators whose vote cannot possibly reach any voter. In general, our problem is closely related to minimizing congestion for confluent flow as studied by Chen et al. [68]. Not only does this connection suggest an optimal algorithm based on mixed integer linear programming, but we also get a polynomial-time $(1 + \ln |V|)$-approximation algorithm, where V is the set of voters. In addition, we show that approximating our problem to within a factor of $\frac{1}{2}\log_2|V|$ is NP-hard.

In Section 1.2.4, to evaluate the benefits of allowing multiple delegations, we propose a probabilistic model for delegation behavior — inspired by the well-known *preferential attachment* model [25] — in which we add agents successively. With a certain probability d, a new agent delegates; otherwise, she votes herself. If she delegates, she chooses k many delegation options among the previously inserted agents. A third parameter γ controls the bias of this selection towards agents who already receive many delegations. Assuming $\gamma = 0$, i.e., that the choice of delegates is unbiased, we prove that allowing two choices per delegator ($k = 2$) asymptotically leads to dramatically lower maximum weight than classical liquid democracy ($k = 1$) using an argument based on the "power of choice." In the latter case, with high probability, the maximum weight is at least $\Omega(t^\beta)$ for some $\beta > 0$, whereas the maximum weight in the former case is only $\mathcal{O}(\log\log t)$ with high probability, where t denotes simultaneously the time step of the process and the number of agents. Our analysis draws on a phenomenon called the *power of choice* that can be observed in many different load balancing models. In fact, even a greedy mechanism that selects a delegation option to locally minimize the maximum weight as agents arrive exhibits this asymptotic behavior, which upper-bounds the maximum weight for optimal resolution.

In Section 1.2.5, we complement our theoretical findings with empirical results. Our simulations demonstrate that our approach continues to outperform classical preferential attachment for higher values of γ. We also show that the most substantial improvements come from increasing k from one to two, i.e., that increasing k even further only slightly reduces the maximum weight. We continue to see these improvements in terms of maximum weight even if just some fraction of delegators give two options while the others specify a single delegate. Finally, we compare the optimal maximum weight with the maximum weight produced by the approximation algorithm and greedy heuristics.

1.2.2 Related Work

Kling et al. [147] conduct an empirical investigation of the existence and influence of super-voters. The analysis is based on daily data dumps, from 2010 until 2013, of the German Pirate Party installation of LiquidFeedback. As noted above, Kling et al. find that super-voters exist, and have considerable power. The results do suggest that super-voters behave responsibly, as they "do not fully act on their power to change the outcome of votes, and they vote in favour of proposals with the majority of voters in many cases." Of course, this does not contradict the idea that a balanced distribution of power would be desirable.

In recent years, there has been an increasing number of theoretical analyses of liquid democracy. In the field of political theory, Blum and Zuber [38] give a normative justification of liquid democracy. They consider two accounts of democracy, which differ in the stated goal of a democratic system. In the *epistemic* framework, the success of a democratic system should lead to good decisions with respect to some objective notion of quality, whereas, in the *egalitarian* framework, a democratic system should allow each individual to impose her particular interests to the same degree. Blum and Zuber conclude that liquid democracy improves upon purely representative democracy with respect to both metrics. They see unequal voting weights as problematic and suggest public deliberation before a vote to attenuate this problem.

In the spirit of the egalitarian framework, Green-Armytage [121] justifies liquid democracy in a spatial model of political preferences similar to facility placement. When an agent has incomplete information about a topic, transitive delegations can help to express her preferences more accurately by harnessing the expertise of like-minded, more qualified agents.

In this paper, we consider a single delegation network. Other works allow agents to specify different delegations for multiple interconnected issues, where the binary preferences and outcomes are restricted to satisfy a propositional formula [72] or to correspond to binary comparisons in a ranking [51]. Both papers propose ways of reconciling contradictory choices made by different delegates.

We also highlight related work that considers models of network formation and influence attenuation in the context of liquid democracy. Bloembergen et al. [37] introduce a game-theoretic model of delegation in order to study rational delegation behavior in liquid-democracy networks. In their model, delegation networks might be formed by a best-response dynamic or as Nash equilibria of a delegation game. Escoffier et al. [98] study a similar delegation game with different incentives.

Boldi et al. [40] study a variant of liquid democracy in which a voter's weight decreases by a discount factor every time her vote is transitively delegated, penalizing long delegation chains. They argue that this variant is more appropriate in online communities, where trust relationships are typically less deep than in the real world. While not intended as such, this variant of liquid democracy can also reduce the weight of super-voters, at least of those who receive most of their delegations indirectly. However, such a variant violates the principle of "one person, one vote" and incentivizes delegation outside of the system [26]. By contrast, our approach reduces the weight of super-voters while preserving each voter's individual influence.

1.2.3 Algorithmic Model and Results

Let us consider a delegative voting process where agents may specify multiple potential delegations. This gives rise to a directed graph, whose nodes represent agents and whose edges represent potential delegations. In the following, we will conflate nodes and the agents they represent. A distinguished subset of nodes corresponds to agents who have voted directly, the *voters*. Since voters forfeit the right to delegate, the voters are a subset of the sinks of the graph. We call all non-voter agents *delegators*.

Each agent has an inherent voting weight of 1. When the delegations will have been resolved, the weight of every agent will be the sum of weights of her delegators plus her inherent weight. We aim to choose a delegation for every delegator in such a way that the maximum weight of any voter is minimized.

This task closely mirrors the problem of congestion minimization for confluent flow (with infinite edge capacity): There, a flow network is also a finite directed graph with a distinguished set of graph sinks, the *flow sinks*. Every node has a non-negative *demand*. If we assume unit demand, this demand is 1 for every node. Since the flow is confluent, for every non-sink node, the algorithm must pick exactly one outgoing edge, along which the flow is sent. Then, the *congestion* at a node n is the sum of congestions at all nodes who direct their flow to n plus the demand of n. The goal in congestion minimization is to minimize the maximum congestion at any flow sink. (We remark that the close connection between our problem and confluent flow immediately suggests a variant corresponding to splittable flow; we discuss this variant at length in Section 1.3.)

In spite of the similarity between confluent flow and resolving potential delegations, the two problems differ when a node has no path to a voter / flow sink. In confluent flow, the result would simply be that no flow exists. In our setting however, this situation can hardly be avoided. If, for example, several friends assign all of their potential delegations to each other, and if all of them rely on the others to vote, their weight cannot be delegated to any voter. Our mechanism cannot simply report failure as soon as a small group of voters behaves in an unexpected way. Thus, it must be allowed to leave these votes unused. At the same time, of course, our algorithm should not exploit this power to decrease the maximum weight, but must primarily maximize the number of utilized votes. We formalize these issues in the following section.

Problem Statement

All graphs $G = (N, E)$ mentioned in this section will be finite and directed. Furthermore, they will be equipped with a set V of distinguished sinks in the graph. For the sake of brevity, these assumptions will be implicit in the notion "graph G with V".

Some of these graphs represent situations in which all delegations have already been resolved and in which each vote reaches a voter: We call a graph (N, E) with V a *delegation graph* if it is acyclic, its sinks are exactly the set V, and every other vertex has outdegree one. In such a graph, define the *weight* $w(n)$ of a node $n \in N$ as

$$w(n) \coloneqq 1 + \sum_{(m,n) \in E} w(m).$$

This is well-defined because E is a well-founded relation on N.

Resolving the delegations of a graph G with V can now be described as the MinMaxWeight problem: Among all delegation subgraphs (N', E') of G with voting vertices V of maximum $|N'|$, find one that minimizes the maximum weight of the voting vertices.

Connections to Confluent Flow

We recall definitions from the flow literature as used by Chen et al. [68]. We slightly simplify the exposition by assuming unit demand at every node.

Given a graph (N, E) with V, a *flow* is a function $f : E \to \mathbb{R}_{\geq 0}$. For any node n, set $in(n) \coloneqq \sum_{(m,n) \in E} f(m, n)$ and $out(n) \coloneqq \sum_{(n,m) \in E} f(n, m)$. At every node $n \in N \setminus V$, a flow must satisfy *flow conservation*:

$$out(n) = 1 + in(n).$$

Note that all nodes in V are sinks in the graph, and thus have no outflow. The congestion at any node n is defined as $1 + in(n)$. A flow is *confluent* if every node has at most one outgoing edge with positive flow. We define MinMaxCongestion as the problem of finding a confluent flow on a given graph such that the maximum congestion is minimized.

To relate the two presented problems, we need to refer to the parts of a graph (N, E) with V from which V is reachable: The *active* nodes $\mathit{active}_V(N, E)$ are all $n \in N$ such that there exists a path from n to a sink $v \in V$ using edges in E. The *active subgraph* is the restriction of (N, E) to $\mathit{active}_V(N, E)$. In particular, V is part of this subgraph.

Lemma 1.2.1. *Let $G = (N, E)$ with V be a graph. Its delegation subgraphs (N', E') that maximize $|N'|$ are exactly the delegation subgraphs with $N' = \mathit{active}_V(N, E)$. At least one such subgraph exists.*

Proof. First, we show that all nodes of a delegation subgraph are active. Indeed, consider any node n_1 in the subgraph. By following outgoing edges, we obtain a sequence of nodes $n_1\, n_2 \ldots$ such that n_i delegates to n_{i+1}. Since the graph is finite and acyclic, this sequence must end with a vertex n_j without outgoing edges. This must be a voter; thus, n_1 is active.

Furthermore, there exists a delegation subgraph of (N, E) with nodes exactly $\mathit{active}_V(N, E)$. Indeed, the shortest-paths-to-set-V forest (with edges pointed in the direction of the paths) on the active subgraph is a delegation graph.

By the first argument, all delegation subgraphs must be subgraphs of the active subgraph. By the second argument, to have the maximum number of nodes, they must include all nodes of this subgraph. $\square$

Lemma 1.2.2. *Let (N, E) with V be a graph and let $f : E \to \mathbb{R}_{\geq 0}$*

Proof. Without the restriction on the outdegree, the problem is NP-hard [106]. We reduce the general case to our special case.

Let G' be an arbitrary directed graph; let s'_1, s'_2, t'_1, t'_2 be distinguished vertices. To restrict the outdegree, replace each node n with outdegree d by a binary arborescence (directed binary tree with edges facing away from the root) with d sinks. All incoming edges into n are redirected towards the root of the arborescence; outgoing edges from n instead start from the different leaves of the arborescence. Call the new graph G, and let s_1, s_2, t_1, t_2 refer to the roots of the arborescences replacing s'_1, s'_2, t'_1, t'_2, respectively.

Clearly, our modifications to G' can be carried out in polynomial time. It remains to show that there are vertex-disjoint paths from s_1 to t_1 and from s_2 to t_2 in G iff there are vertex-disjoint paths from s'_1 to t'_1 and from s'_2 to t'_2 in G'.

If there are disjoint paths in G', we can translate these paths into G by visiting the arborescences corresponding to the nodes on the original path one after another. Since both paths visit disjoint arborescences, the new paths must be disjoint.

Suppose now that there are disjoint paths in G. Translate the paths into G' by visiting the nodes corresponding to the sequence of visited arborescences. Since each arborescence can only be entered via its root, disjointness of the paths in G implies disjointness of the translated paths in G'. □

Now, we can strengthen the hardness of approximation for MinMaxCongestion by Chen et al. [68]. We believe the lemma is of independent interest, as it shows a surprising separation between the case of outdegree 1 (where the problem is moot) and outdegree 2, and that the asymptotically optimal approximation ratio is independent of degree. But it also allows us to prove Theorem 1.8 almost directly.

Lemma 1.2.5. *It is NP-hard to approximate the* MinMaxCongestion *problem to a factor of $\frac{1}{2}\log_2 k$, where k is the number of sinks, even when each node has unit demand and outdegree at most 2.*

Proof of Lemma 1.2.5. We adapt the proof of Theorem 1 of Chen et al. [68].

Let $G = (V, E), s_1, s_2, t_1, t_2$ be given as in Lemma 1.2.4. Without loss of generality, G only contains nodes from which t_1 or t_2 is reachable, t_1 and t_2 are sinks and all four vertices are distinct. Let $\ell = \lceil \log_2 |V| \rceil$ and $k = 2^\ell$. Build the same auxiliary network as that built by Chen et al. [68], which consists of a binary arborescence whose $k-1$ nodes are copies of G. The construction is illustrated in Figure 1.2. For more details, refer to [68].

For ease of exposition, we describe our reduction as returning a flow network with polynomially-bounded positive integer demands. Implicitly, the described network is subsequently translated into one with unary demand; to express a demand of d at a node n in our unit-demand setting, add $d-1$ fresh nodes with a single outgoing edge to n.

Denote the number of nodes in the network by $\phi \coloneqq (k-1)\cdot|V|+k$, and set $\Phi \coloneqq \ell\cdot\phi+1$. In [68], every copy of s_2 and t_2 has demand 1, the copy of s_1 at the root has demand 2, and all other nodes have demand 0. Instead, we give these nodes demands of Φ, 2Φ and 1, respectively. Note that the size of the generated network[4] is polynomial in the size of G and that the outdegree of each node is at most 2. From every node, one of the sinks

[4]Even after unfolding our non-unitary-demand nodes.

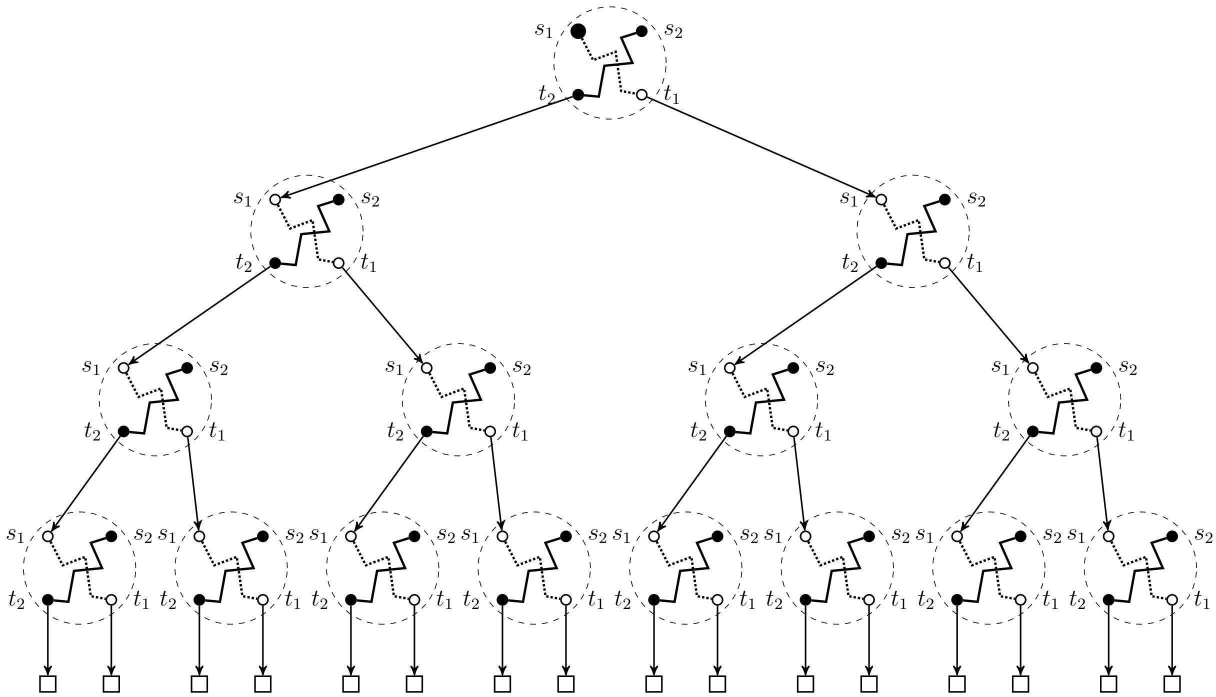


Figure 1.2: Auxiliary network generated from G, here for $k = 16$. Recreation of [68, Fig. 2].

S displayed as rectangles in Figure 1.2 is reachable. Since the minimum-distance-to-S spanning forest describes a flow, a flow in the network exists.

Suppose that G contains vertex-disjoint paths P_1 from s_1 to t_1 and P_2 from s_2 to t_2. In each copy of G in the network, route the flow along these paths. We can complete the confluent flow inside of this copy in such a way that the demand of every node is routed to t_1 or t_2: By assumption, each of the nodes can reach one of these two path endpoints. Iterate over all nodes in order of ascending distance to the closest endpoint and make sure that their flow is routed to an endpoint. For the endpoints themselves, there is nothing to do. For positive distance, a node might be part of a path and thus already connected to an endpoint. Else, look at its successor in a shortest path to an endpoint. By the induction hypothesis, all flow from this successor is routed to an endpoint, so route the node's flow to this successor. If we also use the edges between copies of G and between the copies and the sinks, we obtain a confluent flow. Each sink except for the rightmost one can only collect the demand of two nodes with demand Φ plus a number of nodes with demand 1. The rightmost sink collects the demand from the single node with demand 2Φ plus some unitary demands. Thus, the congestion of the system can be at most $2\Phi + \phi$.

Now, consider the case in which G does not have such vertex-disjoint paths. In every confluent flow and in every copy, there are three options:

- the flow from s_1 flows to t_2 and the flow from s_2 flows to t_1,
- the flow from s_1 and s_2 flows to t_1, or
- the flow from s_1 and s_2 flows to t_2.

In each case, the flow coming in through s_1 is joined by additional demand of at least

Φ. Consider the path from the copy of s_1 at the root to a sink. By a simple inductive argument, the congestion at the endpoint of the ith copy of G on this path is at least $(i+1) \cdot \Phi$. Thus, the total congestion at the sink must be at least $(\ell+1) \cdot \Phi$. The lemma now follows from the fact that

$$\frac{\log_2 k}{2}(2\Phi + \phi) = \frac{\ell}{2}(2\Phi + \phi) < (\ell + 1) \cdot \Phi. \qquad \square$$

Proof of Theorem 1.8. We reduce (gap) MinMaxCongestion with unit demand and outdegree at most 2 to (gap) MinMaxWeight with outdegree at most 2. First, we claim that if there are inactive nodes, there is no confluent flow. Indeed, let n_1 be an inactive node. For the sake of contradiction, suppose that there exists a flow f. Follow the positive flow to obtain a sequence $n_1\, n_2 \ldots$. By definition, none of the nodes reachable from n_1 can be a voter. Since, by flow conservation and unit demand, each node must delegate, the sequence must be infinite. As detailed in the proof of Lemma 1.2.2, a confluent flow with unit demand cannot contain cycles. Thus, the sequence contains infinitely many different nodes, which contradicts the finiteness of G.

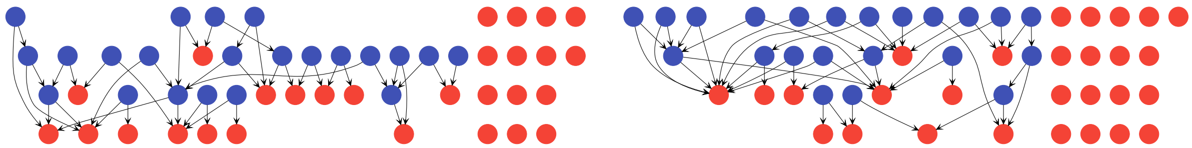
Therefore, we can assume without loss of generality that in the given instance of MinMaxCongestion, all nodes are active (as the problem is still NP-hard). The reduction creates an instance of MinMaxWeight that has the same graph as the given instance of MinMaxCongestion. Using an argument analogous to the proof of Theorem 1.7 (reversing the roles of Lemma 1.2.2 and Lemma 1.2.3 in its proof), we see that this is a strict approximation-preserving reduction. $\square$

1.2.4 Probabilistic Model and Results

Our generalization of liquid democracy to multiple potential delegations aims to decrease the concentration of weight. Accordingly, the success of our approach should be measured by its effect on the maximum weight in real elections. Since, at this time, we do not know of any available datasets,[5] we instead propose a probabilistic model for delegation behavior, which can serve as a credible proxy. Our model builds on the well-known preferential attachment model, which generates graphs possessing typical properties of social networks.

The evaluation of our approach will be twofold: In Sections 1.2.4 and 1.2.4, for a certain choice of parameters in our model, we establish a striking separation between traditional liquid democracy and our system. In the former case, the maximum weight at time t is $\Omega(t^\beta)$ for a constant β with high probability, whereas in the latter case, it is in $\mathcal{O}(\log \log t)$ with high probability, even if each delegator only suggests two options. For other parameter settings, we empirically corroborate the benefits of our approach in Section 1.2.5.

[5] There is one relevant dataset that we know of, which was analyzed by Kling et al. [147]. However, due to stringent privacy constraints, the data privacy officer of the German Pirate Party was unable to share this dataset with us.



(a) $\gamma = 0$ (b) $\gamma = 1$

Figure 1.3: Example graphs generated by the preferential delegation model for $k = 2$ and $d = 0.5$.

The Preferential Delegation Model

Many real-world social networks have degree distributions that follow a power law [153; 180]. Additionally, in their empirical study, Kling et al. [147] observed that the weight of voters in the German Pirate Party was "power law-like" and that the graph had a very unequal indegree distribution. In order to meld the previous two observations in our liquid democracy delegation graphs, we adapt a standard preferential attachment model [25] for this specific setting. At a high level, our *preferential delegation* model is characterized by three parameters: $0 < d < 1$, the probability of delegation; $k \geq 1$, the number of delegation options from each delegator; and $\gamma \geq 0$, an exponent that governs the probability of delegating to nodes based on current weight.

At time $t = 1$, we have a single node representing a single voter. In each subsequent time step, we add a node for agent i and flip a biased coin to determine her delegation behavior. With probability d, she delegates to other agents. Else, she votes independently. If i does not delegate, her node has no outgoing edges. Otherwise, add edges to k many independently selected, previously inserted nodes, where the probability of choosing node j is proportional to $(\mathit{indegree}(j)+1)^{\gamma}$. Note that this model might generate multiple edges between the same pair of nodes, and that all sinks are voters. Figure 1.3 shows example graphs for different settings of γ.

In the case of $\gamma = 0$, which we term *uniform delegation*, a delegator is equally likely to attach to any previously inserted node. Already in this case, a "rich-get-richer" phenomenon can be observed, i.e., voters at the end of large networks of potential delegations will likely see their network grow even more. Indeed, a larger network of delegations is more likely to attract new delegators. In traditional liquid democracy, where $k = 1$ and all potential delegations will be realized, this explains the emergence of super-voters with excessive weight observed by Kling et al. [147]. We aim to show that for $k \geq 2$, the resolution of potential delegations can strongly outweigh these effects. In this, we profit from an effect known as the "power of two choices" in load balancing described by Azar et al. [13].

For $\gamma > 0$, the "rich-get-richer" phenomenon additionally appears at the degrees of nodes. Since the number of received potential delegations is a proxy for an agent's competence and visibility, new agents are more likely to attach to agents with high indegree. In total, this is likely to further strengthen the inherent inequality between voters. For increasing γ, the graph becomes increasingly flat, as a few super-voters receive nearly all delegations. This matches observations from the LiquidFeedback dataset [147] that "the delegation network is slowly becoming less like a friendship network, and more like a bipartite networks of super-voters connected to normal voters." The special case of $\gamma = 1$ corresponds to preferential attachment as described by Barabási and Albert [25].

The most significant difference we expect to see between graphs generated by the preferential delegation model and real delegation graphs is the assumption that agents always delegate to more senior agents. In particular, this causes generated graphs to be acyclic, which need not be the case in practice. It does seem plausible that the majority of delegations goes to agents with more experience on the platform. Even if this assumption should not hold, there is a second interpretation of our process if we assume — as do Kahng et al. [137] — that agents can be ranked by competence and only delegate to more competent agents. Then, we can think of the agents as being inserted in decreasing order of competence. When a delegator chooses more competent agents to delegate to, her choice would still be biased towards agents with high indegree, which is a proxy for popularity.

It may be useful to note that the MinMaxWeight approach based on confluent flow does not require the underlying delegation graph to be acyclic, as the objective tries to minimize the maximum weight of any voter over all possible delegation choices that maximize the total number of utilized votes. In this sense, unavoidable cycles result in lost voting power.

In our theoretical results, we focus on the cases of $k = 1$ and $k = 2$, and assume $\gamma = 0$ to make the analysis tractable. The parameter d can be chosen freely between 0 and 1. Note that our upper bound for $k = 2$ directly translates into an upper bound for larger k, since the resolution mechanism always has the option of ignoring all outgoing edges except for the first two. Therefore, to understand the effect of multiple delegation options, we can restrict our attention to $k = 2$. This crucially relies on $\gamma = 0$, where potential delegations do not influence the probabilities of choosing future potential delegations. Based on related results by Malyshkin and Paquette [167], it seems unlikely that increasing k beyond 2 will reduce the maximum weight by more than a constant factor.

Lower Bounds for Single Delegation ($k = 1$, $\gamma = 0$)

As mentioned above, we first assume uniform delegation and a single delegation option per delegator, and derive a lower bound on the maximum weight. To state our results rigorously, we say that a sequence $(\mathcal{E}_m)_m$ of events happens *with high probability* if $\Pr[\mathcal{E}_m] \to 1$ for $m \to \infty$. Since the parameter going to infinity is clear from the context, we omit it.

Theorem 1.9. *In the preferential delegation model with $k = 1$, $\gamma = 0$, and $d \in (0, 1)$, with high probability, the maximum weight of any voter at time t is in $\Omega(t^\beta)$, where $\beta > 0$ is a constant that depends only on d.*

Proof. It suffices to show that, with high probability, there exists a voter at every time t whose weight is bounded from below by a function in $\Omega(t^\beta)$.

For ease of exposition, we pretend that $i_{max} \coloneqq \log_2 \frac{t}{\ln t}$ is an integer.[6] We divide the t agents into $i_{max} + 1$ blocks $B_0, \ldots, B_{i_{max}}$. The first block B_0 contains agents 1 to $\tau \coloneqq \ln t$, and every subsequent block B_i contains agents $(\tau\, 2^{i-1}, \tau\, 2^i]$.

[6] The same argument works for $i_{max} \coloneqq \left\lfloor \log_2 \frac{t}{\ln t} \right\rfloor$ if we appropriately bound the term.

We keep track of the total weight S_i of all voters in B_0 after the entirety of block B_i has been added. Furthermore, we define an event X_i saying that a high enough number of agents in block B_i transitively delegate into B_0. If all X_i hold, $S_{i_{max}}$ scales like a power function. Then, we show that, as t increases, the probability of any X_i failing goes to zero. Thus, our bound on $S_{i_{max}}$ holds with high probability. The total weight of B_0 and the weight of the maximum-weight voter in B_0 can differ by at most a factor of τ, which is logarithmic in t. Thus, with high probability, there is a voter in B_0 whose weight is a power function.

In more detail, let $\varepsilon \coloneqq \frac{1}{2}$ and let $d' \coloneqq (1-\varepsilon)\, d = \frac{d}{2}$. For each $i \geq 0$, let Y_i denote the number of votes from block i transitively going into B_0. Clearly, $S_i = \sum_{j=0}^{i} Y_i$. For $i > 0$, let X_i denote the event that

$$Y_i > d' \, \frac{\tau \left(1 + \frac{d'}{2}\right)^{i-1}}{2}.$$

Bounding the Expectation of Y_i We first prove by induction on i that, if X_1 through X_i hold, then

$$S_i \geq \tau \left(1 + \frac{d'}{2}\right)^{i}. \tag{1.41}$$

For $i = 0$, $S_0 = \tau$ and the claim holds. For $i > 0$, by the induction hypothesis, $S_{i-1} \geq \tau \left(1 + \frac{d'}{2}\right)^{i-1}$. By the assumption X_i,

$$Y_i > d' \, \frac{\tau \left(1 + \frac{d'}{2}\right)^{i-1}}{2}.$$

Thus,

$$S_i = S_{i-1} + Y_i \geq \tau \left(1 + \frac{d'}{2}\right)^{i-1} + d' \, \frac{\tau \left(1 + \frac{d'}{2}\right)^{i-1}}{2} = \tau \left(1 + \frac{d'}{2}\right)^{i-1} \left(1 + \frac{d'}{2}\right) = \tau \left(1 + \frac{d'}{2}\right)^{i}.$$

This concludes the induction and establishes Equation (1.41).

Now, for any agent j in B_i, the probability of transitively delegating into B_0 is

$$d \, \frac{\sum_{v \in V \cap B_0} w_{j-1}(v)}{j-1} \geq d \, \frac{S_{i-1}}{\tau \, 2^i}.$$

Conditioned on $X_1, \ldots, X_{i-1}$, we can thus lower-bound Y_i by a binomial variable $Bin\left(\tau \, 2^{i-1}, d \, \frac{S_{i-1}}{\tau \, 2^i}\right)$ to obtain

$$\mathbb{E}[Y_i \mid X_1, \ldots, X_{i-1}] \geq \tau \, 2^{i-1} \, d \, \frac{S_{i-1}}{\tau \, 2^i} = d \, \frac{S_{i-1}}{2} \geq d \, \frac{\tau \left(1 + \frac{d'}{2}\right)^{i-1}}{2}.$$

Denoting the right hand side by

$$\mu :- d \, \frac{\tau \left(1 + \frac{d'}{2}\right)^{i-1}}{2},$$

note that X_i holds if $Y_i > (1-\varepsilon)\, \mu$.

Failure Probability Goes to 0 Now, we must show that, with high probability, all X_i hold. By underapproximating the probability of delegation by a binomial random variable as before and by using a Chernoff bound, we have for all $i > 0$

$$\Pr[X_i \mid X_1, \ldots, X_{i-1}] \geq \Pr\left[Bin\left(\tau\, 2^{i-1}, d\, \frac{\tau\, (1 + d'/2)^{i-1}}{\tau\, 2^i}\right) > (1 - \varepsilon)\, \mu\right] \geq 1 - e^{-\frac{\varepsilon^2 \mu}{2}}.$$

By the union bound,

$$\Pr[\exists i, 1 \leq i \leq i_{max} \text{ such that } X_i \text{ fails}] \leq \sum_{i=1}^{i_{max}} e^{-\frac{\varepsilon^2\, d\, \tau\, (1+d'/2)^{i-1}}{4}}.$$

We wish to show that the right hand side goes to 0 as t increases. We have

$$\begin{aligned} \sum_{i=1}^{i_{max}} e^{-\frac{\varepsilon^2\, d\, \tau\, (1+d'/2)^{i-1}}{4}} &\leq i_{max} \left(e^{-\frac{\varepsilon^2\, d\, \tau}{4}}\right) && \text{(by monotonicity)} \\ &= \left(\log_2 \frac{t}{\ln t}\right) \left(t^{-\frac{\varepsilon^2\, d}{4}}\right), && \text{(by definitions of } i_{max},\ \tau) \end{aligned}$$

which indeed approaches 0 as t increases.

Bounding the Maximum Weight Note that the weight of B_0 at time t is exactly $S_{i_{max}}$. Set $x \coloneqq 1 + d'/2 > 1$, which is a constant. With high probability, by Equation (1.41),

$$\frac{S_{i_{max}}}{\tau} \geq \left(1 + \frac{d'}{2}\right)^{i_{max}} = x^{\log_2 \frac{t}{\ln t}} = \left(\frac{t}{\ln t}\right)^{\log_2 x}.$$

Since $x > 1$, $\log_2 x > 0$. For any $0 < \beta < \log_2 x$, $\frac{S_{i_{max}}}{\tau} \in \Omega(t^\beta)$ with high probability. Since B_0 has weight $S_{i_{max}}$ and contains at most τ voters, with high probability there is some voter in B_0 with that much weight. □

Before proceeding to the upper bound and showing the separation, we would like to point out that — with a minor change to our model — these lower bounds also hold for $\gamma = 1$. Consider a model in which the probability of attaching to a delegator n remains proportional to $(1 + \mathit{indegree}(n))^\gamma$, but the probability for voters n is now proportional to $(2 + \mathit{indegree}(n))^\gamma$.[7] If we represent voters with a self-loop edge, both terms just equal $\mathit{degree}(n)^\gamma$, which arguably makes this implementation of preferential attachment cleaner to analyze (e.g., [41]). Thus, we can interpret preferential attachment for $\gamma = 1$ as uniformly picking an edge and then flipping a fair coin to decide whether to attach the new node to the edge's start or endpoint. Since every node has exactly one outgoing edge, this is equivalent to uniformly choosing a node and then, with probability $\frac{1}{2}$, instead picking its successor. This has the same effect on the distribution of weights as just uniformly choosing a node in uniform delegation, so Theorem 1.9 also holds for $\gamma = 1$ in our modified setting. Real-world delegation networks, which we suspect to resemble the case of $\gamma = 1$, should therefore exhibit similar behavior.

[7] Clearly, our results for $\gamma = 0$ hold for both variants.

Upper Bound for Double Delegation ($k = 2$, $\gamma = 0$)

Analyzing cases with $k > 1$ is considerably more challenging. One obstacle is that we do not expect to be able to incorporate optimal resolution of potential delegations into our analysis, because the computational problem is hard even when $k = 2$ (see Theorem 1.8). Therefore, we give a pessimistic estimate of optimal resolution via a greedy delegation mechanism, which we can reason about alongside the stochastic process. Clearly, if this stochastic process can guarantee an upper bound on the maximum weight with high probability, this bound must also hold if delegations are optimally resolved to minimize maximum weight.

In more detail, whenever a new delegator is inserted into the graph, the greedy mechanism immediately selects one of the delegation options. As a result, at any point during the construction of the graph, the algorithm can measure the weight of the voters. Suppose that a new delegator suggests two delegation options, to agents a and b. By following already resolved delegations, the mechanism obtains voters a^* and b^* such that a transitively delegates to a^* and b to b^*. The greedy mechanism then chooses the delegation whose voter currently has lower weight, resolving ties arbitrarily.

This situation is reminiscent of a phenomenon known as the "power of choice." In its most isolated form, it has been studied in the *balls-and-bins* model, for example by Azar et al. [13]. In this model, n balls are to be placed in n bins. In the classical setting, each ball is sequentially placed into a bin chosen uniformly at random. With high probability, the fullest bin will contain $\Theta(\log n / \log \log n)$ balls at the end of the process. In the choice setting, two bins are independently and uniformly selected for every ball, and the ball is placed into the emptier one. Surprisingly, this leads to an exponential improvement, where the fullest bin will contain at most $\Theta\left(\log \log n\right)$ balls with high probability.

We show that, at least for $\gamma = 0$ in our setting, this effect outweighs the "rich-get-richer" dynamic described earlier:

Theorem 1.10. *In the preferential delegation model with $k = 2$, $\gamma = 0$, and $d \in (0, 1)$, the maximum weight of any voter at time t is $\log_2 \ln t + \Theta(1)$ with high probability.*

Because the proof of Theorem 1.10 is quite intricate and technical, we only present a sketch of its structure here. In our proof we build on work by Malyshkin and Paquette [167], who study the maximum *degree* in a graph generated by preferential attachment with the power of choice. In addition, we incorporate ideas by Haslegrave and Jordan [124]. Proofs for the individual lemmas can be found in the full version of the paper [120].

For our analysis, it would be natural to keep track of the number of voters v with a specific weight $w_j(v) = k$ at a specific point j in time. In order to simplify the analysis, we instead keep track of random variables

$$F_j(k) \coloneqq \sum_{\substack{v \in V \\ w_j(v) \geq k}} w_j(v),$$

i.e., we sum up the weights of all voters with weight at least k. Since the total weight increases by one in every step, we have

$$\forall j.\ F_j(1) = j, \text{ and} \tag{1.42}$$

$$\forall j, k.\ F_j(k) \leq j. \tag{1.43}$$

If $F_j(k) < k$ for some j and k, the maximum weight of any voter must be below k.

If we look at a specific $k > 1$ in isolation, the sequence $(F_j(k))_j$ evolves as a Markov process initialized at $F_1(k) = 0$ and then governed by the rule

$$F_{m+1}(k) - F_m(k) = \begin{cases} 1 & \text{with probability } d\left(\frac{F_m(k)}{m}\right)^2 \\ k & \text{with probability } d\left(\left(\frac{F_m(k-1)}{m}\right)^2 - \left(\frac{F_m(k)}{m}\right)^2\right). \\ 0 & \text{else} \end{cases} \tag{1.44}$$

In the first case, both potential delegations of a new delegator lead to voters who already had weight at least k. We must thus give her vote to one of them, increasing $F_m(k)$ by one. In the second case, a new delegator offers two delegations leading to voters of weight at least $k-1$, at least one of which has exactly weight $k-1$. Our greedy algorithm will then choose a voter with weight $k-1$. Because this voter is counted in the definition of $F_j(k)$, $F_m(k)$ increases by k. Finally, if a new voter appears, or if a new delegator can transitively delegate to a voter with weight less than $k-1$, then $F_m(k)$ does not change.

In order to bound the maximum weight of a voter, we first need to get a handle on the general distribution of weights. For this, we define a sequence of real numbers $(\alpha_k)_k$ such that, for every $k \geq 1$, the sequence $\frac{F_j(k)}{j}$ converges in probability to α_k. Set $\alpha_1 \coloneqq 1$. For every $k > 1$, let α_k be the unique root $0 < x < \alpha_{k-1}$ of the polynomial

$$a_k(x, p) \coloneqq d\, x^2 + k\, d\, (p^2 - x^2) - x \tag{1.45}$$

for p set to α_{k-1}.[8] Since $a_k(0, \alpha_{k-1}) > 0$ and $a_k(\alpha_{k-1}, \alpha_{k-1}) < 0$, such a solution exists by the intermediate value theorem. Because the polynomial is quadratic, such a solution must be unique in the interval. It follows that the α_k form a strictly decreasing sequence in the interval $(0, 1]$.

The sequence $(\alpha_k)_k$ converges to zero, and eventually does so very fast. However, this is not obvious from the definition and, depending on d, the sequence can initially decrease slowly. In the full proof, we demonstrate convergence to zero, as well as show that the sequence decreases at a rate in $\mathcal{O}(k^{-2})$. Based on this, we choose an integer k_0 such that the sequence decreases very fast from there. In the same lemma, we define a more nicely behaved sequence $(f(k))_{k \geq k_0}$ that is a strict upper bound on $(\alpha_k)_{k \geq k_0}$ and that is contained between two doubly-exponentially decaying functions.

Lemma 1.2.6. *For all $k \geq 1$, $\varepsilon > 0$ and functions $\omega(m)$ such that $\omega(m) \to \infty$ and $\omega(m) < m$ (for sufficiently large m),*

$$\Pr\left[\exists j,\ \omega(m) \leq j \leq m \text{ s.t. } \frac{F_j(k)}{j} > \alpha_k + \varepsilon\right] \to 0.$$

[8] The equation $0 = a_k(x, p)$ can be obtained from Equation (1.44) by naïvely assuming that $\frac{F_j(k-1)}{j}$ converges to a value p and $\frac{F_j(k)}{j}$ converges to x, then plugging these values in the expectation of the recurrence.

Proof sketch. The proof proceeds by induction on k. For $k = 1$, the claim directly holds. For larger k, we use a suitably chosen δ in place of ε and ω_0 in place of ω for the induction hypothesis. With the induction hypothesis, we bound the $\frac{F_m(k-1)}{m}$ term in the recurrence in Equation (1.44). Furthermore, all increments $F_j(k) - F_{j-1}(k)$ where $\frac{F_{j-1}(k)}{j-1} \geq \alpha_k$ holds can be dominated by independent and *identically distributed* random variables η'_j.

Denote by π the first point $j \geq \omega_0(m)$ such that $\frac{F_j(k)}{j} \leq \alpha_k + \frac{\varepsilon}{2}$. The η'_j then dominate all increments $F_j(k) - F_{j-1}(k)$ for $\omega_0(m) < j \leq \pi$. Using Chernoff's bound and suitably chosen δ and ω_0, we show that, with high probability, $\pi \leq \omega(m)$.

Because of this, if $\frac{F_j(k)}{j} > \alpha_k + \varepsilon$ for some $j \geq \omega(m)$, the sequence $\left(\frac{F_j(k)}{j}\right)_j$ must eventually cross from below $\alpha_k + \frac{\varepsilon}{2}$ to above $\alpha_k + \varepsilon$ without in between falling below α_k. On this segment, we can overapproximate the sequence by a random walk with increments distributed as η'_j. Since the sequence might previously decrease below α_k an arbitrary number of times, we overapproximate the probability of ever crossing $\alpha_k + \varepsilon$ for $j \geq \omega(m)$ by a sum over infinitely many random walks. This sum converges to 0 for $m \to \infty$, which shows our claim. □

The above lemma gives us a good characterization of the behavior of $(F_j(k))_j$ for any fixed k (and large enough j). To prove an upper bound on the maximum weight, however, we are ultimately interested in statements about $F_j(k(m))$, where $k(m) \in \Theta(\log_2 \ln m)$ and the range of j varies with m. In order to obtain such results, we will first show in Lemma 1.2.7 that whole ranges of k simultaneously satisfy bounds with high probability.

As in the previous lemma, we can only show our bounds with high probability for j past a certain period of initial chaos. We will define a function, $\phi(m, k)$, that takes a role similar to $\omega(m)$ in Lemma 1.2.6. The function $\phi(m, k)$ gives each k a certain amount of time to satisfy the bounds, depending on m: Let $\rho(m) := (\ln \ln m)^{\frac{1}{3}}$ and define $\phi(m, k) := \rho(m)\, C^{2^{k+1}}$, where C is an integer that is sufficiently large to satisfy

$$\ln C > \max\left(1,\ c_1,\ \ln\left(\frac{2}{1-d}\right) + \frac{c_1}{2}\right). \tag{1.46}$$

In the above, c_1 is a postive constant defining the lower bound on $f(k)$.

Additionally, let $k_*(m)$ be the smallest integer such that

$$C^{2^{k_*(m)+1}} \geq \sqrt{m}. \tag{1.47}$$

Note that $C^{2^{k_*(m)+1}} < m$ because increasing the double exponent in increments of 1 is equivalent to squaring the term. By applying logarithms to $C^{2^{k_*(m)+1}} \geq \sqrt{m}$ and $C^{2^{k_*(m)+1}} < m$, we obtain $\log_2 \log_C m - 2 \leq k_*(m) < \log_2 \log_C m - 1$, from which it follows that $k_*(m) = \log_2 \ln m + \Theta(1)$.

Lemma 1.2.7. *With high probability, for all $k_0 \leq k \leq k_*(m)$, and for all $\phi(m, k) \leq j \leq m$,*

$$\frac{F_j(k)}{j} \leq f(k).$$

Proof sketch. Let $\mathcal{G}_k$ be the event

$$\mathcal{G}_k := \left\{ \forall j, \phi(m,k) \le j \le m.\ \frac{F_j(k)}{j} \le f(k) \right\}.$$

Our goal is to show that $\mathcal{G}_k$ holds for all k in our range. In the spirit of an inductive argument, we begin by showing $\mathcal{G}_{k_0}$ with high probability and then give evidence for how, under the assumption $\mathcal{G}_k$, $\mathcal{G}_{k+1}$ is likely to happen. Instead of an explicit induction, we piece together these parts in a union bound.

The base case $\mathcal{G}_{k_0}$ follows from Lemma 1.2.6 with $\omega(m) := \phi(m, k_0)$ and $\varepsilon := f(k_0) - \alpha_{k_0}$.

For the step, fix some $k \ge k_0$, and assume $\mathcal{G}_k$. We want to give an upper bound on the probability that $\mathcal{G}_{k+1}$ happens. We split this into multiple substeps: First, we prove that, given $\mathcal{G}_k$, some auxiliary event $\mathcal{E}(k+1)$ happens only with probability converging to 0. Then, we show that $\overline{\mathcal{E}(k+1)} \subseteq \mathcal{G}_{k+1}$ where $\overline{\mathcal{E}}$ denotes the complement of an event $\mathcal{E}$. This means that, whenever the unlikely event does not take place, $\mathcal{G}_{k+1}$ holds. This allows the step to be repeated.

If $\mathcal{G}_k$ does not hold for any $k_0 \le k \le k_*(m)$, then $\overline{\mathcal{G}_{k_0}}$ or one of the $\mathcal{E}(k)$ must have happened. The union bound converges to zero for $m \to \infty$, proving our claim. □

As promised, the last lemma enables us to speak about the behavior of $F_j(k(m))$. We will use a sequence of such statements to show that, with high probability, $F_j(k(m))$ for some $k(m)$ does not change over a whole range of j:

Lemma 1.2.8. *There exists $M > 0$ and an integer $r > 0$ such that, for $j_0(m) := (\ln \ln m)^M$, $F_m(k_*(m) + r) = F_{j_0(m)}(k_*(m) + r)$ holds with high probability. In addition, there is $\beta > \frac{1}{2}$ such that, with high probability,*

$$F_{j_0(m)}(k_*(m) + r - 1) \le j_0(m)^{1-\beta}. \tag{1.48}$$

Proof sketch. We finally get a statement about $F_j(k_*(m))$: By choosing different k for different j in Lemma 1.2.7, we obtain a constant $\beta_0 > 0$ such that, with high probability,

$$\forall j, \ln \ln m \le j \le m.\ \frac{F_j(k_*(m))}{j} \le j^{-\beta_0}.$$

We now increase β_0 until it is larger than $\frac{1}{2}$. Set $r'_0 := 0$ and $M_0 := 1$. In fact, we obtain a stronger proposition of the form

$$\forall j, (\ln \ln m)^{M_i} \le j \le m.\ \frac{F_j(k_*(m) + r'_i)}{j} \le j^{-\beta_i}$$

holding with high probability to obtain, for some $M_{i+1} > 0$ and with high probability,

$$\forall j, (\ln \ln m)^{M_{i+1}} \le j \le m.\ \frac{F_j(k_*(m) + r'_i + 1)}{j} \le j^{-\frac{3}{2}\beta_i}.$$

If we set $r'_{i+1} := r'_i + 1$ and $\beta_{i+1} := \frac{3}{2}\beta_i$, we can repeatedly apply this argument until some $\beta_i > \frac{1}{2}$. Let M, r' and β denote M_i, r'_i and β_i, respectively, for this i. If, furthermore, $r := r' + 1$, Equation (1.48) follows as a special case.

We then simply union-bound the probability of $F_j(k_*(m) + r)$ increasing for any j between $j_0(m)$ and m. Using the above over-approximation in Equation (1.44) gives us an over-harmonic series, whose value goes to zero with $m \to \infty$. $\square$

We can now prove Theorem 1.10. Let Q_i denote the maximum weight after i time steps.

Proof of Theorem 1.10. By Lemma 1.2.8, with high probability, $F_m(k_*(m)+r) = F_{j_0(m)}(k_*(m)+r)$. Therefore, we have that with high probability

$$\begin{aligned}
F_m(k_*(m) + r) &= F_{j_0(m)}(k_*(m) + r) \\
&\leq F_{j_0(m)}(k_*(m) + r - 1) && \text{(by monotonicity)} \\
&\leq j_0(m)^{1-\beta} && \text{(by Equation (1.48))} \\
&= \left((\ln \ln m)^M\right)^{1-\beta} \\
&\leq (\ln \ln m)^{M+1}.
\end{aligned}$$

For any j and k, $Q_j \leq \max\{k, F_j(k)\}$. Since, for large enough m, $k_*(m) + r < (\ln \ln m)^{M+1}$, the maximum weight Q_m is at most $(\ln \ln m)^{M+1}$ with high probability. This result holds for general m, so we are allowed to plug in $j_0(m)$ for m. Then, $Q_{j_0(m)} \leq (\ln \ln j_0(m))^{M+1}$. Moreover, $(\ln \ln j_0(m))^{(M+1)^2} < j_0(m)$ for sufficiently large m because M is a constant and polylogarithmic terms grow asymptotically slower than polynomial terms. Rewriting this yields

$$Q_{j_0(m)} \leq (\ln \ln j_0(m))^{M+1} < j_0(m)^{1/(M+1)}. \tag{1.49}$$

Now, note that $k_*(m) + r \geq \left(j_0(m)^{1/(M+1)}\right)$ for large enough m. Therefore, Equation (1.49) implies that, with high probability, a graph generated in $j_0(m)$ time steps has no voters of weight $k_*(m) + r$ or higher. In other words, with high probability, $F_{j_0(m)}(k_*(m)+r) = 0$, so with high probability $F_m(k_*(m)+r) = 0$ (again by Lemma 1.2.8). This means that the maximum weight after m time steps is also upper-bounded by $k_*(m)+r = \log_2 \ln m + \Theta(1)$. $\square$

1.2.5 Simulations

In this section, we present our simulation results, which support the two main messages of this paper: that allowing multiple delegation options significantly reduces the maximum weight, and that it is computationally feasible to resolve delegations in a way that is close to optimal. x Our simulations were performed on a MacBook Pro (2017) on MacOS 10.12.6 with a 3.1 GHz Intel Core i5 and 16 GB of RAM. All running times were measured with at most one process per processor core. Our simulation software is written in Python 3.6 using Gurobi 8.0.1 to solve MILPs. All of our simulation code is open-source and available at https://github.com/pgoelz/fluid.

Multiple vs. Single Delegations

For the special case of $\gamma = 0$, we have established a doubly exponential, asymptotic separation between single delegation ($k = 1$) and two delegation options per delegator ($k = 2$). While the strength of the separation suggests that some of this improvement will carry over to the real world, we still have to examine via simulation whether improvements are visible for realistic numbers of agents and other values of γ.

To this end, we empirically evaluate two different mechanisms for resolving delegations. First, we optimally resolve delegations by solving the MILP for confluent flow with the Gurobi optimizer. Our second mechanism is the greedy "power of choice" algorithm used in the theoretical analysis and introduced in Section 1.2.4.

In Figure 1.4, we compare the maximum weight produced by a single-delegation process to the optimal maximum weight in a double-delegation process, for different values of γ and d. Since our theoretical analysis used a greedy over-approximation of the optimum, we also run the greedy mechanism on the double-delegation process.

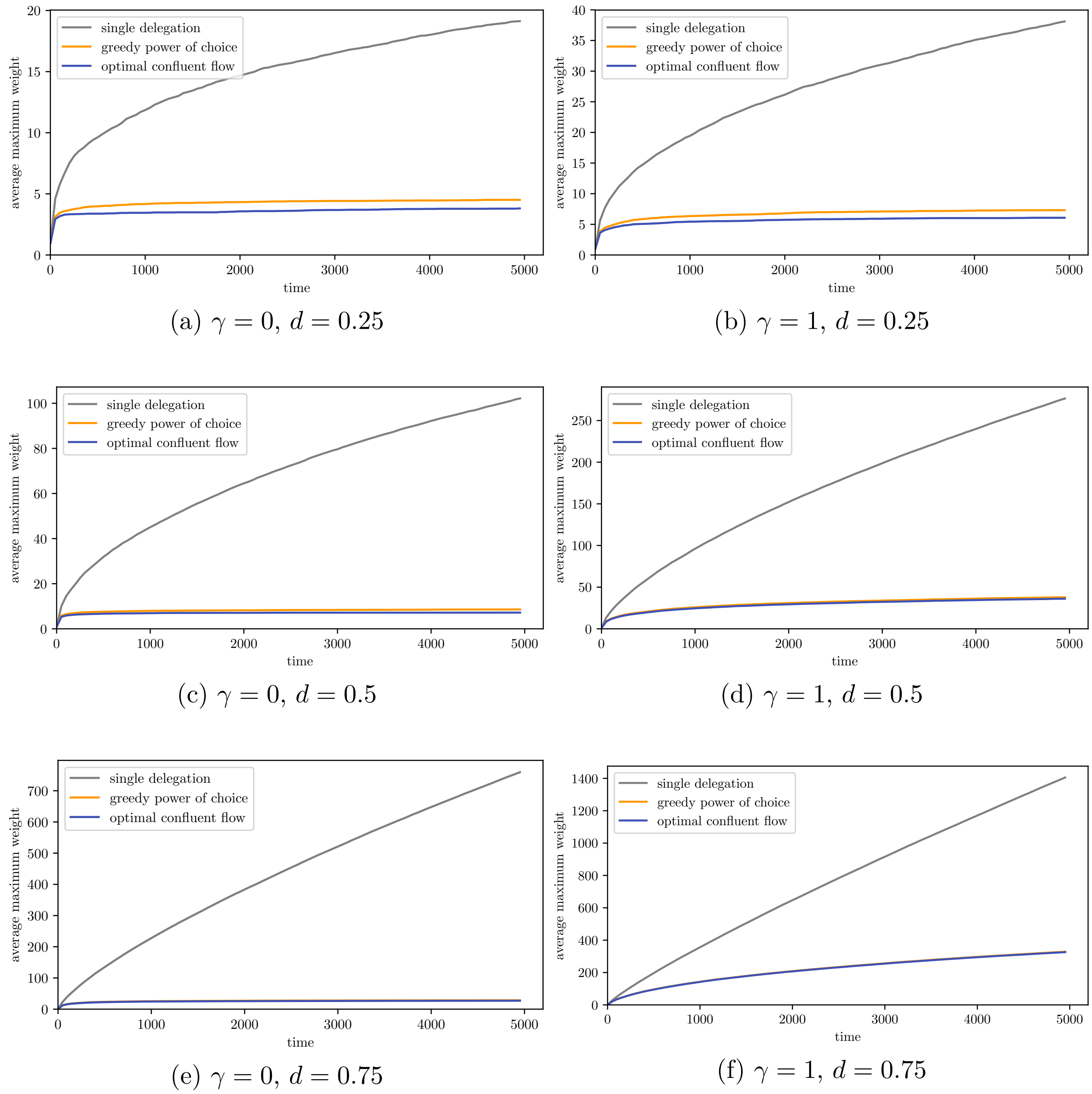
These simulations show that our asymptotic findings translate into considerable differences even for small numbers of agents, across different values of d. Moreover, these differences remain nearly as pronounced for values of γ up to 1, which corresponds to classical preferential attachment. This suggests that our mechanism can outweigh the social tendency towards concentration of votes; however, evidence from real-world elections is needed to settle this question. Lastly, we would like to point out the similarity between the graphs for the optimal maximum weight and the result of the greedy algorithm, which indicates that a large part of the separation can be attributed to the power of choice.

If we increase γ to large values, the separation between single and double delegation disappears. In Figure 1.5a, for $\gamma = 2$, all three curves are hardly distinguishable from the linear function $d \cdot \mathit{time}$, meaning that one voter receives nearly all the weight. The reason is simple: In the simulations used for that figure, 99% of all delegators give two identical delegation options, and 99.8% of these delegators (98.8% of all delegators) give both potential delegations to the heaviest voter in the graph. There are even values of $\gamma > 1$ and d such that the curve for single delegation falls below the ones for double delegation. This can be seen in Figure 1.5b, where 87.7% of voters give two identical delegation options. Since adding two delegation options per step makes the indegrees grow faster, the delegations concentrate toward a single voter more quickly, and again lead to a wildly unrealistic concentration of weight. Thus, it seems that large values of γ do not actually describe our scenario of multiple delegations.

As we have seen, switching from single delegation to double delegation greatly improves the maximum weight in plausible scenarios. It is natural to wonder whether increasing k beyond 2 will yield similar improvements. As Figure 1.6 shows, however, the returns to increasing k quickly diminish, which is common to many incarnations of the power of choice [13].

Evaluating Mechanisms

Already the case of $k = 2$ appears to have great potential; but how easily can we tap it?



(a) $\gamma = 0,\ d = 0.25$

(b) $\gamma = 1,\ d = 0.25$

(c) $\gamma = 0,\ d = 0.5$

(d) $\gamma = 1,\ d = 0.5$

(e) $\gamma = 0,\ d = 0.75$

(f) $\gamma = 1,\ d = 0.75$

Figure 1.4: Maximum weight averaged over 100 simulations of length 5 000 time steps each. Maximum weight has been computed every 50 time steps.

(a) $= 2$, $d = 0.5$ (b) $= 1.5$, $d = 0.5$

Figure 1.5: Maximum weight averaged over 100 simulations, computed every 50 time steps.

Figure 1.6: Optimal maximum weight for different k averaged over 100 simulations, computed every 10 steps. $= 1$, $d = 0.5$.

Figure 1.7: Optimal maximum weight averaged over 100 simulations. Voters give two delegations with probability p; else one. $= 1$, $d = 0.5$.

We have observed that, on average, the greedy power of choice mechanism comes surprisingly close to the optimal solution. However, this greedy mechanism depends on seeing the order in which our random process inserts agents and on the fact that all generated graphs are acyclic, which need not be true in practice. If the graphs were acyclic, we could simply first sort the agents topologically and then present the agents to the greedy mechanism in reverse order. On arbitrary active graphs, we instead proceed through the strongly connected components in reverse topological order, breaking cycles and performing the greedy step over the agents in the component. To avoid giving the greedy algorithm an unfair advantage, we use this generalized greedy mechanism throughout this section. Thus, we compare the generalized greedy mechanism, the optimal solution, the $(1 + \ln |V|)$-approximation algorithm[9] and a random mechanism that chooses a uniformly chosen option per delegator.

At a high level, we find that both the generalized greedy algorithm and the approximation algorithm perform comparably to the optimal confluent flow solution, as shown in

[9] For one of their subprocedures, instead of directly optimizing a convex program, Chen et al. [68] reduce this problem to finding a lexicographically optimal maximum flow in $O(n^5)$. We choose to directly optimize the convex problem in Gurobi, hoping that this will increase efficiency in practice.

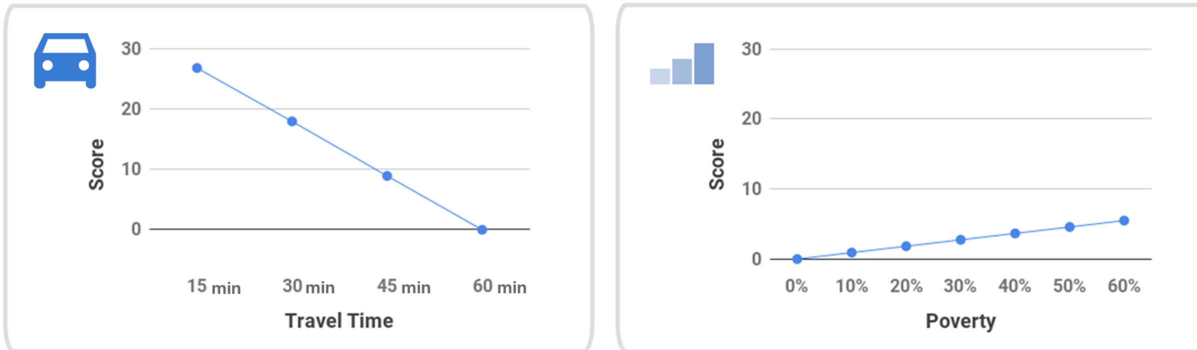


Figure 4.4: Model explanations. Both machine learning and explicit-rule models were represented by graphs that assigned scores according to the varying levels of input features.

Final Individual Models

In total, we trained 23 machine learning models[13] and obtained 15 explicit-rule models. Of the 15 participants who completed all studies and were asked to choose models that better represented their belief, 10 of them chose the machine learning models trained on their pairwise comparisons; the other five chose the models that they explicitly specified.

The machine learning models had higher overall agreement with participant's survey answers than the explicit rule models when tested on 50 new pairwise comparisons provided by each participant, as seen in Table 4.3. However, as our sample size is small, we do not aim to make general claims on which model has better accuracy. In addition, for many, the machine learning model was the one they had built last and therefore reflected their current thinking at the time of comparison; we further elaborate on this in Section 8.1. We also note that we did not observe any differences in participants' perceived accountability in the creation of these models. Both models took an equal amount of participants' time and attention, and participants told us that they felt responsible when making choices and assigning scores.

4.2.6 Collective aggregation

Our framework uses a voting method to aggregate individuals' beliefs. When presented with a new donation decision, each individual's model generates a complete ranking of all possible recipient organizations. The Borda rule aggregates these rankings to derive a consensus ranking and suggest recommendations. We conducted a workshop and interviews to understand participants' perceptions of this method.

[13]We note that there were 8 participants who participated in the first stage of the study but not subsequent stages (Table 4.1). The average cross-validation accuracy of their linear models was quite high, at 0.819.

Method (Workshop)

In an early stage of our research, we conducted a workshop in order to gauge participants' perceptions of the Borda aggregation method and determine the method's appropriateness from a social perspective. Five participants (Table 4.1) who had built their individual models at that time attended the one-hour workshop. All stakeholder groups were represented. We prepared a handout that showed individuals' and stakeholders' average models at the time, and a diagram that explained how the Borda rule worked. The description of the Borda rule given to participants was: "Individuals rank options according to belief. Each option receives a number of points determined by its ranking, with higher-ranked options receiving more points. The points are added up, and the winner is whichever option has the greatest number of points." The words "democratic" or "equal" were not used to avoid potential biases. We facilitated a discussion of how individuals reacted to the similarities and differences between their model and other groups' models, and had individuals discuss whether all the stakeholders' opinions should be weighted equally or differently. For participants who joined our research after this workshop, we asked the same questions about the Borda rule and stakeholder opinion weight in the interview in Session 4.

Varying Stakeholders' Voting Influence

All participants but one believed that the weight given to different stakeholders in the final algorithm should depend on their roles. On average, participants assigned 46% of the voting power to 412 Food Rescue, 24% to recipient organizations, 19% to volunteers, and 11% to donors.[14] Nearly all participants weighted 412 Food Rescue staff as the highest group (n=13 out of 15), as people recognized that they manage the operation and have the most knowledge of the whole system. Donors were weighted the least (or tied for least) by nearly all participants (n=14 out of 15) including the donors themselves, as they are not involved in the process once the food leaves their doors. Recipients and volunteers were weighted similarly because participants recognized that recipient opinions are important to the acceptance of donations, and volunteer drivers have valuable experience interacting with both donors and recipients. In order to translate these weights to Borda aggregation, we allocated each stakeholder group a total number of votes that was commensurate with their weight, and divided up the votes evenly within each group. For example, 412 Food Rescue employees are assigned 46% of the weight; this translates to allocating them 46 votes out of 100 total as a group, where each employee's vote is "replicated" 46/3 times because three 412 Food Rescue employees participated in our study.

4.2.7 Explanation and decision support

Once recommendations are generated, the decision support interface presents the top twelve organizations, accompanied by explanations, to support the human decision-maker who matches incoming donations to recipients. We used the explanations to demonstrate to participants how their opinions had been incorporated into the algorithm's decision-making.

[14] This is based on the input from participants that participated in the workshop and/or Session 4.

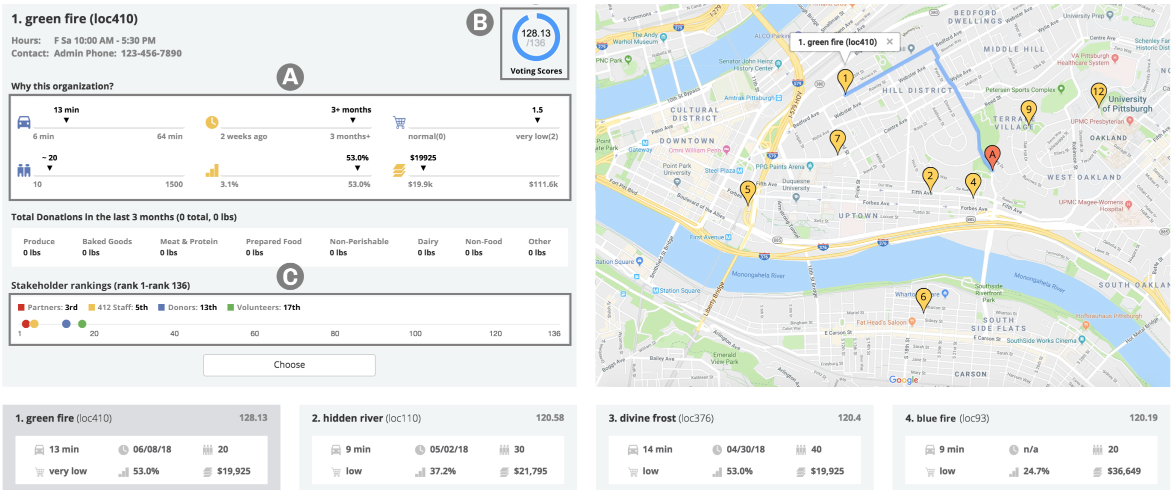


Figure 4.5: The decision support tool explains algorithmic recommendations, including the nature of stakeholder participation, stakeholder voting results, and characteristics of each recommendation. The interface highlights the features of the recommended option that led to its selection (marked by A), the Borda scores given to the recommended options in relation to the maximum possible score (marked by B), and how each option was ranked by stakeholder groups (marked by C). All recipient information and locations are fabricated for the purpose of anonymization.

We also explained average stakeholder models to participants so that they could learn about others' models.

Design of Decision-Support Tool

The interface of our decision support tool is shown in Figure 4.5. The tool was designed with other considerations, such as choice architecture [217], but they are beyond the scope of this paper. We focus instead on the explanation of decisions made by collectively-built algorithms.

- Decision Outcome Explanation (marked by A in Figure 4.5): We used an "input influence" style explanation [33]. Features are highlighted in yellow when an organization is in the top 10% of recipient organizations ranked by that factor. For example, poverty rate is highlighted because the selected organization is in the top 10% of recipients when ranked from highest to lowest poverty rate.
- Voting Score (marked by B in Figure 4.5): The Borda score for each organization is displayed. It shows this option's score in relation to the maximum possible score that an option could receive (i.e., scores when every individual model picks this option as its first choice). This voting score can indicate the degree of consensus among participants.
- Stakeholder Rankings (marked by C in Figure 4.5): Stakeholder rankings show how each stakeholder group ranked the given organization on average. It is a visual re-

minder that all stakeholder groups are represented in the final algorithm and gives the decision-maker additional information about the average opinion of each stakeholder group.

We implemented the interface by integrating it into a customer relations management system currently in use at 412 Food Rescue. Algorithms were coded in Ruby on Rails, the front-end interface used Javascript and Bootstrap, and the database was built with Postgres. The distances and travel times between donors and recipients were pre-computed using the Google Maps API and Python. We used donor and recipient information from the past five months of donation records in the database. On average, the algorithm produced recommendations for each donation in five seconds.

Method (Session 4)

We conducted a one-hour study with each participant to understand how the decision support and explanation influenced their perceptions of the matching algorithms and their attitude toward 412 Food Rescue. In order to generate summary beta vectors for each stakeholder group, we normalized the beta vectors for all stakeholders in the group and took the pointwise average. This yields a summary beta vector where the value of each feature roughly reflects the average weight that stakeholders in the same group give to that feature.

We first showed participants the graphs of their individual models and graphs of the averaged models for each stakeholder group, and asked participants to examine similarities and differences among these models. We next had participants interact with the decision support tool run on a researcher's laptop. The researcher walked participants through the interface, explaining the information and recommendations, and asked them to review the recommendations and pick one to receive the donation. After each donation, participants were asked their opinions of the recommendations, the extent to which they could see their models reflected in the results, and their general experience. We concluded with a 30 minute semi-structured interview in which we asked how participation influenced their attitude toward algorithms and 412 Food Rescue. We also asked participants to reflect on the overall process of giving feedback throughout our studies.

4.2.8 Findings: the impact of participatory algorithm design

In the previous sections, we described how stakeholders used the WeBuildAI framework to build the matching algorithm for 412 Food Rescue over multiple sessions and a workshop. We now report the qualitative findings from observations, think-alouds, and interviews to describe the impacts of the WeBuildAI framework and participation.

Participants' Experience with the WeBuildAI Framework

Overall, the individual belief-elicitation step of the framework—using both machine learning and explicit rule specification methods and visualizing the learned models—successfully enabled participants to build an individual model that represented their beliefs on how the algorithm should make a matching decision. Participants perceived the automatic aggregation method based on the Borda rule as a nuanced, democratic approach; the decision support tool and explanation allowed them to understand how algorithmic recommendations were made. **Effects of Individual Model Building Methods on Elicited Beliefs**:

Participants told us that performing pairwise comparisons and subsequently specifying explicit rules helped them develop and consolidate their beliefs into a set of principles that they could apply consistently in different decision contexts. Answering pairwise comparison questions helped familiarize participants with the problem setting; however, some participants commented that they felt like they were applying internal rules inconsistently, particularly in their first few questions. Explicitly specifying scores for each feature helped them reconcile their conflicting beliefs. For example, V1 told us that she originally used organization size inconsistently, sometime favoring smaller organizations or bigger organizations, but when creating a rule, she determined that organization size should not matter. When she answered the new set of pairwise comparison questions to retrain the machine learning model, she further evaluated whether she could consistently apply her belief, i.e., that the organization size does not matter, to different contexts and whether she encountered any new situations in which she would need to further refine her rule.

In choosing the model to use in the final matching algorithm, the most important factor for all participants, except one, was how closely each model represented their beliefs. In Session 3, 10 out of 15 participants chose their machine learning models. For many, this was the model they had built last and therefore that reflected their thinking at the time of comparison. Others felt that the machine learning model had more nuance in the way different factors were weighted, and some valued the linearity of the model compared to their manual rules, which were often step-wise functions. Explicit-rule models were chosen by five participants. For four of these participants, their explicit-rule model did a better job of weighing all of the factors that mattered to them and screening off unimportant factors. In other words, machine learning models learned rules that they disagreed with—for example, a machine learning model may give linearly increasing weight to larger organization sizes.

On the contrary, for one participant, the procedural difference in the two methods was why he chose the explicit-rule model. R2 trusted the reflective process of specifying a model and did not trust his pairwise answers nor the machine learning model built from them, even though the accuracy of the machine learning model was 90%, compared to 76% for the model that he created. He believed that determining policy should be based on defining principles, rather than case-by-case decisions; for this reason, he wanted to build a rule and follow the outcomes from the rule.

An unexpected finding was that the methods' procedural differences seemed to influence which aspects participants focused on at the time of decision-making and, in some cases, the rules that participants made. Creating a scoring model from a top-down approach seemed to evoke a higher level of construal [220], eliciting an abstract level of thinking that was absent when answering pairwise comparisons. Many participants stated answering pairwise comparisons felt emotional because it made them think of real-world organizations. For example, V1 said that developing explicit scoring rules felt "robotic"; R3 said that he felt that creating the scoring model was easier than the pairwise comparisons because it took the emotion out of the decision-making process. For an administrative decision-maker, F3, answering pairwise questions made her focus on day-to-day operational issues like travel time because she related the questions to real-world decision-making. This contrasted with her explicit-rule model, which favored equity-related factors like income and poverty. When comparing the models in Session 3, she told us that she focused on idealistic matching that prioritized equity when she was specifying scoring rules. In the end, she chose her machine learning model, stating that while her explicit-rule model was appealing as a way of pushing herself beyond her operational thinking, she deemed travel time and last donation date most important in practice.

Responses to the Borda-Based Aggregation: Participants appreciated that the Borda method gave every recipient organization a score (n=5) and that it embodied democratic values (n=4).[15] In the workshop, F1 felt that giving every organization a score captured the subtleties of her thinking better than other methods, such as considering only the top-ranked organization: *"I appreciate the adding up [of] scores. Recognize the subtleties."* V3 also stated that being able to rank all recipients is *"more true to...[being] able to express your beliefs."* R1 approved of the method, saying, *"It's very democratic,"* relating it to a form of human governance. Two other individuals, D2 and D4, also related the method to voting systems in the US. D4 recognized that some US cities in California recently used a similar voting method for their mayoral election. It is worthwhile to note that, when we asked about potential alternatives, participants expressed difficulty thinking of them (n=3). For example, R2 said, *"I guess I don't know what the alternative way to do it would be, so I'm okay with it."*

Responses to the Decision Support Interface: Participants were almost universally appreciative of the fact that the system keeps a human dispatcher in the loop to make the final decision rather than automating the decision entirely. While some participants (F1 and R5) acknowledged that full automation could be more efficient than a human-in-the-loop process, most participants expressed that having a human dispatcher overseeing the process was important as they might have knowledge of additional decision factors outside the scope of the algorithm. F3 expressed that the combination of human and computer decision-making elements was "magical" in that it combined the objective data of an algorithm with human elements *"that the computer will never know... like so and so at this place loves peaches and they make peach pies."* Others (e.g., R2) expressed that the algorithm could enable human decision-making in a way that reduces bias or favoritism on the part of the dispatcher, thereby making the decisions of the organization more fair and objective.

Participants were interested in the stakeholder rankings and asked to see more information. Given that the top twelve results often did not show the first choice for any stakeholder group, several participants wanted to see the first choice for each stakeholder group in addition to the voting aggregation scale (n=7). Participants appreciated that the stakeholder rankings showed opinions that might differ from those of 412 Food Rescue dispatchers (n=4). V6, who was concerned that 412 Food Rescue staff did not heavily weight

[15] We note that the description of Borda given to participants described a scoring process and did not include words such as "voting" and "democracy" as reported in Section 6.1.

factors that were important to her, was pleased that the voter preference scale illustrated the difference between her stakeholder group's average model and 412 Food Rescue's average model. She hoped that the staff would see that their thinking differed from other stakeholders and perhaps reconsider their decisions in order to be more inclusive of other groups' opinions. 412 Food Rescue staff were interested in the information as well and F3 mentioned that, while she would not solely base her decisions on stakeholder ranking information, she might use it as a tiebreaker between two similar organizations.

Participation and Perceptions of Algorithmic Governance

In a manner consistent with theories on procedural justice [160; 159] and participatory policy-making [111], participants believed that having control over the algorithm through participatory algorithm design made the resulting algorithm fair, and this process improved their attitudes toward the organization as a whole.

Procedural Fairness in Participatory Algorithm Design: All participants mentioned that the fact that the organization was putting a priority on fairness, being open to new ideas, and including multiple stakeholder groups improved their perceived fairness and trust of both the matching algorithm and the organization itself. For example, one participant said, *"These are everybody's brain power who were deemed to be important in this decision... it should be the most fair that you could get."* Some expressed that participation expanded the algorithm's assumptions beyond those of the organization and developers (n=6). V6 noted that it is easy for organizations to remain isolated in their own viewpoints and that building an algorithm based on collective knowledge was more trustworthy to her than *"412 [Food Rescue] in a closed bubble coming up with the algorithm for themselves."* V3 echoed this sentiment, stating that participation was *"certainly more fair than somebody sitting at a desk trying to figure it out on their own."* At 412 Food Rescue, F2 stated that *"getting input from everyone involved is important"* to challenge organizational assumptions and increase the effectiveness of their work. Other participants noted that all stakeholders have limited viewpoints that can be overcome with collective participation (n=3). R1 felt the algorithm would be fair only *"if you took the average of everybody. ...[My model] is only my experience. And I view my experience differently than the next place down the road. And my experience is subjective."*

Empathetic Stance toward the Governing Organization: Participation in algorithm design led many participants to increase the degree to which they viewed 412 Food Rescue positively and develop a more empathetic stance toward the organization (n=8). For some participants, this happened because participation exposed the difficulty of making donation matching decisions and made them realize that there might not be a perfect solution, which in turn made them thankful for the work of the organization (n=4). For example, after experiencing the burden of making the matching decision and seeing how similar the recommended recipients can be in the interface, D2 and V3 both expressed gratitude for 412 Food Rescue. Participants also expressed appreciation for the organization's concern for fairness and the effort needed to continually make such decisions. This shift in perception is particularly important because it can improve people's tolerance for and understanding of tradeoffs in governance decisions.

The participatory algorithm design also increased some participants' motivation to engage with the organization (n=4). Many participants appreciated that their opinions were valued by the organization enough to be considered in the algorithm building process and expressed that they may increase their involvement with the organization in the future either through increased volunteer work (V3 and V6) or donation acceptance (R2).

Reactions to Other Stakeholders' Models: While sharing other stakeholders' models is not a requirement of our framework, in this work, we showed the models to participants in order to get feedback on the fully transparent implementation of our framework.[16] We report how participants responded to similarities and differences in stakeholder models.

In individual models, all participants considered efficiency and equity factors. For example, all stakeholder group models valued distance as one of the top three factors and favored organizations that were deemed to be in greater need. Reviewing the models, participants expressed feeling assured that they shared these guiding principles with other participants (n=8). For example, all prioritized higher as opposed to lower poverty, and lower as opposed to higher food access. R7 was pleased to note that all participants were *"on the same page"* and concluded that *"no matter what group or individuals we're feeding, [we] have the same regard for the food and the individuals that we're serving."*

A main source of disagreement among models was how the factors were balanced. 412 Food Rescue Staff tended to weight travel time and last donation significantly more than the other factors. Donors, recipients, and volunteers tended to give all factors other than organization size relatively equal importance. Participants also had divided views on organization size, arguing for larger or smaller organizations, and did not prioritize this factor compared to others. In responses, participants acknowledged these differences and sought to make sense of others' assumptions. For example, R1, referencing how important travel time was to her, mentioned that hers is more of a *"business model"* whereas others were more altruistic, more heavily weighting factors like income and food access. Some participants were even pleased to see differences in the models (n=3). R3 was pleased that other participants were considering unique viewpoints. Likewise V4 and R1 both stated that it was natural to expect differences between stakeholders, as everyone has unique experiences, and that *"this is the point of democracy"* (V4).

However, one participant, V6, was concerned that 412 Food Rescue staff did not weight heavily her most important factors such as food access, income, and poverty. While she said that the algorithm was "fair" as it was collectively created, her trust in the organization was lowered as a result, because she inferred that they believe in different principles. She also raised a concern about other participants' input qualities. It took her significant effort to develop a model that accurately represented her views, and she could not judge whether other participants were *"thoughtful enough to really put the effort into their models and capture their own emotions with it."* She concluded that she still trusted the algorithm, but appreciated having human oversight of the final decision.

[16] Participants also told us that they were curious about other stakeholders' beliefs.

Participation and Awareness of Algorithms and Organizational Decision-Making

Our findings suggest that participating in algorithm design improved algorithmic awareness at an individual level, as well as awareness of inconsistencies in decision-making practices at an organizational level.

Increased Algorithmic Awareness: At an individual level, participating in algorithm design changed participants' attitudes toward algorithms.[17] They felt they better understood what an algorithm was and had more appreciation for the kinds of decisions that algorithms could make. For some participants, seeing how the two models predicted their answers in our study session made them rethink their initial skepticism and begin to trust the algorithm. V1, who in earlier studies expressed doubt that an algorithm could be of any use in such a complex decision space, stated at the end of Session 3 that he now "wholeheartedly" trusted the algorithm, a change brought about by seeing the work that went into developing his models and how they performed. F3 expressed that before participating, *"the process of building an algorithm seemed horrible"* given the complexities of allocation decisions. Seeing how the process of building the algorithm was broken down *"into steps ... and just taking each one at a time"* made the construction of an algorithm seem much more attainable. For D2, interacting with the researchers who were building the algorithm gave him an awareness of the role human developers play in determining algorithms. He said that, after this process, his judgment of an algorithm's fairness in other algorithmic systems would be based on *"how it was developed and who's behind it and programmed [it] and how it's influenced."* D2 felt that the final algorithm for 412 Food Rescue was fair because he came to know and trust the researchers over the course of his participation.

Improved Awareness of Inconsistency in Organizational Decision-Making: The process of eliciting individual models allowed participants from the governing organization to be more aware of internal inconsistencies in decision-making within their organization, and provided an opportunity for them to revisit their own assumptions about other stakeholders. Guided only by the broad goals of the organization's mission, the employees previously made matching decisions according to their own criteria and interpretations of that mission. By externalizing their decision-processes into computational belief models, the employees were able to formalize their own decision-making processes, and see how their models meshed with or differed from other employees' processes, which brought hidden assumptions to the surface. For example, after seeing other employees' models, they discovered that some employees prioritized mid-sized organizations whereas others prioritized larger organizations, and employees differed in the ways they weighted poverty, income, and food access.

Moreover, seeing other stakeholders' models allowed employees to compare their assumptions about other stakeholders with the models actually made by the stakeholders. One common assumption held by the staff was that volunteers would prioritize travel time, but our volunteer stakeholders had diverse models, varying from one that predominantly weighted travel time to one that gave equal weights to travel time and recipient organizations' needs. When F2 saw that volunteers did not weight travel time as highly as she

[17] None of our participants had a background in programming.

had thought, she questioned her evaluation of travel time: *"Maybe [volunteers] don't care as much. I think you end up hearing from the people who care... It's like that saying with customer service: Only complain when something's happened."* This reflection opens up the possibility that the organization could seek to appeal to diverse volunteer motivations and tailor recruiting methods accordingly.

4.2.9 Evaluation of Algorithmic Outcomes

Our qualitative findings in the previous section show the procedural effect of participatory algorithm design, but what outcomes do collectively-built algorithms produce? In this section, we evaluate the algorithm's performance on various metrics.

Evaluation Goal

In the literature on policy-related algorithmic systems, the status quo—current human decision-making practice—is deemed to be an appropriate baseline for comparison to measure the algorithmic tool's efficacy; thus, we compare our algorithm with current human decision-making at 412 Food Rescue. One major reason that the organization wanted to introduce the algorithmic allocation system was to improve equity in donation allocation made by organizational staff and distribute the donations to a larger set of recipients. Indeed, the skewness of their current distribution of donations (i.e., 20% of the organizations receiving 70% of the donations (Figure 4.6a)) is not the result of conscious strategy, but rather the result of, for example, the memory bias of human decision-makers selecting recipients that they have given donations to recently.

Dataset

The final matching algorithm included 23 individual models (Section 4.2.5) that generated complete rankings of possible recipients for each incoming donation; the rankings were then aggregated using the Borda method with the stakeholder weights provided in Section 6. We ran this collectively-built algorithm on historical allocation data from 412 Food Rescue containing a total of 1,760 donations from 169 donors over the course of five months (March–August 2018).[18] There were 380 eligible recipient organizations in the database, and 277 of those received donations in the timeframe we considered.[19] We compared our algorithm (AA) with two benchmarks: human allocations recorded in historical data (HA), and a random algorithm that selected a recipient uniformly at random (RA). In the simulations for our algorithm and the random algorithm, we applied some of the real-world constraints that influenced human dispatchers' decisions: for any given donation, we filtered out recipients that did not handle the donation type or were not open for at least 2 hours between the incoming donation time and 6 pm.

[18] The original data set had 1,862 donations from 177 donors given to 305 recipient organizations. 412 Food Rescue staff told us that 28 of the recipient organizations were either backup recipient organizations or became inactive at the time of the evaluation, thus we excluded them from the data.

[19] 46 recipients were added during the course of the five months, and for each day, we filtered out organizations based on the date when the recipient organizations were added in algorithm testing.

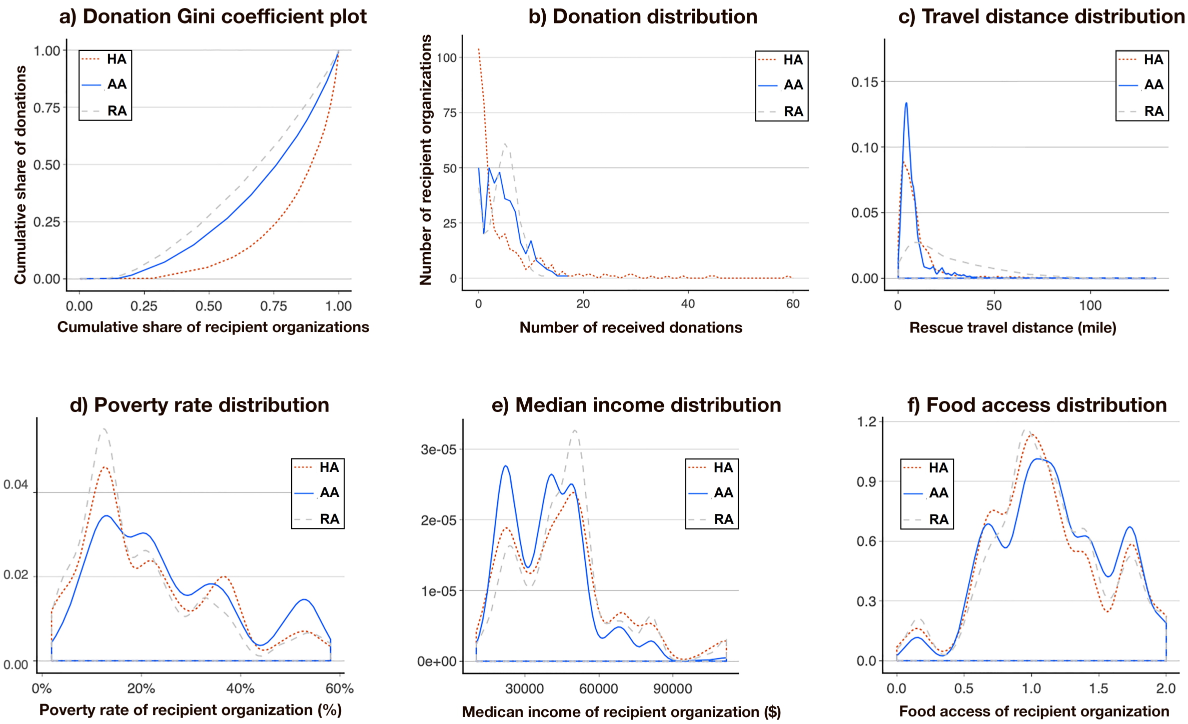


Figure 4.6: The performance of our algorithm (AA) versus the human allocation (HA) and a uniformly random allocation (RA), on various metrics.

Results

The results indicate that our algorithm can make donation allocations more equitable compared to human allocation without hurting efficiency (Figure 4.6).

Number of Donations Allocated to Recipient Organizations: Our algorithm resulted in a more equal donation distribution compared to human allocation, as illustrated in Figure 4.6b. As the human donation distribution is skewed, we conducted a Mann-Whitney U test, a nonparametric test that does not require the data to be normally distributed, to compare the number of donations allocated to recipient organizations.[20] The results show that algorithmic allocation was significantly more equally distributed than human allocation (AA Median = 4 donations (SD = 3.73), Min-Max:0-20; HA Median = 2 donations (SD = 7.26) Min-Max:0-59, U = 57814, $p < .00000001$).

We also conducted a Gini coefficient analysis, a standard economic inequality measure of income [117] or other kinds of resources [12]. A Gini index of zero means perfect equality, with everyone getting the same number of donations, and an index of 100 means maximum inequality, with one organization receiving all donations. Algorithmic allocation resulted in a Gini index of 42, which was lower than the Gini index of 68 in human allocation; this indicates that the algorithmic allocation was more equal. The random allocation algorithm achieved a Gini index of 32, which intuitively is close to the minimum possible, subject

[20] The convention is to report medians as the data is not normally distributed.

to the constraints. Graphically, as seen in Figure 4.6a, the closer the allocation line is to the diagonal line $y = x$, the fairer the allocation. Additionally, the x-axis is ordered from lowest to highest, so, for instance, our results show that the lowest 50% of all recipient organizations received about 5% of all donations under a human dispatcher, but received about 20% of all donations under our algorithm.

Poverty, Income, and Food Access of Recipients: When considering poverty, income, and food access levels, random allocation can be seen as uniformly sampling from the poverty, median income, and food access rates of all recipients because these features are completely recipient-specific. As illustrated in Figure 4.6d, Figure 4.6e, and Figure 4.6f, the human dispatcher's decisions closely followed the underlying population distributions, but our algorithm donated to recipients with higher poverty rates, lower median incomes, and worse food access. A Mann-Whitney U test shows that the algorithmic allocation gave donations to areas with higher poverty rates (Median = 21.6%, SD = 14.44%) significantly more than human allocation (Median = 18.3%, SD = 13.73%, U = 1303400, $p < .00000001$). Indeed, Figure 4.6d shows that the human and the random algorithm gave more donations to areas with 10%-15% poverty rates, whereas our algorithms gave more donations to areas with about 50% poverty rates. Algorithmic allocation also gave more donations to recipients with lower income (Median = $40,275, SD = $16,312) than human allocation did (Median = $42,255, SD = $22,037, U = 1773200, $p < .00000001$), and the same pattern is observed to a lesser degree in the recipients' access to food levels (AA Median: 1.15 (SD=0.42), HA Median: 1.06 (SD=0.44), U=1414400, $p = .0002$; 0=Normal access, 2=Extremely low access).

Distance and Efficiency: One of the concerns of the organization was that distributing the donations more equitably could lead to longer and less efficient donation allocation. Our simulation results suggest that algorithmic allocation did not increase rescue distance, as illustrated in Figure 4.6c. A Mann-Whitney U test shows that the distance of rescues under algorithmic allocation, whose median is 5.5 miles, is significantly shorter than under human allocation, whose median is 6.15 miles (U = 1646900, $p = 0.001$).

4.2.10 Discussion

In this paper, we envision a future in which people are empowered to build algorithmic governance mechanisms for their own communities. Our framework, WeBuildAI, represents one way to realize this goal. We have implemented and evaluated a system of collective algorithmic decision-making, contributing to the emerging research agenda on algorithmic fairness and governance by advancing understanding of the effects of participation.

Summary of the Research Questions and Results

We summarize our results in response to the research questions raised in the introduction.

What socio-technical methods will effectively elicit individual and collective beliefs about policies and translate them into computational algorithms? How should the resulting algorithms be explained so that participants understand their roles and administrators understand their decisions?

- The WeBuildAI framework successfully enabled participants to build models that they felt confident represented their own decision-making patterns. Participants understood graphical representations of individual models (Figure 4.4) and felt that collective aggregation via the Borda rule was fair. The decision support helped organizational administrators and other stakeholders understand how the final recommendations were made.
- Our findings suggest the elicitation method design could influence elicited beliefs. The top-down explicit-rule method may have promoted idealistic beliefs, while the bottom-up pairwise comparison-based machine learning method may have promoted realistic beliefs that accounted for emotions and constraints associated with tasks.

How does participation influence participants' perceptions of and interactions with algorithmic governance?

- Participation not only resulted in new technology design but also affected participating individuals and organizations [227; 111]. Our participants reported greater trust in and perceived fairness of the matching algorithm, the governing institution, and administrative decisions after participating. Some participants were more motivated to use the services, felt respected and empowered by the governing institution, and reported a greater empathy for difficulties in the organization's decision-making process.
- Our participatory algorithm design, particularly the individual model building method, increased participants' algorithmic awareness and literacy. Through the process of translating their judgments into algorithms, they gained a new understanding and appreciation of algorithms. The method also revealed inconsistencies in employee decision-making in the governing organization, and made employees revisit their assumptions of other stakeholders.

How does the resulting collectively-built algorithm perform?

- The comparisons of the collectively-built matching algorithm and human allocation, using five months of historic data, suggest that the matching algorithm makes donation allocations more even, and gives more donations to recipient organizations in areas with higher poverty, lower income, and low access to food, without increasing the transportation distance.

Contributions to Research on Human-Centered Algorithmic Systems

Fairness and Moral Behavior in AI: In response to recent scholarly and journalistic work that has pointed out the need for "fair" algorithms, much research has been done to devise computational techniques that guarantee fairness in algorithmic outcomes. Our work offers a method for building procedurally-fair governing algorithms [159]. Our findings also offer empirical evidence of the effects of procedural fairness from the perspectives of both those who are affected by algorithms and those who use algorithms; the framework not only increased perceived fairness and trust of the algorithm but also influenced the organization by making the disparate effects of the algorithm more salient in their daily operation.

Our work also suggests that ongoing research seeking to understand people's moral concepts for algorithms and AI needs to be more cognizant of the design of the stimuli. (Some studies use more illustrative, vivid descriptions, whereas others use abstract textual descriptions.) Previous work in experimental moral psychology suggests that the vividness and realism of stimuli influences participants' answers. Consistent with this literature, our work suggests that the top-down versus bottom-up approach of building an algorithm may elicit different levels of construal, resulting in qualitatively different algorithmic models. It is important to choose an elicitation method and level of abstraction appropriate for the task context, and to take a reflective approach so that people can be aware of those situational effects and build a model in accordance with their beliefs.

Community Engagement in Algorithm Design: Our work contributes to recent research that calls for community engagement in AI design by offering a method to leverage varying stakeholders' participation directly in the design of the algorithm. By working with real-world stakeholders with various educational and economic backgrounds to build an algorithm that operates a service, we demonstrate the feasibility and potential of community involvement in algorithm design. At the outset of our research, we were unsure whether participants would feel confident and comfortable enough to express their beliefs on algorithms, and were concerned they might mistrust AI due to negative representations in popular media. It has been a rewarding experience to see participants not only expressing their beliefs, but also gaining trust in and becoming empowered through algorithmic systems. AI systems should be designed to facilitate these changes.

Levels of Participation in Algorithmic Governance

In this section, we define levels of participation in algorithmic governance. We discuss the upsides and downsides of different forms of governance and when collective participation is appropriate, reflecting on our research.

Closed, Non-Participatory Governance: Institutions can design a governing algorithm without involving stakeholders by drawing from their existing data and assumptions. This form of governance is cost-effective compared to participatory governance, which requires effort and resources in soliciting and synthesizing participation. Closed governance is appropriate when there are legitimate metrics for algorithm design. For example, it might be appropriate if the goal is solely to minimize the volunteers' travel time. In our research, the organization was open to stakeholder participation because the staff were unclear on how to balance efficiency and equity in their daily operations. Additionally, closed governance may not inherently earn stakeholders' trust; it works best when the governing institution has already established trust with those being governed. Otherwise the algorithmic decisions may be challenged, mistrusted, or not adopted.

Mediated, Indirect Participatory Governance: Another form of governance is the mediated use of participants' input, resulting in participants' indirect influence on final algorithmic policy. In this form, stakeholders provide input to inform the designers and policymakers, who later design and implement the governing algorithms. The input can be collected through interviews or tools such as individual belief modeling, as in our framework. This form allows the governing organizations to operate on more accurate stakeholder assumptions, and communicating about the stakeholders' involvement can cultivate trust and increase the chances of adoption by those who are governed. This form is most appropriate when the organization seeks to use participatory feedback while retaining full control of the algorithm's design.

Direct Participatory Governance: In fully participatory algorithmic governance, stakeholders' participation is directly implemented in the final algorithm. In this form, participants feel most empowered and responsible, according to both existing literature and our work. However, the governing organization has less control over the final algorithm design. Direct participatory governance is most appropriate in contexts where stakeholders' trust and motivation to participate in the governing organization are critical, when a high level of procedural fairness is required, or in organizations and communities that are already self-governed, such as Reddit.

Extension of the WeBuildAI Framework and Future Work

Our application of the framework to 412 Food Rescue is a case study that implements participatory governance in one context. Our framework can be used and extended to support both mediated and direct participatory governance, and potentially for other algorithmic governance situations that involve normative design decisions and associated tradeoffs. For example, our framework could be used to create governing algorithms that allocate public resources or contribute to smart planning services, placement algorithms in school districts or online education forums, or hiring recommendation algorithms that balance candidate merit with equity issues. Extending our framework to new contexts requires addressing several challenges.

Individual Model Building as a Design Tool: Our findings suggest that the process of building individual models of algorithmic policy has many benefits. Externalized models provide a concrete place for starting a conversation about similarities and differences among the stakeholders or staff members of the organization. Designers and policymakers can use the models to inform algorithm design, or as an auditing or evaluation metric to assess the algorithm's effects from diverse stakeholders' perspectives. However, our research only used about 8-10 features that people could understand. Further research will be needed to apply the individual modeling method to algorithms with hundreds of features or more complex features. New techniques will be needed to explain and combine the features into a set that people can process.

Collectively Aggregated Decisions for Direct Participatory Governance: Our framework can be applied to enable direct participatory governance, particularly in contexts in which trust, motivation, and perceived fairness matter, and, in its current implementation, contexts that do not require instantaneous decisions (within, say, less than a second).

One challenge, though, is to determine who participates and whether participation needs to be regulated. Opening up an algorithm to participation means that some participants may potentially hold opinions that are not socially acceptable. One way to avoid this is to limit participation so that democratic control of algorithms is subject to the constraints of public reason [195; 31]. This ensures that the behavior of algorithms is justified by a universally agreed-upon subset of principles. Future work would need to investigate how to broaden participation while respecting diversity within public reason, and devise an ethical way to determine the boundaries of participation.

Another challenge is ensuring the quality of participation, particularly when participation occurs at scale. Techniques used in crowdsourcing for quality assurance could be adopted to judge the quality of participation based on the amount of time and number of iterations people use in creating their models. Anecdotally, in our study, we observed that the machine learning model's accuracy was low when participants told us that they were applying rules inconsistently. Further work needs to investigate whether model accuracy can be another metric.

When people participate in building systems, those systems become more transparent to them and they gain a deeper understanding of how the systems work. While this is one of the main sources of trust, one potential concern is that people will use this knowledge to game and strategically manipulate the system. To clarify, we do not mean that the potential manipulation of the systems by the disempowered is a risk. We aim to create benefits for all those in need, and we believe the system could be at risk if some individual parties skew the results to maximize their own benefits when all participating individuals have a similar level of need. Indeed, one of the main topics of research in computational social choice [48] is the design of voting rules that discourage strategic behavior—situations where voters report false preferences in order to sway the election towards an outcome that is more favorable according to their true preferences. However, this is not likely to be an issue for our framework because each individual does not have direct control over the final algorithm behavior. One may try to manipulate one's pairwise comparisons or specify preferences to obtain a model that might lead to preferred outcomes in very specific situations, but the same model would play a role in multiple, unpredictable decisions. The relation between their models and future outcomes is so indirect that it is virtually impossible for individuals to benefit by behaving strategically. That said, future work would need to evaluate this question in the real world.

Promoting Representative Participation: One of our goals in designing this participatory framework is to empower stakeholders who typically do not have a say in the algorithms that govern their services, communities, or organizations. By empowering, we mean providing a method or tool that allows people to influence and control a system that they themselves use or an institution to which they belong [75; 97]. This shared power between users and developers, or individuals and governing parties, could increase the self-efficacy [24] and motivation [75] of stakeholders. Empowerment is one of the traditional values of HCI research and practice [198].

However, recently scholars have also pointed out that "material empowerment," or the technical tool itself [198] is not enough to enable people to make positive effects on social problems; one needs to devise solutions that also account for legal, social, and economic constraints [198; 175]. Our framework provides a tool that can enable stakeholders to participate in algorithm design, but it in and of itself will not necessarily result in equal empowerment of all stakeholders. Including representation from communities that are underserved or disadvantaged is a critically important challenge to address in future work. While many in these communities may technically have the opportunity to participate, they may face barriers like time or resource constraints that limit their access to participation. For our context with 412 Food Rescue, we acknowledge for example that volunteers must have access to at least two relatively scarce commodities: access to a private vehicle and free time. Furthermore, recipient organizations often do not have reliable contact information for their clients, who may not have regular access to email or cell phone service. This poses a practical barrier to participant recruitment. In addition to technological design interventions like those we put forward in this paper, social and economic infrastructure will be necessary to ensure equal participation of all stakeholders.

Limitations

Like any study, our work has limitations that readers should consider. Our study evaluated people's experiences with participation, as well as their attitudes toward and perceptions of the resulting algorithmic systems. As our next step, we will deploy the system in the field in order to understand long-term effects and behavioral responses. In the deployment, we will also consider additional evaluation measures for the algorithm, such as stakeholder satisfaction. Additionally, in developing our framework, we intentionally used a focused group of participants to get in-depth insights and feedback on our tools and framework. As we implement our next version, we will examine participation with a larger group of people, including recipient organizations' clients, by developing an educational component and targeted recruiting methods. We will also explore the possibility of running an open system, where people can join at any time or update their models by providing more data. We also acknowledge that despite our best efforts to base our design choices on participants' input gained through interviews (for example, who the stakeholders are, what factors to use), our views might have influenced our analysis of participants' inputs. Our plan to have an online system where participants can further comment on the selected features, stakeholders, and evaluation measures may mitigate this in the future. Finally, our framework needs to be tested with other contexts and tasks that involve different cultures and group dynamics. We are particularly interested in the effects of participation when collective opinions are polarized. On the one hand, it might be the case that a participatory, voting-based approach would be the only way to find a consensus solution. On the other hand, additional techniques—such as public deliberation through an open forum—might be needed to bring together polarized parties to ensure the efficacy of the resulting algorithms. Future work would need to investigate this question further.

4.3 Conclusions

4.3.1 Virtual Democracy, in Theory

Our theoretical and empirical results identify Borda count as an especially attractive voting rule for virtual democracy, from a statistical viewpoint. However, Borda count is also compelling in terms of usability and explainability.

In more detail, in our implemented donor-recipient matching system, clicking on a recommended alternative displays an explanation for why it was ranked highly by Borda count, which consists of two components. First, we show the alternative's average position in the predicted preferences of each of the four stakeholder groups. Note that this information determines the Borda score of the alternative, given the weight of each stakeholder group.[21] Second — this is the more novel component — we show specific features in which the recommended alternative stands out. This is interesting because classic social choice theory does not have features for alternatives, and we are able to give this type of explanation precisely because our alternatives are represented as vectors of features (which is crucial for the application of learning-to-rank algorithms).

Based on the results presented in this paper, as well as these additional insights, we use Borda count in our implemented virtual-democracy-based system.

4.3.2 Virtual Democracy, in Practice: WeBuildAI

Increasingly, algorithms make decisions influencing multiple stakeholders in government institutions, private organizations, and community services. We envision a future in which people are empowered to build algorithmic governance mechanisms for their own communities. Toward this goal, we proposed the WeBuildAI framework. In this framework, stakeholders build an algorithmic model that represents their beliefs about ideal algorithm operation. For each decision task, each individual's model votes on alternatives, and the votes are aggregated to reach a final decision.

As a case study, we designed a matching algorithm that operates 412 Food Rescue's on-demand transportation service, implementing the framework with their stakeholders: donors, volunteers, recipient organizations, and 412 Food Rescue's staff. We then evaluated the resulting algorithm with historical donation data, which showed that our algorithm leads to a more even donation distribution that prioritizes organizations with lower income, higher poverty rate, and lower food access clients compared to human allocation decisions. Our findings suggest that the framework improved the perceived fairness of the allocation method. It also increased individuals' awareness of algorithmic technology as well as the organization's awareness of the algorithm's impact and employee decision-making inconsistencies.

Our study demonstrates the value and promise of using the WeBuildAI framework as a design tool in order to achieve human-centered algorithmic governance. Future work needs to investigate mechanisms to expand the application of the framework and its boundary conditions, as well as ways to overcome existing socioeconomic and institutional barriers to enabling wider participation.

However, there are some limitations to our case study. First, although the staff appreciated the implementation of our virtual democracy framework with the Borda count, it is unclear whether this is because of the well-designed interface or the use of the Borda count as the aggregation method. While the Borda count is easy to explain in certain ways (e.g., in the Borda count, the sum of scores is all that matters, and we can easily compute the average Borda scores for each group of stakeholders individually), perhaps using another reasonable algorithm would also lead to an explainable and practicable interface.

[21] These weights were decided by the stakeholders themselves.

Additionally, we noticed that some participants viewed features non-monotonically. In particular, multiple participants had single-peaked preferences with respect to recipient size, where they valued mid-size organizations the highest due to the amount of food (one carload) that would be delivered, and their utilities tapered off on both sides. This suggests that linear models are inappropriate. Indeed, because of this feedback, we explored polynomial models as well as decision trees, SVMs, and neural networks, but found that linear models performed the best due to the relative scarcity of data we collected.

One other area for improvement is in the data collection stage. We allowed participants to only choose their more preferred destination in each pairwise comparison, but it could be more instructive to allow them to express a wider range of preferences. For instance, allowing them to specify the strength of their preference (strong, weak, indifferent) could significantly reduce cognitive load and result in more informative and consistent preferences.

Finally, we note that although we included recipient organizations in our stakeholder groups, we did not source comparisons from the people who consume the food donations. It would be interesting and valuable to include them in future work to get a more holistic set of opinions from all stakeholders in the 412 Food Rescue ecosystem.

Role		Studies Involved
412 Food Rescue.*		
F1		Sessions 1-4
F2		Sessions 1-4
F3		Sessions 1-4, w
Recipient organizations. (Clients served monthly, client neighborhood poverty rate)		
R1	Human services program manager (N=150, 13%)	Sessions 1-4
R2	Shelter & food pantry center director (N=50, 20%)	Sessions 1-4
R3	Food pantry employee (N=200, 53%)	Sessions 1-4
R4	Animal shelter staff	Session 1
R5	Food pantry staff (N=500, 5%)	Sessions 1-4
R6	After-school program employee (N=20, 33%)	Session 1, w
R7	Home-delivered meals delivery manager (N=50, 11%)	Sessions 1-4
R8	Food pantry director (N=200, 14%)	Sessions 1-2
Volunteers.		
V1	White male, 60s	Sessions 1-4, w
V2	White female, 30s	Session 1
V3	White female, 70s	Sessions 1-4, w
V4	White female, 70s (V4a), white male, 70s (V4b) †	Sessions 1-4
V5	White female, 60s	Sessions 1-4
V6	White female, 20s	Sessions 1-4
Donor organizations.		
D1	School A dining service manager	Session 1
D2	School B dining service manager	Sessions 1-4
D3	Produce company marketing coordinator	Session 1
D4	Grocery store manager	Sessions 1-4
D5	Manager at dining and catering service contractor	Session 1
D6	School C dining service employee	Session 1, w

Table 4.1: Participants. Sessions indicate the study sessions that they participated in: *w* represents a workshop study. *Info excluded for anonymity. † A couple participated together.

Factor	Explanation
Travel Time	The expected travel time between a donor and a recipient organization. Indicates time that volunteers would need to spend to complete a rescue. (0-60+ minutes)
Recipient Size	The number of clients that a recipient organization serves every month. (0-1000 people; AVG: 350)
Food Access	USDA-defined food access level in the client neighborhood that a recipient organization serves. Indicates clients' access to fresh and healthy food. (Normal (0), Low (1), Extremely low(2)) [225]
Income Level	The median household income of the client neighborhood that a recipient organization serves (0-100K+, Median=$41,283) [224]. Indicates access to social and institutional resources [200].
Poverty Rate	Percentage of people living under the US Federal poverty threshold in the client neighborhood that a recipient organization serves. (0-60 %; AVG=23% [224])
Last Donation	The number of weeks since the organization last received a donation from 412 Food Rescue. (1 week–12 weeks, never)
Total Donations	The number of donations that an organization has received from 412 Food Rescue in the last three months. (0-12 donations) A unit of donation is a carload of food (60 meals).
Donation Type	Donation types were common or uncommon. Common donations are bread or produce and account for 70% of donations. Uncommon donations include meat, dairy, prepared foods, etc.

Table 4.2: Factors of matching algorithm decisions. The ranges of the factors are based on their real-world distributions.

	D2	D4	F2	F3	R1	R2	R3	R5	R7	V1	V3	V4	V5	V6
ML	**0.86**	0.78	**0.92**	**0.92**	**0.90**	0.90	**0.78**	**0.94**	0.74	**0.90**	**0.92**	**0.78**	0.56	0.68
ER	0.68	**0.68**	0.68	0.86	0.80	**0.76**	0.70	0.92	**0.74**	0.76	0.82	0.82	**0.80**	**0.88**

Table 4.3: Accuracy of the Machine Learning (ML) model and the Explicit-Rule (ER) model. Bold denotes the model the participant chose as the one that better represented their belief after seeing both models' explanations (Figure 4.4) and their predictions on the 50 evaluation pairwise comparisons. F1 chose the machine learning model but did not complete additional survey questions to calculate model agreement, so the result is not included in this table.

Human beings, who are almost unique in having the ability to learn from the experience of others, are also remarkable for their apparent disinclination to do so.

Douglas Adams.

5 Impartial Ranking

In this chapter, we study rank aggregation algorithms that take as input the opinions of players over their peers, represented as rankings, and output a social ordering of the players (which reflects, e.g., relative contribution to a project or fit for a job). To prevent strategic behavior, these algorithms must be *impartial*, i.e., players should not be able to influence their own position in the output ranking. We design several randomized algorithms that are impartial and closely emulate given (non-impartial) rank aggregation rules in a rigorous sense. Experimental results further support the efficacy and practicability of our algorithms.

Based on our theoretical findings, we also develop *HirePeer*, a novel alternative approach to hiring at scale. HirePeer leverages peer assessment to elicit honest assessments of fellow workers' job application materials, which it then aggregates using an impartial ranking algorithm. We perform three studies that investigate both the costs and the benefits to workers and employers of impartial peer-assessed hiring. We find, to solicit honest assessments, algorithms must be communicated in terms of their impartial effects. Second, in practice, peer assessment is highly accurate, and impartial rank aggregation algorithms incur a small accuracy cost for their impartiality guarantee. Third, workers report finding peer-assessed hiring useful for receiving targeted feedback on their job materials.

5.1 Impartial Ranking, in Theory

We now turn to designing voting rules that satisfy certain axiomatic desiderata. First, we examine a setting in which the set of voters exactly corresponds to the set of alternatives; that is, voters must evaluate themselves in order to return a complete ranking. In this setting, we wish to design voting rules that do not incentivize voters to strategically misreport their preferences in order to alter the results of the aggregation in their favor.

Our work is primarily motivated by online labor markets, such as Upwork or Freelancer. In the bigger markets, employers typically receive dozens of applications for a job, but employers do not have the knowledge required to accurately evaluate applicants. For the past year we have been building a prototype of a new online labor market, where applicants for a job—who are well-suited to evaluate applications for that same job—rank each other. We would like to implement a mechanism that aggregates these rankings into a single ranking that is then shown to the employer.

However, the foregoing application has a clear problem, which gets in the way of applying standard rank aggregation rules: strategic behavior. Specifically, in these relatively high-stakes scenarios, it is likely that a player would try to improve his own position in the output ranking by manipulating his reported ranking. For example, he might weaken a strong contender for the top position by ranking him last. Our goal, therefore, is to design rank aggregation rules that are *impartial*, in the sense that the position of a player in the output ranking is completely independent of the report of that player. However, it is easy to prove that there are no deterministic rank aggregation rules that are both impartial (according to our definition) and Pareto efficient [177].[1] Therefore, we restrict our design space to randomized algorithms.

5.1.1 Our Approach and Results

On a high level, our approach is to design *randomized* rank aggregation rules that are impartial and closely emulate standard rank aggregation rules that are not impartial. Specifically, we focus on providing impartial approximations to the important class of *pairwise* rules, which, as input, only require information about the fraction of players ranking any one player above another. Our theoretical results crucially depend on the notion of approximation—or measure of error—in question.

In Section 5.1.5, we introduce the k-PARTITE algorithm, which, in a nutshell, randomly partitions the players into subsets, builds a probability distribution over the positions of members of one subset based on the aggregate opinion of members of other subsets, and then generates a distribution over rankings that is consistent with these distributions over positions. We prove that k-PARTITE is impartial, and, when used in conjunction with any pairwise rule, it provides small *backward error* with respect to that rule: With high probability, k-PARTITE places each player in the same position that the given pairwise rule would have placed him had the input rankings been slightly perturbed.

In Section 5.1.6, we present the COMMITTEE algorithm. It randomly chooses a subset of players, who serve as the eponymous committee. Each committee member is positioned based on the opinions of other committee members, and then all other players are ordered by the committee. The key idea is that, to avoid conflicts and achieve impartiality, each committee member has slots that are reserved for him, and he is inserted into the reserved

[1] The latter property means that if everyone ranks one player above another, so does the output ranking.

slot that most closely matches the aggregate opinion of other committee members. We prove that Committee provides *mixed error* guarantees with respect to any given pairwise rule, that is, with high probability, Committee places each player in a position that is *close* to where the given pairwise rule would have placed him had the input rankings been slightly perturbed. Taking on some *forward error*—a mismatch between the positions—allows for improved backward error compared to k-partite.

In Section 5.1.7, we empirically measure the performance of our impartial algorithms with respect to the popular *Kemeny rule*, which is defined via a natural optimization objective. The experimental results demonstrate that our impartial algorithms, when coupled with the Kemeny rule, output near-optimal rankings with respect to the Kemeny objective, despite the impartiality constraint.

5.1.2 Related Work

At this point there is a significant body of work on the design of impartial mechanisms [87; 6; 128; 44; 214; 29; 104; 164; 35], including several papers in major AI conferences [154; 17]. We only elaborate on the papers that are most closely related to ours.

The paper of de Clippel et al. [87] introduced the notion of impartiality, in the context of dividing *credit* for a joint project. Specifically, the output of their mechanism is the fraction of the total credit each player receives, and impartiality means that a player cannot affect his own share of the credit. This mechanism is deployed on the fair division website `Spliddit.org`, where one of the suggested applications is ordering authors on scientific papers. However, an impartial credit division mechanism does *not* induce an impartial ranking mechanism, because, when players are sorted by credit, a player can improve his own position by decreasing another player's share.

Berga and Gjorgjiev [29] study the impartial rank aggregation problem from an axiomatic viewpoint, but focus on deterministic rules and a stronger notion of impartiality. Their results suggest that deterministic impartial rank aggregation methods are severely limited, and support our focus on randomized algorithms.

On a technical level, our k-partite algorithm is reminiscent of an algorithm of Alon et al. [6], in that it randomly partitions the players into subsets, and the outcome of players in one subset is only determined by players in other subsets. But the details of the algorithm, and its analysis, are completely different.

5.1.3 Preliminaries

In this section we introduce terminology and notations that are standard in computational social choice [47], as well as the formal instantiation of the concept of impartiality in our setting.

Rankings and Aggregation

For any $k \in \mathbb{N}$, let $[k] = \{1, \ldots, k\}$. Our setting involves a set of players $[n] = \{1, \ldots, n\}$. The opinions of players are represented as rankings over $[n]$, which we think of as *permu-*

tations. Let Π represent the set of all permutations of $[n]$, and let Π^n represent the set of all *input profiles*. For any $\sigma \in \Pi$, let $\sigma(j)$ be the player at position j in σ and let $\sigma^{-1}(i)$ be the position of player i in the ranking σ (where position 1 is the highest and position n is the lowest).

A *deterministic rank aggregation rule* (also known as a *social welfare function*) is a function $f : \Pi^n \to \Pi$, which takes in an input profile and returns a ranking. A *randomized* rank aggregation rule returns a probability distribution over rankings. We sometimes find it convenient to slightly abuse notation and think of the domain of a rank aggregation rule as $\Pi^n \times 2^{[n]}$—for $\vec{\sigma} = (\sigma_1, \ldots, \sigma_n)$ and $X \subseteq [n]$, $f(\vec{\sigma}, X)$ is the application of the rule to the input profile $(\sigma_i)_{i \in X}$.

Pairwise Rank Aggregation Rules

An input profile $\vec{\sigma} = (\sigma_1, \ldots, \sigma_n)$ induces a *pairwise comparison matrix* $A(\vec{\sigma})$, where

$$A(\vec{\sigma})_{ij} = \frac{|\{k \in [n] : \ \sigma_k^{-1}(i) < \sigma_k^{-1}(j)\}|}{n}.$$

In words, the (i, j) entry is the fraction of players who rank i above j. Let Ω be the set of pairwise comparison matrices. Therefore, we can think of $A : \Pi^n \to \Omega$ as a function that takes in an input profile and returns its associated pairwise comparison matrix. As before, we will also use the notation $A(\vec{\sigma}, X)$, for a subset of players $X \subseteq [n]$, to denote the pairwise comparison matrix associated with the rankings of the players in X.

Some rank aggregation rules only require the information encoded in the pairwise comparison matrix to compute their output. Formally, a deterministic *pairwise* rank aggregation rule is a function $f : \Omega \to \Pi$. We denote the class of all deterministic pairwise rules by $\mathcal{P}$.[2]

We pay special attention to two popular pairwise rules:

- *The Borda Rule:* Given $\vec{\sigma} \in \Pi^n$, the score of each player i is $\sum_{j=1}^{n}(n - \sigma_j^{-1}(i))$ (that is, each player awards $n - k$ points to the player in position k), and the players are ranked by non-increasing score. It may not be immediately apparent that Borda is a pairwise rule—proving this well-known fact is left to the curious reader as an easy exercise.
- *The Kemeny Rule:* The *Kendall tau distance* d_{KT} between two rankings $\sigma, \tau \in \Pi$ is the number of pairs of players on which the two rankings disagree. Given $\vec{\sigma} \in \Pi^n$, the Kemeny rule returns a ranking in $\text{argmin}_{\tau \in \Pi} \sum_{i=1}^{n} d_{KT}(\tau, \sigma_i)$. Computing the output of the Kemeny rule is hard, but can be done in practice using integer programming or heuristic algorithms [77].

Other well-known rules, such as Copeland and Maximin, are also pairwise.

[2]We do not consider randomized pairwise rules in this section. Strictly speaking, we do not require this determinism, but we assume it as it makes the proofs more transparent.

Impartiality

Recall that we are interested in designing rank aggregation rules that are impartial, that is, no player i can affect his probability of being ranked in position j, for all $i, j \in [n]$. Formally:

Definition 5.1. *A (possibly randomized) rank aggregation rule f is* impartial *if for all $i \in [n]$, all input profiles $(\sigma_1, \ldots, \sigma_n) \in \Pi^n$, and all $\tilde{\sigma}_i \in \Pi$, it holds that $\vec{x} = \vec{y}$, where x_j is the probability i is ranked in position j in $f(\sigma_1, \ldots, \sigma_{i-1}, \sigma_i, \sigma_{i+1}, \ldots, \sigma_n)$, and y_j is the probability i is ranked in position j in $f(\sigma_1, \ldots, \sigma_{i-1}, \tilde{\sigma}_i, \sigma_{i+1}, \ldots, \sigma_n)$.*

In an alternative model, we may assume that each player i has a value v_{ij} for being ranked in position j, and then impartiality would mean no player can affect his *expected* value for the outcome, regardless of his value function. Although this definition may seem weaker than Definition 5.1 at first glance, it is easy to verify that the two definitions are, in fact, equivalent.

5.1.4 Measures of Error

Given that our goal is to approximate rank aggregation rules, the measure of error is critical to the statement of the formal problem. To define appropriate notions, we adapt concepts that are standard in scientific computing (e.g., in numerical stability analysis): forward error, backward error, and mixed error. We view these imported definitions as part of our conceptual contribution.

Definition 5.2. *Let f be a rank aggregation rule. A rank aggregation rule g is said to have* (Δ_P, Δ_F) forward error *with respect to f if for every input profile $\vec{\sigma} \in \Pi^n$, the probability that for all $i \in [n]$ it holds that*

$$\frac{|f(\vec{\sigma})^{-1}(i) - g(\vec{\sigma})^{-1}(i)|}{n} < \Delta_F$$

is at least $1 - \Delta_P$.

Intuitively, a low amount of forward error implies that every player i is placed near his correct rank (as determined by f) with high probability. Unfortunately, as the next theorem states, impartial rank aggregation rules cannot approximate the Borda rule. Since Borda is a pairwise rule, the theorem rules out the possibility of approximating all pairwise rules.

Theorem 5.3. *For all $n \geq 2$ and $\varepsilon > 0$, there exists no impartial rank aggregation rule g that gives a $(1/2 - \varepsilon, 1/3)$ forward error with respect to the Borda rule f.*

Proof. For $n = 2$, a direct analysis (which we omit) gives the result. Let us therefore consider only the case $n \geq 3$. Let g be an impartial rank aggregation rule.

Suppose we have the input profile $\vec{\sigma}$ where $i \neq 2$ gives the ranking $(i-1, \ldots, n, 1, \ldots, i-2)$. Note that if player 2 continued this trend and gave the ranking $1, \ldots, n$ then all players would have the same Borda score.

Now let us consider player 2 in more depth, and define the probability vector $\vec{x} \in [0, 1]^n$, where x_i denotes the probability player 2 will be in position i when g determines the ranking. By impartiality we know that $\vec{x}$ does not depend on the ranking of player 2. As $\vec{x}$ is a probability vector, we must have one of the following.

Case 1: The first $\lfloor n/2 \rfloor$ entries of $\vec{x}$ sum to at most $1/2$. In this case, if player 2 has the ranking $(2, 1, 3, 4, 5, \ldots, n)$ in $\vec{\sigma}$, then $f(\vec{\sigma})^{-1}(2) = 1$.

Case 2: The last $\lfloor n/2 \rfloor$ entries of $\vec{x}$ sum to at most $1/2$. In this case, if player 2 has the ranking $(1, 3, 2, 4, 5, \ldots, n)$ in $\vec{\sigma}$, then $f(\vec{\sigma})^{-1}(2) = n$.

In either case, we find that with probability at least $1/2$, g will place 2 in a position at distance at least $\lfloor n/2 \rfloor$ from f's placement. That is, with probability at least $1/2$ we have $|f(\vec{\sigma})^{-1}(2) - g(\vec{\sigma})^{-1}(2)| \geq \lfloor n/2 \rfloor \geq n/3$, giving at best a forward error of $(1/2, 1/3)$. □

With this impossibility in hand, we set our sights on an alternate error measure, which is well defined only with respect to *pairwise* rank aggregation rules. For this definition and throughout this section, we use the Frobenius norm and denote $\|A\|_\infty = \max_{i,j} |A_{i,j}|$.

Definition 5.4. *Let $f \in \mathcal{P}$. A rank aggregation rule g is said to have (Δ_P, Δ_B)* backward error *with respect to f if for every input profile $\vec{\sigma} \in \Pi^n$ the probability that for all $i \in [n]$ there exists a matrix $\tilde{A} \in \Omega$ such that*

1. $\left\|A(\vec{\sigma}) - \tilde{A}\right\|_\infty < \Delta_B$, *and*
2. $f(\tilde{A})^{-1}(i) = g(\vec{\sigma})^{-1}(i)$,

is at least $1 - \Delta_P$.

Intuitively, a low amount of backward error implies that every player i is placed in a rank that had the players altered their opinions slightly, i would be in the correct rank (according to f) with high probability.

Finally, we define a third measure of error, which, in a sense, is a union of the two previous notions.

Definition 5.5. *Let $f \in \mathcal{P}$. A rank aggregation rule g is said to have $(\Delta_P, \Delta_B, \Delta_F)$* mixed error *with respect to f if for every input profile $\vec{\sigma} \in \Pi^n$, the probability that for all $i \in [n]$ there exists a matrix $\tilde{A} \in \Omega$ such that*

1. $\left\|A(\vec{\sigma}) - \tilde{A}\right\|_\infty < \Delta_B$, *and*
2. $\frac{\left|f(\tilde{A})^{-1}(i) - g(\vec{\sigma})^{-1}(i)\right|}{n} < \Delta_F$,

is at least $1 - \Delta_P$.

5.1.5 The k-Partite Algorithm

We now introduce and analyze our first impartial rule, k-PARTITE, which is formally given as Algorithm 2. As it appears somewhat opaque, it is best to understand its ideas when we assume that all the X_i are the same size, i.e., k divides n, $|X_i| = n/k$, and $\gamma_i = k$ for all $i \in [k]$. Slight adjustments are made when this is not the case, which for purposes of intuition can be safely ignored.

First, players are randomly split into k groups of equal size $X_1, \ldots, X_k$, and then each such group separately ranks all n players producing rankings τ_i. The crux of the algorithm is the construction of the matrix Z, which, in turn, is the sum of $Z^{(i)}$ matrices. Intuitively, the $Z^{(i)}$ matrix represents X_i's contribution to Z, and its (a, b) entry indicates the probability that a should be placed in position b overall. Specifically, each player not in X_i is placed in his exact position dictated by τ_i with probability $1/k$, and in all positions that the players in X_i themselves were assigned to in τ_i with probability $1/(n(k-1))$. This information is encoded as the only non-zero entries in $Z^{(i)}$—each column then sums to $1/k$, each row representing a player in X_i is zero, and all other rows sum to $1/(k-1)$. As we show in the full version of the paper, Z is doubly stochastic (its rows and columns sum to 1); hence we can apply the Birkhoff-von Neumann Theorem [34; 228] to sample from this distribution and remain faithful to the probabilities.

input: $f \in \mathcal{P}$ and $\vec{\sigma} \in \Pi^n$

1: Randomly split all n players into k groups $X_1, \ldots, X_k$ where $|X_i| \in \{\lfloor n/k \rfloor, \lceil n/k \rceil\}$
2: **for** $i = 1, \ldots, k$ **do**
3: $\quad \tau_i \leftarrow f(\vec{\sigma}, X_i)$
4: $\quad \gamma_i \leftarrow n/|X_i|$
5: $\quad$ Let $Z^{(i)} \in \mathbb{R}^{n \times n}$ where

$$Z^{(i)}_{a,b} \leftarrow \begin{cases} \frac{1}{\gamma_i} & \text{if a} \notin X_i \text{ and } \tau_i(b) = a \\ \frac{1}{\gamma_i(\gamma_i - 1)|X_i|} & \text{if a} \notin X_i \text{ and } \tau_i(b) \in X_i \\ 0 & \text{otherwise} \end{cases}$$

6: **end for**
7: $Z \leftarrow \sum_{i \in [k]} \frac{\frac{n}{|X_i|} - 1}{k-1} Z^{(i)}$
8: Sample a ranking σ such that a is ranked in position b with probability $Z_{a,b}$
9: **return** σ

Algorithm 2: k-PARTITE

Our goal is to prove the following theorem, which states the guarantees of k-PARTITE.

Theorem 5.6. *k-PARTITE is impartial, and, for every $f \in \mathcal{P}$ and $\vec{\sigma} \in \Pi^n$, if $k = \lfloor (n/\ln n)^{1/3} \rfloor$, it gives at most*

$$(4/k, 4/k) \in \left(O\left(\left(\frac{\ln n}{n} \right)^{1/3} \right), O\left(\left(\frac{\ln n}{n} \right)^{1/3} \right) \right)$$

backward error with respect to f.

Note that, in particular, the error goes to 0 as n grows. Turning to the proof, it is clear that the algorithm is impartial because of the inability of any player i to affect the i^{th} row of the Z matrix. We therefore focus on establishing the error bound. To this end, we first prove several lemmas.

Lemma 5.7. *If t players $X = \{x_1, \ldots, x_t\}$ are sampled without replacement from $[n]$ with input profile $\vec{\sigma} \in \Pi^n$, then*

$$\mathbb{P}\left[\| A(\vec{\sigma}) - A(\vec{\sigma}, X) \|_\infty \geq \varepsilon \right] < n^2 \exp\left(-\frac{t\varepsilon^2}{2} \right).$$

Proof.

$$\begin{aligned}
&\mathbb{P}\left[\|A(\vec{\sigma}) - A(\vec{\sigma}, X)\|_\infty \geq \varepsilon\right] \\
&\qquad \leq \sum_{i<j} \mathbb{P}\left[|A(\vec{\sigma})_{i,j} - A(\vec{\sigma}, X)_{i,j}| \geq \varepsilon\right] \\
&\qquad \leq \sum_{i<j} 2\exp\left(-\frac{t\varepsilon^2}{2}\right) = 2\binom{n}{2}\exp\left(-\frac{t\varepsilon^2}{2}\right) \\
&\qquad < n^2 \exp\left(-\frac{t\varepsilon^2}{2}\right),
\end{aligned}$$

where the first transition follows from the union bound, and the second transition follows from Hoeffding's Inequality. □

Lemma 5.8. *For every $f \in \mathcal{P}$, $\vec{\sigma} \in \Pi^n$, and $\varepsilon > 0$, k-PARTITE gives at most*

$$\left(1 - \left(\frac{k-2}{k-1}\right)\left(1 - n^2 k \exp\left(-\frac{\lfloor n/k \rfloor \varepsilon^2}{2}\right)\right), \varepsilon\right)$$

backward error to f.

Proof. Observe that

$$\begin{aligned}
&\mathbb{P}\left[\exists i \in [k] \text{ s.t. } \|A(\vec{\sigma}) - A(\vec{\sigma}, X_i)\|_\infty \geq \varepsilon\right] \\
&\qquad \leq \sum_{i=1}^{k} \mathbb{P}\left[\|A(\vec{\sigma}) - A(\vec{\sigma}, X_i)\|_\infty \geq \varepsilon\right] \\
&\qquad \leq \sum_{i=1}^{k} n^2 \exp\left(-\frac{|X_i|\,\varepsilon^2}{2}\right) \\
&\qquad \leq \sum_{i=1}^{k} n^2 \exp\left(-\frac{\lfloor n/k \rfloor \varepsilon^2}{2}\right) \\
&\qquad = n^2 k \exp\left(-\frac{\lfloor n/k \rfloor \varepsilon^2}{2}\right),
\end{aligned}$$

where the second inequality follows from Lemma 5.7.

Further observe that, by Lines 5 and 7 of k-PARTITE, for any player a he is placed directly where one of the X_i places him with probability

$$\begin{aligned}\sum_{i\in[k]:\ a\notin X_i} \frac{\frac{n}{|X_i|}-1}{k-1}\frac{1}{\gamma_i} &= \sum_{i\in[k]:\ a\notin X_i} \frac{\frac{n}{|X_i|}-1}{k-1}\frac{1}{\frac{n}{|X_i|}} \\ &= 1-\frac{1}{n(k-1)}\sum_{i\in[k]:\ a\notin X_i}|X_i| \\ &\geq 1-\frac{1}{n(k-1)}\sum_{i=1}^{k}|X_i| \\ &= 1-\frac{1}{n(k-1)}n \\ &= \frac{k-2}{k-1}.\end{aligned}$$

Now, if for all $i \in [n]$, $\|A(\vec{\sigma}) - A(\vec{\sigma}, X_i)\|_\infty < \varepsilon$, and each player a is placed in the position that some X_{i_a} places him, then we can set $\tilde{A} = A(\vec{\sigma}, X_{i_a})$ for all $a \in [n]$ to satisfy the conditions of backward error. Moreover, these events are independent. We conclude that k-PARTITE gives at most

$$\left(1-\left(\frac{k-2}{k-1}\right)\left(1-n^2k\exp\left(-\frac{\lfloor n/k\rfloor\,\varepsilon^2}{2}\right)\right),\varepsilon\right)$$

backward error, as stated. □

Proof of Theorem 5.6. From Lemma 5.8 it suffices to show that if we have $\varepsilon = 4/k$,

$$1-\left(\frac{k-2}{k-1}\right)\left(1-n^2k\exp\left(-\frac{\lfloor n/k\rfloor\,\varepsilon^2}{2}\right)\right) \leq \frac{4}{k}.$$

Observe that

$$\begin{aligned}&n^2k\exp\left(-\frac{\lfloor n/k\rfloor\,\varepsilon^2}{2}\right) \\ &\quad\leq n^2k\exp\left(-\frac{(n/k-1)\varepsilon^2}{2}\right) \\ &\quad= n^2k\exp\left(-\frac{(n/k-1)(4/k)^2}{2}\right) \\ &\quad= n^2k\exp\left(\frac{8}{k^2}\right)\exp\left(-\frac{8n}{k^3}\right) \\ &\quad\leq n^2\left(n^{1/3}\right)\exp\left(\frac{8}{2^2}\right)\exp\left(-\frac{8n}{\frac{n}{\ln n}}\right) \\ &\quad= e^2n^{-17/3} \\ &\quad\leq n^{-2}.\end{aligned}$$

Thus we see that

$$
\begin{aligned}
1 - &\left(\frac{k-2}{k-1}\right)\left(1 - n^2 k \exp\left(-\frac{\lfloor n/k \rfloor \varepsilon^2}{2}\right)\right) \\
&= \frac{1}{k-1} + \left(\frac{k-2}{k-1}\right)\left(n^2 k \exp\left(-\frac{\lfloor n/k \rfloor \varepsilon^2}{2}\right)\right) \\
&\leq \frac{1}{k-1} + n^{-2} \\
&\leq 2/k + 2/k \\
&= 4/k.
\end{aligned}
$$

□

A natural question is why we insist on what appears to be a somewhat convoluted algorithm, instead of a more natural approach such as the impartial NAIVE-BIPARTITE, formally given as Algorithm 3. The reason is that this algorithm does not even guarantee tolerable *mixed* error in general. Indeed, consider $f \in \mathcal{P}$ that is defined as follows. Let $X \subseteq \{2, \ldots, n\}$ be the set of players such that at least one player ranks i above 1; return the ranking starting with the players of X ordered lexicographically, followed by the players of $[n] \setminus (\{1\} \cup X)$ ordered lexicographically, and player 1 inserted into position $\lfloor n/3 \rfloor$ overall (shifting appropriately). Now consider the input profile where i reports the ranking $(i, 1, 2, \ldots, i-1, i+1, \ldots, n)$. Then NAIVE-BIPARTITE will always return a ranking where player 1 is placed first or second—as he will always top his set. This means that the algorithm cannot even provide a mixed error of $(1/2, 1, 1/4)$.

input: $f \in \mathcal{P}$ and $\vec{\sigma} \in \Pi^n$

1: Randomly split the n players into two sets X and Y where $|X| = \left\lceil \frac{n}{2} \right\rceil$ and $|Y| = \left\lfloor \frac{n}{2} \right\rfloor$
2: $\tau_1 \leftarrow f(\vec{\sigma}, X)$ restricted to the players only in Y
3: $\tau_2 \leftarrow f(\vec{\sigma}, Y)$ restricted to the players only in X
4: σ interlaces τ_1 and τ_2, that is,

$$
\sigma(i) \leftarrow \begin{cases} \tau_1\left((i+1)/2\right) & \text{if } i \text{ is odd} \\ \tau_2\left(i/2\right) & \text{if } i \text{ is even} \end{cases}
$$

5: **return** σ

Algorithm 3: NAIVE-BIPARTITE

5.1.6 The Committee Algorithm

k-PARTITE demonstrates that there exist impartial rules that accurately imitate any $f \in \mathcal{P}$. Observe, however, that the algorithm is somewhat hamstrung by the fact that a player must be (with high probability) ranked in *exactly* the location that a small perturbation of the input rankings would give.

To allow more flexibility, we focus on mixed error, and consider COMMITTEE, given as Algorithm 4. Intuitively, this algorithm selects a random committee $X = \{x_1, \ldots, x_k\}$, which then determines the entire ranking. First, for each committee member x_i, we determine their rank using only the rankings given by the remaining $k-1$ members. However, as directly placing each committee member in this fashion may cause collisions (i.e., multiple members may be assigned the same rank) we restrict placement of x_i to only the positions $i, i+k, i+2k, \ldots$ Specifically, we assign x_i to the closest such position to the rank given to x_i by the other committee members. There are then k of the n positions assigned. Second, the committee ranks all of the n players, and the non-committee members are placed in the order ranked by the committee in the remaining $n-k$ slots.

input: $f \in \mathcal{P}$ and $\vec{\sigma} \in \Pi^n$

Randomly select a subset $X = \{x_1, \ldots, x_k\} \subseteq [n]$
for $i = 1, \ldots, k$ **do**
    $c \leftarrow \arg\min_{j \in \{i, i+k, \ldots\}} \left| j - f\left(\vec{\sigma}, X \setminus \{x_i\}\right)^{-1}(x_i) \right|$
    $\sigma(c) \leftarrow x_i$
end for
$\tau \leftarrow f(\vec{\sigma}, X)$
$j \leftarrow 1$
for $i = 1, \ldots, n$ **do**
    if $\tau(i) \notin X$ **then**
        while $\sigma(j)$ is occupied **do**
            $j \leftarrow j + 1$
        end while
        $\sigma(j) \leftarrow \tau(i)$
    end if
end for
return σ

Algorithm 4: COMMITTEE

The algorithm yields the following guarantees

Theorem 5.9. COMMITTEE *is impartial, and, for every* $f \in \mathcal{P}$, $\vec{\sigma} \in \Pi^n$, *and* $\varepsilon > 0$, *if* $k = 1 + \frac{2}{\varepsilon^2} \ln\left(\frac{n^3}{\varepsilon}\right)$, *it gives at most* $(\varepsilon, \varepsilon, (k+1)/n)$ *mixed error with respect to* f.

Importantly, this theorem allows for an incomparable error to Theorem 5.6. That is, we can reduce the backward error so long as we are willing to take on some forward error. For example, setting ε appropriately gives at most $\left(n^{-2/5}, n^{-2/5}, 2/n + (34/5)n^{-1/5} \ln n\right)$ mixed error.

As in the case of k-PARTITE, the impartiality of COMMITTEE is obvious, because the position of each committee member is only determined by other committee members, and the position of non-committee members is determined by committee members. The proof of Theorem 5.9 therefore focuses on establishing the stated mixed error guarantee; it is relegated to the full version of the paper.

5.1.7 Experiments

Our theoretical results indicate that impartial rules like k-PARTITE and COMMITTEE are likely to yield rankings that are close, in a sense (backward error or mixed error), to the output of a given rank aggregation rule. In this section we investigate a more natural metric, which is beyond the reach of our theory, and empirically demonstrate that our rules perform well with respect to this metric, too.

In our experiments, we focus on the Kemeny rule (see Section 5.1.3), as it is defined via an optimization problem, so we can use its objective function as our measure. Specifically, we interpret the Kemeny rule as maximizing the number of *agreements* with the input rankings, that is, given $\vec{\sigma} \in \Pi^n$, it chooses a ranking $\tau \in \Pi$ that maximizes

$$\text{Kem}(\tau, \vec{\sigma}) = \sum_{i=1}^{n} \left(\binom{m}{2} - d_{KT}(\tau, \sigma_i) \right). \tag{5.1}$$

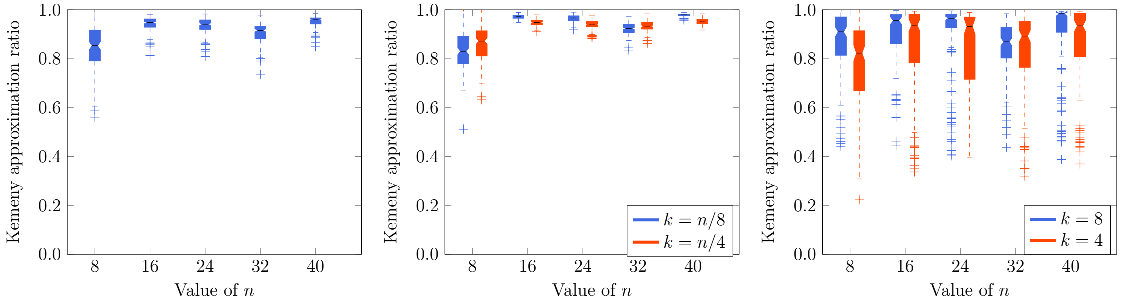
We quantify the error of an impartial rule by comparing how well it does with respect to measure (5.1) with the performance of the optimal ranking returned by the Kemeny rule. In more detail, let f be the Kemeny rule, and let g be an impartial rule; given an input profile $\vec{\sigma}$, we are interested in the *Kemeny approximation ratio* $\text{Kem}(g(\vec{\sigma}), \vec{\sigma})/\text{Kem}(f(\vec{\sigma}), \vec{\sigma})$; this ratio is upper-bounded by 1 due to the definition of the Kemeny rule. We use the number of agreements, instead of the number of disagreements, as our measure because in cases where the number of disagreements is very small, the ratio would be misleadingly large.

The input profiles are generated according to the popular Mallows model [166]. In this model, there is a base ranking of the alternatives τ^*, and rankings are drawn i.i.d. from a probability distribution over Π, defined by

$$\Pr[\sigma \mid \tau^*] = \frac{\phi^{d_{KT}(\sigma, \tau^*)}}{\sum_{\sigma' \in \Pi} \phi^{d_{KT}(\sigma', \tau^*)}},$$

for a *dispersion* parameter $\phi \in [0, 1]$. Note that $\phi = 1$ corresponds to uniformly random rankings (and therefore input rankings disagree on pairs of alternatives with probability $1/2$), whereas $\phi = 0$ means, by convention, that all rankings coincide with the base ranking, that is, there is unanimous agreement. We empirically study the Kemeny approximation ratio of the impartial rules NAIVE-BIPARTITE, COMMITTEE, and k-PARTITE, for multiple values of ϕ, each of which represents a different level of agreement.

Throughout our experiments, we let $k = n/4, n/8$ for COMMITTEE and let $k = 4, 8$ for k-PARTITE. The intuition behind these choices is that the size of the initial committee and each subset in the partition should grow with n, and choosing these values of k works reasonably in practice. We ran experiments with $n \in \{8, 16, 24, 32, 40\}$ players and $\phi \in \{0, 0.1, 0.3, 0.5, 0.7, 0.9, 1\}$.



(a) Naive-bipartite (b) Committee (c) k-partite

Figure 5.1: Kemeny approximation ratio of three impartial mechanisms for $\phi = 0.3$. The median of each boxplot is marked with a black line, the edges of each box denote the quartile values, and the whiskers extend to data within 1.5 times the interquartile range from the edges of each box.

Our results for the three impartial rules—Naive-bipartite, Committee, and k-partite–and $\phi = 0.3$ are shown in Figures 5.1a, 5.1b, and 5.1c, respectively; the results for other values of ϕ, which can be found in the full version of the paper, are qualitatively similar. In each figure, the x axis shows the values of n, and the y axis shows the Kemeny approximation ratio obtained by the impartial rule. All three impartial rules perform well as n increases. As a baseline, a straightforward calculation shows that with $\phi = 0.3$, a ranking drawn from the Mallows model would agree with the base ranking (which is typically the output of Kemeny when the input profile is drawn from Mallows) with probability 0.77 on any given pair of alternatives, and that probability is 0.5 if the latter ranking is replaced with a uniformly random ranking, so the (impartial) rule that chooses a ranking uniformly at random gives a Kemeny approximation ratio of $0.5/0.77 = 0.65$.

On a high level, the three impartial rules achieve excellent Kemeny approximations despite their very different theoretical guarantees. But k-partite has by far the highest variance. This phenomenon is due to the fact that if the position of a player is chosen from a setting in which he was not placed in his exact place as prescribed by players in some partition, he is essentially placed in a random location in the final ranking. In Naive-bipartite and Committee, this is not an issue, as players are always placed in some sense close to a position in which a subset of players believes they belong. As can be seen in the full version of the paper, this phenomenon is more pronounced for lower values of ϕ because placing players in a random location is penalized more heavily when the population is generally sharply clustered around a certain ranking.

Furthermore, the performance of our impartial algorithms depends on the specific choice of k (except for Naive-bipartite, which does not depend on any parameter k). The following observations can be seen in the full version of the paper. Committee performs better with the smaller value of k, but this effect lessens as ϕ increases (i.e., it helps more when players have generally similar opinions). With small ϕ, because most committees will agree on a consistent ranking, the additional error from inserting players into larger buckets leads to a noticeable difference. However, as players start to disagree more, the benefit of getting better estimates from larger committees counteracts the insertion error. k-PARTITE acts similarly: the larger value of k leads to better performance for low ϕ, but this effect again lessens as ϕ increases. This is because each player is placed exactly where one of the groups places him with probability $(k-2)/(k-1)$, which increases with k, but being placed in one of these positions is most beneficial when players generally agree. As opinions become increasingly random and diffuse, groups disagree more strongly about where to place a specific player.

5.2 Impartial Ranking, in Practice: HirePeer

Expert crowdsourcing (e.g., Upwork.com) provides promising benefits such as productivity improvements for employers, and flexible working arrangements for workers. However, to realize these benefits, a key persistent challenge is effective hiring at scale. Current approaches, such as reputation systems and standardized competency tests, develop weaknesses such as score inflation over time, thus degrading market quality. In conjunction with the theoretical work described above, we develop *HirePeer*, a novel alternative approach to hiring at scale that leverages peer assessment to elicit honest assessments of fellow workers' job application materials, which it then aggregates using an impartial ranking algorithm [148]. We perform three studies that investigate both the costs and the benefits to workers and employers of impartial peer-assessed hiring. We find, to solicit honest assessments, algorithms must be communicated in terms of their impartial effects. Second, in practice, peer assessment is highly accurate, and impartial rank aggregation algorithms incur a small accuracy cost for their impartiality guarantee. Third, workers report finding peer-assessed hiring useful for receiving targeted feedback on their job materials.

5.2.1 Introduction

Expert crowdsourcing is on the rise. From 2009 and 2013, one of the largest platforms for expert crowdsourcing, Upwork.com (previously oDesk), witnessed an 800% increase in the number of paying employers [3]. Yet as more employers and workers move to expert crowdsourcing, a critical challenge remains: employers struggle to hire effectively and efficiently at scale. On Upwork, for instance, it takes employers approximately three days to screen, interview, and hire every candidate [223]. Relative to the duration of an expert crowdsourcing task, this cost in time and effort is enormous, encouraging employers to adopt a satisficing strategy (i.e., hiring workers who are "good enough" instead of finding the most qualified candidate overall) [216]. This cost also damages workers' prospects: when employers cannot confidently identify qualified applicants, they offer lower wages to offset their risk of low-quality results; such depressed wages consequently turn away qualified workers, or workers may respond to lower payment with lower quality work [207]. Indeed, in other markets, such large costs for hiring have been shown to dissuade employers from hiring workers entirely [212]. This may cause online crowdsourcing markets to degrade over time. This section investigates a new scalable method to hire expert workers quickly and accurately.

Perhaps the most widely adopted method today to address the large costs of screening applicants is reputation systems. These systems aggregate a candidate's prior task performance, as assessed by past employers, into a score. Although reputation systems are widely adopted by platforms, they bring with them their own set of challenges to effective hiring at scale, which worsen over time. For instance, online reputations become inflated over time: the (social) cost of giving negative feedback is higher than positive feedback [129]. As a result, norms shift over time, and reputation inflation worsens, reducing reliability.

While ongoing work continues to improve existing approaches to address some of these limitations, this section instead presents an entirely new approach to hiring at scale. Our approach is based on a widely used technique to address the need for accurate assessments of open-ended material at massive scale: peer assessment. To date, peer assessment remains the gold standard of review, as seen in its use to assess quality in top-tier academic conferences [229], grant reviewing [73], and more recently massive online classrooms [151]. We investigate: can crowd experts peer-assess each others' job materials to identify qualified candidates? Specifically, we investigate if peer assessment can generate a ranked list of all job applicants from which the employer can make final hiring decisions.

As might be apparent, the conflicts of interest that arise in a hiring setting are the central challenge in realizing this approach. Specifically, because all crowd experts applying to a task presumably would like to take the job, they have an incentive to rate other applications *strategically*, to make themselves look more attractive to the employer. This section describes a system, HirePeer, that overcomes these conflicts. Overcoming conflicts requires both algorithms that can aggregate judgments such that participants derive no benefit from strategic assessments (*impartial* algorithms), and a careful consideration of human-centered components of this process.

First, we investigate whether automatic impartial aggregation of worker assessments of open-ended work is necessary in real-world hiring settings with conflicts of interest. Our first study creates an environment within Amazon Mechanical Turk with conflicts of interest through carefully designed incentives. It then demonstrates the need for impartial algorithms, and the necessity of communicating the presence of such impartial algorithm to participants. We find an effective introduction does not need rely on explaining a complicated randomized algorithm, but rather on the psychology of choice. In a between-subjects randomized experiment ($n = 170$), we find a *consequence explanation* results in the least amount of strategic behavior [172]. On the other hand, we find communication based on a "policing" framing to be ineffective.

Second, we investigate HirePeer's real-world implications for employers. Importantly, we find peer assessment is feasible for hiring in expert crowdsourcing, with accuracies of more than 90% compared to non-conflicted expert judgments (such as those made by employers). We then examine the cost of impartial peer assessment by analyzing the accuracy of three impartial aggregation algorithms [136] and find that, in practice, impartiality comes at a small price. In a between-subjects randomized experiment ($n = 150$), we find impartial peer assessment, in a setting that utilizes the consequence explanation introduced in this section, results in a 8% decrease in accuracy compared to peer assessment where impartiality is not guaranteed.

Finally, we explore worker-oriented implications of peer-assessed hiring. Specifically we look at, if, and how, expert crowdworkers might benefit from peer assessment and feedback. We conduct a case study to deploy HirePeer in a real-world expert crowd hiring setting, where crowd experts complete an open-ended, complex task. This case study suggests peer-assessed hiring benefits crowd experts by a) exposing them to how other applicants assembled resumés and applications, b) introducing them to new skills to develop in the future, and c) giving them targeted feedback on their job materials.

In short, this section has three contributions. First, it introduces peer assessment as a new, scalable, and accurate approach to hiring in expert crowdsourcing marketplaces, instantiated in a system *HirePeer*. Second, through a real-world deployment of three impartial mechanisms, it quantifies the tradeoff between guaranteeing impartiality and accurate ranking. Third, it presents a brief exploration of how workers may benefit from peer-assessed hiring.

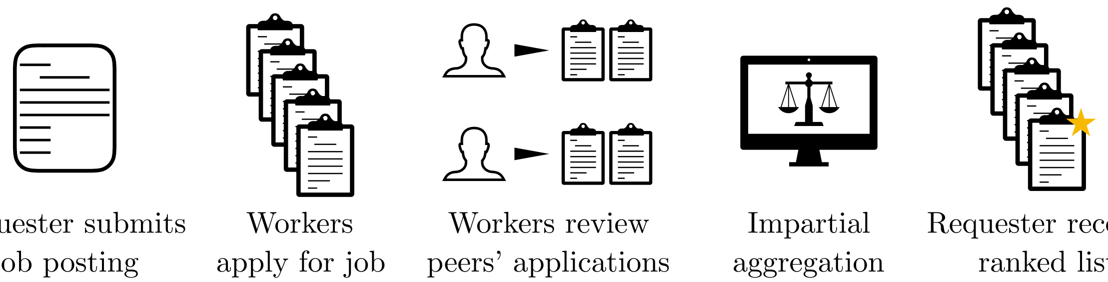


Figure 5.2: HirePeer's workflow of impartial peer-assessed hiring for expert crowdsourcing

5.2.2 Related Work

This section primarily draws on three bodies of literature: a) existing interventions for large-scale hiring on online marketplaces, b) online peer assessment in education, and c) impartial mechanism design.

Platform-specific reputation systems are perhaps the most widely-adopted approach to facilitate hiring in expert crowdsourcing. Although reputation systems are intended to signal worker trustworthiness and facilitate transactions between strangers, they suffer from reputation inflation [129]—eventually, employers almost always award high feedback scores to employees.

Peer review remains the gold standard for assessing open-ended materials, as evinced by its wide adoption in academia to judge paper submissions [229] and by the NSF to review grants [73]. More recently, *online* peer assessment has been introduced in educational settings; in both massive online open courses (MOOCs) and in large physical classrooms, peer assessment has proved to be an effective way to scale accurate assessments of open-ended complex work [70; 226]. However, applications of scalable online peer assessment outside of the classroom remain limited.

Realizing peer-assessed hiring requires careful consideration for how to effectively handle conflicts of interest at scale (all workers who apply to a task would like to be chosen for the task). Recently, Kahng et al. presented three impartial[3] algorithms (called NAIVE-BIPARTITE, COMMITTEE, and k-PARTITE) which aggregate pairwise comparisons to generate a ranked list [136]. While all three impartial mechanisms have strong theoretical guarantees, we explore their performance in a real-world setting.

5.2.3 HirePeer: System description

A requester using HirePeer posts her task to the labor platform (e.g., Upwork) as usual. However, instead of applying to the job directly, workers who are interested in the task are notified to apply to the task on the HirePeer website (see Figure 5.2). When applicants have completed their job application, they are then asked to review a machine-selected set of other applications. To reduce inadvertent biases in evaluation, reviewing is double-blind [151]. Before workers start reviewing, they are notified their assessments will be aggregated with an impartial mechanism.

Because prior work shows pairwise comparisons encourage attention to non-superficial features and lead to more accurate assessment [63], workers conduct pairwise comparisons of peers' anonymized job materials. An expert-generated rubric for the specific task type guides evaluation—our current system has rubrics for web design, data visualization, etc. The rubric contains a) domain-specific criteria, b) more general criteria that are important in an expert crowdsourcing context like communication and timeliness of task completion, and c) qualitative textual feedback on job materials. An expert rubric allows us to collect accurate assessments from both novice and expert workers [54]. Feedback on application materials is later shown to both the task requester and to the applicant.

Once peer assessments have been collected, they are aggregated by the impartial mechanism. Importantly, our mechanisms aggregate assessments into a ranked list (rather than merely choosing a subset of qualified candidates). Armed with this ranked list and the qualitative feedback on each application, the requester can hire the best suited applicant on the crowdsourcing platform.

5.2.4 Study 1: Is an impartial algorithm necessary?

While there have been many theoretical papers on the design of impartial algorithms [87; 136], little work has been done on effectively communicating the presence of impartial algorithms to users. Such an introduction is not only important given increased calls for algorithmic transparency across the community, but also because participants may behave strategically (i.e., attempt to boost their own position) if they do not realize their assessments are aggregated impartially.

If participants behave non-strategically in general, then it may be unnecessary to communicate the impartial mechanism at all (in fact, the mechanism itself may be unnecessary except to thwart the occasional strategic behavior). But if participants engage in strategic behavior, it is important to investigate:

[3] A ranking mechanism is *impartial* if no participant can affect her position in the final ranking [136].

Research Question 1: For accurate assessments, should the presence of an impartial algorithm be communicated to participants?

If strategic behavior is commonplace, then communicating an impartial mechanism may discourage it if participants believe that strategic behavior has no benefit to them. It is likely that different ways of communicating impartial mechanisms may differ in their effectiveness at discouraging strategic behavior; so our study also investigates:

Research Question 2: Which framing of impartial algorithms best discourages strategic participant behavior?

Changing behavior without technical explanations

If impartial mechanisms are to be deployed widely to non-experts, it would be desirable for explanations to not rely on mathematical understanding. We consider two ways of doing so: a) by describing consequences, and b) by leveraging psychological theories of choice to nudge behavior. In particular, we leverage the effects of different "framings," or methods to describe the same situation, that emphasize different attributes. Different framings of game-theoretic tasks result in drastically different outcomes: Tversky and Kahneman found basic tenets of rational behavior can be violated with simple word changes in task instructions [222]. These results were later corroborated in diverse, real-world applications on Amazon Mechanical Turk [184]. Thus, we investigate whether using a framing approach is even more beneficial than describing potential consequences, as it not only it does not require participants to have knowledge of algorithms or mathematics, but also it relies on fundamental and systematic human biases.

Three Ways to Communicate Impartiality

We consider three different ways to communicate impartiality. First, we consider a *consequential* explanation. To discourage strategic behavior, we describe the consequences of using an impartial algorithm: "The ranking you generate will not affect the final aggregated ranking of your item as we use an impartial algorithm." Note that prior work suggests that such an approach may not completely prevent strategic behavior, but may reduce it. For example, Mazar et al. suggest that when consequences of "dishonest" (i.e., strategic) actions are well-known, such as while claiming exaggerated income tax exemptions, people only behave dishonestly to a small extent, as doing so allows them to preserve their positive self-image [172].

We also consider two framing-based approaches. First, we consider a *policing* approach, which is the most common technique in the related literature [57]. Participants in this condition were told, "To prevent you from cheating, we implemented an impartial algorithm." Second, we consider an *responsibility externalization* framing, based on Greenwald's theory of the totalitarian ego, specifically *beneffectance* [122]. This theory suggests while people perceive themselves to be responsible for desired outcomes (such as performing a kind act), but responsibility for undesired outcomes is externalized to others (e.g., traffic leading to aggressive driving). As such, this theory suggests participants see themselves to be honest, but may be concerned that others may behave strategically. Participants in this setting were told, “For your protection, we prevent other workers from cheating using an impartial algorithm.”

Participants and experimental setup

We conducted a between-subjects randomized experiment on Amazon Mechanical Turk (AMT) to test which of three communications of an impartial mechanism minimized strategic behavior compared to our control condition ($n = 170$). We used AMT as an experimental setting for two reasons: first, it can be challenging to discern strategic behavior from low quality work on AMT [131], providing a rich experimental setting to evaluate decision making; second, AMT is a representative sample of a typical online labor market, and has been shown to be a reliable environment for behavioral studies [170].

Participants were randomly assigned to one of four between-subjects conditions. The control condition made no mention of an impartial mechanism, and instead simply reminded participants to read instructions carefully (this has been shown in previous crowd work to have no effect). The other three conditions described the algorithm as above (with consequences, policing, or responsibility externalization). We displayed each in a reminder (in bold) at the bottom of the task instructions on AMT, depending on which condition a participant was randomly assigned. We also included this reminder a second time, immediately before the task.

Task structure and strategic behavior

The experiment used a simple task with known ground truth, to simplify evaluation, while still leaving room for well-defined strategic behavior.

Task We collected eight product reviews from Amazon for the bestselling mobile phone when this study was conducted: the Samsung Galaxy. The reviews were collected to have large differences in quality (the numbers of upvotes for the reviews differed by orders of magnitude). We then introduced typos into each review. Unbeknownst to the participants, all participants edited the same review across all conditions, which was at position #6 in ground-truth (where product review #8 was lowest in quality).

Participants were first asked to proof-read these reviews, and fix typos. Each participant then ranked eight product reviews from the Amazon product page (i.e., without introduced typos), and their edited review, in terms of quality. The product reviews, including their own, were presented to participants in order of true quality, measured by the number of upvotes on Amazon.

The task took at most 15 minutes, and participants were paid $10 USD per hour (before bonuses, described below).

Incentives for strategic behavior Participants were notified the rankings provided by all study participants would be aggregated (similar to peer-assessed hiring), and they would receive a bonus if their review landed in one of the top five positions in the aggregated ranking (a similar incentive structure to peer-assessed hiring). Specifically, the bonus structure was $5 USD if their review landed in position 1, $4 USD for position 2, and so on, and bonuses were awarded as promised. Because most workers in AMT’s labor pool participate to earn money, this task's incentive structure aligns with participant motivations, and is therefore an ecologically valid way to create a similar incentive structure to peer-assessed hiring [131]. Each participant edited the same review, compared it to the the same ground truth ranking of reviews, and had the same incentive to manipulate their report.

This incentive structure also allows for only one kind of strategic behavior: exaggerating the ranking for the edited review, by placing it above position #6. It also allows for a measure of strategic behavior: how much higher than position #6 they placed their review (as reviews differed in quality by orders of magnitude).

Comparison to peer-assessed hiring This task design has critical similarities to hiring. First, ranking edited reviews is similar to ranking job materials, e.g., resumes; and the ranking is similarly subjective, allowing for strategic interpretation. Similarly, there is a strong incentive to rank oneself higher.

The task differs from peer-assessed hiring in that participants are only comparing one artifact, instead of the multiple used in hiring, such as resumes, work experience, etc. Such a comparison would be even more subjective, but allows for similar strategic behavior. Second, our task has bonuses for even small strategic behaviors. The hiring scenario would be more analogous to having a very large bonus for position #1 (i.e., being hired), and vanishing bonuses for other positions. Our task design is necessary because we seek to measure the degree of strategic behavior.

Result: Need for introduction of impartial algorithm

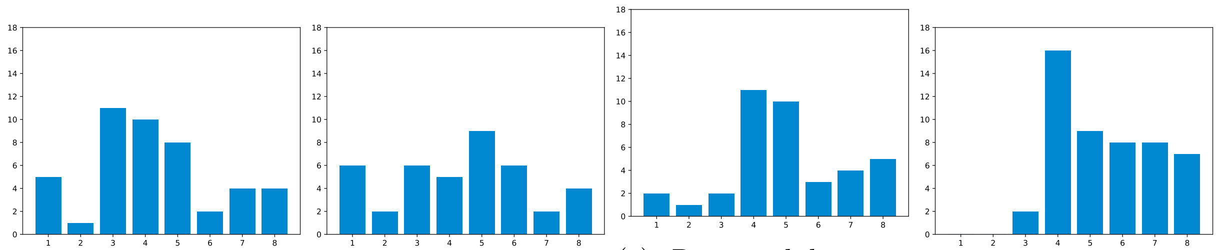
Participants spent a median duration of 9.5 minutes to complete this task. In the control condition, participants had a significantly lower average rank (mean = 4.2, ground truth = 6, $F(1, 166) = 15.3, p < 0.001$). In other words, control participants exaggerated their assessment by 30%, suggesting an impartial algorithm (and its effective communication) are necessary.

Result: Consequence explanation most effective

As shown in Figure 5.3, participants exposed to the consequence explanation exaggerated the ranking of their product review an average of 10% ($p < 0.01$), far less than the total possible, and lower than both the control and other framing-based explanations. This is similar to the reuslts of Mazar et al., where participants engaged only in limited strategic behavior when consequences were known.

5.2.5 Study 2: Is peer assessment for hiring accurate?

Study 1 demonstrated the need to communicate an impartial framing, and an effective way to do so. Study 2 investigates the real-world performance of impartial ranking. Impartial rank-aggregation methods guarantee their outcomes are resilient to strategic assessments (i.e., artificially inflating a worker's own position), but in theory, impartiality comes at a cost to accuracy [136]. This is because an impartial aggregation algorithm, by design,



(a) Control (b) Police (c) Responsibility externalization (d) Consequence

Figure 5.3: From Study 1, histogram of review placement for each framing condition; x: position, y: frequency. A skew to the right suggests less strategic behavior. Consequence explanation resulted in the least strategic behavior.

ignores some information (for instance, Naive-bipartite disregards 75% of comparisons in expectation to ensure impartiality).

In practice, the effect on overall accuracy is context dependent. On the one hand, the final ranking may be more accurate if participants report more accurate assessments (because manipulation is no longer beneficial). However, if the strategic manipulation without such a mechanism is small enough, the loss of information may result in lower real-world accuracy. Furthermore, if participant outcomes are not dependent on their own assessments, some participants may put in less effort in creating accurate assessments. In this study, we investigate:

> **Research Question 3:** Does peer assessment result in more accurate ranking of applicants in an expert hiring setting?

> **Research Question 4:** What is the net cost in accuracy for impartial guarantees of ranked aggregation?

Coefficients	β	F	p-value
Intercept (control)	4.2667	15.336	<2e-16
Police	0.1083	0.267	0.78971
Responsibility Externalization	0.7333	1.783	0.07630
Consequence	1.2333	3.216	0.00156

Table 5.1: From Study 1, consequence description leads to the least amount of strategic behavior. β coefficients are the average difference in rank from control condition (positive is less strategic behavior).

Participants and recruitment

We conducted a two-condition between-subjects experiment on AMT with 50 participants per condition. Workers who had previously taken part in our studies were not allowed to participate. This study was conducted on AMT because the platform allowed us to readily hire a large number of workers, as required for our experimental design below.

Task structure

Study 1 used a simplified task structure to make strategic behavior readily apparent. This study uses our HirePeer system introduced before, and asks for multiple paired-comparisons, instead of a ranking task.

Since multiple comparisons can be composed into a (partial) ranking, the two tasks are similar in the strategic behavior they support. (However, we acknowledge that participants may not see as readily how best to behave strategically while comparing two artifacts created by peers.)

To simulate the hiring scenario, we wanted a "job" that most AMT workers would believe they were qualified for, and had subjective selection criteria but did not require specific domain skills. Furthermore, because AMT is a micro-task market where workers are not looking for long-term employment, we wanted tasks that did not require workers to commit to long-term work, yet offered a significant monetary reward.

Therefore, our task asks crowd-workers to write feedback to newcomers to AMT. This is a task that is subjective, does not require specialized domain skills, and is something expert AMT workers might believe they are qualified for. To ensure participants felt they were qualified, participants were required to have a Master's Qualification on AMT: an indication of consistently high quality performance and familiarity with AMT. Along with potential bonuses, the task paid up to \$20, which is a significant monetary reward on the platform.

Task design: Participants were first asked to write several paragraphs of advice for AMT newcomers. The task instructions stated, "In your advice paragraphs, share tips on how to be successful, mistakes you made that you recommend they avoid, and other information you think a new Turker would find helpful."

Then, they assessed a randomly selected subset of other peers' work (i.e., their peers' advice). Concretely, at most four hours after the first phase, participants completed 50 randomly-generated pairwise comparisons among pieces of advice written by peers in the same condition, deciding which piece of advice in each comparison was higher quality (where quality was defined as more actionable and specific). Repeated pairwise comparisons were permitted (and outputs were used for quality control). At the end of both phases, participants were asked to complete a 13-question survey to understand perceptions of trust, fairness, and effort. We also captured how long they spent writing advice.

Incentive structure: Participants received a bonus if their advice piece placed in the top ten spots of the overall ranking, out of 50 total spots (\$10 USD for position one, \$9 USD for position two, and so on). There were two conditions. The *impartial* condition used the *consequence explanation* from Study 1. The control condition did not include this explanation, and instead reminded participants to pay attention to instructions (as in Study 1).

Collecting ground truth: Ground-truth ranking for each condition's advice was generated by asking 50 non-conflicted workers—25 per condition—to compare pieces of advice. This is similar to ground-truth collection in other peer assessment evaluations [151]. Non-conflicted participants were both Master Turkers and completed over 10,000 accepted human intelligence tasks (HITs) to establish a high level of expertise in the task. Non-conflicted participants conducted 50 pairwise comparisons for which piece of advice (generated from conflicted participants) was of higher quality, where quality was defined as actionable and specific. All non-conflicted participants evaluated the same 50 comparisons to generate ground-truth. Note that this method yields *ground truth comparisons*, rather than a ground-truth ranking. Ranking the 50 pieces of advice would be a prohibitively time- and effort-intensive task.

Data analysis

First, we read all responses to ensure they were sensible; all but three responses across conditions were grammatically correct and included actionable advice. These responses were kept for the following analysis. The quality of advice was similar across conditions: 1,044 characters in control vs. 1,143 characters in impartial; length is correlated with quality [150]. Median time spent writing advice (9.5 minutes control vs. 6.5 minutes impartial) did not differ in a statistically significant manner. This suggests no differences in participant recruitment across conditions.

To create rankings, we used jackknife resampling, similar to other peer assessment evaluation work [151]. In each condition, first we chose 35 of the 50 conflicted participants without replacement and sampled 25 of their pairwise comparisons, also without replacement. Because impartial algorithms are *randomized*, we ran each impartial rank-aggregation algorithm 50 times on each set of assessments to capture the variability of results. Similarly, we repeated the process of choosing participants and assessments 25 times for each condition to capture the variability caused by choosing particular assessments. This process as a whole resulted in 1250 bootstrapped rankings across conditions. We then used bootstrap significance tests introduced by Politis and Romano [190] for accuracy measures.

To evaluate the accuracy of our ranking mechanisms, we measured the agreement between the complete ranking output by each mechanism and the non-conflicted comparisons. First, given the output of a ranking mechanism, we extracted the 50 pairwise comparisons seen by non-conflicted participants from the output of the peer assessment process. Then, we assigned the output a score that measures how well the ranking agrees with the non-conflicted comparisons. The score is equal to the total number of non-conflicted participants who agree with the relative ordering of the 50 pairwise comparisons in the output ranking divided by the total number of non-conflicted participants in the majority opinion for all 50 pairwise comparisons. Note that the score is calculated relative to the majority of non-conflicted participants; this allows us to penalize mechanisms less for confusing the order of alternatives that non-conflicted participants are less sure about (i.e., which have only a slim majority among expert opinions) and to penalize mechanisms more for disagreeing with the order of alternatives that non-conflicted participants heavily agree with (i.e., alternatives with a solid majority consensus among non-conflicted participants).

Result: Peer assessment with conflicts of interest is accurate

To generate rankings without guaranteeing impartiality, we used the Kemeny rule [140], a standard method to generate rankings from an incomplete set of comparisons. Overall, the aggregated peer assessed ranking was highly similar to non-conflicted participant judgements. Even without aggregating peer assessments in an impartial manner, the accuracy was 96.6% using our metric above; see Table 5.2. This suggests peer assessed hiring could form the basis for scalable expert hiring.

Result: Guaranteeing impartiality leads to a modest loss in ranking accuracy

We compared the performance of the Kemeny rule with no framing (the *control* condition) to rankings generated from data from the *impartial* framing condition with impartial aggregation. The accuracy of ranked aggregation decrease by 8% (96.6% in control/non-impartial, vs. 88.8% in impartial); see Table 5.2. In other words, the theoretical guarantees of impartiality come at a cost of 8% in accuracy in our experimental setup.

What is an 8% loss in practice? If non-conflicted participants generating ground-truth assessments are 75% in agreement on average, as was the case in our study, and perform 20 comparisons each, then with 20 candidates a 6.67% loss in accuracy corresponds to two switches in the true ranking (e.g., switching candidates in the 10th and 11th position with each other, and the third and fourth positions with each other), and a 10% loss is equivalent to three such switches. Depending on the stakes, this loss in accuracy (and the resulting increase in employer time to hire) may be acceptable.

Aggregation Mechanism	Average Accuracy
Kemeny	0.9665*
Naive-bipartite	0.8884
Committee	0.8044
k-partite	0.7831

Table 5.2: From Study 2, (Naive-bipartite) aggregation led to a reduction of accuracy by 8%, as compared to aggregation of assessments from control condition with the Kemeny rule; each entry represents average accuracy for each condition and related aggregation. All other rows represent aggregations of assessments from experimental (i.e., impartial) conditions.

Question	Average Likert Score
I will make changes to resumé	4.6
I will learn a new skill	3.6
The feedback helped me	5.0
I put in effort	4.2
I was honest	4.8
My peers put in effort	3.6
My peers were honest	4.0
My effort affects my ranking	4.0
I enjoyed the process	5.0

Table 5.3: From Study 3, average Likert scores from post-use survey; 1: strongly disagree, 5: strongly agree. Even in a competitive hiring setting, expert crowd workers perceived peer assessment to be helpful, enjoyable, and were inclined to iterate on their job materials.

Result: Consequence explanations catalyze beliefs that assessment effort is unrelated to final ranking

Participants in the impartial condition were significantly more likely to believe their effort did not impact the final ranking of their advice piece (Control median: 4, Impartial median rating: 2, 7-point Likert scale; Wilcoxon $Z = 612.5$, $p < 0.01$) (No other survey responses differed significantly across conditions). This is interesting because the impartial framing makes no mention of how effort affects ranking. In fact, to be effective, the impartial mechanism relies on worker assessments to be honest and effort-full. It seems likely that because of this belief, participants in the impartial condition put in less effort into comparisons, slightly decreasing accuracy.

In sum, Study 2 suggests peer assessment is an accurate alternative to hiring based on expert assessment. The benefits to employers, such as decreased time to hire, and lesser reliance on worker reputations are potentially enormous. Employers can also guard themselves against individual strategic assessment at a small cost (8%) to accuracy. Next, we turn to how peer-assessed hiring may affect workers.

5.2.6 Do workers benefit from peer-assessed hiring?

In the classroom, peer assessment improves students' self-reflection [151], iteration on work [152], and development of criteria for goodness that are better aligned with experts [63]. Do these benefits transfer to workers in peer-assessed hiring? Furthermore, what reactions do expert crowd workers have to peer-assessed hiring more generally? In short, we investigate:

> **Research Question 5**: What benefits of peer assessment in education transfer to peer-assessed hiring?

To address this research question, we conducted a case study for hiring on Upwork.com; an expert crowdsourcing platform for programmers, designers, and other expert professions. Note this case study is meant to be suggestive, rather than evaluative. If participants reported none of the benefits of classroom peer-assessment, then this may not be an aspect to study further in future work. On the other hand, if participants reported some benefits (as we found), these findings may better inform and focus further research. First, to inform the design of this study, we ran two small pilots: hiring for a data visualization project and a Django development task. For this present case study, we hired expert crowd workers for the task of creating a banner ad for one of our research group's software tools, and included details about this study alongside the job description. Eleven Upwork professionals applied to this task. We describe results from the five participants who completed every stage in our protocol. We acknowledge that because of the high attrition rate, collected feedback may be biased.

Consenting participants submitted their anonymized applications to HirePeer (witnessing the impartial framing). Then, they conducted three randomly generated pairwise comparisons among their peers' job application materials. Since our system asks for comparisons, we modified the comparison-based user interface developed by Cambre et al., to ensure that assessment was scaffolded effectively [63]. After submitting these comparisons, each participant filled out a post-use survey similar to Study 2 to gather their feedback on HirePeer. The survey consisted of Likert questions to measure perceptions of effort and truthfulness of both themselves and their peers and free-response questions about overall experiences from the process.

Additionally, workers were rewarded for ranking their peers, and we ran impartial algorithms on their comparisons in order to select a winner who was invited to the task and paid for it separately.

Result: Feedback generation and reception helpful to identify new skills and improve job materials

Consistent with peer assessment literature in the classroom, multiple participants stressed the peer assessment process made them more mindful about writing a coherent and convincing application [204]. One participant stated HirePeer "helped me a lot to organize my mind and write the right things," and another wrote HirePeer "was a good exercise in application writing." Interestingly, all participants were receptive to feedback received from peers (selection bias may factor into this). Concretely, participants reported they "liked comparing proposals,", that "receiving feedback of other freelancers is a great one", and also noted no other platforms integrate this feature. One participant reported "topics that were included on the proposal [peer's resumés]...helped me a lot." Additionally, participants were slightly more likely to want to learn a new skill after this process (table 5.3).

Result: Not all participants completed assessment

Five of the 11 participants completed all steps of the review process and the post-use survey. This attrition rate is similar to peer-assessment in large-scale online courses [149], yet could be explored in more detail in future work. In addition, while our sample size is too small to draw statistical conclusions, participants who did complete our task "somewhat agreed" their effort did in fact impact their final placement (average Likert 4.0). We explore the emergent relationship between effort and impartiality in the discussion section, and how future work might rigorously investigate this.

5.3 Conclusions

5.3.1 Impartial Peer Ranking, in Theory

Pairwise comparisons.

In practice, players would often be asked to compare pairs of players, rather than giving a complete ranking. Crucially, our results seamlessly extend to that setting. Indeed, standard measure concentration inequalities show that a small number of random caparisons are sufficient to accurately estimate the pairwise comparison matrix (see, e.g., Procaccia and Shah 2016). Because we focus on pairwise rank aggregation rules, this means that the input to the rule is qualitatively unaffected by the transition from complete rankings to pairwise comparisons.

Strong impartiality.

A natural notion of impartiality that is stronger than ours (call it *strong impartiality*) requires that a player would not be able to affect the *subset* of players that are ranked above himself. In the context of online labor markets, for example, the rationale is that an employer is more likely to select the applicant in position k if the applicants in positions $1, \ldots, k-1$ are relatively weak, so it is not just the applicant's position that determines his chances of getting the job. Unfortunately, strong impartiality seems too stringent to admit reasonable rules. In fact, we can prove that no strongly impartial rule can give a $(1/3, 1/7, 1/10)$ *mixed* error with respect to Borda, but this statement requires the additional assumption that strongly impartial randomized rules are distributions over strongly impartial deterministic rules.[4] We believe that a similar statement holds without the additional assumption.

Flipping the quantifiers.

Our definitions of backward error and mixed error include the words "for all $i \in [n]$ there exists a matrix $\tilde{A} \in \Omega$." That is, for each player we can find a pairwise comparison matrix close to the original one, such that our rule puts i in a position that is identical or close to that in which the given rule would put i on the input $\tilde{A}$. The definitions would be even more compelling if the quantifiers were flipped, i.e., "there exists a matrix $\tilde{A} \in \Omega$ such that for all $i \in [n]$." Under this alternative formulation, we are not allowed to tailor $\tilde{A}$ to i, but rather there is one pairwise comparison matrix that achieves the desired property for every i. It remains an open problem whether our theorems (or variants thereof) still hold under

[4] Note that Committee and Naive-bipartite can be represented as distributions over impartial (not strongly impartial) deterministic rules, but k-partite cannot.

these more demanding notions of error, and, if not, whether these notions are feasible at all.

5.3.2 Impartial Peer Ranking, in Practice: HirePeer

Peer-assessed hiring in expert crowdsourcing is a novel alternative approach to hiring that is likely to engender many emergent effects that future work could investigate.

Practical peer-assessed hiring of experts

Even in the conflicted setting of hiring, we found scalable peer assessment can be accurate. While Study 1 shows that workers are likely to inflate their own assessment without impartial framing, Study 2 shows that the aggregated assessment of peers is highly correlated with non-conflicted expert assessors, even without using impartial aggregation. With such high agreement (96%), it seems reasonable to suggest that peer-assessed hiring can offer an alternate, scalable method to hiring crowd-experts. In particular, peer-assessed hiring can even empower non-expert employers to accurately hire qualified employees.

Collusion and privacy concerns

This work is limited in its notion of strategic behavior: although impartial mechanisms ensure any participant cannot affect her final position, it is still possible to manipulate the order of *other* applicants by reporting strategically. For instance, a coalition of workers (e.g., friends) could collectively manipulate their final placement by always selecting each others' proposals. Future work may investigate mechanisms that are resilient to collusion in their guarantees.

Another salient concern is that of anonymity. When the pool of applicants is small enough, participants may be able to identify competitors from their de-identified profiles. However, these concerns are less applicable to the expert crowdsourcing space, where the applicant pool is large, and typically has no means of communicating with each other.

Amplifying pedagogical benefits of peer review in hiring

We present initial observations that peer assessment benefits from the classroom may transfer to expert crowdsourcing. Future work may incorporate several existing interventions to improve feedback quality, such as providing tiered rubrics and banks of exemplar feedback to reuse. Furthermore, while the small sample for the case study allowed initial, qualitative observations, future work could study these benefits at larger scale with a more diverse population and investigate if pedagogical benefits evolve over time: if a crowd expert is not selected for a job, can peer-assessed hiring help them land the next job?

Incentivizing Effort

Importantly, impartial algorithms do not necessarily incentivize effort on the part of participants. Because any agent's report does not affect the distribution of her ranking, purely self-interested agents do not have any incentive to put effort into evaluating pairwise comparisons and could, theoretically, be just as well off reporting random noise as their true opinions. We did not observe this phenomenon in our (small) pilot study, and agents in general invested effort into their comparisons, but this requires further study.

One potential approach to incentivizing effort in impartial mechanisms is to offer small monetary bonuses for accurate comparisons, where accuracy is evaluated based on a small set of ground-truth comparisons performed by an impartial and knowledgeable party. As long as these monetary bonuses are not too large (i.e., the maximum bonus is less than the perceived difference between positions in the final ranking), agents will be incentivized to put in effort and impartial guarantees would still hold. However, implementing this scheme requires additional funds on the part of the mechanism designer as well as specialized expertise on the part of the "spot-checker."

Democracy cannot be static. Whatever is static is dead.

Eleanor Roosevelt.

6 Conclusion

The word "democratize" has multiple meanings. One potential definition is "introduce democratic principles to;" another is "make broadly accessible." In this thesis and beyond, I hope to conduct research that encapsulates both of these meanings. This involves exploring how computer science can help make democracy more efficient, fair, and robust; it also includes incorporating democratic ideas in new settings in order to address difficult problems in computer science and beyond.

Broadly, all projects presented in this thesis can be thought of as partial answers to the question, "How can we give people more direct and provably robust control over decision-making processes?" For instance, liquid democracy allows voters to choose delegates whose opinions, at least in theory, align more directly with their own, participatory budgeting allows citizens to allocate a portion of their city's budget toward public projects they deem to be personally important to them, and multiwinner elections allow citizens to elect more broadly representative committees instead of choosing one person to represent a heterogeneous electorate. Virtual democracy and impartial peer ranking each provide a framework for incorporating peoples' opinions in arenas where they were previously ill-suited due to a lack of infrastructure or misaligned incentives.

The technical contributions throughout this thesis can be viewed as an attempt to design democratic systems with theoretical properties that align with traditional desiderata in computer science. For liquid democracy, we explore how an increased flexibility of delegation changes the robustness and weight distribution of the voting system; for participatory budgeting, we examine how to ensure fairness and efficiency in a more flexible budgeting system; in the case of multiwinner elections, we attempt to ensure proportionality while allowing for variable-sized committees that are responsive to voters' opinions. Likewise, for virtual democracy, we seek to find a voting rule that is robust to the sort of machine learning errors that are inevitable when trying to predict peoples' opinions, and for impartial peer ranking, we design a system that balances impartiality and accuracy in order to convince rational and self-interested parties to report actionable information for peer ranking.

It is also important to note that none of the democratic paradigms exist in a vacuum. While we have focused our attention on each individually, there is significant room for synergistic interactions among them. For instance, using liquid democracy in conjunction with participatory budgeting could allow for additional flexibility in allocating funds toward public projects, and the general approach central to virtual democracy—namely, learning models of people and letting the models vote—is broadly applicable in many different contexts where voting is expensive but individual opinions are the best source of information, which is a theme in participatory design in artificial intelligence.

Limitations and Future Work

However, there exists a gap between theory and practice, and the work in this thesis is not yet implementable in the real world. Although our work in virtual democracy and impartial peer ranking has a significant HCI component that shows promise, further work must be done in order to implement and deploy our results. Part of this divide stems from the assumptions necessary for theoretical approaches. In order to prove theorems, we must assume a well-defined model of human behavior and identify clear objectives to optimize; in practice, humans are often irrational, do not have clear objective functions in mind, or may not even know their own utility function for various alternatives. However, theoretical results are still a useful tool for directional analysis, and I hope to continue working on theoretical projects that will have a tangible impact on the world.

In a sense, there is also a gap between theory and theory. Computational social choice provides a useful and robust theoretical framework in which to study human decision-making that intuitively operates in two stages. In the first stage, voters are allowed to express preferences over a predetermined, fixed set of alternatives in a specific format (e.g., approval votes, complete rankings, or partial rankings). In the second stage, a predetermined, fixed mechanism aggregates the reported preferences and produces an output. However, computational social choice assumes that voters' utilities are fixed throughout the process (i.e., it cannot take deliberation into account), provides no guidance about how the set of alternatives is to be chosen, and only allows voters to communicate their opinions to the mechanism in very specific formats. While some simplification is necessary in order to obtain tractable and clean mathematical results, the scope of computational social choice is quite limited. In particular, I am particularly interested in exploring models of deliberation that take into account the dynamics of changing opinions based on social interaction among voters, which traditional computational social choice approaches does not encapsulate.

Finally, I would like to stress that the distinction between using computer science to help democracy and using democracy to help computer science is not, by any means, well-defined, and the topics discussed in this thesis are by no means an exhaustive set of topics at the intersection of democracy and computer science. Rather, there are many other democratic processes and problems in computer science that would benefit from cross-pollination, and I am excited to continue exploring this line of work in the future.

Bibliography

[1] 412 Food Rescue Organization Website, 2018. URL https://412foodrescue.org. [On p. 100.]

[2] B. Abramowitz and N. Mattei. Flexible representative democracy: An introduction with binary issues. *arXiv:1811.02921*, 2018. [On p. 10.]

[3] Ajay Agrawal, John Horton, Nicola Lacetera, and Elizabeth Lyons. Digitization and the contract labor market: A research agenda. Technical report, National Bureau of Economic Research, 2013. [On p. 137.]

[4] M. Aleksandrov, H. Aziz, S. Gaspers, and T. Walsh. Online fair division: Analysing a food bank problem. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2540–2546, 2015. [On p. 84.]

[5] D. Alger. Voting by proxy. *Public Choice*, 126(1–2):1–26, 2006. [On p. 9.]

[6] N. Alon, F. Fischer, A. D. Procaccia, and M. Tennenholtz. Sum of us: Strategyproof selection from the selectors. In *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*, pages 101–110, 2011. [On p. 127.]

[7] Oscar Alvarado and Annika Waern. Towards algorithmic experience: Initial efforts for social media contexts. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page Paper 286, 2018. [On p. 93 and 97.]

[8] A. Aneesh. *Virtual Migration: The Programming of Globalization*. Duke University Press, 2006. [On p. 95.]

[9] Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012. [On p. 62.]

[10] K. Arrow. *Social Choice and Individual Values*. Wiley, 1951. [On p. 82 and 84.]

[11] Kenneth J Arrow. A difficulty in the concept of social welfare. *Journal of political economy*, 58(4):328–346, 1950. [On p. 3.]

[12] Yukiko Asada. Assessment of the health of americans: the average health-related quality of life and its inequality across individuals and groups. *Population Health Metrics*, 3(1):article 7, 2005. [On p. 116.]

[13] Y. Azar, A. Z. Broder, A. R. Karlin, and E. Upfal. Balanced allocations. In *Proceedings of the 26th Annual ACM Symposium on Theory of Computing (STOC)*, pages 593–602, 1994. [On p. 39, 42, and 47.]

[14] H. Azari Soufiani, D. C. Parkes, and L. Xia. Random utility theory for social choice. In *Proceedings of the 26th Annual Conference on Neural Information Processing Systems (NIPS)*, pages 126–134, 2012. [On p. 10 and 84.]

[15] H. Azari Soufiani, D. C. Parkes, and L. Xia. Preference elicitation for general random utility models. In *Proceedings of the 29th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 596–605, 2013. [On p. 10 and 84.]

[16] H. Azari Soufiani, D. C. Parkes, and L. Xia. Computing parametric ranking models via rank-breaking. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, pages 360–368, 2014. [On p. 84.]

[17] H. Aziz, O. Lev, N. Mattei, J. S. Rosenschein, and T. Walsh. Strategyproof peer selection: Mechanisms, analyses, and experiments. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, pages 397–403, 2016. [On p. 127.]

[18] H. Aziz, M. Brill, V. Conitzer, E. Elkind, R. Freeman, and T. Walsh. Justified representation in approval-based committee voting. *Social Choice and Welfare*, 48 (2):461–485, 2017. [On p. 57 and 58.]

[19] H. Aziz, E. Elkind, S. Huang, M. Lackner, L. Sánchez-Fernández, and P. Skowron. On the complexity of extended and proportional justified representation. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, pages 902–909, 2018. [On p. 58.]

[20] Haris Aziz and Nisarg Shah. Participatory budgeting: Models and approaches. In Tamás Rudas and Gábor Péli, editors, *Pathways Between Social Science and Computational Social Science: Theories, Methods, and Interpretations*. Springer, 2020. [On p. 55.]

[21] Haris Aziz, Markus Brill, Vincent Conitzer, Edith Elkind, Rupert Freeman, and Toby Walsh. Justified representation in approval-based committee voting. *Social Choice and Welfare*, 48(2):461–485, 2017. [On p. 69 and 70.]

[22] Haris Aziz, Edith Elkind, Shenwei Huang, Martin Lackner, Luis Sánchez-Fernández, and Piotr Skowron. On the complexity of extended and proportional justified representation. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, pages 902–909, 2018. [On p. 69 and 71.]

[23] Madeline Balaam, Stefan Rennick Egglestone, Geraldine Fitzpatrick, Tom Rodden, Ann-Marie Hughes, Anna Wilkinson, Thomas Nind, Lesley Axelrod, Eric Harris, Ian Ricketts, et al. Motivating mobility: Designing for lived motivation in stroke rehabilitation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3073–3082, 2011. [On p. 97.]

[24] Albert Bandura. Self-efficacy. *The Corsini Encyclopedia of Psychology*, pages 1–3, 2010. [On p. 121.]

[25] A. L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999. [On p. 30 and 39.]

[26] J. Behrens, A. Kistner, A. Nitsche, and B. Swierczek. *The Principles of LiquidFeed-*

back. Interaktive Demokratie, 2014. [On p. 30 and 32.]

[27] G. Benade, A. Kahng, and A. D. Procaccia. Making right decisions based on wrong opinions. In *Proceedings of the 18th ACM Conference on Economics and Computation (EC)*, pages 267–284, 2017. [On p. 83.]

[28] Neeli Bendapudi and Robert P Leone. Psychological implications of customer participation in co-production. *Journal of Marketing*, 67(1):14–28, 2003. [On p. 94.]

[29] D. Berga and R. Gjorgjiev. Impartial social rankings. Manuscript, 2014. [On p. 127.]

[30] Dimitris Bertsimas, Vivek F Farias, and Nikolaos Trichakis. On the efficiency-fairness trade-off. *Management Science*, 58(12):2234–2250, 2012. [On p. 96.]

[31] Reuben Binns. Algorithmic accountability and public reason. *Philosophy & Technology*, pages 1–14, 2017. [On p. 120.]

[32] Reuben Binns. Fairness in machine learning: Lessons from political philosophy. *arXiv preprint arXiv:1712.03586*, 2017. [On p. 96.]

[33] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. "It's reducing a human being to a percentage": Perceptions of justice in algorithmic decisions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 377, 2018. [On p. 109.]

[34] G. Birkhoff. Three observations on linear algebra. *Universidad Nacional de Tucumán, Revista A*, 5:147–151, 1946. [On p. 130.]

[35] A. Bjelde, F. Fischer, and M. Klimm. Impartial selection and the power of up to two choices. In *Proceedings of the 11th Conference on Web and Internet Economics (WINE)*, pages 146–158, 2015. [On p. 127.]

[36] D. Bloembergen, D. Grossi, and M. Lackner. On rational delegations in liquid democracy. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI)*, pages 1796–1803, 2019. [On p. 10.]

[37] D. Bloembergen, D. Grossi, and M. Lackner. On rational delegations in liquid democracy. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI)*, pages 1796–1803, 2019. [On p. 31.]

[38] C. Blum and C. I. Zuber. Liquid Democracy: Potentials, Problems, and Perspectives. *Journal of Political Philosophy*, 24(2):162–182, 2016. [On p. 31.]

[39] Keld Bødker, Finn Kensing, and Jesper Simonsen. *Participatory IT Design: Designing for Business and Workplace Realities*. MIT press, 2009. [On p. 97.]

[40] P. Boldi, F. Bonchi, C. Castillo, and S. Vigna. Viscous democracy for social networks. *Communications of the ACM*, 54(6):129–137, 2011. [On p. 31.]

[41] B. Bollobás and O. M. Riordan. Mathematical results on scale-free random graphs. In S. Bornholdt and H. G. Schuster, editors, *Handbook of Graphs and Networks: From the Genome to the Internet*, chapter 1, pages 1–34. Wiley-VCH, 2003. [On p. 42.]

[42] Jean-François Bonnefon, Azim Shariff, and Iyad Rahwan. The social dilemma of

autonomous vehicles. *Science*, 352(6293):1573–1576, 2016. [On p. 93 and 97.]

[43] Alan Borning and Michael Muller. Next steps for value sensitive design. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1125–1134, 2012. [On p. 97.]

[44] N. Bousquet, S. Norin, and A. Vetta. A near-optimal mechanism for impartial selection. In *Proceedings of the 10th Conference on Web and Internet Economics (WINE)*, pages 133–146, 2014. [On p. 127.]

[45] Steven J Brams, D Marc Kilgour, and M Remzi Sanver. A minimax procedure for electing committees. *Public Choice*, 132(3-4):401–420, 2007. [On p. 71.]

[46] Florian Brandl and Dominik Peters. An axiomatic characterization of the Borda mean rule. *Social Choice and Welfare*, 52(4):685–707, 2019. [On p. 71.]

[47] F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors. *Handbook of Computational Social Choice*. Cambridge University Press, 2016. [On p. 2, 10, 82, and 127.]

[48] Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D Procaccia. *Handbook of Computational Social Choice*. Cambridge University Press, 2016. [On p. 96 and 121.]

[49] M. Braverman and E. Mossel. Sorting from noisy information. arXiv preprint arXiv:0910.1191, 2009. [On p. 87.]

[50] M. Brill. Interactive democracy. In *Proceedings of the 17th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 1183–1187, 2018. [On p. 53.]

[51] M. Brill and N. Talmon. Pairwise liquid democracy. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 137–143, 2018. [On p. 10 and 31.]

[52] M. Brill, R. Freeman, S. Janson, and M. Lackner. Phragmén's voting methods and justified representation. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, pages 406–413, 2017. [On p. 58.]

[53] Markus Brill, Rupert Freeman, Svante Janson, and Martin Lackner. Phragmén's voting methods and justified representation. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, pages 406–413, 2017. [On p. 69 and 72.]

[54] Susan M Brookhart. *How to create and use rubrics for formative assessment and grading*. Ascd, 2013. [On p. 139.]

[55] Anna Brown, Alexandra Chouldechova, Emily Putnam-Hornstein, Andrew Tobin, and Rhema Vaithianathan. Toward algorithmic accountability in public services: A qualitative study of affected community perspectives on algorithmic decision-making in child welfare services. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, page Paper 41, 2019. [On p. 97.]

[56] J Randall Brown. The knapsack sharing problem. *Operations Research*, 27(2):341–355, 1979. [On p. 58.]

[57] Christopher J Bryan, Gabrielle S Adams, and Benoît Monin. When cheating would make you a cheater: Implicating the self prevents unethical behavior. *Journal of Experimental Psychology: General*, 142(4):1001, 2013. [On p. 140.]

[58] Eric Budish. The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes. *Journal of Political Economy*, 119(6):1061–1103, 2011. [On p. 59.]

[59] Bill Buxton. *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann, 2010. [On p. 97.]

[60] Y. Cabannes. Participatory budgeting: a significant contribution to participatory democracy. *Environment and Urbanization*, 16(1):27–46, 2004. [On p. 55.]

[61] Y Cabannes. Participatory budgeting in paris: Act, reflect, grow. *Another city is possible with participatory budgeting*, pages 179–203, 2017. [On p. 56.]

[62] Yves Cabannes. Contribution of participatory budgeting to provision and management of basic services. *London: IIED*, 2014. [On p. 55 and 57.]

[63] Julia Cambre, Scott Klemmer, and Chinmay Kulkarni. Juxtapeer: Comparative peer review yields higher quality feedback and promotes deeper reflection. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 294:1–294:13, 2018. [On p. 139 and 146.]

[64] I. Caragiannis, A. D. Procaccia, and N. Shah. Modal ranking: A uniquely robust voting rule. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI)*, pages 616–622, 2014. [On p. 10, 83, and 84.]

[65] I. Caragiannis, A. D. Procaccia, and N. Shah. When do noisy votes reveal the truth? *ACM Transactions on Economics and Computation*, 4(3): article 15, 2016. [On p. 10, 83, 84, and 85.]

[66] Chandra Chekuri and Sanjeev Khanna. A polynomial time approximation scheme for the multiple knapsack problem. *SIAM Journal on Computing*, 35(3):713–728, 2005. [On p. 58 and 59.]

[67] J. Chen, R. Rajaraman, and R. Sundaram. Meet and merge: Approximation algorithms for confluent flows. *Journal of Computer and System Sciences*, 72(3):468–489, 2006. [On p. 35.]

[68] J. Chen, R. D. Kleinberg, L. Lovász, R. Rajaraman, R. Sundaram, and A. Vetta. (Almost) tight bounds and existence theorems for single-commodity confluent flows. *Journal of the ACM*, 54(4): article 16, 2007. [On p. xiii, 30, 33, 35, 36, 37, and 49.]

[69] Yu Cheng, Zhihao Jiang, Kamesh Munagala, and Kangning Wang. Group fairness in committee selection. In *Proceedings of the 20th ACM Conference on Economics and Computation (ACM EC)*, pages 263–279, 2019. [On p. 58 and 76.]

[70] D. Chinn. Peer assessment in the algorithms course. *ACM SIGCSE Bulletin*, 37(3): 69–73, 2005. [On p. 139.]

[71] Alexandra Chouldechova. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, 5(2):153–163, 2017. [On p. 96.]

[72] Z. Christoff and D. Grosssi. Binary voting with delegable proxy: An analysis of liquid democracy. In *Proceedings of the 16th Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*, pages 134–150, 2017. [On p. 9 and 31.]

[73] Daryl E Chubin, Edward J Hackett, and Edward J Hackett. *Peerless science: Peer review and US science policy*. Suny Press, 1990. [On p. 137 and 139.]

[74] G. Cohensius, S. Mannor, R. Meir, E. Meirom, and A. Orda. Proxy voting for better outcomes. In *Proceedings of the 16th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 858–866, 2017. [On p. 9.]

[75] Jay A Conger and Rabindra N Kanungo. The empowerment process: Integrating theory and practice. *Academy of Management Review*, 13(3):471–482, 1988. [On p. 121.]

[76] V. Conitzer and T. Sandholm. Common voting rules as maximum likelihood estimators. In *Proceedings of the 21st Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 145–152, 2005. [On p. 10 and 84.]

[77] V. Conitzer, A. Davenport, and H. Kalagnanam. Improved bounds for computing Kemeny rankings. In *Proceedings of the 21st AAAI Conference on Artificial Intelligence (AAAI)*, pages 620–626, 2006. [On p. 128.]

[78] V. Conitzer, M. Rognlie, and L. Xia. Preference functions that score rankings and maximum likelihood estimation. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI)*, pages 109–115, 2009. [On p. 10 and 84.]

[79] V. Conitzer, W. Sinnott-Armstrong, J. Schaich Borg, Y. Deng, and M. Kramer. Moral decision making frameworks for artificial intelligence. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, pages 4831–4835, 2017. [On p. 83.]

[80] Vincent Conitzer. The maximum likelihood approach to voting on social networks. In *Proceedings of the 51st*. [On p. 10.]

[81] Vincent Conitzer. Should social network structure be taken into account in elections? *Mathematical Social Sciences*, 64(1):100–102, 2012. [On p. 10.]

[82] Juliet Corbin, Anselm Strauss, and Anselm L Strauss. *Basics of Qualitative Research*. Sage, 2014. [On p. 103.]

[83] Stéphane Côté, Paul K Piff, and Robb Willer. For whom do the ends justify the means? social class and utilitarian moral judgment. *Journal of Personality and Social Psychology*, 104(3):490–503, 2013. [On p. 99.]

[84] John Danaher, Michael J Hogan, Chris Noone, Rónán Kennedy, Anthony Behan, Aisling De Paor, Heike Felzmann, Muki Haklay, Su-Ming Khoo, John Morison, et al. Algorithmic governance: Developing a research agenda through the power of collective intelligence. *Big Data & Society*, 4(2), 2017. [On p. 96.]

[85] Norman Daniels. Reflective equilibrium. In *Stanford Encyclopedia of Philosophy*. 2016. [On p. 99.]

[86] Robyn M Dawes and Bernard Corrigan. Linear models in decision making. *Psycho-*

logical Bulletin, 81(2):95–106, 1974. [On p. 99.]

[87] G. de Clippel, H. Moulin, and N. Tideman. Impartial division of a dollar. *Journal of Economic Theory*, 139:176–191, 2008. [On p. 127 and 139.]

[88] Marquis de Condorcet. Essai sur l'application de l'analyse à la probabilité de décisions rendues à la pluralité de voix. Imprimerie Royal, 1785. Facsimile published in 1972 by Chelsea Publishing Company, New York. [On p. 3.]

[89] A. Désir, V. Goyal, S. Jagabathula, and D. Segev. Mallows-smoothed distribution over rankings approach for modeling choice. SSRN preprint, 2018. [On p. 87.]

[90] Carl DiSalvo, Illah Nourbakhsh, David Holstius, Ayça Akin, and Marti Louw. The neighborhood networks project: A case study of critical engagement and creative expression through participatory design. In *Proceedings of the 10th Anniversary Conference on Participatory Design*, pages 41–50, 2008. [On p. 97.]

[91] Jonathan Dodge, Q. Vera Liao, Yunfeng Zhang, Rachel K. E. Bellamy, and Casey Dugan. Explaining models: An empirical study of how explanations impact fairness judgment. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pages 275–285, 2019. [On p. 97.]

[92] Conal Duddy, Ashley Piggins, and William S Zwicker. Aggregation of binary evaluations: a Borda-like approach. *Social Choice and Welfare*, 46(2):301–333, 2016. [On p. 71.]

[93] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (ITCS)*, pages 214–226. ACM, 2012. [On p. 96.]

[94] J. Edmonds and K. Pruhs. Cake cutting really is not a piece of cake. In *Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 271–278, 2006. [On p. 2.]

[95] E. Elkind and N. Shah. Electing the most probable without eliminating the irrational: Voting over intransitive domains. In *Proceedings of the 30th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 182–191, 2014. [On p. 84.]

[96] E. Elkind, P. Faliszewski, and A. Slinko. Good rationalizations of voting rules. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI)*, pages 774–779, 2010. [On p. 10 and 84.]

[97] Marie Ertner, Anne Mie Kragelund, and Lone Malmborg. Five enunciations of empowerment in participatory design. In *Proceedings of the 11th Biennial Participatory Design Conference*, pages 191–194. ACM, 2010. [On p. 121.]

[98] B. Escoffier, H. Gilbert, and A. Pass-Lanneau. The Convergence of Iterative Delegations in Liquid Democracy in a Social Network. 2019. [On p. 31.]

[99] B. Fain, A. Goel, and K. Munagala. The core of the participatory budgeting problem. In *Proceedings of the 12th International Conference on Web and Internet Economics (WINE)*, pages 384–399, 2016. [On p. 58.]

[100] B. Fain, K. Munagala, and N. Shah. Fair allocation of indivisible public goods.

In *Proceedings of the 19th ACM Conference on Economics and Computation (ACM EC)*, pages 575–592, 2018. Extended version arXiv:1805.03164. [On p. 58.]

[101] P. Faliszewski, P. Skowron, A. Slinko, and N. Talmon. Multiwinner voting: A new challenge for social choice theory. In U. Endriss, editor, *Trends in Computational Social Choice*, chapter 2. 2017. [On p. 57.]

[102] Piotr Faliszewski, Arkadii Slinko, and Nimrod Talmon. The complexity of multiwinner voting rules with variable number of winners. arXiv preprint arXiv:1711.06641, 2017. [On p. 70 and 71.]

[103] FATML. Fairness, accountability, and transparency in machine learning workshop, 2018. [On p. 96.]

[104] F. Fischer and M. Klimm. Optimal impartial selection. In *Proceedings of the 15th ACM Conference on Economics and Computation (EC)*, pages 803–820, 2014. [On p. 127.]

[105] Peter C Fishburn and Aleksandar Pekeč. Approval voting for committees: Threshold approaches. Technical Report, 2004. [On p. 71.]

[106] S. Fortune, J. E. Hopcroft, and J. Wyllie. The directed subgraph homeomorphism problem. *Theoretical Computer Science*, 10:111–121, 1980. [On p. 36.]

[107] R. Freedman, J. Schaich Borg, W. Sinnott-Armstrong, J. P. Dickerson, and V. Conitzer. Adapting a kidney exchange algorithm to align with human values. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, pages 1636–1643, 2018. [On p. 83.]

[108] R. Freeman, A. Kahng, and D. M. Pennock. Proportionality in approval-based elections with a variable number of winners. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI)*, 2021. [On p. 5.]

[109] Batya Friedman and Helen Nissenbaum. Bias in computer systems. *ACM Transactions on Information Systems (TOIS)*, 14(3):330–347, 1996. [On p. 96.]

[110] Masako Fujimoto and Takeo Yamada. An exact algorithm for the knapsack sharing problem with common items. *European Journal of Operational Research*, 171(2): 693–707, 2006. [On p. 58.]

[111] Archon Fung. Recipes for public spheres: Eight institutional design choices and their consequences. *Journal of Political Philosophy*, 11(3):338–367, 2003. [On p. 93, 94, 98, 100, 101, 102, 103, 113, and 118.]

[112] Archon Fung. Varieties of participation in complex governance. *Public administration review*, 66:66–75, 2006. [On p. 93.]

[113] Archon Fung. Putting the public back into governance: The challenges of citizen participation and its future. *Public Administration Review*, 75(4):513–522, 2015. [On p. 93.]

[114] D. Gale and L. S. Shapley. College admissions and the stability of marriage. *Americal Mathematical Monthly*, 69(1):9–15, 1962. [On p. 2.]

[115] Michael R Garey and David S Johnson. *Computers and intractability*, volume 174. Freeman San Francisco, 1979. [On p. 59.]

[116] A. Gibbard. Manipulation of voting schemes. *Econometrica*, 41:587–602, 1973. [On p. 3.]

[117] Corrado Gini. Measurement of inequality of incomes. *The Economic Journal*, 31 (121):124–126, 1921. [On p. 116.]

[118] Ashish Goel, Anilesh K Krishnaswamy, Sukolsak Sakshuwong, and Tanja Aitamurto. Knapsack voting for participatory budgeting. *ACM Transactions on Economics and Computation (TEAC)*, 7(2):1–27, 2019. [On p. 57.]

[119] P. Gölz, A. Kahng, S. Mackenzie, and A.D. Procaccia. The fluid dynamics of liquid democracy. In *Proceedings of the 13th Conference on Web and Internet Economics (WINE)*, pages 188–202, 2018. [On p. 4.]

[120] P. Gölz, A. Kahng, and A. D. Procaccia. Paradoxes in fair machine learning. In *Proceedings of the 33rd Annual Conference on Neural Information Processing Systems (NIPS)*, pages 8340–8350, 2019. [On p. 43.]

[121] J. Green-Armytage. Direct voting and proxy voting. *Constitutional Political Economy*, 26(2):190–220, 2015. [On p. 7, 9, and 31.]

[122] Anthony G Greenwald. The totalitarian ego: Fabrication and revision of personal history. *American Psychologist*, 35(7):603, 1980. [On p. 140.]

[123] B. Grofman, G. Owen, and S. L. Feld. Thirteen theorems in search of the truth. *Theory and Decision*, 15(3):261–278, 1983. [On p. 10, 13, and 25.]

[124] J. Haslegrave and J. Jordan. Preferential attachment with choice. *Random Structures and Algorithms*, 48(4):751–766, 2016. [On p. 43.]

[125] D Ellis Hershkowitz, Anson Kahng, Dominik Peters, and Ariel D Procaccia. District-fair participatory budgeting. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*, 2021. [On p. 4.]

[126] Mhand Hifi, Hedi M'Halla, and Slim Sadfi. An exact algorithm for the knapsack sharing problem. *Computers & Operations Research*, 32(5):1311–1324, 2005. [On p. 58.]

[127] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. [On p. 18.]

[128] R. Holzman and H. Moulin. Impartial nominations for a prize. *Econometrica*, 81(1): 173–196, 2013. [On p. 127.]

[129] John J Horton and Joseph M Golden. Reputation inflation: Evidence from an online labor market. 2015. [On p. 137 and 138.]

[130] Lucas D Introna and Helen Nissenbaum. Shaping the web: Why the politics of search engines matters. *The Information Society*, 16(3):169–185, 2000. [On p. 96.]

[131] Panagiotis G Ipeirotis, Foster Provost, and Jing Wang. Quality management on amazon mechanical turk. In *Proceedings of the ACM SIGKDD Workshop on Human*

Computation, pages 64–67, 2010. [On p. 141.]

[132] S. Janson. Phragmén's and Thiele's election methods. Technical report, 2016. arXiv:1611.08826 [math.HO]. [On p. 58.]

[133] Svante Janson. Phragmén's and Thiele's election methods. arXiv preprint arXiv:1611.08826, 2016. [On p. 72.]

[134] A. X. Jiang, L. S. Marcolino, A. D. Procaccia, T. Sandholm, N. Shah, and M. Tambe. Diverse randomized agents vote to win. In *Proceedings of the 28th Annual Conference on Neural Information Processing Systems (NIPS)*, pages 2573–2581, 2014. [On p. 84.]

[135] Zhihao Jiang, Kamesh Munagala, and Kangning Wang. Approximately stable committee selection. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 463–472, 2020. [On p. 58.]

[136] A. Kahng, Y. Kotturi, C. Kulkarni, D. Kurokawa, and A. D. Procaccia. Ranking wily people who rank each other. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, pages 1087–1094, 2018. [On p. 6, 138, 139, and 142.]

[137] A. Kahng, S. Mackenzie, and A. D. Procaccia. Liquid democracy: An algorithmic perspective. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, 2018. Forthcoming. [On p. 4, 39, and 52.]

[138] A. Kahng, M. K. Lee, R. Noothigattu, A. D. Procaccia, and C. Psomas. Statistical foundations of virtual democracy. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pages 3173–3182, 2019. [On p. 5.]

[139] Anson Kahng, Min Kyung Lee, Ritesh Noothigattu, Ariel D Procaccia, and Christos-Alexandros Psomas. Statistical foundations of virtual democracy. In *Proceedings of the 36th International Conference on Machine Learning*, pages 3173–3182, 2019. [On p. 100.]

[140] John G Kemeny. Mathematics without numbers. *Daedalus*, 88(4):577–591, 1959. [On p. 145.]

[141] D Marc Kilgour. Approval elections with a variable number of winners. *Theory and Decision*, 81(2):199–211, 2016. [On p. 71.]

[142] D Marc Kilgour and Erica Marshall. Approval balloting for fixed-size committees. In *Electoral Systems*, pages 305–326. 2012. [On p. 79.]

[143] D Marc Kilgour, Steven J Brams, and M Remzi Sanver. How to elect a representative committee using approval balloting. In *Mathematics and Democracy*, pages 83–95. Springer, 2006. [On p. 71.]

[144] Rob Kitchin. Thinking critically about and researching algorithms. *Information, Communication & Society*, 20(1):14–29, 2017. [On p. 98.]

[145] Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807*, 2016. [On p. 96.]

[146] C. C. Kling, J. Kunegis, H. Hartmann, M. Strohmaier, and S. Staab. Voting behaviour and power in online democracy. In *Proceedings of the 9th International AAAI Conference on Web and Social Media (ICWSM)*, pages 208–217, 2015. [On p. 9.]

[147] C. C. Kling, J. Kunegis, H. Hartmann, M. Strohmaier, and S. Staab. Voting behaviour and power in online democracy. In *Proceedings of the 9th International AAAI Conference on Web and Social Media (ICWSM)*, pages 208–217, 2015. [On p. 30, 31, 38, and 39.]

[148] Y. Kotturi, A. Kahng, A. D. Procaccia, and C. Kulkarni. Hirepeer: Impartial peer-assessed hiring at scale in expert crowdsourcing markets. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*, 2020. [On p. 6 and 136.]

[149] Yasmine Kotturi, Chinmay E Kulkarni, Michael S Bernstein, and Scott Klemmer. Structure and messaging techniques for online peer learning systems that increase stickiness. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*, pages 31–38. ACM, 2015. [On p. 147.]

[150] Yasmine Kotturi, Andrew Du, Scott Klemmer, and Chinmay Kulkarni. Long-term peer reviewing effort is anti-reciprocal. In *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale*, pages 279–282. ACM, 2017. [On p. 144.]

[151] Chinmay Kulkarni, Koh Pang Wei, Huy Le, Daniel Chia, Kathryn Papadopoulos, Justin Cheng, Daphne Koller, and Scott R Klemmer. Peer and self assessment in massive online classes. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(6):33, 2013. [On p. 137, 139, 144, and 146.]

[152] Chinmay E Kulkarni, Michael S Bernstein, and Scott R Klemmer. Peerstudio: rapid peer feedback emphasizes revision and improves performance. In *Proceedings of the Second (2015) ACM Conference on Learning at Scale*, pages 75–84. ACM, 2015. [On p. 146.]

[153] R. Kumar, J. Novak, and A. Tomkins. Structure and evolution of online social networks. In *Proceedings of the 12th International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 611–617, 2006. [On p. 38.]

[154] D. Kurokawa, O. Lev, J. Morgenstern, and A. D. Procaccia. Impartial peer review. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 582–588, 2015. [On p. 127.]

[155] M. K. Lee, D. Kusbit, A. Kahng, J. T. Kim, X. Yuan, A. Chan, R. Noothigattu, D. See, S. Lee, C. Psomas, and A. D. Procaccia. WeBuildAI: Participatory framework for fair and efficient algorithmic governance. In *Proceedings of the 22nd ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW)*, pages 1–35, 2019. [On p. 5 and 93.]

[156] Min Kyung Lee. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1):1–16, 2018. [On p. 97.]

[157] Min Kyung Lee and Su Baykal. Algorithmic mediation in group decisions: Fairness perceptions of algorithmically mediated vs. discussion-based social division. In *Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW)*, pages 1035–1048, 2017. [On p. 93, 96, and 97.]

[158] Min Kyung Lee, Ji Tae Kim, and Leah Lizarondo. A human-centered approach to algorithmic services: Considerations for fair and motivating smart community service management that allocates donations to non-profit organizations. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3365–3376. ACM, 2017. [On p. 93, 97, 102, and 104.]

[159] Min Kyung Lee, Anuraag Jain, Hae Jin Cha, Shashank Ojha, and Daniel Kusbit. Procedural justice in algorithmic fairness: Leveraging transparency and outcome control for fair algorithmic mediation. *Proceedings of the ACM : Human-Computer Interaction*, 3(CSCW):Article 182, 26 pages, 2019. [On p. 97, 103, 113, and 119.]

[160] E Allan Lind and Tom R Tyler. *The Social Psychology of Procedural Justice*. Springer, 1988. [On p. 103 and 113.]

[161] Jarl Waldemar Lindeberg. Eine neue herleitung des exponentialgesetzes in der wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 15(1):211–225, 1922. [On p. 21.]

[162] John Locke. *Locke: Two treatises of government*. Awnsham Churchill, 1689. [On p. 1.]

[163] T. Lu and C. Boutilier. Robust approximation and incremental elicitation in voting protocols. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*, pages 287–293, 2011. [On p. 10 and 84.]

[164] A. Mackenzie. Symmetry and impartial lotteries. *Games and Economic Behavior*, 94:15–28, 2015. [On p. 127.]

[165] Bertram F Malle, Matthias Scheutz, Thomas Arnold, John Voiklis, and Corey Cusimano. Sacrifice one for the good of many? people apply different moral norms to human and robot agents. In *Proceedings of the 10th annual ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 117–124. ACM, 2015. [On p. 93 and 97.]

[166] C. L. Mallows. Non-null ranking models. *Biometrika*, 44:114–130, 1957. [On p. 51, 83, 85, and 135.]

[167] Y. Malyshkin and E. Paquette. The power of choice over preferential attachment. *Latin American Journal of Probability and Mathematical Statistics*, 12(2):903–915, 2015. [On p. 40 and 43.]

[168] Charles F Manski. The structure of random utility models. *Theory and Decision*, 8 (3):229–254, 1977. [On p. 99 and 105.]

[169] A. Mao, A. D. Procaccia, and Y. Chen. Better human computation through principled voting. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence (AAAI)*, pages 1142–1148, 2013. [On p. 10, 83, and 84.]

[170] Winter Mason and Siddharth Suri. Conducting behavioral research on amazon's mechanical turk. *Behavior Research Methods*, 44(1):1–23, 2012. [On p. 141.]

[171] J Nathan Matias and Merry Mou. Civilservant: Community-led experiments in platform governance. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page paper 9. ACM, 2018. [On p. 94.]

[172] Nina Mazar, On Amir, and Dan Ariely. The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45(6):633–644, 2008. [On p. 138 and 140.]

[173] J. C. Miller. A program for direct and proxy voting in the legislative process. *Public Choice*, 7(1):107–113, 1969. [On p. 7 and 9.]

[174] Eyal Mizrachi, Roy Schwartz, Joachim Spoerhase, and Sumedha Uniyal. A tight approximation for submodular maximization with mixed packing and covering constraints. arXiv:1804.10947, 2018. [On p. 64 and 65.]

[175] Jessica Morley and Luciano Floridi. The limits of empowerment: How to reframe the role of mhealth tools in the healthcare ecosystem. *Science and Engineering Ethics*, pages 1–25, 2019. [On p. 121.]

[176] Frederick Mosteller. Remarks on the method of paired comparisons: I. the least squares solution assuming equal standard deviations and equal correlations. In *Selected Papers of Frederick Mosteller*, pages 157–162. Springer, 2006. [On p. 105.]

[177] H. Moulin. Personal communication, 2014. [On p. 126.]

[178] John E Mueller. Voting on the propositions: Ballot patterns and historical trends in california. *American Political Science Review*, 63(4):1197–1212, 1969. [On p. 70.]

[179] Michael J Muller. Participatory design: The third space in hci. In *Human-Computer Interaction*, pages 181–202. CRC press, 2009. [On p. 97.]

[180] M. E. J. Newman. Clustering and preferential attachment in growing networks. *Physical review E*, 64(2):1–13, 2001. [On p. 38.]

[181] R. Noothigattu, S. Gaikwad, E. Awad, S. Dsouza, I. Rahwan, P. Ravikumar, and A. D. Procaccia. A voting-based system for ethical decision making. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, pages 1587–1594, 2018. [On p. 93, 97, and 105.]

[182] R. Noothigattu, S. S. Gaikwad, E. Awad, S. Dsouza, I. Rahwan, P. Ravikumar, and A. D. Procaccia. A voting-based system for ethical decision making. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, pages 1587–1594, 2018. [On p. 82, 83, and 84.]

[183] Arthur M Okun. *Equality and Efficiency: The Big Tradeoff*. Brookings Institution Press, 1975. [On p. 96.]

[184] Gabriele Paolacci, Jesse Chandler, and Panagiotis G Ipeirotis. Running experiments on amazon mechanical turk. 2010. [On p. 140.]

[185] Michael Q Patton. *Qualitative Research and Evaluation Methods*. Sage, 1980. [On

p. 103.]

[186] D. Peters. Proportionality and strategyproofness in multiwinner elections. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, volume 1549–1557, 2018. [On p. 66.]

[187] Dominik Peters and Piotr Skowron. Proportionality and the limits of welfarism. arXiv preprint arXiv:1911.11747, 2019. [On p. 69 and 73.]

[188] Dominik Peters and Piotr Skowron. Proportionality and the limits of welfarism. In *Proceedings of the 21st ACM Conference on Economics and Computation (ACM EC)*, pages 793–794, 2020. [On p. 58.]

[189] Edvard Phragmén. Till frågan om en proportionell valmetod. *Statsvetenskaplig Tidskrift*, 2(2):297–305, 1899. [On p. 72.]

[190] Dimitris N Politis and Joseph P Romano. The stationary bootstrap. *Journal of the American Statistical association*, 89(428):1303–1313, 1994. [On p. 145.]

[191] A. D. Procaccia and N. Shah. Optimal aggregation of uncertain preferences. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, 2016. Forthcoming. [On p. 148.]

[192] A. D. Procaccia, S. J. Reddi, and N. Shah. A maximum likelihood approach for selecting sets of alternatives. In *Proceedings of the 28th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 695–704, 2012. [On p. 10 and 84.]

[193] A. D. Procaccia, N. Shah, and Y. Zick. Voting rules as error-correcting codes. *Artificial Intelligence*, 231:1–16, 2016. [On p. 10 and 83.]

[194] Iyad Rahwan. Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology*, 20(1):5–14, 2018. [On p. 97.]

[195] John Rawls. The idea of public reason revisited. *The University of Chicago Law Review*, 64(3):765–807, 1997. [On p. 120.]

[196] John Rawls. *A Theory of Justice*. Harvard University Press, 2009. [On p. 99.]

[197] J. M. Robertson and W. A. Webb. *Cake Cutting Algorithms: Be Fair If You Can*. A. K. Peters, 1998. [On p. 2.]

[198] David Roedl, Shaowen Bardzell, and Jeffrey Bardzell. Sustainable making? balancing optimism and criticism in hci discourse. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 22(3):article 15, 2015. [On p. 121.]

[199] A. E. Roth and M. Sotomayor. *Two-Sided Mathing: A Study in Game-Theoretic Modelling and Analysis*. Cambridge University Press, 1990. [On p. 2.]

[200] Robert J Sampson, Jeffrey D Morenoff, and Thomas Gannon-Rowley. Assessing “neighborhood effects”: Social processes and new directions in research. *Annual Review of Sociology*, 28(1):443–478, 2002. [On p. 105.]

[201] L. Sánchez-Fernández, E. Elkind, M. Lackner, N. Fernández, J. A. Fisteus, P. Basanta Val, and P. Skowron. Proportional justified representation. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, pages 670–676, 2017. [On p. 58.]

[202] Luis Sánchez-Fernández, Edith Elkind, Martin Lackner, Norberto Fernández, Jesús A Fisteus, Pablo Basanta Val, and Piotr Skowron. Proportional justified representation. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, pages 670–676, 2017. [On p. 69, 70, 71, and 72.]

[203] M. Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10:187–217, 1975. [On p. 3.]

[204] D. Schön. The design studio: An exploration of its traditions and potential. *London: Royal Institute of British Architects*, 1985. [On p. 147.]

[205] Eric Schwitzgebel. Belief. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2019 edition, 2019. [On p. 95.]

[206] Amartya Sen. *Collective Choice and Social Welfare: Expanded edition*. Penguin UK, 2017. [On p. 98 and 99.]

[207] M Silberman, Joel Ross, Lilly Irani, and Bill Tomlinson. Sellers' problems in human computation markets. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, pages 18–21, 2010. [On p. 137.]

[208] P Skowron, M Lackner, M Brill, D Peters, and E Elkind. Proportional rankings. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 409–415, 2017. [On p. 71.]

[209] Piotr Skowron. Proportionality degree of multiwinner rules. arXiv preprint arXiv:1810.08799, 2018. [On p. 71.]

[210] Piotr Skowron, Martin Lackner, Edith Elkind, and Luis Sánchez-Fernández. Optimal average satisfaction and extended justified representation in polynomial time. arXiv preprint arXiv:1704.00293, 2017. [On p. 69.]

[211] H. Steinhaus. The problem of fair division. *Econometrica*, 16:101–104, 1948. [On p. 2.]

[212] George J Stigler. Information in the labor market. In *Investment in Human Beings*, pages 94–105. 1962. [On p. 137.]

[213] Nimrod Talmon and Piotr Faliszewski. A framework for approval-based budgeting methods. In *Proceedings of the 33rd Conference on Artificial Intelligence (AAAI)*, volume 33, pages 2181–2188, 2019. [On p. 57.]

[214] S. Tamura and S. Ohseto. Impartial nomination correspondences. *Social Choice and Welfare*, 43:47–54, 2014. [On p. 127.]

[215] M. Tennenholtz and A. Zohar. The axiomatic approach and the Internet. In F. Brandt, V. Conitzer, U. Endress, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 18. Cambridge University Press, 2016. [On p. 84.]

[216] Marko Terviö. Superstars and mediocrities: Market failure in the discovery of talent. *The Review of Economic Studies*, 76(2):829–850, 2009. [On p. 137.]

[217] Richard H Thaler, Cass R Sunstein, and John P Balz. Choice architecture. Manuscript, 2014. [On p. 109.]

[218] Thorvald N Thiele. Om flerfoldsvalg. *Oversigt over det Kongelige Danske Videnskabernes Selskabs Forhandlinger*, 1895:415–441, 1895. [On p. 72.]

[219] Louis L Thurstone. *The Measurement of Values*. University of Chicago Press, 1959. [On p. 105.]

[220] Yaacov Trope and Nira Liberman. Construal-level theory of psychological distance. *Psychological Review*, 117(2):440–463, 2010. [On p. 111.]

[221] G. Tullock. Computerizing politics. *Mathematical and Computer Modelling*, 16(8–9): 59–65, 1992. [On p. 9.]

[222] Amos Tversky and Daniel Kahneman. The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458, 1981. [On p. 140.]

[223] Upwork. Online work report: Global, 2014 full year data. Technical report, Upwork, 2014. URL http://elance-odesk.com/online-work-report-global. [On p. 137.]

[224] US Census Bureau. American factfinder, 2018. [On p. 105.]

[225] USDA. Food access research atlas, 2017. [On p. 105.]

[226] A. Venables and R. Summit. Enhancing scientific essay writing using peer assessment. *Innovations in Education and Teaching International*, 40(3):281–290, 2003. [On p. 139.]

[227] John Vines, Rachel Clarke, Peter Wright, John McCarthy, and Patrick Olivier. Configuring participation: On how we involve people in design. In *Proceedings of the 2013 CHI Conference on Human Factors in Computing Systems*, pages 429–438. ACM, 2013. [On p. 97 and 118.]

[228] J. von Neumann. A certain zero-sum two-person game equivalent to the optimal assignment problem. In W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games*, volume 2, pages 5–12. Princeton University Press, 1953. [On p. 130.]

[229] Mark Ware. Peer review in scholarly journals: Perspective of the scholarly community–results from an international study. *Information Services & Use*, 28 (2):109–112, 2008. [On p. 137 and 139.]

[230] Max Weber. *The Theory of Social and Economic Organization*. Simon and Schuster, 2009. [On p. 93.]

[231] Meredith Whittaker, Kate Crawford, Roel Dobbe, Genevieve Fried, Elizabeth Kaziunas, Varoon Mathur, and Jason Schultz. Ai now report 2018, 2018. [On p. 96.]

[232] Langdon Winner. Do artifacts have politics? *Daedalus*, pages 121–136, 1980. [On p. 96.]

[233] Allison Woodruff, Sarah E Fox, Steven Rousso-Schindler, and Jeffrey Warshaw. A qualitative exploration of perceptions of algorithmic fairness. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page Paper 656,

2018. [On p. 93 and 97.]

[234] L. Xia. Bayesian estimators as voting rules. In *Proceedings of the 32nd Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, 2016. [On p. 84.]

[235] L. Xia and V. Conitzer. A maximum likelihood approach towards aggregating partial orders. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*, pages 446–451, 2011. [On p. 10 and 84.]

[236] L. Xia, V. Conitzer, and J. Lang. Aggregating preferences in multi-issue domains by using maximum likelihood estimators. In *Proceedings of the 9th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 399–408, 2010. [On p. 10 and 84.]

[237] Takeo Yamada and Mayumi Futakawa. Heuristic and reduction algorithms for the knapsack sharing problem. *Computers & operations research*, 24(10):961–967, 1997. [On p. 58.]

[238] Takeo Yamada, Mayumi Futakawa, and Seiji Kataoka. Some exact algorithms for the knapsack sharing problem. *European Journal of Operational Research*, 106(1): 177–183, 1998. [On p. 58.]

[239] Tal Zarsky. The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology, & Human Values*, 41(1):118–132, 2016. [On p. 96.]

[240] Jiaming Zeng, Berk Ustun, and Cynthia Rudin. Interpretable classification models for recidivism prediction. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(3):689–722, 2017. [On p. 99.]

[241] Haiyi Zhu, Bowen Yu, Aaron Halfaker, and Loren Terveen. Value-sensitive algorithm design: Method, case study, and lessons. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):article 194, 2018. [On p. 93 and 97.]