2021

Algorithmic Processes And Social Values

Zachary Schutzman
University of Pennsylvania

Algorithmic Processes And Social Values

Abstract

In this thesis, we study several problems at the interface of algorithmic decision-making and society, focusing on the tensions that arise between these processes and social values like fairness and privacy. In the first chapter, we examine the design of financial portfolios which adequately serve all segments of the population. In the second, we examine an allocation setting where the allocator wishes to distribute a scarce resource across many groups fairly, but does not know ahead of time which groups have a need for the resource. In the third, we study a game-theoretic model of information aggregation and the effects of individuals acting to preserve the privacy of their personal beliefs on the collective welfare of the population. Finally, we look at some of the issues that arise from the desire to apply automated techniques to problems in redistricting, including fundamental flaws in the definitions and frameworks typically used.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Computer and Information Science

First Advisor

Aaron Roth

Keywords

algorithms, fairness, privacy, redistricting

Subject Categories

Computer Sciences

ALGORITHMIC PROCESSES AND SOCIAL VALUES

Zachary Ian Schutzman

A DISSERTATION

in

Computer and Information Science

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2021

Supervisor of Dissertation

Aaron Roth, Professor, Computer and Information Science

Graduate Group Chairperson

Mayur Naik, Professor, Computer and Information Science

Dissertation Committee:
Michael Kearns, Professor National Center Chair, Department of Computer and Information Science
Sampath Kannan, Henry Salvatori Professor, Department of Computer and Information Science
Kristian Lum, Assistant Research Professor, Department of Computer and Information Science
Ariel Procaccia, Gordon McKay Professor, Department of Computer Science, Harvard University

Acknowledgment

It took 1,800 days to complete my PhD. 1,800 days between my first day of graduate school and the date on my diploma. I am grateful to every single person who made every one of those days one where I learned, succeeded, and grew in some way.

First, to my advisor Aaron Roth, I am grateful for all of his guidance and mentorship and collaboration both past and ongoing as well as for admitting me to Penn in the first place and giving me the opportunity to do so much living and learning over the last five years. I am also thankful to Michael Kearns for being a spectacular unofficial second advisor. Without Aaron's and Michael's teaching, support, and direction, none of the work in this dissertation would have been possible.

Thanks to Sampath Kannan, Kristian Lum, and Ariel Procaccia for their service on my thesis committee. Their insight and perspective have helped steer my research and writing both within this document as well as beyond.

I'm thankful for my parents, Linda and Jeff, my brother, Ethan, my grandparents, Pat, Joe, Peggy, and Peter, and everyone else in my family for supporting me over these last 22 years of education and for politely smiling and nodding when I tried to explain my research.

To my professors and mentors at Colby College, especially Tim Hubbard and Sahan Dissanayake, thank you for, in whatever small way, helping me reach this point in my academic career. To Moon Duchin, Justin Solomon, Aidan Kestigian, and Daryl DeFord, thank you for bringing me to the Voting Rights Data Institute which both advanced my research work by leaps and bounds as well as gave me the opportunity to meet friends and colleagues who will last a lifetime.

To my friends Assaf, Lorenzo, Max, Ruth, and the Assembly of Dogs, thanks for putting up with me.

Finally, this work would not have been possible without the hard work of my coauthors and collaborators throughout graduate school. Tara Abrishami, Emilia Alvarez, Assaf Bar-Natan, Ruth Buck, Aloni Cohen, Daryl DeFord, Emily Diana, Travis Dick, Jinshuo Dong, Seth Drew, Moon Duchin, Hadi Elzayn, Michelle Feng, Patrick Girardet, Nestor Guillen, Natalia Hajlasz, Eugene Henninger-Voss, Eduardo Chavez Heredia, Max Hully, Shahin Jabbari, Amara Jaeger, Matthew Joseph, Christopher Jung, Michael Kearns, Hugo Lavenant, Lorenzo Najt, Seth Neel, Heather Newman, Sloan Nietert, Sasho Nikolov, Aidan Perrault, Aaron Roth, Parker Rule, Saeed Sharifi-Malvajerdi, Justin Solomon, Jon Ullman, Bo Waggoner, Thomas Weighill, Si Wu, Steven Wu, and Juba Ziani. I look forward to this list growing in the years to come.

Chapter 3 is based on *Fair Algorithms for Learning in Allocation Problems* by Elzayn, Jabbari, Jung, Kearns, Neel, Roth, and Schutzman, in the Proceedings of FAT, 2019.
Chapter 2 is based on *Algorithms and learning for Fair Portfolio Design* by Diana, Dick, Elzayn, Kearns, Roth, Schutzman, Sharifi-Malvajerdi, and Ziani, To appear in the Proceedings of EC, 2021.
Chapter 4 is based on *Priceof Privacy in the Keynesian Beauty Contest*, Elzayn and Schutzman, in the Proceedings of EC, 2019.
Chapter 5 is based on *The Gerrymandering Jumble: Map Projections Permute Districts' Compactness Scores*, Bar-Natan, Najt, and Schutzman, in *Cartography and GIS* ,2020 and *Trade-Offs in Fair Redistricting*, Schutzman, in the Proceedings of AIES, 2020.

ABSTRACT

ALGORITHMIC PROCESSES AND SOCIAL VALUES

Zachary Ian Schutzman

Aaron Roth

In this thesis, we study several problems at the interface of algorithmic decision-making and society, focusing on the tensions that arise between these processes and social values like fairness and privacy. In the first chapter, we examine the design of financial portfolios which adequately serve all segments of the population. In the second, we examine an allocation setting where the allocator wishes to distribute a scarce resource across many groups fairly, but does not know ahead of time which groups have a need for the resource. In the third, we study a game-theoretic model of information aggregation and the effects of individuals acting to preserve the privacy of their personal beliefs on the collective welfare of the population. Finally, we look at some of the issues that arise from the desire to apply automated techniques to problems in redistricting, including fundamental flaws in the definitions and frameworks typically used.

Contents

List of Figures

Chapter 1

Introduction

Although the ubiquity of computing has allowed for a rapid increase in both the frequency and the breadth of algorithmic processes making decisions that affect our day-to-day lives, the idea of taking some information from an individual or a group of people, feeding it through an algorithm, and using the output to inform some action or policy is far from a new phenomenon. Shop owners use product popularity information to decide which goods to stock and how to price them. An enormous set of formulas takes as input the amount of money an individual earns in a year and outputs the amount they ought to pay in taxes. Voting systems aggregate voter preferences into decisions about who should lead a country. Of course, concerns that algorithmic decision-making might flout social norms are not new either. What kinds of people face higher or lower tax burdens? Does monitoring a customer's shopping habits violate their privacy? What kinds of views are shut out from political representation?

The landscape of algorithmically-informed decision-making has only become more complex with the rise of computing and 'big data'. Every time you open a webpage, an algorithm decides which ads to display based on some combination of the information about the page in particular, your personal browsing habits, and the advertiser's willingness-to-pay to serve you an ad. Similarly, algorithms tell us which schools to attend, which news to read, and which roads to drive on. Hundreds of times per day, as algorithmic processes make decisions for and about us, we are left to worry about how the algorithm designers and data collectors are respecting *social norms* like fairness and privacy.

In this dissertation, we look at four settings in which we study an algorithmic process together with a question about how to modify it to comport with a social norm. We examine these contexts from a computational perspective, considering the design and analysis of algorithms and developing mathematical proofs to describe what these algorithms formally can and cannot accomplish.

The first setting we examine is *optimal portfolio design*, which is the process of constructing a collection of financial assets for a customer which maximizes the amount of monetary returns they receive while limiting the risk they are exposed to. We study this together with the question of *'How might we prevent the designer from constructing products which only serve a small segment of the population?'* We think about a designer who must offer a small number of investment options for a broad population, which is divided into groups. We might think of these groups as corresponding to demographics or a particular employment status or some other label which is sensible in this context. In a standard profit maximization setup, the designer would probably want to design portfolios which serve wealthier customers well and ignore the segment of the population that is less well-off. In our model, we develop a notion of *minmax fairness*, which asks that no group's access to desirable portfolios is substantially worse than any other group's.

In the United States and many other Western nations, wealth inequalities are magnified by exclusive access to certain classes of assets and investments. For example, in the U.S. one must show a minimum amount of wealth or income in order to invest in hedge funds, startups, and venture capital vehicles. While putatively for the protection of ordinary consumers against risky investments, such restrictions prevent large swaths of society from reaping the often outsized returns that they provide over time. In addition to such explicit barriers, many of the aforementioned asset classes also have implicit and social barriers, requiring personal connections and status to access them. Overall such frictions fall disproportionately on the less wealthy members of society, which itself correlates with demographic factors such as race and ethnicity.

In Chapter 2 we study financial asset design with fairness as an explicit goal. We consider a variation on the classical problem of optimal portfolio design. In this setting, there is a collection of individual financial assets, each with an expected return and a risk, which is the variance of the distribution from which the actual return is drawn. The returns on these assets may be correlated with one another, either positively or negatively. The collection of assets can therefore be abstractly represented as a vector of returns and a covariance matrix of risks. Then, an individual consumer is specified by their *risk tolerance* and the classical portfolio design problem asks to choose a weight for each asset such that the total weight does not exceed one and the risk level of the resulting portfolio does not exceed their risk tolerance.

If the designer can only offer a small number of different portfolios, then a consumer might be assigned to a portfolio which does not achieve as high of a return as a bespoke one designed for them. We can then impose fairness considerations by asking that the suite of portfolios the designer chooses does not result in this gap between realized and bespoke returns falling disproportionately on some group of consumers.

Our main results are algorithms for optimal and near-optimal portfolio design for both social welfare and fairness objectives, both with and without assumptions on the underlying group structure. We describe an efficient algorithm based on an internal two-player zero-sum game that learns near-optimal fair portfolios *ex ante* and show experimentally that it can be used to obtain a small set of fair portfolios *ex post* as well. For the special but natural case in which group structure coincides with risk tolerances (which models the reality that wealthy consumers generally tolerate greater risk), we give an efficient and optimal fair algorithm. We also provide generalization guarantees for the underlying risk distribution that has no dependence on the number of portfolios and illustrate the theory with simulation results.

The next setting is a *resource allocation* problem where an allocator wants to distribute a resource to deserving candidates across multiple groups, which we might think of again as corresponding to demographics or the groups might be geographic, such as neighborhoods. The question we ask is

'If the allocator does not know how many people in each group are deserving of the resource and can only learn about it by making an allocation and observing some outcome of that allocation, what can the allocator do to ensure fair distribution?' The natural concern here is of *feedback loops*, where if the allocator incorrectly believes that one group has a high number of candidates, that group will continue to be overallocated the resource and not enough will be learned about the other groups to eventually correct the mistake. We study this problem from a learning-theoretic perspective and give a natural algorithm to solve the problem.

Many settings can be modelled as a central agent allocating a limited resource among several groups, such as administering loans or scholarships, distributing consumer goods to retail locations, or assigning doctors to hospitals. The allocator typically has a straightforward objective of maximizing the number of individuals who receive the resource: qualified loan applicants who successfully pay back loans, retail sales, and patients given medical care. Often in such settings, *fairness* is a concern. One natural notion of fairness, based on general principles of *equality of opportunity*, asks that conditional on an individual being a candidate for the resource in question, the probability of actually receiving it is approximately independent of the individual's group. For example, equally creditworthy individuals in different racial groups ought to have roughly equal chances of receiving a loan; sick patients who live in different neighborhoods shouldn't face dramatically different waiting times to see a doctor at a clinic. In these settings, the allocator does not know ahead of time the exact number of individuals in each group who will need the resource, and rather operates in a *censored feedback* environment. The allocator only learns the amount of the resource that was used at a particular time step in a particular group.

In Chapter 3, we formalize this general notion of fairness for allocation problems and investigate its algorithmic consequences. Our main technical results include an efficient learning algorithm that converges to an optimal fair allocation even when the allocator does not know the distributions of the frequency of candidates in each group. This algorithm operates in a *censored* feedback model in which only the number of candidates who received the resource in a given allocation can be observed, rather than the true number of candidates in each group.

The third setting is one where we examine privacy concerns in a strategic setting. The Keynesian Beauty Contest is a classical game in which strategic agents seek to both accurately guess the true state of the world as well as the average action of all agents. We study an augmentation of this game where agents are concerned about revealing their private information and additionally suffer a loss based on how well an observer can infer their private signals. We solve for an equilibrium of this augmented game and quantify the loss of social welfare as a result of agents acting to obscure their private information, which we call the 'price of privacy'. We analyze two versions of this this price: one from the perspective of the agents measuring their diminished ability to coordinate due to acting to obscure their information and another from the perspective of an aggregator whose statistical estimate of the true state of the world is of lower precision due to the agents adding random noise to their actions. We show that these quantities are high when agents care very strongly about protecting their personal information and low when the quality of the signals the agents receive is poor.

The final setting is in applying algorithms to *redistricting*, particularly electoral redistricting. Jurisdictions draw geographic districts for all sorts of purposes, including to assign children to schools, fire stations to homes, and legislative seats to voters. Over the last decade, there has been a strong interest in using more automated and computer-supported decision-making in drawing these districts. The arguments in favor of this tend to follow the belief that while human line-drawers might draw the lines for some personal or nefarious reason, a computer algorithm brings no such biases. The question we ask is *'How might proposed frameworks for fair algorithmic redistricting still be susceptible to the influence of human decision-makers and to what degree are these frameworks compatible with each other?'* We explore this question in two parts.

In the first, we look at *compactness scores*, perhaps the most classical way of infusing algorithmic constraints into the redistricting process. In political redistricting, the *compactness* of a district is used as a quantitative proxy for its fairness. Several well-established, yet competing, notions of geographic compactness are commonly used to evaluate the shapes of regions, including the *Polsby-Popper score*, the *convex hull score*, and the *Reock score*, and these scores are used to compare two or more districts or plans. In the first part of Chapter 5, we prove mathematically that any *map projection* from the sphere to the plane reverses the ordering of the scores of some pair of regions for all three of these scores. We evaluate these results empirically on United States congressional districts and demonstrate that this order-reversal does occur in practice with respect to commonly-used projections. Furthermore, the Reock score ordering in particular appears to be quite sensitive to the choice of map projection.

Second, we examine two prototypical frameworks for fair automated redistricting: draw districts to be as compact as possible and draw districts to be as politically representative as possible. What constitutes a 'fair' electoral districting plan is a discussion dating back to the founding of the United States and, in light of several recent court cases, mathematical developments, and the approaching 2020 U.S. Census, is still a fiercely debated topic today. In light of the growing desire and ability to use algorithmic tools in drawing these districts, we discuss two prototypical formulations of fairness in this domain: drawing the districts by a neutral procedure or drawing them to intentionally induce an equitable electoral outcome. We then generate a large sample of districting plans for North Carolina and Pennsylvania and consider empirically how *compactness* and *partisan symmetry*, as instantiations of these frameworks, trade off with each other – prioritizing the value of one of these necessarily comes at a cost in the other.

Chapter 2

Fair Portfolios

2.1 Introduction

In this work, we initiate the study of financial portfolio design with fairness as an explicit goal. In many economies, including (but not limited to) the United States and many other Western nations, financial assets serve as a vehicle that allow consumers to build long-term wealth; unequal access to these assets can greatly exacerbate economic inequality and social immobility. Some assets, such as stocks, bonds, and other publicly-traded securities, yield a relatively small but steady expected return that allows wealth to grow over time, yet are unequally held across groups; for instance, an analysis by the Federal Reserve of the 2019 Survey of Consumer Finances found that while 60 percent of White households had direct or indirect equity, only 33 percent of Black households and 24 percent of Hispanic households did (Bhutta et al., 2020). For riskier assets, in the United States, potential investors must meet minimum wealth or income requirements in order to invest in hedge funds, startups, and venture capital vehicles.

While putatively for the protection of ordinary consumers against risky investments, such restrictions prevent large swaths of society from reaping the often-outsized returns they provide over time. In addition to such explicit barriers, many of the aforementioned asset classes also have social barriers, requiring connections to access them. Overall, such frictions fall disproportionately on the less wealthy, and as such may also be correlated with race and other demographic factors.

While these structural inequalities may not be entirely mitigated by fairer portfolio design alone, explicitly incorporating fairness concerns into this process could have significant impacts on the lives of individual consumers. In this work, we consider algorithmic and learning problems in the fair design of financial products under a simple model. We imagine a large population of individual retail investors or *consumers*, each of whom has her own tolerance for investment risk in the form of a limit on the variance of returns. It is well known in quantitative finance that for any set of financial assets, the optimal expected returns on investment are increasing with risk.

We assume that a large retail investment firm (such as Vanguard or Fidelity) wishes to design portfolios to serve these consumers under the common practice of assigning consumers only to portfolios with lower risks than their tolerances. The firm would like to design and offer only a *small* number of such products — much smaller than the number of consumers — since the execution and maintenance of portfolios is costly and ongoing. The overarching goal is to design the products to minimize consumer *regret* — the loss of returns due to being assigned to lower-risk portfolios compared to the bespoke portfolios saturating tolerances — both with and without fairness considerations. We highlight that fairness concerns are becoming increasingly more prominent among both practitioners and academics in the financial sector - see discussion in, for instance, Klein, Lee and Floridi (2020), Shu and Ucla (2012), and Sarra (2014).

We consider a notion of group fairness adapted from the literature on *fair division.* Consumers belong to underlying groups that may be defined by standard demographic features such as race, gender or age, or they may be defined by the risk tolerances themselves, as higher risk appetite is generally correlated with higher wealth. We study *minmax group fairness*, in which the goal is to minimize the maximum regret across groups — i.e. to optimize for the *least well-off* group. Compared to the approach of constraining regret to be equal across groups (which is not even always feasible in our setting), minmax optimal solutions have the property that they Pareto-dominate regret-equalizing solutions; under the minmax-optimal suite of portfolios, every group experiences regret less than or equal to the all-groups-equal-optimal suite.

2.1.1 Related Work

Our work generally falls within the literature on fairness in machine learning, which is too broad to survey in detail here; see Chouldechova and Roth (2020) for a recent overview. We are perhaps closer in spirit to research in fair division or allocation problems Barman and Krishnamurthy (2017); Procaccia and Wang (2014); Budish (2011), in which a limited resource must be distributed across a collection of players so as to maximize the utility of the *least* well-off; here, the resource in question is the small number of portfolios to design. However, we are not aware of any technical connections between our work and this line of research. While our work is inspired and motivated by the context of and applications to the portfolio design problem in finance, there are similarities between the problem we study and *one-dimensional clustering* and *one-dimensional facility location* problems Nielsen and Nock (2014); Bellman (1973); Chen and Wang (2011); Hassin and Tamir (1991). In these problems, there are a collection of individuals along the real line, and the algorithm designer must choose a collection of points on the line to serve as 'centers' so as to minimize some cost function associated with the individuals accessing the centers. The algorithm we give in Section 2.3 for the problem absent fairness is a standard dynamic programming formulation for these settings. There is some recent literature on *fair facility location* problems Jung et al. (2019); Mahabadi and Vakilian (2020), which attempt to choose a small number of "centers" to minimize a cost function together with constraints informed by the need to serve a large and diverse population. Our work also has similarities to work on designing algorithms for "fair allocation" problems that arise in contexts such as loan administration or predictive policing Ensign et al. (2017); Elzayn et al. (2019); Donahue and Kleinberg (2020).

There seems to have been relatively little explicit consideration of fairness issues in quantitative finance generally and optimal portfolio design specifically. An exception is Iancu and Trichakis

(2014), in which the interest is in fairly amortizing transaction costs of a single portfolio across investors rather than designing multiple portfolios to meet a fairness criterion.

2.1.2 Our Results and Techniques

In Section 2.3, we provide a dynamic program to find p products that optimize the average regret of a single population. In Section 2.4, we divide the population into different groups and develop techniques to guarantee minmax group fairness: in Section 2.4.1, we show a separation between deterministic solutions and randomized solutions (i.e. distributions over sets of p products) for minmax group fairness; in Section 2.4.2, we leverage techniques for learning in games to find a distribution over products that optimizes for minmax group fairness; in Section 2.4.3, we focus on deterministic solutions and extend our dynamic programming approach to efficiently optimize for the minmax objective when the number of groups is constant. In Section 2.5 we give an approximate greedy algorithm leveraging the submodularity of the regret function. Section 2.6 discusses a generalization of our regret notion where we allow users to be assigned to products with a *higher* level of risk than their tolerance. Finally, in Section 2.7, we provide experiments to complement our theoretical results.

2.2 Model and Preliminaries

We aim to create products (portfolios) consisting of weighted collections of assets with differing means and standard deviations (risks). Each consumer is associated with a real number $\tau \in \mathbb{R}_{\geq 0}$, which is the *risk tolerance* of the consumer — an upper bound on the standard deviation of returns. We assume that consumers' risk tolerances are bounded.

2.2.1 Bespoke Problem

We adopt the standard Markowitz framework Markowitz (1952) for portfolio design. Given a set of m assets with mean $\mu \in \mathbb{R}_+^m$ and covariance matrix $\Sigma \in \mathbb{R}^{m \times m}$, and a consumer risk tolerance τ, the *bespoke* portfolio achieves the maximum expected return that the consumer can realize by assigning weights $\mathbf{a}$ over the assets (where the weight of an asset represents the fraction of the portfolio allocated to said asset) subject to the constraint that the overall risk, quantified as the standard deviation of the mixture $\mathbf{a}$ over assets, does not exceed their tolerance τ. Finding a consumer's bespoke portfolio can be written down as an optimization problem. We formalize the bespoke problem in Equation (2.1) below, and call the solution $r(\tau)$.

$$r(\tau) = \max_{\mathbf{a} \in \mathbb{R}^m} \left\{ \mathbf{a}^\top \mu \,|\, \mathbf{a}^\top \Sigma \, \mathbf{a} \leq \tau^2, \mathbb{1}^\top \mathbf{a} = 1 \right\} \tag{2.1}$$

Here we note that optimal portfolios are summarized by $r(\tau)$, which is non-decreasing. Since consumer tolerances are bounded, we let B denote the maximum value of $r(\tau)$ across all consumers.

2.2.2 A Regret Notion

To frame our problem in terms of standard algorithmic formulations, we need a notion of *regret*. Two considerations inform our particular choice of regret definition. First, we would like the solution to our problem to correspond to a good outcome for consumers, and this requires that the regret of a choice of products measures how well the consumers could have done under some other choice (as opposed to thinking about how well the portfolio designer could have done, which may or may not align with consumer interests). The solution to the bespoke problem is, by definition, the best the consumer could have done given their risk tolerance, and thus is an appropriate benchmark. Second, the *realized* outcome of any given asset or portfolio may include significant randomness, and so it is more useful to consider the *expected value* of instruments. Hence, the notions of regret we will consider will use the difference in expected return between a consumer's best choice under our selected portfolio and their bespoke portfolio.

Suppose there are n consumers to whom we want to offer products. Our goal is to design $p \ll n$ products that minimize a notion of *regret* for a given set of consumers. A product has a risk (standard deviation) which we will denote by c. We assume throughout that in addition to the selected p products, there is always a risk-free product (say cash) available that has zero return; we will denote this risk-free product by $c_0 \equiv 0$ throughout the paper ($r(c_0) = 0$). For a given consumer with risk threshold τ, the regret of the consumer with respect to a set of products $\mathbf{c} = (c_1, c_2, \ldots, c_p) \in \mathbb{R}^p_{\geq 0}$ is the difference between the return of her bespoke product and the maximum return of any product with risk that is *less than or equal to* her risk threshold. To formalize this, the regret of products $\mathbf{c}$ for a consumer with risk threshold τ is defined as $\mathrm{R}_\tau(\mathbf{c}) = r(\tau) - \max\limits_{c_j \leq \tau} r(c_j)$. Note since $c_0 = 0$ always exists, the $\max_{c_j \leq \tau} r(c_j)$ term is well defined. Now for a given set of consumers $S = \{\tau_i\}_{i=1}^n$, the regret of products $\mathbf{c}$ on S is simply defined as the *average regret* of $\mathbf{c}$ on S:

$$\mathrm{R}_S(\mathbf{c}) \triangleq \frac{1}{n} \sum_{i=1}^{n} \mathrm{R}_{\tau_i}(\mathbf{c}) \tag{2.2}$$

When S includes the entire population of consumers, we call $\mathrm{R}_S(\mathbf{c})$ the *population regret.* The following notion for the *weighted regret* of $\mathbf{c}$ on S, given a vector $\mathbf{w} = (w_1, \ldots, w_n)$ of weights for each consumer, will be useful in Sections 2.3 and 2.4:

$$\mathrm{R}_S(\mathbf{c}, \mathbf{w}) \triangleq \sum_{i=1}^{n} w_i \, \mathrm{R}_{\tau_i}(\mathbf{c}) \tag{2.3}$$

Absent any fairness concern, our goal is to design efficient algorithms to minimize $\mathrm{R}_S(\mathbf{c})$ for a given set of consumers S and target number of products p: $\min_{\mathbf{c}} \mathrm{R}_S(\mathbf{c})$. This will be the subject of Section 2.3. We can always find an optimal set of products as a subset of the n consumer risk thresholds $S = \{\tau_i\}_i$, because if any product c_j is not in S, we can replace it by $\min\{\tau_i \mid \tau_i \geq c_j\}$ without decreasing the return for any consumer. We let $C_p(S)$ represent the set of all subsets of size p for a given set of consumers S: $C_p(S) = \{\mathbf{c} = (c_1, c_2, \ldots, c_p) \subseteq S\}$.

We can reduce our regret minimization problem to the following problem:

$$\mathcal{R}\left(S, p\right) \triangleq \min_{\mathbf{c} \in C_p(S)} \mathrm{R}_S(\mathbf{c}) \tag{2.4}$$

Similarly, we can reduce the weighted regret minimization problem to the following problem:

$$\mathcal{R}\left(S, \mathbf{w}, p\right) \triangleq \min_{\mathbf{c} \in C_p(S)} \mathrm{R}_S(\mathbf{c}, \mathbf{w}). \tag{2.5}$$

Remark 2.1. *Later, we consider a more general regret framework that allows consumers to be assigned to products with risk higher than their tolerance. In this case, it is not necessarily optimal to always place products on the consumer risk thresholds; in turn, the techniques we use are of independent interest and different from those used in the main body of the paper.*

2.2.3 Group Fairness: *ex post* and *ex ante*

Like much of the literature, we focus on group fairness, but our setting is somewhat unique. First, while a designer with access to demographic information may well wish to enforce fairness with respect to traditional protected statuses like race, gender, and age, other consideration may also be reasonable and important. For instance, partitions of risk tolerances themselves - e.g. high risk tolerance, medium risk tolerance, and low risk-tolerance - are likely correlated with wealth, socioeconomic status, or race and ethnicity. If this data is not available to the designer, enforcing fairness with respect to classes of risk tolerance may be a reasonable way to proxy for demographic fairness. Our framework is agnostic to what the groups represent.

Additionally, in settings such as designing portfolios for large organizations, enforcing fairness with respect to job groups (for instance, executives versus support staff) may be another natural concern. We also highlight that our methods provide a form of *individual fairness* if groups are taken to be the individuals themselves, at the cost of runtime and sample size increases accordingly.

Our approach also engenders nuance in the *timing* of fairness considerations. That is, as we use randomized algorithms, one may consider the regret that will be achieved in expectation *before* the randomness of the algorithm is realized, or first realize the randomness and then measure fairness with respect to the portfolios ultimately selected. Both notions are useful: *ex ante* fairness is easier to satisfy computationally, and may also capture repeated settings where the portfolios are modified frequently and the risk tolerances of consumers are not known in advance. On the other hand, the stronger notion of *ex post* fairness captures settings where redesign opportunities may be few and far between or the consumers' tolerances are fixed in advance (such as providing retirement investment products to a class of employees in a single large firm).

Suppose consumers are partitioned into g groups: $S = \{G_k\}_{k=1}^{g}$, e.g. based on their attributes like race or risk levels. Each G_k consists of the consumers of group k represented by their risk thresholds. We will often abuse notation and write $i \in G_k$ to denote that consumer i has threshold $\tau_i \in G_k$. Given this group structure, minimizing the regret of the whole population absent any constraint might lead to some groups incurring much higher regret than others. With fairness concerns in mind, we turn to the design of efficient algorithms to minimize the maximum regret over groups (we call this maximum "group regret"):

$$\mathcal{R}_{\text{fair}}\left(S, p\right) \triangleq \min_{\mathbf{c} \in C_p(S)} \left\{ \max_{1 \le k \le g} \mathrm{R}_{G_k}(\mathbf{c}) \right\} \tag{2.6}$$

The set of products $\mathbf{c}$ that solves the above minmax problem will be said to satisfy *ex post* minmax fairness (for brevity, we call this "fairness" throughout). One can relax Program (2.6) by allowing the designer to *randomize* over sets of p products and output a *distribution* over $C_p(S)$ (as opposed to one deterministic set of products) that minimizes the maximum *expected* regret of groups:

$$\widehat{\mathcal{R}}_{\text{fair}}\left(S, p\right) \triangleq \min_{\mathcal{C} \in \Delta(C_p(S))} \left\{ \max_{1 \le k \le g} \mathbb{E}_{\mathbf{c} \sim \mathcal{C}}\left[\mathrm{R}_{G_k}(\mathbf{c})\right] \right\} \tag{2.7}$$

where $\Delta(A)$ represents the set of probability distributions over the set A, for any A. The distribution $\mathcal{C}$ that solves the above minmax problem will be said to satisfy *ex ante* minmax fairness — meaning fairness is satisfied in expectation *before* realizing any set of products drawn from the distribution $\mathcal{C}$ — but there is no fairness guarantee on the realized draw from $\mathcal{C}$. Such a notion of fairness is useful in settings in which the designer has to make repeated decisions over time and has the flexibility to offer different sets of products in different time steps. In Section 2.4, we provide algorithms that solve both problems cast in Programs (2.6) and (2.7). We note that while there is a simple integer linear program (ILP) that solves Programs (2.6) and (2.7), such an ILP is often intractable to solve. We use it in our experiments on small instances to evaluate the quality of our efficient algorithms.

2.3 Regret Minimization Absent Fairness

In this section, we provide an efficient dynamic programming algorithm for finding the set of p products that minimizes the (weighted) regret for a collection of consumers (absent fairness constraints). This dynamic program will be used as a subroutine in our algorithms for finding optimal products for minmax fairness. A firm uninterested in fairness could use such a program (choosing weights increasing in customer profitability) to minimize regret for profitable customers and consequently, maximize its own profits; the fact that this approach is only an ingredient, not sufficient alone, in finding a fair portfolio highlights the fact that the fairness-constrained problem is considerably more technically challenging.

Let $S = \{\tau_i\}_{i=1}^{n}$ be a collection of consumer risk thresholds and $\mathbf{w} = (w_i)_{i=1}^{n}$ be their weights, such that $\tau_1 \leq \cdots \leq \tau_n$ (without loss of generality). The key idea is as follows: suppose that consumer index z defines the riskiest product in an optimal set of p products. Then all consumers $z, \ldots, n$ will be assigned to that product and the consumers $1, \ldots, z-1$ will not be. Therefore, if we knew the highest risk product in an optimal solution, we would be left with a smaller subproblem in which the goal is to optimally choose $p-1$ products for the first $z-1$ consumers. Our dynamic programming algorithm finds the optimal p' products for the first n' consumers for all values of $n' \leq n$ and $p' \leq p$.

More formally for any $n' \leq n$, let $S[n'] = \{\tau_i\}_{i=1}^{n'}$ and $\mathbf{w}[n']$ denote the n' lowest risk consumers and their weights. For any $n' \leq n$ and $p' \leq p$, let $T(n', p') = \mathcal{R}(S[n'], \mathbf{w}[n'], p')$ be the optimal weighted regret achievable in the subproblem using p' products for the first n' weighted consumers. We make use of the following recurrence relations:

Lemma 2.2. *The function T defined above satisfies the following properties:*

1. *For any $1 \leq n' \leq n$, we have $T(n', 0) = \sum_{i=1}^{n'} w_i r(\tau_i)$.*

2. For any $1 \le n' \le n$ and $0 \le p' \le p$, we have

$$T(n', p') = \min_{z \in \{p', \ldots, n'\}} \left(T(z-1, p'-1) + \sum_{i=z}^{n'} w_i \big(r(\tau_i) - r(\tau_z) \big) \right).$$

Proof. The first property is immediate, because when $p' = 0$, we do not choose any products, and there is only one valid solution that assigns all consumers to the zero-risk cash product. For this solution, the weighted regret is simply the sum of the weighted returns for each consumer's bespoke portfolio.

We now turn to proving the second property. For any consumer index $z \in [n']$, define $T(n', p', z)$ to be the optimal weighted regret for the first n' consumers using p' products subject to the constraint that the highest risk product has risk threshold set to τ_z. That is,

$$T(n', p', z) = \min_{\substack{\mathbf{c} = (c_1, \ldots, c_{p'}) \subset S[n'] \\ c_1, \ldots, c_{p'} \le \tau_z \\ c_{p'} = \tau_z}} \mathrm{R}_{S[n']}(\mathbf{c}, \mathbf{w}[n']).$$

The products achieving weighted regret $T(n', p', z)$ must choose $c_{p'} = \tau_z$ and choose $c_1, \ldots, c_{p'-1}$ to be optimal products for the consumers indexed $1, \ldots, z-1$, who are not served by the product $c_{p'}$ because it is too high risk. On the other hand, the weighted regret of consumers $z, \ldots, n'$ when assigned to a product with risk limit τ_z is given by $\sum\limits_{i=z}^{n'} w_i \cdot \big(r(\tau_i) - r(\tau_z) \big)$. Together, this implies that

$$T(n', p', z) = T(z-1, p'-1) + \sum_{i=z}^{n'} w_i \cdot \big(r(\tau_i) - r(\tau_z) \big).$$

On the other hand, for any n' and p', we have that $T(n', p') = \min\limits_{z \in [n']} T(n', p', z)$, because the optimal p' products for the first n' consumers has a largest risk threshold equal to some consumer risk threshold. Combining these equalities gives

$$\begin{aligned} T(n', p') &= \min_{z \in [n']} T(n', p', z) \\ &= \min_{z \in [n']} T(z-1, p'-1) + \sum_{i=z}^{n'} w_i \cdot \big(r(\tau_i) - r(\tau_z) \big), \end{aligned}$$

as required. □

The running time of our dynamic programming algorithm, which uses the above recurrence relations to solve all subproblems, is summarized below.

Theorem 2.3. *There exists an algorithm that, given a collection of consumers* $S = \{\tau_i\}_{i=1}^n$ *with weights* $\mathbf{w} = (w_i)_{i=1}^n$ *and a target number of products* p*, outputs a collection of products* $\mathbf{c} \in C_p(S)$ *with minimal weighted regret:* $\mathrm{R}_S(\mathbf{c}, \mathbf{w}) = \mathcal{R}(S, \mathbf{w}, p)$*. This algorithm runs in time* $O(n^2 p)$*.*

Proof. The algorithm computes a table containing the values $T(n', p')$ for all values of $n' \leq n$ and $p' \leq p$ using the above recurrence relations. The first column, when $p' = 0$, is computed using property 1 from Lemma 2.2, while the remaining columns are filled using property 2. By keeping track of the value of the index z achieving the minimum in each application of property 2, we can also reconstruct the optimal products for each subproblem.

To bound the running time, observe that the sums appearing in both properties can be computed in $O(1)$ time, after pre-computing all partial sums of the form $\sum_{i=1}^{n'} w_i r(\tau_i)$ and $\sum_{i=1}^{n'} w_i$ for $n' \leq n$. Computing these partial sums takes $O(n)$ time. With this, the first property can be evaluated in $O(1)$ time, and the second can be evaluated in $O(n)$ time by iterating over the values of z. In total, we can fill out all $O(np)$ table entries in $O(n^2 p)$ time. Reconstructing the optimal set of products takes $O(p)$ time given the completed table. □

2.4 Regret Minimization with Group Fairness

In this section, we consider the problem of choosing a set of p products in a way that satisfies *group fairness.* As noted before, groups can correspond to protected classes such as race, gender, age, and so on, or can simply correspond to partitions of risk tolerances (which may be correlated[1] with protected status). Importantly, in our notion of minmax fairness, our objective is to optimize the *welfare of the worst-off group* under our choice. Thus we are implicitly enforcing fairness to protect *all* groups from inequity, while still achieving good overall performance. This may not be appropriate in certain settings - for instance, if the designer wished to prioritize *specific* groups

[1] It is important to note that correlation is not causation. Indeed, while some research has found differences in risk aversion by race, other work has argued that these differences are driven by financial status or other circumstantial, rather than preferential, factors. See discussion in, for instance, Yao et al. (2005) and Fang et al. (2013).

to remedy historical inequity, they may prefer to solve a weighted regret-minimization using the dynamic program described in Section Section 2.3 and weight the groups they wish to protect.

Formally, we study the problem of choosing p products when the consumers can be partitioned into g groups and we want to optimize *minmax fairness* across groups, for both the *ex post* minmax fairness Program (2.6) and the *ex ante* minmax fairness Program (2.7).

We start the discussion of minmax fairness by showing a separation between the *ex post* objective in Program (2.6) and the *ex ante* objective in Program (2.7). More precisely, we show in Section 2.4.1 that the objective value of Program (2.6) can be $\Omega(g)$ times higher than the objective value of Program (2.7).

In the remainder of the section, we provide algorithms to solve Programs (2.6) and (2.7). In Section 2.4.2, we provide an algorithm that solves Program (2.7) to any desired additive approximation factor via no-regret dynamics. In Section 2.4.3, we provide a dynamic program approach that finds an approximately optimal solution to Program (2.6) when the number of groups g is small.

2.4.1 Separation Between Randomized and Deterministic Solutions

The following theorem shows a separation between the minmax (expected) regret achievable by deterministic versus randomized strategies (as per Programs (2.6) and (2.7)); in particular, the regret $\mathcal{R}_{\text{fair}}$ of the best deterministic strategy can be $\Omega(g)$ times worse than the regret $\widehat{\mathcal{R}}_{\text{fair}}$ of the best randomized strategy:

Theorem 2.4. *For any g and p, there exists an instance S consisting of g groups such that*

$$\frac{\widehat{\mathcal{R}}_{fair}\left(S,p\right)}{\mathcal{R}_{fair}\left(S,p\right)} \leq \frac{1}{p+1}\left\lceil\frac{p+1}{g}\right\rceil$$

Proof. Note there is a one-to-one relation (on some domain $[0,a]$ for risk thresholds) between any risk threshold τ and its corresponding return given by $r(\tau)$ by Equation (2.1) — we will therefore (only for simplicity of exposition) construct our instance by defining a set of returns instead of risk thresholds. Let $A \triangleq \{r, 2r, \ldots, (p+1)r\}$ be the set of all possible returns in our instance for some constant $r > 0$. We will take $r \equiv 1$ for simplicity but our proof extends to any $r > 0$.

Our instance construction is simple. If $g \leq p+1$, we partition A into g subsets of size at most $\lceil \frac{p+1}{g} \rceil$, and we let each group be defined as one of the partition elements. In this case, there will be one consumer for every return value in A. For e.g., if $p = 4$ and $g = 2$, we can define $G_1 = \{1, 2, 3\}$ and $G_2 = \{4, 5\}$. If $g > p+1$ (allowing consumers having the same return) we let each group be defined by a single return value in A. For e.g., if $p = 2$ and $g = 4$, we can define $G_1 = \{1\}$, $G_2 = \{2\}$, and $G_3 = G_4 = \{3\}$. To formalize this construction, define $s \triangleq \min\{g, p+1\}$ and let $\{P_i\}_{i=1}^{s}$ be a s-sized partition of R such that $\max_i |P_i| = \lceil \frac{p+1}{g} \rceil$. Instance S of size $n = \max\{g, p+1\}$ is defined as follows.

$$S = \{G_k\}_{k=1}^{g} \quad \text{where} \quad \forall\, k \in [g]: \quad G_k = \begin{cases} P_k & k \leq s \\ P_s & k > s \end{cases}$$

Let $A_p = \{B \subseteq A : |B| = p\}$ and observe that $|A_p| = p+1$. We have that

$$\begin{aligned} \mathcal{R}_{\text{fair}}\left(S, p\right) &= \min_{B \in A_p} \left\{ \max_{1 \leq k \leq g} \mathrm{R}_{G_k}(B) \right\} \\ &= \frac{1}{\max_k |G_k|} \\ &= \frac{1}{\lceil \frac{p+1}{g} \rceil} \end{aligned}$$

where the first equality follows from the definition of $\mathcal{R}_{\text{fair}}\left(S, p\right)$ in this specific instance that all consumer returns are specified by the set A of size $p+1$. The second follows from the fact that for any set of products $B \in A_p$, all groups G_k that have a consumer with return $A \setminus B$ will incur an average regret of $1/|G_k|$. Finally, the third holds because $\max_k |G_k| = \max_i |P_i| = \lceil \frac{p+1}{g} \rceil$. Next, by looking at the uniform distribution over A_p,

$$\begin{aligned} \widehat{\mathcal{R}}_{\text{fair}}\left(S, p\right) &\leq \max_{1 \leq k \leq g} \frac{1}{p+1} \sum_{B \in A_p} \mathrm{R}_{G_k}(B) \\ &= \max_{1 \leq k \leq g} \frac{1}{p+1} \sum_{r \in G_k} \frac{1}{|G_k|} = \frac{1}{p+1} \end{aligned}$$

where the inequality in the first line follows from the definition of $\widehat{\mathcal{R}}_{\text{fair}}\left(S, p\right)$. The equality in the second line follows from the fact that for every group k and every $r \in G_k$, there is one (and only one) set of products, namely $B = A \setminus \{r\}$, that makes G_k incur a regret of $1/|G_k|$. We therefore have that

$$\frac{\widehat{\mathcal{R}}_{\text{fair}}(S,p)}{\mathcal{R}_{\text{fair}}(S,p)} \leq \frac{1}{p+1}\left\lceil \frac{p+1}{g} \right\rceil .$$

□

In the following theorem, we show that for any instance of our problem, by allowing a multiplicative factor g blow-up in the target number of products p, the optimal deterministic minmax value will be at least as good as the randomized minmax value with p products.

Theorem 2.5. *We have that for any instance S consisting of g groups, and any p,*

$$\mathcal{R}_{fair}(S, gp) \leq \widehat{\mathcal{R}}_{fair}(S,p)$$

Proof. Fix any instance $S = \{G_k\}_{k=1}^{g}$ and any p. Let $\mathbf{c}_k^* \triangleq \underset{\mathbf{c}\in C_p(S)}{\operatorname{argmin}} \mathrm{R}_{G_k}(\mathbf{c})$ which is the best set of p products for group G_k. We have that

$$\begin{aligned}\mathcal{R}_{\text{fair}}(S, gp) &= \min_{\mathbf{c}\in C_{gp}(S)} \max_{1\leq k\leq g} \mathrm{R}_{G_k}(\mathbf{c})\\ &\leq \max_{1\leq k\leq g} \mathrm{R}_{G_k}(\cup_k \mathbf{c}_k^*)\\ &\leq \max_{1\leq k\leq g} \mathrm{R}_{G_k}(\mathbf{c}_k^*)\\ &\leq \max_{1\leq k\leq g} \mathbb{E}_{\mathbf{c}\sim\mathcal{C}}\left[\mathrm{R}_{G_k}(\mathbf{c})\right]\end{aligned}$$

where the last inequality follows from the definition of $\mathbf{c}_k^*$ and it holds for any distribution $\mathcal{C} \in \Delta(C_p(S))$. □

2.4.2 An Algorithm to Optimize for *ex ante* Fairness

In this section, we provide an algorithm to solve the *ex ante* Program (2.7). Recall that the optimization problem is given by

$$\widehat{\mathcal{R}}_{\text{fair}}(S,p) \triangleq \min_{\mathcal{C}\in\Delta(C_p(S))} \left\{ \max_{1\leq k\leq g} \mathbb{E}_{\mathbf{c}\sim\mathcal{C}}\left[\mathrm{R}_{G_k}(\mathbf{c})\right] \right\}.$$

Algorithm 2.1 relies on the dynamics introduced by Freund and Schapire (1996) to solve Program (2.7). The algorithm interprets this minmax optimization problem as a zero-sum game between the designer, who wants to pick products to minimize regret, and an adversary, whose goal is to pick the highest regret group. This game is played repeatedly, and agents update their strategies at every time step based on the history of play. In our setting, the adversary uses the multiplicative weights algorithm to assign weights to groups (as per Freund and Schapire (1996)) and the designer best-responds using the dynamic program from Section 2.3 to solve Equation (2.8) to choose an optimal set of products, noting that

$$\begin{aligned}\mathbb{E}_{k\sim D(t)}\left[\mathrm{R}_{G_k}(\mathbf{c})\right] &= \sum_{k\in[g]} D_k(t)\sum_{i\in G_k}\frac{R_{\tau_i}(\mathbf{c})}{|G_k|}\\ &= \sum_{i=1}^{n} R_{\tau_i}(\mathbf{c})\sum_{k\in[g]}\frac{D_k(t)\mathbb{I}\{i\in G_k\}}{|G_k|} = \mathrm{R}_S\left(\mathbf{c},\mathbf{w}(t)\right)\end{aligned}$$

where $w_i(t) \triangleq \sum\limits_{k\in[g]} \frac{D_k(t)}{|G_k|}\mathbb{I}\{i\in G_k\}$ denotes the weight assigned to agent i, at time step t.

Algorithm 2.1 2-Player Dynamics for the *Ex Ante* Minmax Problem

```
**Input:** $p$ target number of products, consumers $S$ partitioned in groups $G_1,\ldots,G_g$, and $T$.
**Initialization:** The no-regret player picks the uniform distribution $D(1)=\left(\frac{1}{g},\ldots,\frac{1}{g}\right)\in\Delta([g])$,
**for** $t=1,\ldots,T$ **do**
    The best-response player chooses $\mathbf{c}(t)=(c_1(t),\ldots,c_p(t))\in C_p(S)$ so as to solve
        $\mathbf{c}(t)=\underset{\mathbf{c}\in C_p(S)}{\operatorname{argmin}}\,\mathbb{E}_{k\sim D(t)}\left[\mathrm{R}_{G_k}(\mathbf{c})\right].$        (2.8)
    The no-regret player observes $u_k(t)=\mathrm{R}_{G_k}(\mathbf{c}(t))/B$ for all $k\in[g]$.
    The no-regret player sets $D(t+1)$ via multiplicative weight update with $\beta=\frac{1}{1+\sqrt{2\frac{\ln g}{T}}}\in(0,1)$,
as follows:
        $D_k(t+1)=\frac{D_k(t)\beta^{u_k(t)}}{\sum_{h=1}^{g}D_h(t)\beta^{u_h(t)}}\ \forall k\in[g],$
**Output:** $\mathcal{C}_T$: the uniform distribution over $\{\mathbf{c}(t)\}_{t=1}^{T}$.
```

Theorem 2.6 shows that the time-average of the strategy of the designer in Algorithm 2.1 is an approximate solution to minmax Program (2.7).

Theorem 2.6. *Suppose that for all $i\in[n]$, $r(\tau_i)\le B$. Then for all $T>0$, Algorithm 2.1 runs in time $O(Tn^2p)$ and the output distribution $\mathcal{C}_T$ satisfies*

$$\max_{k\in[g]}\mathbb{E}_{\mathbf{c}\sim\mathcal{C}_T}\left[\mathrm{R}_{G_k}(\mathbf{c})\right]\le\widehat{\mathcal{R}}_{fair}(S,p)+B\left(\sqrt{\frac{2\ln g}{T}}+\frac{\ln g}{T}\right). \tag{2.9}$$

Proof. Note that the action space of the designer $C_p(S)$ and the action set of the adversary $\{G_k\}_{k=1}^g$ are both finite, so our zero-sum game can be written in normal form. Further, $u_k(t) \in [0,1]$, noting that the return of each agent is in $[0, B]$ — so must be the average return of a group. Therefore, our minmax game fits the framework of Freund and Schapire (1996), and we have that

$$\begin{aligned}
&\max_{k\in[g]} \mathbb{E}_{\mathbf{c}\sim\mathcal{C}_T}\left[\mathrm{R}_{G_k}(\mathbf{c})\right] \\
&\leq \min_{\mathcal{C}\in\Delta(C_p(S))} \max_{\mathcal{G}\in\Delta([g])} \mathbb{E}_{k\sim\mathcal{G},\,\mathbf{c}\sim\mathcal{C}}\left[\mathrm{R}_{G_k}(\mathbf{c})\right] \\
&+ B\left(\sqrt{\frac{2\ln g}{T}} + \frac{\ln g}{T}\right).
\end{aligned} \tag{2.10}$$

The approximation statement is obtained by noting that for any distribution $\mathcal{G}$ over groups,

$$\max_{\mathcal{G}\in\Delta([g])} \mathbb{E}_{k\sim\mathcal{G},\,\mathbf{c}\sim\mathcal{C}}\left[\mathrm{R}_{G_k}(\mathbf{c})\right] = \max_{1\leq k\leq g} \mathbb{E}_{\mathbf{c}\sim\mathcal{C}}\left[\mathrm{R}_{G_k}(\mathbf{c})\right].$$

With respect to running time, note that at every time step $t \in [T]$, the algorithm first solves

$$\mathbf{c}(t) = \operatorname*{argmin}_{\mathbf{c}\in C_p(S)} \mathbb{E}_{k\sim D(t)}\left[\mathrm{R}_{G_k}(\mathbf{c})\right].$$

We note that this can be done in time $O(n^2p)$ as per Section 2.3, remembering that

$$\mathbb{E}_{k\sim D(t)}\left[\mathrm{R}_{G_k}(\mathbf{c})\right] = \mathrm{R}_S\left(\mathbf{c}, \mathbf{w}(t)\right)$$

i.e., Equation (2.8) is a weighted regret minimization problem. Then, the algorithm computes $u_k(t)$ for each group k, which can be done in time $O(n)$ (by making a single pass through each customer i and updating the corresponding u_k for $i \in G_k$). Finally, computing the g weights takes $O(g) \leq O(n)$ time. Therefore, each step takes time $O(n^2p)$, and there are T such steps, which concludes the proof. □

Importantly, Algorithm 2.1 outputs a *distribution* over p-sets of products; Theorem 2.6 shows that this distribution $\mathcal{C}_T$ approximates the *ex ante* minmax regret $\widehat{\mathcal{R}}_{\text{fair}}$. $\mathcal{C}_T$ can be used to construct a *deterministic* set of products with good regret guarantees: while each p-set in the support of $\mathcal{C}_T$ may have high group regret, the union of all such p-sets must perform at least as well as $\mathcal{C}_T$ and therefore meet benchmarks $\widehat{\mathcal{R}}_{\text{fair}}$ and $\mathcal{R}_{\text{fair}}$. However, this union may lead to an undesirable blow-up in the number of deployed products. The experiments in Section 2.7 show how to avoid such a blow-up in practice.

2.4.3 An *ex post* Minmax Fair Strategy for Few Groups

In this section, we present a dynamic programming algorithm to find p products that approximately optimize the *ex post* maximum regret across groups, as per Program (2.6). Recall, the optimization problem is given by

$$\mathcal{R}_{\text{fair}}\left(S,p\right) = \min_{\mathbf{c}\in C_p(S)} \left\{ \max_{1\le k\le g} \mathrm{R}_{G_k}(\mathbf{c}) \right\} \tag{Program (2.6)}$$

The algorithm aims to build a set of all g-tuples of average regrets $(\mathrm{R}_{G_1},\dots,\mathrm{R}_{G_g})$ that are simultaneously achievable for groups $G_1,\dots,G_g$. However, doing so may be computationally infeasible, as there can be as many regret tuples as there are ways of choosing p products among n consumer thresholds, i.e. $\binom{n}{p}$ of them. Instead, we discretize the set of possible regret values for each group and build a set that only contains rounded regret tuples via recursion over the number of products p. The resulting dynamic program runs efficiently when the number of groups g is a small constant and can guarantee an arbitrarily good additive approximation to the minmax regret.

Given bound B on the maximum group regret and a step size of α, we let

$$N^{(\alpha)} \triangleq \left\{ i\alpha :\ i \in \left\{0,\dots,\left\lceil \frac{B}{\alpha} \right\rceil \right\} \right\}$$

be a net of discretized regret values in $[0,B]$ with discretization size α. Given any regret R, we let $\mathrm{ceil}_\alpha(\mathrm{R})$ be the regret obtained by rounding R up to the closest higher regret value in $N^{(\alpha)}$. I.e., $\mathrm{ceil}_\alpha(\mathrm{R})$ is uniquely defined so as to satisfy $\mathrm{R} \le \mathrm{ceil}_\alpha(\mathrm{R}) < \mathrm{R}+\alpha$ and $\mathrm{ceil}_\alpha(\mathrm{R}) \in N^{(\alpha)}$. Our dynamic program implements the following recursive relationship:

$$\begin{aligned}\mathcal{F}^{(\alpha)}(n',p') \triangleq \Bigg\{ &\left(\mathrm{ceil}_\alpha \left(\mathrm{R}_{G_k} + \sum_{i=z}^{n'} \frac{\mathbb{1}\left\{i \in G_k\right\}}{|G_k|} \mathrm{R}_{\tau_i}(\tau_z) \right) \right)_{k=1}^{g} \\ &\text{s.t. } z \le n',\ \ (\mathrm{R}_{G_k})_{k=1}^{g} \in \mathcal{F}^{(\alpha)}(z-1,p'-1) \Bigg\}\end{aligned} \tag{2.11}$$

where for all $n' \le n$,

$$\mathcal{F}^{(\alpha)}(n',0) \triangleq \left\{ \left(\sum_{i=1}^{n'} \frac{\mathbb{1}\left\{i \in G_k\right\}}{|G_k|} r(\tau_i) \right)_{k=1}^{g} \right\}. \tag{2.12}$$

Note that $\mathcal{F}^{(\alpha)}(n',0)$ contains a single regret tuple, whose k-th coordinate is the weighted regret of agents $G_k \cap \{1,\dots,n'\}$ when using weight $1/|G_k|$ and offering no product.

Intuitively, $\mathcal{F}^{(\alpha)}(n', p')$ keeps track of a rounded up version of the feasible tuples of weighted group regrets when using weight $1/|G_k|$ in group G_k and when offering p' products and considering the regret of consumers 1 to n' only. The corresponding set of products used to construct the regret tuples in $\mathcal{F}^{(\alpha)}(n', p')$ can be kept in a hash table whose keys are the regret tuples in $\mathcal{F}^{(\alpha)}(n', p')$; we denote such a hash table by $\mathcal{I}^{(\alpha)}(n, p)$. While there can be several p'-tuples of products that lead to the same rounded regret tuple in $\mathcal{F}^{(\alpha)}(n', p')$, the dynamic program only stores one of them in the corresponding entry in the hash table at each time step. The program terminates after computing $\mathcal{F}^{(\alpha)}(n, p)$, $\mathcal{I}^{(\alpha)}(n, p)$, and outputting the product vector in $\mathcal{I}^{(\alpha)}(n, p)$ corresponding to the regret tuple with smallest regret for the worst-off group in $\mathcal{F}^{(\alpha)}(n, p)$. The sets $\mathcal{F}^{(\alpha)}(n, p)$, $\mathcal{I}^{(\alpha)}(n, p)$ satisfy the following guarantee:

Lemma 2.7. *Fix any $\alpha > 0$. Let $(\mathrm{R}_{G_1}, \ldots, \mathrm{R}_{G_g})$ be any regret tuple that can be achieved using p products. There exist consumer indices $(z_1, \ldots, z_p) \in \mathcal{I}^{(\alpha)}(n, p)$ such that the corresponding regret tuple $(\mathrm{R}^{(\alpha)}_{G_1}, \ldots, \mathrm{R}^{(\alpha)}_{G_g}) \in \mathcal{F}^{(\alpha)}(n, p)$ satisfies*

$$\mathrm{R}_{G_k}(\mathbf{c}) \leq \mathrm{R}^{(\alpha)}_{G_k} \leq \mathrm{R}_{G_k} + p\alpha, \ \forall k \in [g],$$

where $\mathbf{c} = (\tau_{z_1}, \ldots, \tau_{z_p})$.

Proof. We let $\mathcal{F}(n', p')$ be the set of weighted regret tuples that are achievable using p' thresholds when only agents 1 to n' are considered, and the regret of agents in group G_k is weighted by $1/|G_k|$ – importantly, this reweighting is independent of the choice of n' and computes the average regret of a group as if all of its consumers were present. The set $\mathcal{F}(n, p)$ contains all feasible tuples of average regret given groups $\{G_k\}_{k=1}^{g}$. Importantly, $\mathcal{F}(n, p)$ is different from $\mathcal{F}^{(\alpha)}(n, p)$: $\mathcal{F}(n, p)$ contains all achievable tuples of regret, even those that do not belong to the net $N^{(\alpha)}$; in turn $\mathcal{F}(n, p)$ may contain up to $\binom{n}{p}$ regret tuples. In comparison, $\mathcal{F}^{(\alpha)}(n, p)$ is a smaller set of size at most $\left(\left\lceil \frac{B}{\alpha} \right\rceil + 1\right)^g$ that only contains regret tuples with values in the net $N^{(\alpha)}$. $\mathcal{F}^{(\alpha)}(n, p)$ is built with the intent of approximating the true set of achievable regret tuples, $\mathcal{F}(n, p)$.

The proof idea is to show that $\mathcal{F}^{(\alpha)}(n, p)$ is a good approximation of $\mathcal{F}(n, p)$, as intended. More precisely, we want to show that for every feasible regret tuple $(\mathrm{R}_{G_k})_{k=1}^{g}$ in $\mathcal{F}(n, p)$, the discretized set $\mathcal{F}^{(\alpha)}(n,p)$ contains a regret tuple $\left(\mathrm{R}^{(\alpha)}_{G_k}\right)_{k=1}^{g}$ that is at most $p\alpha$ away from $(\mathrm{R}_{G_k})_{k=1}^{g}$; further, the corresponding product vector $\mathbf{c}^{(\alpha)} \in \mathcal{I}^{(\alpha)}(n,p)$ has true, *unrounded* regret also within $p\alpha$ of $(\mathrm{R}_{G_k})_{k=1}^{g}$.

We start by showing that $\mathcal{F}(n,p)$ obeys the following recursive relationship:

Claim 2.8. *$\mathcal{F}(n,p)$ satisfies*

$$\begin{aligned}\mathcal{F}(n',p') = \Bigg\{ &\left(\mathrm{R}_{G_k} + \sum_{i=z}^{n'} \frac{\mathbb{1}\left\{i \in G_k\right\}}{|G_k|}\,\mathrm{R}_{\tau_i}(\tau_z)\right)_{k=1}^{g} \\ &\quad s.t.\ z \le n',\ \ (\mathrm{R}_{G_k})_{k=1}^{g} \in \mathcal{F}(z-1,p'-1)\Bigg\}\end{aligned} \tag{2.13}$$

where

$$\mathcal{F}(n',0) \triangleq \mathcal{F}^{(\alpha)}(n',0) = \left\{\left(\sum_{i=1}^{n'} \frac{\mathbb{1}\left\{i \in G_k\right\}}{|G_k|} r(\tau_i)\right)_{k=1}^{g}\right\}. \tag{2.14}$$

Proof. When $p' = 0$, the result is immediate, noting that agents 1 to n' are assigned to the cash option c_0 with 0 return and incur regret $r(\tau_i)$ each. The total regret in group G_k, considering agents 1 to n' and reweighting regret by $1/|G_k|$, is given by

$$\sum_{i=1}^{n'} \frac{\mathbb{1}\left\{i \in G_k\right\}}{|G_k|} r(\tau_i).$$

Now, take $p' > 0$. Fix the highest offered product to be τ_z, corresponding to consumer z. First, agents $z,\ldots,n'$ are assigned to product τ_z corresponding to consumer z; the weighted regret incurred by these agents, limited to those in group G_k, is exactly

$$\sum_{i=z}^{n'} \frac{\mathbb{1}\left[i \in G_k\right]}{|G_k|}\left(r(\tau_i) - r(\tau_z)\right) = \sum_{i=z}^{n'} \frac{\mathbb{1}\left[i \in G_k\right]}{|G_k|}\,\mathrm{R}_{\tau_i}(\tau_z).$$

The remaining agents are 1 to $z-1$ and have $p'-1$ products available to them; hence, a regret tuple $(\mathrm{R}_{G_1},\ldots,\mathrm{R}_{G_g})$ can be feasibly incurred by these agents if and only if $(\mathrm{R}_{G_1},\ldots,\mathrm{R}_{G_g}) \in \mathcal{F}(z-1,p'-1)$, by definition of $\mathcal{F}(z-1,p'-1)$. To conclude the proof, it is enough to note that the total regret incurred by agents in group G_k is the sum of the regrets of agents $\{1,\ldots,z-1\} \cap G_k$ and the regret of agents $\{z,\ldots,n'\} \cap G_k$. □

We now show that for any $\alpha > 0$, $\mathcal{F}^{(\alpha)}(n,p)$ provides a $p\alpha$-additive approximation to the true set of possible regret tuples $\mathcal{F}(n,p)$:

Lemma 2.9. *For all $p \in \mathcal{N}$, for all $z \in [n]$, and for any regret tuple $(\mathrm{R}_{G_1}, \ldots, \mathrm{R}_{G_g}) \in \mathcal{F}(n,p)$, there exists a regret tuple $(\mathrm{R}^{(\alpha)}_{G_1}, \ldots, \mathrm{R}^{(\alpha)}_{G_g}) \in \mathcal{F}^{(\alpha)}(n,p)$ such that $\mathrm{R}^{(\alpha)}_{G_k} \leq \mathrm{R}_{G_k} + p\alpha$ for all $k \in [g]$.*

Proof. The proof follows by induction on p. At step $p' \leq p$, the induction hypothesis states that for all n', for any regret tuple $(\mathrm{R}_{G_1}, \ldots, \mathrm{R}_{G_g}) \in \mathcal{F}(n',p')$, there exists a regret tuple $(R^{(\alpha)}_{G_1}, \ldots, R^{(\alpha)}_{G_g}) \in \mathcal{F}^{(\alpha)}(n',p')$ such that $R^{(\alpha)}_{G_k} \leq \mathrm{R}_{G_k} + p'\alpha$ for all $k \in [g]$.

First, when $p' = 0$, the induction hypothesis holds immediately: by definition, $\mathcal{F}^{(\alpha)}(n',0) = \mathcal{F}(n',0)$. Now, suppose the induction hypothesis holds for $p' - 1$. Pick any regret tuple $(\mathrm{R}_{G_1}(n',p'), \ldots, \mathrm{R}_{G_g}(n',p')) \in \mathcal{F}(n',p')$; we will show that the induction hypothesis holds for this tuple. First, note that there exists z and $(\mathrm{R}_{G_1}(z-1,p'-1), \ldots, \mathrm{R}_{G_g}(z-1,p'-1)) \in \mathcal{F}(z-1,p'-1)$ such that

$$\mathrm{R}_{G_k}(n',p') = \mathrm{R}_{G_k}(z-1,p'-1) + \sum_{i=z}^{n'} \frac{\mathbb{1}\left[i \in G_k\right]}{|G_k|} \mathrm{R}_{\tau_i}(\tau_z) \;\forall k \in [g]$$

by definition of $\mathcal{F}(n',p')$. Further, by induction hypothesis, there exists a g-tuple of rounded regret $(\mathrm{R}^{(\alpha)}_{G_1}(z-1,p'-1), \ldots, \mathrm{R}^{(\alpha)}_{G_g}(z-1,p'-1))$ in $\mathcal{F}^{(\alpha)}(z-1,p'-1)$ such that

$$\mathrm{R}^{(\alpha)}_{G_k}(z-1,p'-1) \leq \mathrm{R}_{G_k}(z-1,p'-1) + (p'-1)\alpha \;\forall k \in [g].$$

Combining the above two equations, we get that

$$\begin{aligned}
&\mathrm{R}^{(\alpha)}_{G_k}(z-1,p'-1) + \sum_{i=z}^{n'} \frac{\mathbb{1}\left[i \in G_k\right]}{|G_k|} \mathrm{R}_{\tau_i}(\tau_z) \\
&\leq \mathrm{R}_{G_k}(z-1,p'-1) + \sum_{i=z}^{n'} \frac{\mathbb{1}\left[i \in G_k\right]}{|G_k|} \mathrm{R}_{\tau_i}(\tau_z) + (p'-1)\alpha \\
&= \mathrm{R}_{G_k}(n',p') + (p'-1)\alpha \;\forall k \in [g].
\end{aligned}$$

Now, let

$$\begin{aligned}
&\mathrm{R}^{(\alpha)}_{G_k}(n',p') \\
&= \mathrm{ceil}_\alpha \left(\mathrm{R}^{(\alpha)}_{G_k}(z-1,p'-1) + \sum_{i=z}^{n'} \frac{\mathbb{1}\left[i \in G_k\right]}{|G_k|} \mathrm{R}_{\tau_i}(\tau_z) \right).
\end{aligned} \tag{2.15}$$

First, $(\mathrm{R}^{(\alpha)}_{G_1}(n',p'),\dots,\mathrm{R}^{(\alpha)}_{G_g}(n',p'))$ is in $\mathcal{F}^{(\alpha)}(n',p')$ by definition. Second,

$$\begin{aligned}\mathrm{R}^{(\alpha)}_{G_k}(n',p') &= \mathrm{ceil}_\alpha\left(\mathrm{R}^{(\alpha)}_{G_k}(z-1,p'-1)+\sum_{i=z}^{n'}\frac{\mathbb{1}\left[i\in G_k\right]}{|G_k|}\,\mathrm{R}_{\tau_i}(\tau_z)\right)\\ &\le \mathrm{R}^{(\alpha)}_{G_k}(z-1,p'-1)+\sum_{i=z}^{n'}\frac{\mathbb{1}\left[i\in G_k\right]}{|G_k|}\,\mathrm{R}_{\tau_i}(\tau_z)+\alpha\\ &\le \mathrm{R}_{G_k}(n',p')+p'\alpha.\end{aligned}$$

This concludes the induction. □

To conclude the proof, let $(\mathrm{R}_{G_1},\dots,\mathrm{R}_{G_g})$ be a tuple of regret that can be achieved using p products. The tuple belongs to $\mathcal{F}(n,p)$, by definition of $\mathcal{F}(n,p)$. Therefore, by Lemma 2.9, there exists a regret tuple $(\mathrm{R}^{(\alpha)}_{G_1},\dots,\mathrm{R}^{(\alpha)}_{G_g})\in\mathcal{F}^{(\alpha)}$ such that

$$\mathrm{R}^{(\alpha)}_{G_k}\le\mathrm{R}_{G_k}+p\alpha\ \forall k\in[g].$$

Let $\mathbf{c}\triangleq\{c_1,\dots,c_p\}\in\mathcal{I}^{(\alpha)}(n,p)$ be the product vector that was used to construct regret tuple $\left(\mathrm{R}^{(\alpha)}_{G_k}\right)_{k=1}^{g}$. Since at each step p', the dynamic program rounds regret tuples to higher values, a simple induction shows that

$$\mathrm{R}_{G_k}(\mathbf{c})\le\mathrm{R}^{(\alpha)}_{G_k}\ \ \forall k\in[g].$$

Combining the two above equations, we get the result:

$$\mathrm{R}_{G_k}(\mathbf{c})\le\mathrm{R}^{(\alpha)}_{G_k}\le\mathrm{R}_{G_k}+p\alpha\ \forall k\in[g].$$

□

In particular, Lemma 2.7 implies that the dynamic program run with discretization parameter α approximately minimizes the maximum regret across groups (i.e., the optimal value of Program (2.6)) within an additive approximation factor of $p\alpha$. Letting $\alpha=\frac{\varepsilon}{p}$ yields an ε-approximation to the minmax regret.

The running time of our dynamic program is summarized below:

Theorem 2.10. *Fix any $\alpha>0$. There exists a dynamic programming algorithm that, given a collection of consumers $S=\{\tau_i\}_{i=1}^{n}$, groups $\{G_k\}_{k=1}^{g}$, and a target number of products p, computes*

$\mathcal{F}^{(\alpha)}(n,p)$ and $\mathcal{I}^{(\alpha)}(n,p)$ in time $O\left(n^2 p \left(\lceil \frac{B}{\alpha} \rceil + 1\right)^g\right)$.

When the desired accuracy is ε, the dynamic program uses $\alpha = \frac{\varepsilon}{p}$ and has running time $O\left(n^2 p \left(\left\lceil \frac{Bp}{\varepsilon} \right\rceil + 1\right)^g\right)$.

This running time is efficient when g is small, with a much better dependency in parameters p and n than the brute force approach that searches over all $\binom{n}{p}$ ways of picking p products.

Proof. Note that each set $\mathcal{F}^{(\alpha)}(n',p')$ and $\mathcal{I}^{(\alpha)}(n',p')$ built by the dynamic program has size at most $\left(\lceil \frac{B}{\alpha} \rceil + 1\right)^g$, as $\mathcal{F}^{(\alpha)}(n',p')$ only contains regret values in $N^{(\alpha)}$ by construction and $\mathcal{I}^{(\alpha)}(n',p')$ contains one product tuple per regret tuple in $\mathcal{F}^{(\alpha)}(n',p')$.

Each sum used in the dynamic program can be computed in time $O(1)$, given that $\sum_{i=1}^{n'} \frac{\mathbb{1}\{i \in G_k\}}{|G_k|} r(\tau_i)$ and $\sum_{i=1}^{n'} \frac{\mathbb{1}\{i \in G_k\}}{|G_k|}$ have been precomputed for all $n' \in [n], k \in [g]$ and stored in a hash table. The precomputation and storage of these partial sums can be done in $O(gn)$ time.

Now, each time step of the dynamic program corresponds to the p'-th product with $p' \leq p$. In each time step p', we construct $O(n)$ sets $\mathcal{F}^{(\alpha)}\,(n',p')$, one for each value of n'. For each value of n', the dynamic program searches over i) $z \in [n]$ and ii) at most $\left(\lceil \frac{B}{\alpha} \rceil + 1\right)^g$ tuples of regret in $\mathcal{F}^{(\alpha)}(z, p'-1)$. Therefore, building $\mathcal{F}^{(\alpha)}(n,p)$ can be done in time $O\left(n^2 p \left(\lceil \frac{B}{\alpha} \rceil + 1\right)^g\right)$

Finally, finding the tuple with the smallest maximum regret in $\mathcal{F}^{(\alpha)}(n,p)$ requires searching over at most $\left(\lceil \frac{R}{\alpha} \rceil + 1\right)^g$ product tuples in $\mathcal{I}^{(\alpha)}(n,p)$. Therefore, running the dynamic program requires time $O\left(n^2 p \left(\lceil \frac{B}{\alpha} \rceil + 1\right)^g\right)$. □

2.4.4 Optimizing for Population vs. Least Well-Off Group: An Example

In this section, we show that optimizing for population regret may lead to arbitrarily bad maximum group regret, and optimizing for maximum group regret may lead to arbitrarily bad population regret. To do so, we consider the following example: there is a set S of n consumers, divided into two groups G_1 and G_2. We let $|G_1| = 1$ and $|G_2| = n-1$ and assume the single consumer in group G_1 has risk threshold τ_1, and the $n-1$ consumers in group G_2 all have the same risk threshold

$\tau_2 > \tau_1$. We let $r_1 < r_2$ be the returns corresponding to risk thresholds τ_1, τ_2. Let $p = 1$, i.e. the designer can pick only one product; either $\mathbf{c} = \tau_1$ or $\mathbf{c} = \tau_2$.

When picking $\mathbf{c} = \tau_1$, we have that the average group and population regrets are given by:

$$\mathrm{R}_{G_1}(\tau_1) = 0,\ \ \mathrm{R}_{G_2}(\tau_1) = r_2 - r_1,\ \ \mathrm{R}_S(\tau_1) = \frac{(n-1)(r_2 - r_1)}{n}.$$

When picking $\mathbf{c} = \tau_2$ instead, we have

$$\mathrm{R}_{G_1}(\tau_2) = r_1,\ \ \mathrm{R}_{G_2}(\tau_2) = 0,\ \ \mathrm{R}_S(\tau_2) = \frac{r_1}{n}.$$

Suppose $0 < r_2 - r_1 < r_1$ and $n - 1 > \frac{r_1}{r_2 - r_1}$. Then, the optimal product to optimize for maximum group regret is $\mathbf{c}^{grp} = \tau_1$, and the optimal product to optimize for population regret is $\mathbf{c}^{pop} = \tau_2$. Then, we have that

1. The ratio of population regret using $\mathbf{c}^{grp}$ over that of $\mathbf{c}^{pop}$ is given by:

$$\frac{\mathrm{R}_S(\mathbf{c}^{grp})}{\mathrm{R}_S(\mathbf{c}^{pop})} = \frac{(n-1)(r_2 - r_1)/n}{r_1/n} = \frac{(n-1)(r_2 - r_1)}{r_1}.$$

   This ratio can be made arbitrarily large by letting $n \to +\infty$, at $\frac{r_2 - r_1}{r_1}$ constant.

2. The ratio of maximum group regret using $\mathbf{c}^{pop}$ over that of $\mathbf{c}^{grp}$ is given by:

$$\frac{\max\left(\mathrm{R}_{G_1}(\mathbf{c}^{pop}),\ \ \mathrm{R}_{G_2}(\mathbf{c}^{pop})\right)}{\max\left(\mathrm{R}_{G_1}(\mathbf{c}^{grp}),\ \ \mathrm{R}_{G_2}(\mathbf{c}^{grp})\right)} = \frac{r_1}{r_2 - r_1}.$$

   This ratio can be made arbitrarily large by letting $r_2 - r_1 \to 0$ at r_1 constant.

2.5 Approximate Population Regret Minimization via Greedy Algorithm

Recall that given $S = \{\tau_i\}_{i=1}^n$ and set $\mathbf{c} \subseteq S$ of products, the population regret of S is given by

$$\mathrm{R}_S(\mathbf{c}) = \frac{1}{n}\sum_{i=1}^n \mathrm{R}_{\tau_i}(\mathbf{c}) = \frac{1}{n}\sum_{i=1}^n \left(r(\tau_i) - f_{\tau_i}(\mathbf{c})\right) = \frac{1}{n}\sum_{i=1}^n r(\tau_i) - f_S(\mathbf{c})$$

where for any τ, $f_\tau(\mathbf{c}) \triangleq \max\limits_{c_j \leq \tau} r(c_j)$, and

$$f_S : 2^S \to \mathbb{R}_{\geq 0}, \quad f_S(\mathbf{c}) \triangleq \frac{1}{n}\sum_{i=1}^n f_{\tau_i}(\mathbf{c}).$$

First, note that when no product is offered, consumers pick the cash option c_0 and get return $r(c_0) = 0$:

Fact 2.11 (Centering)**.** *For any S, $f_S(\emptyset) = 0$.*

Second, f_S is immediately monotone non-decreasing, as consumers deviate to an additional product only when they get higher return from doing so:

Fact 2.12 (Monotonicity)**.** *For any S, if $\mathbf{c} \subseteq \mathbf{d}$, then $f_S(\mathbf{c}) \leq f_S(\mathbf{d})$.*

Finally, f_S is submodular:

Claim 2.13 (Submodularity)**.** *For any S, f_S is submodular.*

Proof. We first show that for every i, $f_{\tau_i}(\cdot)$ is submodular: for any $\mathbf{c}, \mathbf{d} \subseteq S$, we have that

$$f_{\tau_i}(\mathbf{c} \cup \mathbf{d}) + f_{\tau_i}(\mathbf{c} \cap \mathbf{d}) \leq f_{\tau_i}(\mathbf{c}) + f_{\tau_i}(\mathbf{d}).$$

If $\mathbf{c} = \emptyset$ or $\mathbf{d} = \emptyset$, the claim trivially holds. So assume $\mathbf{c}, \mathbf{d} \neq \emptyset$. Let c be such that $f_{\tau_i}(\mathbf{c} \cup \mathbf{d}) = r(c)$. If $c = c_0 = 0$, the claim holds because all four terms above will be zero. If $c \in \mathbf{c}$, the claim holds because $f_{\tau_i}(\mathbf{c} \cup \mathbf{d}) = f_{\tau_i}(\mathbf{c})$ and $f_{\tau_i}(\mathbf{d}) \geq f_{\tau_i}(\mathbf{c} \cap \mathbf{d})$ by Fact 2.12. The same argument holds when $c \in \mathbf{d}$ by symmetry. The proof is complete by noting that any simple average of submodular functions (in general, any linear combination with non-negative coefficients) is submodular. □

Remark 2.14. *Claim 2.13 extends to any weighted set of consumers: fix any set $S = \{\tau_i\}_{i=1}^n$ of consumers and any nonnegative weight vector $\mathbf{w} \in \mathbb{R}^n_{\geq 0}$. Then the function $f_S : 2^S \to \mathbb{R}$ given by $f_S(\mathbf{c}) = \sum_{i=1}^n w_i f_{\tau_i}(\mathbf{c})$ is submodular.*

We therefore have that for any S and any target number of products p,

$$\min_{\mathbf{c} \in C_p(S)} \mathrm{R}_S(\mathbf{c}) = \frac{1}{n} \sum_{i=1}^{n} r(\tau_i) - \max_{\mathbf{c} \subseteq S,\, |\mathbf{c}| = p} f_S(\mathbf{c})$$

where by Facts 2.11 and 2.12 and Claim 2.13, the maximization problem on the right hand side is the maximization of a nonnegative monotone submodular function with a cardinality constraint. Using a greedy algorithm (that runs in $O(np)$ time), we get p products represented by $\mathbf{c}^{grd}$ such that

$$R_S(\mathbf{c}^{grd}) \leq \frac{1}{n}\sum_{i=1}^{n} r(\tau_i) - \left(1 - e^{-1}\right) \cdot \max_{\mathbf{c}\subseteq S,\, |\mathbf{c}|=p} f_S(\mathbf{c}).$$

2.6 A More General Regret Notion

In this section we propose a family of regret functions parametrized by a positive real number that captures the regret notion we have been using throughout the paper as a special case. We will show how minor tweaks allow us to extend the dynamic program for whole population regret minization of Section 2.3, as well as the no-regret dynamics for *ex ante* fair regret minimization of Section 2.4.2, to this more general notion of regret. For a given consumer with risk threshold τ, and for any $\alpha \in (0, \infty]$, the regret of the consumer when assigned to a single product with risk threshold c is defined as follows:

$$\mathrm{R}_\tau^\alpha(c) = \begin{cases} r(\tau) - r(c) & c \leq \tau \\ \alpha\,(c - \tau) & c > \tau \end{cases} \tag{2.16}$$

When offering more than one product, say p products represented by $\mathbf{c} = (c_1, c_2, \ldots, c_p)$, the regret of a consumer with risk threshold τ is the best regret she can get using these products. Concretely,

$$\mathrm{R}_\tau^\alpha(\mathbf{c}) = \min_{1\leq i\leq p} \mathrm{R}_\tau^\alpha(c_i) \tag{2.17}$$

We note that our previous regret notion is recovered by setting $\alpha = \infty$. When $\alpha \neq \infty$, a consumer may be assigned to a product with higher risk threshold, and the consumer's regret is then measured by the difference between her desired risk and the product's risk, scaled by a factor of α. We note that in practice, consumers may have different behavior as to whether they are willing to accept a higher risk product, i.e. different consumers may have different values of α; in this section, we use a unique value of α for all consumers for simplicity of exposition and note that our insights generalize to different consumers having different α's. The regret of a set of consumers $S = \{\tau_i\}_{i=1}^n$ is defined as the average regret of consumers, as before. In other words,

$$\mathrm{R}_S^\alpha(\mathbf{c}) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{R}_{\tau_i}^\alpha(\mathbf{c}) \tag{2.18}$$

We first show in Lemma 2.16 how we can find one single product to minimize the regret of a set of consumers represented by a set $S = \{\tau_i\}_{i=1}^n$. Note this general formulation of regret will allow choosing products that do not necessarily fall onto the consumer risk thresholds. In Lemma 2.16 we will show that given access to the derivative function $r'(\tau)$, the problem $\min_{c\in\mathbb{R}_+} \mathrm{R}_S^\alpha(c)$ can be reduced to an optimization problem that can be solved exactly in $O(n)$ time. Before that, we first observe the following property of function $r(\tau)$ Equation (2.1) which will be used in Lemma 2.16:

Claim 2.15. *The function r is concave.*

Proof. Let X be the vector of random variables representing m assets and recall μ is the mean of X and Σ is its covariance matrix. Fix $\tau_1, \tau_2 \geq 0$, and $\beta \in (0,1)$. For $i \in \{1,2\}$, let $\mathbf{w}_i$ be an optimal solution to the optimization problem for $r(\tau_i)$, i.e., $\mathbf{w}_i$ is such that $r(\tau_i) = \mathbf{w}_i^\top \mu$. We want to show that

$$r\left(\beta\tau_1 + (1-\beta)\,\tau_2\right) \geq \beta r(\tau_1) + (1-\beta) r(\tau_2) = \left(\beta\mathbf{w}_1 + (1-\beta)\,\mathbf{w}_2\right)^\top \mu \tag{2.19}$$

So all we need to show is that $\mathbf{w} \triangleq \beta\mathbf{w}_1 + (1-\beta)\mathbf{w}_2$ is feasible in the corresponding optimization problem for $r\left(\beta\tau_1 + (1-\beta)\,\tau_2\right)$. Then, by definition, Equation (2.19) holds. We have that

$$\mathbb{1}^\top \mathbf{w} = \beta(\mathbb{1}^\top \mathbf{w}_1) + (1-\beta)(\mathbb{1}^\top \mathbf{w}_2) = 1$$

and

$$\begin{aligned}
\mathbf{w}^\top \Sigma\, \mathbf{w} &= \beta^2 \mathbf{w}_1^\top \Sigma\, \mathbf{w}_1 + (1-\beta)^2 \mathbf{w}_2^\top \Sigma\, \mathbf{w}_2 + 2\beta(1-\beta)\mathbf{w}_1^\top \Sigma\, \mathbf{w}_2 \\
&\leq \beta^2\tau_1^2 + (1-\beta)^2\tau_2^2 + 2\beta(1-\beta)\cdot Cov\left(\mathbf{w}_1^\top X, \mathbf{w}_2^\top X\right) \\
&\leq \beta^2\tau_1^2 + (1-\beta)^2\tau_2^2 + 2\beta(1-\beta)\cdot \sqrt{Var(\mathbf{w}_1^\top X) Var(\mathbf{w}_2^\top X)} \\
&= \beta^2\tau_1^2 + (1-\beta)^2\tau_2^2 + 2\beta(1-\beta)\cdot \sqrt{(\mathbf{w}_1^\top \Sigma\, \mathbf{w}_1)(\mathbf{w}_2^\top \Sigma\, \mathbf{w}_2)} \\
&\leq \beta^2\tau_1^2 + (1-\beta)^2\tau_2^2 + 2\beta(1-\beta)\tau_1\tau_2 \\
&= \left(\beta\tau_1 + (1-\beta)\tau_2\right)^2
\end{aligned}$$

where the second inequality follows from Cauchy-Schwarz inequality.

Note also that $Cov\left(\mathbf{w}_1^\top X, \mathbf{w}_2^\top X\right) = \mathbf{w}_1^\top \Sigma\, \mathbf{w}_2$ and $Var(\mathbf{w}_i^\top X) = \mathbf{w}_i^\top \Sigma\, \mathbf{w}_i$, for $i \in \{1,2\}$. □

Lemma 2.16. *Let $S = \{\tau_i\}_{i=1}^n$ where $\tau_1 \le \tau_2 \le \ldots \le \tau_n$. We have that $\mathrm{R}_S^\alpha(c)$ is a convex function and*

$$\begin{aligned}&\min_{c \in \mathbb{R}_{\ge 0}} \mathrm{R}_S^\alpha(c) \\ &= \min\left\{\min_{1\le i\le n} \mathrm{R}_S^\alpha(\tau_i), \min_{1\le i\le n-1} \mathrm{R}_S^\alpha(c_i)\cdot \mathbb{I}_\infty\{\tau_i < c_i < \tau_{i+1}\}\right\}\end{aligned} \tag{2.20}$$

where for every $1 \le i \le n-1$, $c_i = (r')^{-1}\left(\frac{\alpha i}{n-i}\right)$ ($c_i = \infty$ if $(r')^{-1}\left(\frac{\alpha i}{n-i}\right)$ does not exist) and

$$\mathbb{I}_\infty\{\tau_i < c_i < \tau_{i+1}\} \triangleq \begin{cases} 1 & \tau_i < c_i < \tau_{i+1} \\ \infty & \textit{otherwise}\end{cases}$$

Proof. First observe that Claim 2.15 implies for every τ, $\mathrm{R}_\tau^\alpha(c)$ defined in Equation (2.16) is convex. Hence, $\mathrm{R}_S^\alpha(c)$ is convex because it is an average of convex functions. We have that for a single product c,

$$\mathrm{R}_S^\alpha(c) = \frac{1}{n}\sum_{j=1}^n \{(r(\tau_j) - r(c))\,\mathbb{I}\{c \le \tau_j\} + \alpha\,(c - \tau_j)\,\mathbb{I}\{c > \tau_j\}\}$$

Note that $\mathrm{R}_S^\alpha(\tau_1) \le \mathrm{R}_S^\alpha(c)$ for every $c < \tau_1$ and $\mathrm{R}_S^\alpha(\tau_n) \le \mathrm{R}_S^\alpha(c)$ for every $c > \tau_n$. We can therefore focus on the domain $[\tau_1, \tau_n]$ to find the minimum. The function $\mathrm{R}_S^\alpha(c)$ is differentiable everywhere except for the points given by consumers' risk thresholds: $S = \{\tau_i\}_{i=1}^n$. This justifies the first term appearing in the $\min\{\cdot,\cdot\}$ term of Equation (2.20). For every $1 \le i \le n-1$, the function $\mathrm{R}_S^\alpha(c)$ on domain (τ_i, τ_{i+1}) is differentiable and can be written as:

$$\mathrm{R}_S^\alpha(c) = \frac{1}{n}\left[\alpha \sum_{j \le i}(c - \tau_j) + \sum_{j \ge i+1}(r(\tau_j) - r(c))\right]$$

The minimum of $\mathrm{R}_S^\alpha(c)$ on domain (τ_i, τ_{i+1}) is achieved on points c where

$$\frac{d}{dc}\mathrm{R}_S^\alpha(c) = \frac{1}{n}\left[\alpha i - r'(c)(n-i)\right] = 0 \quad\Longrightarrow\quad r'(c_i) = \frac{\alpha i}{n-i}$$

We note that $\mathrm{R}_S^\alpha(c)$ is a convex function by the first part of this Lemma implying that c_i (if belongs to the domain (τ_i, τ_{i+1})) is a local *minimum*. This justifies the second term appearing in the $\min\{\cdot,\cdot\}$ term of Equation (2.20) and completes the proof. □

Remark 2.17. *Given any set of weights $\mathbf{w} \in \mathbb{R}_{\ge 0}^n$ over consumers, Lemma 2.16 can be easily extended to optimizing the weighted regret of a set of consumers given by:*

$$\mathrm{R}_S^\alpha(\mathbf{c}, \mathbf{w}) = \sum_{i=1}^n w_i \, \mathrm{R}_{\tau_i}^\alpha(\mathbf{c})$$

In fact, for any S and $\mathbf{w}$, we have that $\mathrm{R}_S^\alpha(c, \mathbf{w})$ is a convex function and

$$\begin{aligned}
&\min_{c \in \mathbb{R}_{\geq 0}} \mathrm{R}_S^\alpha(c, \mathbf{w}) \\
&= \min \left\{ \min_{1 \leq i \leq n} \mathrm{R}_S^\alpha(\tau_i, \mathbf{w}), \min_{1 \leq i \leq n-1} \mathrm{R}_S^\alpha(c_i, \mathbf{w}) \cdot \mathbb{I}_\infty\{\tau_i < c_i < \tau_{i+1}\} \right\}
\end{aligned} \tag{2.21}$$

where for every $1 \leq i \leq n-1$,

$$c_i = (r')^{-1} \left(\frac{\alpha \sum\limits_{j \leq i} w_j}{\sum\limits_{j \geq i+1} w_j} \right)$$

(we take ($c_i = \infty$ if the right hand side does not exist) and

$$\mathbb{I}_\infty\{\tau_i < c_i < \tau_{i+1}\} \triangleq \begin{cases} 1 & \tau_i < c_i < \tau_{i+1} \\ \infty & \textit{otherwise} \end{cases}$$

We now provide the idea behind a dynamic programming approach for choosing p products that minimize the weighted regret of a population $S = \{\tau_i\}_{i=1}^n$. The approach relies on the simple observation that there exists an optimal solution such that if the consumers in set $S(p')$ are assigned to the p'-th product $c_{p'}$, $c_{p'} \in \operatorname{argmin}_{c \in \mathbb{R}^+} \mathrm{R}_{S(p')}(c)$ (for a single product c). On the one hand, for all p' and given $S(p')$, picking $c_{p'}$ in such a manner provides a lower bound on the achievable average regret. On the other hand, $c_{p'}$ yields $S(p')$ as a set of consumers assigned to $c_{p'}$ in an optimal solution. Indeed, if consumers in $S(p')$ strictly preferred picking a different product, this could only be because they would get strictly better regret from doing so: by picking a different product, they would then decrease the population regret below our lower bound, which is a contradiction..

Therefore, to find an optimal choice of products, it suffices to i) correctly guess which subset $S(p')$ of consumers are assigned to the p'-th product, then ii) optimize the choice of product for $S(p')$, which can be done using Lemma 2.16. Noting that $S(p')$ is an interval for all p', $S(p')$ is entirely characterized by z, the first agent assigned to $c_{p'}$, and n', the last agent assigned to $c_{p'}$. In turn, as in Section 2.3, our dynamic program can be characterized by a recursive relationship of the form

$$T(n', p') = \min_{z \in \{1,\ldots,n'\}} \left(T(z-1, p'-1) + \min_{c \in \mathbb{R}^+} \sum_{i=z}^{n'} w_i \, \mathrm{R}_{\tau_i}^{\alpha}(c) \right) , \tag{2.22}$$

where $T(n', p')$ represents the minimum weighted regret that can be achieved by providing p' products to consumers 1 to n'. Our dynamic program will implement this recursive relationship.

The results of Section 2.4.2 immediately extends to this more general regret notion, given that dynamic Equation (2.22) can be used as an optimization oracle for the problem solved by the best-response player in Algorithm 2.1.

2.7 Experiments

In this section, we present experiments that complement our theoretical results.

2.7.1 Data

The underlying data for our experiments consists of time series of daily closing returns for 50 publicly traded U.S. equities over a 15-year period beginning in 2005 and ending in 2020; the equities chosen were those with the highest liquidity during this period. From these time series, we extracted the average daily returns and covariance matrix, which we then annualized by the standard practice of multiplying returns and covariances by 252, the number of trading days in a calendar year. The mean annualized return across the 50 stocks is 0.13, and all but one are positive due to the long time period spanned by the data. The correlation between returns and risk (standard deviation of returns) is 0.29 and significant at $P = 0.04$. The annualized returns and covariances are then the basis for the computation of optimal portfolios given a specified risk limit τ as per Section 2.2.[2]

In Figure 2.1, we show a scatter plot of risk vs. returns for these 50 stocks. For sufficiently small risk values, the optimal portfolio has almost all of its weight in cash, since all of the equities have higher risk and insufficient independence. At intermediate values of risk, the optimal portfolio

[2] For ease of understanding, here we consider a restriction of the Markowitz portfolio model of Markowitz (1952) and Equation (2.1) in which short sales are not allowed, i.e. the weight assigned to each asset must be non-negative.

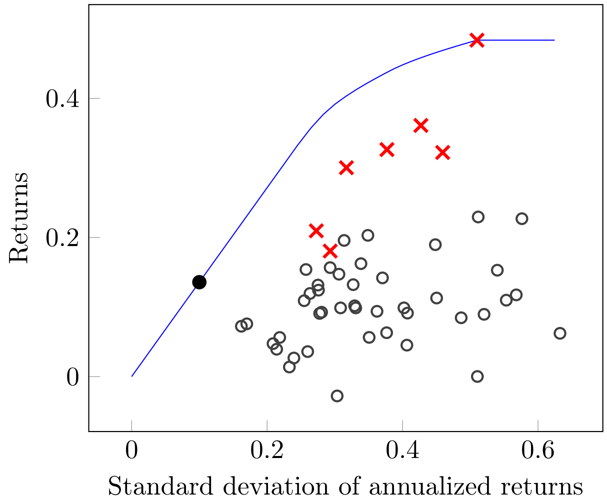
concentrates its weight on just the seven stocks highlighted in red in Figure 2.1 and listed in the table in Figure 2.1. This figure also plots the optimal risk-return frontier, which generally lies to the northwest (lower risk and higher return) of the stocks themselves, due to the optimization's exploitation of independence. The black dot highlights the optimal return for risk tolerance 0.1, for which we show the optimal portfolio weights in Figure 2.1. Note that once the risk tolerance reaches that of the single stock with highest return (the red point lying on the optimal frontier, representing Netflix), the frontier becomes flat, since at that point the optimal portfolio is one fully invested in this single stock.

2.7.2 Algorithms and Benchmarks

We consider both population and group regret and a number of algorithms: the integer linear program (ILP) for optimizing group regret; an implementation of the no-regret dynamics (NR) for group regret described in Algorithm 2.1; two strategies for "sparsifying" NR (described below); the dynamic program (DP) of Section 2.3 for optimizing population regret; and a greedy heuristic, which iteratively chooses the product that reduces population regret the most.The average population return $f_S(\mathbf{c}) = \frac{1}{n}\sum_i \max_{c_j \leq \tau_i} r(c_j)$ is submodular, and thus the greedy algorithm, which has the advantage of $O(np)$ running time compared to the $O(n^2p)$ of the DP, also enjoys the standard approximate submodular performance guarantees.

Note that the NR algorithm (Algorithm Algorithm 2.1) outputs a *distribution* over sets of p products; a natural way of extracting a fixed set of products is to take the union of the support, but in principle this could lead to far more than p products. We propose two sparsification techniques that extract p products from the support, both of which try to remove redundant products (i.e., products with very similar returns). Our first strategy, which we refer to as NR-Sparse-G, is a greedy algorithm that repeatedly finds the pair of products with closest returns and removes the higher risk one until only p products remain. The second algorithm is based on a version of one-dimensional asymmetric k-center clustering, which we refer to as NR-Sparse-C. NR-Sparse-C optimally selects p

products from the NR support such that the maximum difference in return from any product in the NR support to the closest selected product with lower risk is minimized. Details for NR-Sparse-C are given in the extended version.



Company	Allocation
Apple	10.6%
Amazon	4.3%
Gilead Sciences	2.3%
Monster Beverage	8.3%
Netflix	8.7%
NVIDIA	0.3%
Ross Stores	6%
Cash reserves	59.5%

Risk vs. Returns: Scatterplot and Optimal Frontier.

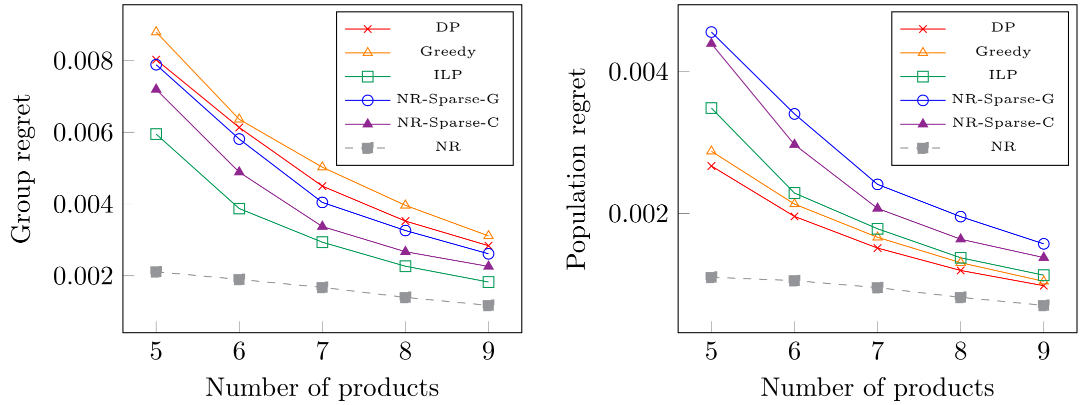
Optimal Portfolio for Risk = 0.1.

Figure 2.1: Asset Risks and Returns, Optimal Frontier and Portfolio Weights

2.7.3 Experimental Design and Results

Our first experiment compares the population and minimax group regrets of the algorithms we have discussed on two different population distributions as we vary the number of products, p. Both population distributions have $n = 100$ consumers divided into $g = 5$ groups and the consumers belonging to group $i \in \{1, \ldots, 5\}$ have risk thresholds drawn from a Gaussian distribution with mean $\mu_i = i/10$ and standard deviation $\sigma_i = 0.01$. Any negative risk thresholds are set to 0. In the first population distribution, which we call the unbalanced distribution, each consumer belongs to group 2 with probability 5/9, and each of the remaining groups with probability 1/9. In the second population distribution, called the balanced distribution, consumers belong to each group with equal probability. We compare the NR algorithm run for $T = 200$ iterations, both NR sparsification algorithms, the dynamic program from Section 2.3, and the greedy algorithm. Results are averaged over 100 populations sampled from the population distribution. Further experimental details are given in the extended version.

Unbalanced Distribution. Figure 2.2 shows the performance of each algorithm averaged over 100 samples from the unbalanced population distribution for $p \in \{5, \dots, 9\}$ products. The NR algorithm achieves the lowest population and minimax group regret. However, we are plotting the regret of the union of the products in the support of the output distribution, which contains many more than p products: for $p \in \{5, 6, 7, 8, 9\}$, NR used an average of $\{10.31, 10.3, 10.75, 11.62, 12.53\}$ products on the unbalanced distribution, respectively. The ILP algorithm achieves the optimal minimax group regret with exactly p products, but is significantly more computationally expensive than the other methods; for $p = 5$, the average solve time for the ILP was 17.64 seconds, while the dynamic program ran in just 0.0008 seconds on average. For larger sets of consumers, solving the ILP becomes intractable. The two NR sparsification algorithms have the next best minimax group regret, and for $p \in \{5, 6, 7\}$, using NR-Sparse-C with $p+1$ products results in minimax group regret lower than the ILP solution with p products. Finally, the dynamic program and greedy algorithm for minimizing population regret have the worst minimax group regret across all values of p. As expected, the dynamic program achieves the optimal average population regret, followed closely by the greedy algorithm.

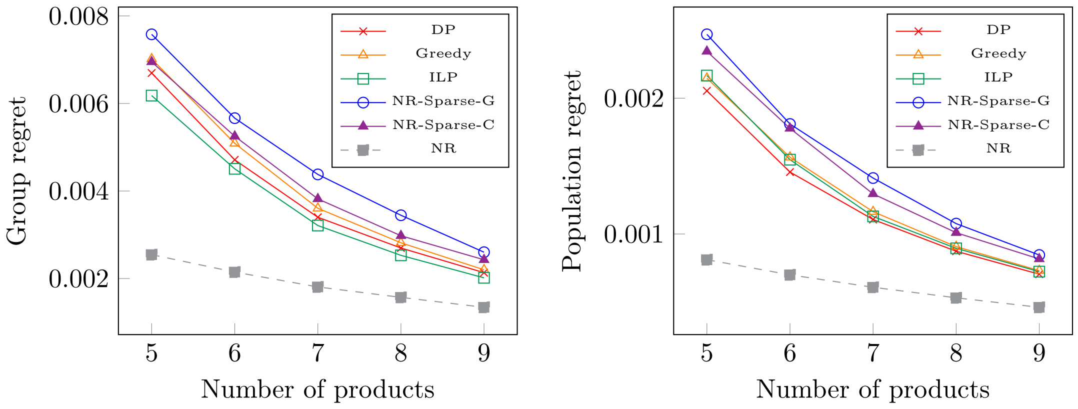


Group Regret

Population Regret

Figure 2.2: Algorithm Performance on the Unbalanced Population Distribution

Balanced Distribution. Figure 2.3 plots the same performance measures for the balanced distribution. As in the unbalanced distribution, the NR algorithm achieves the lowest population and minimax group regret, but this is again due to it using more than p products. Unlike the unbalanced distribution, there is less variation in the performance of the algorithms, and our NR sparsification strategies no longer outperform the dynamic program and greedy algorithms on the minimax group regret. One possible explanation is that on the balanced distribution, solutions that minimize the overall population regret also do a reasonably good job at minimizing the minimax group regret, making this an easy case for the DP algorithm. Consider the case where $p = 5 = g$. Intuitively, since the groups form mostly non-overlapping risk intervals, the optimal minimax group regret solution will tend to select one product from each group. On the other hand, since the groups are all equally sized, minimizing the average population regret also requires that we choose products to serve each group (by symmetry, there is no reason for the algorithm to focus on any one group more than any other). When these intuitions hold, an optimal set of products for minimizing the average population regret may be nearly optimal for minimizing the minimax group regret as well. This is consistent with our empirical results, where we see that the gaps in both population and minimax group regret between the ILP and DP solutions are small.

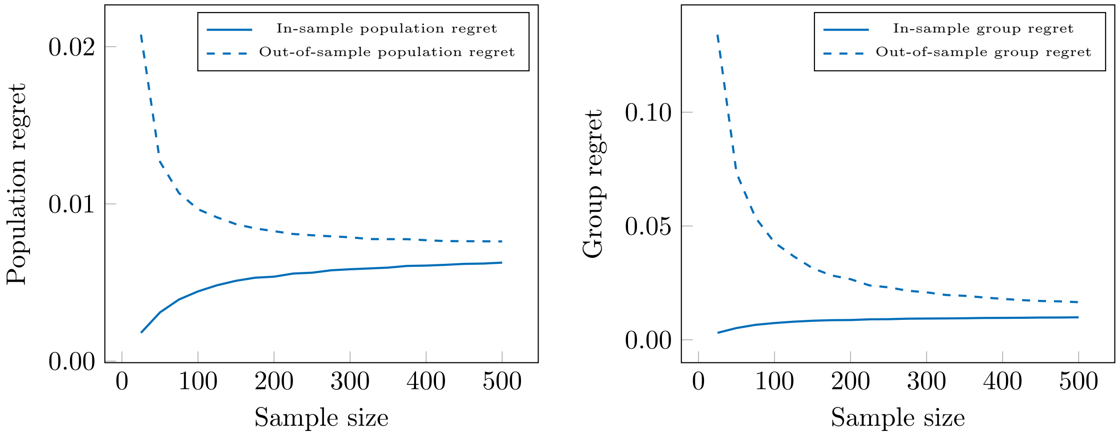


Group Regret

Population Regret

Figure 2.3: Algorithm Performance on the Balanced Population Distribution

Our second experiment explores generalization. We fix $p = 5$ and use the NR-Sparse-C algorithm to choose products. We draw a test set of 5000 consumers from the unbalanced distribution described above. For sample sizes of $\{25, 50, ..., 500\}$ consumers, we obtain product sets using NR-Sparse-C and calculate the incurred regret using these products on both the training and test sets. We repeat this process 1000 times for each sample size and average them. This is plotted in Figure 2.4; we observe that measured both by population regret as well as by group regret, the test regret converges to the training regret as the sample size increases. The decay rate is roughly $1/\sqrt{n}$, as suggested by theory, but our theoretical bound is loose due to sub-optimal constants.[3] Training regret increases with sample size because, for a fixed number of products, it is harder to satisfy a larger number of consumers.



Population Regret

Group Regret

Figure 2.4: Generalization for NR-Sparse-C with $p = 5$ on the unbalanced distribution.

2.8 Conclusion

2.8.1 Future Directions

Our work is the first (to our knowledge) to study portfolio design from the perspective of fairness, and examining other consumer-facing applications of these classical optimization problems is a rich area for future study. Additionally, we develop a particular definition of group fairness which constrains

[3] The theoretical bound is roughly an order of magnitude higher than the experimental bound; we do not plot it as it makes the empirical errors difficult to see.

the maximum regret across all groups, but there may be other notions of fairness to consider in this setting which may lead to different algorithmic results. Finally, we note that because of the similarity of our problem to a variant of a one-dimensional facility location problem, our work may point to interesting extensions in the *fair facility location* domain, which is also a relatively new and exciting area of study.

2.8.2 Broader Impacts

As with many papers in the fairness in machine learning literature, this paper is written with positive broader impacts in mind but also is at risk of providing a technical facade of “fairness” that might mask more serious underlying issues. Our aim in this paper is to provide efficient algorithms to allow financial companies to choose products not just to optimize for the overall benefit for their customers *on average* (which naturally favors majority groups over minority groups) but to instead optimize for the *least well-off* subgroup. Providing algorithms for this task, as well as simply clearly laying out the goal, helps lessen the institutional frictions that might otherwise prevent companies from even investigating fairness motivated objectives. This is particularly salient in the finance domain where algorithmic processes such as quantitative trading and so-called ‘robo-advisors’ are already widely used in consumer-facing settings. We therefore see fair algorithms such as those in this work as demonstrating that these practices can be improved for the benefit of those users who may be harmed by a standard profit-maximization approach to algorithm design.

On the other hand, we stress that we are guaranteeing only a particular technical notion of fairness that does not aim to address many underlying financial inequities. There is always a risk with such technologies that they will be applied and then used to justify doing nothing more, because some notion of “fairness” has been satisfied. We additionally recognize that the similarities between the finance context we study and certain kinds of one-dimensional facility location problems may lead to future work that lifts our algorithms into that context. We aim to avoid these outcomes with a clear discussion of what our methods can (and cannot) promise and we urge researchers building on this work to make similar considerations.

Bibliography

S. Barman and S. K. Krishnamurthy. Approximation algorithms for maximin fair division. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 647–664, 2017.

R. Bellman. A note on cluster analysis and dynamic programming. *Mathematical Biosciences*, 18 (3-4):311–312, 1973.

N. Bhutta, A. C. Chang, L. J. Dettling, and J. W. Hsu. Disparities in wealth by race and ethnicity in the 2019 survey of consumer finances. FEDS Notes.Washington: Board of Governors of the Federal Reserve System, 2020. URL `https://doi.org/10.17016/2380-7172.2797`.

E. Budish. The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes. *Journal of Political Economy*, 119(6):1061–1103, 2011.

D. Z. Chen and H. Wang. New algorithms for 1-d facility location and path equipartition problems. In F. Dehne, J. Iacono, and J.-R. Sack, editors, *Algorithms and Data Structures*, pages 207–218, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

A. Chouldechova and A. Roth. A snapshot of the frontiers of fairness in machine learning. *Communications of the ACM*, 63(5):82–89, 2020.

K. Donahue and J. Kleinberg. Fairness and utilization in allocating resources with uncertain demand. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 658–668, 2020.

H. Elzayn, S. Jabbari, C. Jung, M. Kearns, S. Neel, A. Roth, and Z. Schutzman. Fair algorithms for learning in allocation problems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 170–179, 2019.

D. Ensign, S. A. Friedler, S. Neville, C. Scheidegger, and S. Venkatasubramanian. Runaway feedback loops in predictive policing. *arXiv preprint arXiv:1706.09847*, 2017.

M.-C. Fang, S. D. Hanna, and S. Chatterjee. The impact of immigrant status and racial/ethnic group on differences in responses to a risk aversion measure. *Journal of Financial Counseling & Planning*, 24(2), 2013.

Y. Freund and R. E. Schapire. Game theory, on-line prediction and boosting. In *Proceedings of the ninth annual conference on Computational learning theory*, pages 325–332, 1996.

R. Hassin and A. Tamir. Improved complexity bounds for location problems on the real line. *Operations Research Letters*, 10(7):395–402, 1991.

D. A. Iancu and N. Trichakis. Fairness and efficiency in multiportfolio optimization. *Operations Research*, 62(6):1285–1301, 2014.

C. Jung, S. Kannan, and N. Lutz. A center in your neighborhood: Fairness in facility location. *arXiv preprint arXiv:1908.09041*, 2019.

A. Klein. Reducing bias in ai-based financial services. URL `https://www.brookings.edu/research/reducing-bias-in-ai-based-financial-services/`. Accessed: October 06, 2020.

M. S. A. Lee and L. Floridi. Algorithmic fairness in mortgage lending: from absolute conditions to relational trade-offs. *Available at SSRN*, 2020.

S. Mahabadi and A. Vakilian. (individual) fairness for k-clustering. *arXiv preprint arXiv:2002.06742*, 2020.

H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952. doi: 10.1111/j.1540-6261.1952.tb01525.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1952.tb01525.x`.

F. Nielsen and R. Nock. Optimal interval clustering: Application to bregman clustering and statistical mixture learning. *IEEE Signal Processing Letters*, 21(10):1289–1292, 2014.

A. D. Procaccia and J. Wang. Fair enough: Guaranteeing approximate maximin shares. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pages 675–692, 2014.

J. Sarra. Fairness in financial markets, reconceiving market norms. In *Forum on Public Policy: A Journal of the Oxford Round Table*. Forum on Public Policy, 2014.

S. B. Shu and S. M. Ucla. 1 applying fairness theories to financial decision making. 2012.

R. Yao, M. S. Gutter, and S. D. Hanna. The financial risk tolerance of blacks, hispanics and whites. *Journal of Financial Counseling and Planning*, 16(1), 2005.

Chapter 3

Fair Allocations

3.1 Introduction

The bulk of the literature on algorithmic fairness has focused on classification and regression problems (see e.g. Hardt et al. (2016); Joseph et al. (2016); Kleinberg et al. (2017); Jabbari et al. (2017); Dwork et al. (2012); Zafar et al. (2017); Corbett-Davies et al. (2017); Woodworth et al. (2017); Zemel et al. (2013); Liu et al. (2018); Chierichetti et al. (2017); Calders et al. (2013); Berk et al. (2018, 2017) for a collection of recent work), but fairness concerns also arise naturally in many resource allocation settings. Informally, a resource allocation problem is one in which there is a limited supply of some *resource* to be distributed across multiple groups with differing needs. Resource allocation problems arise in financial applications (e.g. allocating loans), disaster response (allocating aid), and many other domains. In this work we use the running example of a public health authority distributing medical tests for diagnosing a particular condition to neighborhood clinical sites. We can think of the service area of each site as being a group, and each group has a different prevalence of the condition which may change over time. We model this by considering the number of individuals with the condition in each group to be drawn from some distribution. The allocator's goal (absent fairness concerns) is to distribute tests in a way that maximizes the number of cases of the condition detected.

Since the allocator does not know the number of candidates for the resource in each group or even the distribution from which these numbers are drawn ahead of time, they must run some sort of learning algorithm to form estimates of these distributions. Furthermore, the allocator only learns about these distributions by making allocations and receives feedback in the form of the number of candidates in each group who received the resource. A key challenge is therefore for the allocator to disentangle observing a small number of candidates in a group who received the resource as a result of there simply not being a large number of candidates in that group from that feedback being the result of underallocating to that group.

The primary fairness concern in this setting is falling into *feedback loops*, where an underallocation to a particular group leads the allocator to believe that there are not many candidates for the resource in that group and therefore to continue to underallocate. This problem is particularly salient in the domain of *predictive policing*, and a line of work on algorithmic feedback loops in that setting inspired the technical basis of this work. In the predictive policing problem, the resource to be distributed is police officers, which can be dispatched to different districts. Each district has a different crime distribution, and the goal (absent additional fairness constraints) might be to maximize the number of crimes caught. Of course, fairness concerns abound in this setting, and recent work (see e.g. Lum and Isaac (2016); Ensign et al. (2018a,b)) has highlighted the extent to which algorithmic allocation might exacerbate those concerns. For example, Lum and Isaac (2016) show that if predictive policing algorithms such as PredPol are trained using past arrest data to predict future crime, then these pernicious feedback loops can arise, which misestimate the true crime rates in certain districts, leading to an overallocation of police. The model in Ensign et al. (2018a) incorporates notions of algorithmic fairness and implicitly considers an allocation to be *fair* if police are allocated across districts in direct proportion to the district's true crime rate, and examines the implications of falsely believing that a particular neighborhood has a higher crime rate than another. Generally extended, this definition asks that units of a resource are allocated according to the group's share of the total candidates for that resource.

In our work, we study a different notion of allocative fairness that has a similar motivation to the notion of *equality of opportunity* proposed by Hardt et al. (2016) in classification settings. Informally speaking, it asks that the probability that a candidate for a resource be allocated a resource should be independent of their group.

3.1.1 Our Results

To define the extent to which an allocation satisfies our fairness constraint, we must model the specific mechanism by which resources deployed to a particular group reach their intended targets. We study two such *discovery models*, and we the explicit framing of this modeling step is one of the contributions of our work; the implications of a fairness constraint depend strongly on the details of the discovery model. Selecting and articulating such a model is an important step in making one's assumptions transparent.

We study two discovery models which capture two extremes of targeting ability, and the difference in both the qualitative and quantitative aspects of the technical results in these two models highlights how critical these modelling choices are. In the *random* discovery model, however many units of the resource are allocated to a given group, all individuals within that group are equally likely to be assigned a unit, regardless of whether they are a candidate for the resource or not. Units of the resource assigned to non-candidates are considered 'wasted' and do not contribute to the allocator's utility. In other words, the probability that a candidate receives a resource is equal to the ratio of the number of units of the resource assigned to their group to the size of the group (*independent* of the number of candidates in the group).

At the other extreme, in the *precision* discovery model, units of the resource are given only to actual candidates within a group, as long as there is sufficient supply of the resource. In this model, the only units of the resource that are 'wasted' are those allocated to a particular group beyond the number of candidates in that group. In other words, the probability that a candidate receives a resource is equal to the ratio of the number of units of the resource assigned to their group to the number of *candidates* within that group.

In the context of medical testing, these two models can be viewed as two (unrealistic) extremes for targeting individuals with the condition. The random model corresponds to simple random testing. If a neighborhood is allocated v tests, clinicians draw the names of v residents from a hat and test them. In the precision model, perhaps testing for the disease occurs further downstream in the patient care process, and therefore tests are only administered to individuals who doctors are already nearly certain have the condition. This highlights the importance of specifying the discovery model in analyzing the fairness implications of algorithms in this context. For highly contagious diseases with community spread, a testing strategy closer to the random model may be the one preferred by health experts, and for other kinds of conditions, the precision model may be more realistic.

Since these different discovery models have different implications for fairness, we analyse them separately. In the random model, fairness constrains resources to be distributed in amounts proportional to *group sizes*, regardless of the distribution of candidates, and so is uninteresting from a learning perspective, since we take the population of each group to be known. On the other hand, the precision model yields an interesting fairness-constrained learning problem when the distribution of the number of candidates in each group must be learned via observation, and what counts as a 'fair' allocation depends greatly on these distributions.

Formally, we study learning in a censored feedback setting: each round, the algorithm can choose a feasible deployment of resources across groups. Then the number of candidates for the current round in each group is drawn from a fixed, but unknown group-dependent distribution (which might be not be independent from the distributions in other groups). The algorithm does not observe the number of candidates present in each group, but only the number of candidates that received the resource. Thus, the extent to which the algorithm can learn about the distribution in a particular group is limited by the number of resources it deploys there. The goal of the algorithm is to converge to an optimal fairness-constrained allocation, where both the objective value of the solution and the constraints imposed on it depend on the unknown distributions.

One trivial solution to the learning problem is to sequentially deploy *all* of the resources to each group in turn for a sufficient amount of time to accurately learn the candidate distributions. This would reduce the learning problem to an offline constrained optimization problem, which we show can be efficiently solved by a greedy algorithm. But this algorithm is unreasonable: it has a large exploration phase in which it uses nonsensical deployments, vastly overallocating to some groups and underallocating to others.

A more natural approach would be for the allocator to maintain estimates of the candidate distributions for each group, greedily deploys an allocation to each group that is fair with respect to these estimates, and then updates its beliefs according to both the feedback from this current allocation as well as its past history of actions. However, we show that this algorithm will not converge to an allocation which is fair with respect to the true underlying distributions without more assumptions on the structure of these distributions, and we carefully develop this impossibility result in Section 3.3.3.

This impossibility result motivates us to consider the learning problem in which the unknown distributions are from a known parametric family in Section 3.3.4. The natural algorithm uses an optimal fair deployment at each round given the maximum likelihood estimates of candidate distributions given its (censored) observations so far; for concreteness, we analyze this algorithm in case of the Poisson distribution, and show that it converges to an optimal fair allocation, but our analysis generalizes for any single-parameter Lipschitz-continuous family of distributions. We characterize the *price of fairness* of this algorithm, which is the loss in utility to the allocator for meeting the fairness constraint rather than simply allocating in a way that maximizes the number of individuals (irrespective of their groups) who receive the resource.

We conclude with a brief discussion of future directions in this area as well as some of the modelling gaps between our work and a real-world deployment of these ideas.

3.1.2 Further Related Work

Our precision discovery model is inspired by and has technical connections to the model of Ganchev et al. (2009), which models the *dark pool problem* from quantitative finance, in which a trader wishes to execute a specified number of trades across a set of exchanges of unknown but independently distributed liquidity. Ganchev et al. (2009) design an optimal allocation algorithm under the censored feedback of the precision model. It is straightforward to map their setting onto ours, but they assume independence between different exchanges, while the candidate distributions in our setting need not be independent. Regardless, we show that their allocation algorithm can be used to compute an optimal allocation (ignoring fairness) even when the independence assumption is relaxed (see Remark 3.2). Later, Agarwal et al. (2010) extended the dark pool problem to an adversarial (rather than distributional) setting. This is quite closely related to the work of Ensign et al. (2018b) who also consider the precision model (under a different name) in an adversarial predictive policing setting. They provide no-regret algorithms for this setting by reducing the problem to learning in a partial monitoring environment. Since their setting is equivalent to that of Agarwal et al. (2010), the algorithms in that work can be directly applied to the problem studied by Ensign et al. (2018b).

Our desire to study the natural greedy algorithm rather than an algorithm which uses "unreasonable" allocations during an exploration phase is an instance of a general concern about exploration in fairness-related problems, as in Bird et al. (2016). There has been much recent work on the performance of greedy algorithms in different settings for this reason, such as Bastani et al. (2017); Kannan et al. (2018); Raghavan et al. (2018). Lastly, the term *fair allocation* appears in the *fair division* literature (see e.g. Procaccia (2013) for a survey), but that body of work is technically quite distinct from the problem we study here.

3.2 Setting

We study an *allocator* with $\mathcal{V}$ units of a resource and tasked with distributing them across a population partitioned into $\mathcal{G}$ groups. Each group is divided into *candidates*, who are the individuals the allocator would like to receive the resource, and *non-candidates*, who are the remaining individuals. We let m_i denote the total number of individuals in group i. The number of candidates c_i in group i is a random variable drawn from a fixed but unknown distribution $\mathcal{C}_i$ called the *(marginal) candidate distribution.* We do not make any assumptions about the relationship between the candidate distributions across different groups and in particular these distributions need not be independent. We use M to denote the total size of all groups (i.e., $M = \sum\limits_{i \in [\mathcal{G}]} m_i$). An allocation $\mathbf{v} = (v_1, \ldots, v_{\mathcal{G}})$ is a partitioning of these $\mathcal{V}$ units, where $v_i \in \{0, \ldots, \mathcal{V}\}$ denotes the units of resources allocated to group i. Every allocation is bound by a *feasibility* constraint which requires that $\sum\limits_{i \in [\mathcal{G}]} v_i \leq \mathcal{V}$.

A *discovery model* is a (possibly randomized) function $\operatorname{disc}(v_i, c_i)$ mapping the number of units v_i allocated to group i and the number of candidates c_i in group i to the number of candidates discovered in group i. In the learning setting, upon fixing an allocation $\mathbf{v}$, the learner will get to observe (a realization of) $\operatorname{disc}(v_i, c_i)$ for the realized value of c_i for each group i. Fixing an allocation $\mathbf{v}$, a discovery model $\operatorname{disc}(\cdot)$ and candidate distributions for all groups $\mathcal{C} = \{\mathcal{C}_i : i \in [\mathcal{G}]\}$, we define the total expected number of discovered candidates, $\chi(\mathbf{v}, \operatorname{disc}(\cdot), \mathcal{C})$, as

$$\chi\left(\mathbf{v}, \operatorname{disc}(\cdot), \mathcal{C}\right) = \sum_{i \in [\mathcal{G}]} \mathop{\mathbb{E}}_{c_i \sim \mathcal{C}_i} \left[\operatorname{disc}(v_i, c_i)\right], \tag{3.1}$$

where the expectation is taken over $\mathcal{C}_i$ and any randomization in the discovery model $\operatorname{disc}(\cdot)$. When the discovery model and the candidate distributions are fixed, we will simply write $\chi(\mathbf{v})$ for brevity. We also use the total expected number of discovered candidates and *(expected) utility* interchangeably. We refer to an allocation that maximizes the expected number of discovered candidates over all feasible allocations as an *optimal allocation* and denote it by $\mathbf{w}^*$.

3.2.1 Allocative Fairness

We take as our definition of *fairness* that an allocation is *fair* if it satisfies *approximate equality of candidate discovery probability* across groups. We call this *discovery probability* for brevity. This formalizes the intuition that it is unfair if candidates in one group have an inherently higher or lower probability of receiving the resource than candidates in another. Formally, we define our notion of *allocative fairness* as follows.

Definition 3.1 (α-fair allocation)**.** Fix a discovery model disc$(\cdot)$ and the candidate distributions $\mathcal{C}$. For an allocation $\mathbf{v}$, let

$$f_i\left(v_i, \mathrm{disc}(\cdot), \mathcal{C}_i\right) = \mathop{\mathbb{E}}_{c_i \sim \mathcal{C}_i}\left[\frac{\mathrm{disc}\left(v_i, c_i\right)}{c_i}\right],$$

denote the expected probability that a random candidate from group i receives a unit of the resource at allocation $\mathbf{v}$ (i.e. the discovery probability in group i). Then for any $\alpha \in [0,1]$, $\mathbf{v}$ is α-fair if

$$\left| f_i\left(v_i, \mathrm{disc}(\cdot), \mathcal{C}_i\right) - f_j\left(v_j, \mathrm{disc}(\cdot), \mathcal{C}_j\right) \right| \leq \alpha,$$

for all pairs of groups i and j.

When it is clear from the context, for brevity, we write $f_i(v_i)$ for the discovery probability in group i. We emphasize that this definition depends crucially on the chosen discovery model, and requires and implies nothing about the treatment of non-candidates. We think of this as a *minimal* definition of fairness where natural extensions might include constraining the treatment of non-candidates or enriching the space of candidates.

Since discovery probabilities $f_i(v_i)$ and $f_j(v_j)$ are in $[0,1]$, the absolute value of their difference is always in $[0,1]$. Therefore, taking $\alpha = 1$ imposes no fairness constraints whatsoever on the allocations and $\alpha = 0$ corresponds to *exact* fairness.

We write the allocator's utility as $\chi(\cdot)$ and refer to a feasible allocation $\mathbf{v}$ that maximizes $\chi(\mathbf{v})$ subject to α-fairness as an *optimal α-fair allocation* $\mathbf{w}^\alpha$. In general, $\chi(\mathbf{w}^\alpha)$ is a non-increasing quantity in α, since as α increases towards one, the set of feasible α-fair allocations only becomes larger.

Remark 3.2. *Both the utility and discovery probabilities can be written solely in terms of the* marginal *candidate distributions in each of the groups, even when these distributions are not independent. This is because the number of candidates discovered in a group depends only on the number of candidates in the group and the allocation to that group, regardless of those values in other groups. This assumption together with the linearity of expectation allows us to write the expected utility as in the right hand side of Equation* (3.1)*.*

3.3 The Precision Discovery Model

We begin by describing the *precision model* of discovery. Allocating v_i units to group i in the precision model results in the discovery of $\mathrm{disc}(c_i, v_i) \triangleq \min(c_i, v_i)$ candidates. This models the ability to perfectly discover and reach candidates in a group with resources deployed to that group, limited only by the number of deployed resources and the number of candidates present.

The precision model results in *censored* observations that have a particularly intuitive form. Recall that in general, a learning algorithm at each round gets to choose an allocation $\mathbf{v}$ and then observes $\mathrm{disc}(v_i, c_i)$ for each group i. In the precision model, this results in the following kind of observation: when v_i is larger than c_i, the allocator learns the number of candidates c_i present on that day exactly. We refer to this kind of feedback as an *uncensored observation.* When v_i is smaller than c_i, all the allocator learns is that the number of candidates is *at least* v_i. We refer to this kind of feedback as a *censored observation.*

The rest of this section is organized as follows. In Sections 3.3.1 and 3.3.2 we characterize optimal and optimal fair allocations for the precision model when the candidate distributions are known. In Section 3.3.3 we focus on learning an optimal fair allocation when these distributions are unknown. We show that any learning algorithm that is guaranteed to find a fair allocation in the *worst case* over candidate distributions must have the undesirable property that at some point, it must allocate a vast number of its resources to each group individually. To bypass this hurdle, in Section 3.3.4 we show that when the candidate distributions have a parametric form, a natural greedy algorithm which always uses an optimal fair allocation for the current maximum likelihood estimates of the candidate distributions converges to an optimal fair allocation. The analysis in this section is with respect to the marginal candidate distributions being *Poisson*, although the results immediately generalize to any family of Lipschitz-continuous, single-parameter distributions. The Poisson distribution arises naturally when examining the expected number of times an independent event with a particular mean rate occurs over a given period. For example, if every day, each person independently develops a particular medical condition with probability p, then the number of people who develop the condition over the course of a week is drawn from a Poisson distribution with parameter $\lambda = 7p$.

3.3.1 Optimal Allocation

We first describe how an optimal allocation (absent fairness constraints) can be computed efficiently when the marginal candidate distributions $\mathcal{C}_i$ are known.

We can use the same algorithm as in Ganchev et al. (2009) to compute an optimal allocation in our setting; this is because, as stated in Remark 3.2, the utility in both settings can be written solely in terms of the (marginal) candidate distributions even when the candidate distributions are not independent across different groups. Here, we present the high level ideas of their algorithm in the language of our model. While they assume that the distributions of shares across different dark pools are independent, our formulation does not require this assumption of independence.

Let $\mathcal{T}_i(c) = \Pr_{c_i \sim \mathcal{C}_i}[c_i \geq c]$ denote the probability that there are at least c candidates in group i. We refer to $\mathcal{T}_i(c)$ as the *tail probability of* $\mathcal{C}_i$ *at* c. Recall that the value of the *cumulative distribution function* (CDF) of $\mathcal{C}_i$ at c is defined to be

$$\mathcal{F}_i(c) = \sum_{c' \leq c} \Pr_{c_i \sim \mathcal{C}_i} \left[c_i = c'\right].$$

So $\mathcal{T}_i(c)$ can be written in terms of CDF values as $\mathcal{T}_i(c) = 1 - \mathcal{F}_i(c-1)$.

We first show how the expected number of discovered candidates in a group in the precision model can be written as a function of the tail probabilities of the group's candidate distribution.

Lemma 3.3 (Ganchev et al. (2009))**.** *The expected number of discovered candidates in the precision model when allocating v_i units of resource to group i can be written as* $\mathbb{E}_{c_i \sim \mathcal{C}_i}[\min(c_i, v_i)] = \sum_{c=1}^{v_i} \mathcal{T}_i(c)$.

Proof. We have

$$\begin{aligned}
\mathop{\mathbb{E}}_{c_i \sim \mathcal{C}_i}[\min(c_i, v_i)] &= \sum_{c=1}^{m_i} \Pr_{c_i \sim \mathcal{C}_i}[c_i = c] \cdot \min(c, v_i) \\
&= \sum_{c=1}^{v_i - 1} \Pr_{c_i \sim \mathcal{C}_i}[c_i = c] \cdot c + v_i \mathcal{T}_i(v_i) \\
&= \sum_{c=1}^{v_i - 2} \Pr_{c_i \sim \mathcal{C}_i}[c_i = c] \cdot c + (v_i - 1)\mathcal{T}_i(v_i - 1) + \mathcal{T}_i(v_i) \\
&= \mathcal{T}_i(1) + \cdots + \mathcal{T}_i(v_i - 1) + \mathcal{T}_i(v_i) \\
&= \sum_{c=1}^{v_i} \mathcal{T}_i(c).
\end{aligned}$$

Note that we can perform the telescoping in the 3rd and 4th lines by observing that

$$\Pr_{c_i \sim \mathcal{C}_i}[c_i = c - 1] + \mathcal{T}_i(c) = \mathcal{T}_i(c - 1).$$

□

Using this, we can give a greedy algorithm to find an optimal allocation in the precision model when the candidate distributions are known.

Lemma 3.4 (Theorem 1 in Ganchev et al. (2009))**.** *The allocation returned by greedily allocating the next unit of resource to a group in*

$$\operatorname*{argmax}_{i \in [\mathcal{G}]} \left(\mathcal{T}_i(v_i^t + 1) - \mathcal{T}_i(v_i^t) \right),$$

where v_i^t is the current allocation to group i at the t^{th} round maximizes the expected number of candidates discovered.

Proof. Since the tail probability functions $\mathcal{T}_i(c)$ are all non-increasing (that is, for $c \leq c'$, we have $\mathcal{T}_i(c') \leq \mathcal{T}_i(c)$), the greedy allocation returns an allocation $\mathbf{v}$ which maximizes

$$\chi(\mathbf{v}) = \sum_{i \in [\mathcal{G}]} \sum_{c=1}^{v_i} \mathcal{T}_i(c) \text{ such that } \sum_{i \in [\mathcal{G}]} v_i = \mathcal{V}.$$

Using Lemma 3.3 we have that

$$\sum_{i \in [\mathcal{G}]} \mathop{\mathbb{E}}_{c_i \sim \mathcal{C}_i} [\min(c_i, v_i)] = \sum_{i \in [\mathcal{G}]} \sum_{c=1}^{v_i} \mathcal{T}_i(c).$$

So the above double-summation is exactly equal to the expected number of discovered candidates. To see that the greedy solution is optimal, notice that any solution which does not allocate the marginal resource to the tail with the highest remaining probability can be improved by reallocating the final allocated resource in some lower tail probability group to the one in the higher tail probability. Finally since each term in the objective function is non-negative an optimal allocation would use all the $\mathcal{V}$ units of resource (so the feasibility constraint is tight). □

3.3.2 Optimal Fair Allocation

We next show how to compute an optimal α-fair allocation in the precision model when the candidate distributions are known and do not need to be learned. To build intuition for how the algorithm works, imagine that the group i has the highest discovery probability in $\mathbf{w}^\alpha$, and the allocation w_i^α to that group is somehow known to the algorithm ahead of time. The constraint of α-fairness then implies that the discovery probability for each other group j in $\mathbf{w}^\alpha$ must satisfy $f_j \in [f_i - \alpha, f_i]$. This in turn implies upper and lower bounds on the feasible allocations w_j^α to group j. The algorithm is then simply a constrained greedy algorithm: subject to these implied constraints, it iteratively allocates the next unit so as to maximize their marginal probability of reaching another candidate. Since the group i maximizing the discovery probability in $\mathbf{w}^\alpha$ and the corresponding allocation w_i^α are not known ahead of time, the algorithm simply iterates through each possible choice of i. Pseudocode is given in Algorithm 3.2. We prove that this algorithm returns an optimal α-fair allocation in Theorem 3.5.

Theorem 3.5. *Algorithm 3.2 computes an optimal α-fair allocation for the precision model in time*

Algorithm 3.2 Computing an optimal fair allocation in the precision model

```
Input: $\alpha$, $\mathcal{C}$ and $\mathcal{V}$.
Output: An optimal $\alpha$-fair allocation $\mathbf{w}^{\alpha}$.
  $\mathbf{w}^{\alpha} \leftarrow \vec{0}$.                                  ▷ Initialize the output.
  $\chi_{\max} \leftarrow 0$.                                 ▷ Keep track of the utility of the output.
  for $i \in [\mathcal{G}]$ do                    ▷ Guess for group with the highest probability of discovery.
      $\mathbf{v} \leftarrow \vec{0}$.
      for $v_i \in \{0, \dots \mathcal{V}\}$ do                         ▷ Guess for the allocation to that group.
          Set $v_i$ in $\mathbf{v}$ and compute $f_i(v_i)$.
          $ub_i \leftarrow v_i$.                              ▷ Upper bound on allocation to group $i$.
          $lb_i \leftarrow v_i$.                              ▷ Lower bound on allocation to group $i$.
          for $j \neq i, j \in [\mathcal{G}]$ do                   ▷ Upper and lower bounds for other groups.
              Update $lb_j$ and $ub_j$ using $f_i(v_i)$, $\alpha$ and $\mathcal{C}_j$.
              $v_j \leftarrow lb_j$.                        ▷ Assign the lower bound allocation to group $j$.
          if $\Sigma_{i \in [\mathcal{G}]} v_i > \mathcal{V}$ then
              continue.                                           ▷ Allocation is not feasible.
          $\mathcal{V}_r = \mathcal{V} - \Sigma_{i \in [\mathcal{G}]} v_i$
          for $t = 1, \dots, \mathcal{V}_r$ do    ▷ Allocate the remaining resources greedily while obeying fairness.
              $j^* \in \underset{j \in [\mathcal{G}]}{\operatorname{argmax}} \left( \mathcal{T}_j(v_j + 1) - \mathcal{T}_j(v_j) \right)$ s.t. $v_j < ub_j$.
              $v_{j^*} \leftarrow v_{j^*} + 1$.
          $\chi(\mathbf{v}) = \Sigma_{i \in [\mathcal{G}]} \Sigma_{c=1}^{v_i} \mathcal{T}_i(c)$.                      ▷ Compute the utility of $\mathbf{v}$.
          if $\chi(\mathbf{v}) > \chi_{\max}$ then                  ▷ Update the best $\alpha$-fair allocation found so far.
              $\chi_{\max} \leftarrow \chi(\mathbf{v})$.
              $\mathbf{w}^{\alpha} \leftarrow \mathbf{v}$.
  return $\mathbf{w}^{\alpha}$.
```

$O(\mathcal{G}\mathcal{V}(\mathcal{G}\mathcal{V}+M))$.

Proof. Fix an optimal α-fair allocation $\mathbf{w}^\alpha$. In $\mathbf{w}^\alpha$, some group i has the highest f_i and receives allocation w_i^α. Suppose we know i and w_i^α; we will relax this assumption at the end of the proof. Using knowledge of $\mathcal{C}_i$, we can compute $f_i(w_i^\alpha)$. This implies that $f_j \in [f_i - \alpha, f_i]$ for every other group j, which in turn can be used to derive the set of all possible allowable allocations w_j^α which do not violate α-fairness.

We claim that if we initialize the allocation w_j^α to be the lower bound of the interval corresponding to group j, then greedily assign the surplus units with the added restriction that w_j^α is always inside of its respective interval, we achieve an optimal α-fair allocation.

Since we assume we know the allocation to group i in an optimal fair allocation w_i^α, this allocation must be achievable by picking some value from each of the intervals, thus initializing the allocation to the lower bound of each interval certainly cannot assign more than $\mathcal{V}$ units in total. By the same argument as for the unconstrained greedy algorithm, since the objective function is concave (recall that the tail probabilities are non-increasing) and increasing in each argument w_j^α, a greedy search over this feasible region finds the desired allocation.

The algorithm does not know *a priori* the group i which has the maximum f_i or w_i^α, so it must search over these options. There are $\mathcal{G}$ guesses for group i and $\mathcal{V}+1$ guesses for the allocation to the group. So there are a total of $\mathcal{G}(\mathcal{V}+1)$ guesses that need to be considered. For each guess, it takes $O(M)$ time to compute the upper and lower bounds on the allocation to each of the groups and $O(\mathcal{G}\mathcal{V})$ time to run the greedy algorithm. So the running time of Algorithm 3.2 is $O(\mathcal{G}\mathcal{V}(\mathcal{G}\mathcal{V}+M))$.

□

3.3.3 Learning May Require Brute-Force Exploration

In Sections 3.3.1 and 3.3.2 we assumed the candidate distributions were known. When the candidate distributions are unknown, learning algorithms intending to converge to optimal α-fair allocations must learn a sufficient amount about the distributions in question to certify the fairness of the allocation they finally output. Because learners must deal with feedback in the censored observation model, this places constraints on how they can proceed. Unfortunately, as we show in this section, in the worst case, the candidate distributions might force the learner to engage in "brute-force exploration" — the iterative deployment of a large fraction of the resources to each subgroup in turn. This is formalized in Theorem 3.6.

Theorem 3.6. *Define $m^* = \max_{i \in [\mathcal{G}]} m_i$ to be the size of the largest group and assume $m_i > 6$ for all i and $\mathcal{G} \geq 2$. Let $\alpha \in [0, 1/(2m^*))$, $\delta \in (0, 1/2)$, and $\mathcal{A}$ be any learning algorithm for the precision model which runs for a finite number of rounds and outputs an allocation. Suppose that there is some group i for which $\mathcal{A}$ has not allocated at least $m_i/2$ units for at least $k \ln(1/\delta)/(\alpha m_i)$ rounds upon termination, where k is an absolute constant. Then there exists a candidate distribution such that, with probability at least δ, $\mathcal{A}$ outputs an allocation that is not α-fair.*

Proof. Let i denote the group to which $\mathcal{A}$ has not allocated at least

$m_i/2$ units for at least $k \ln(1/\delta)/(\alpha m_i)$ rounds upon its termination. We fix an arbitrary allocation $\mathbf{v}$ and design two candidate distributions for group i such that the discovery probabilities given v_i computed under the two different distributions are at least 2α-apart.[1] Any algorithm guaranteeing α-fairness must distinguish between these two distributions with high probability, or $\mathbf{v}$ could have higher unfairness than α. We then show that to distinguish between these two candidate distributions, with probability of at least $1 - \alpha$, any algorithm is *required* to send $m_i/2$ units to group i for at least $k \ln(1/\delta)/(\alpha m_i)$ rounds.

Consider two candidate distributions $\mathcal{C}_i$ and $\mathcal{C}'_i$ for group i. We use the shorthand $p_i(c) = \Pr_{c_i \sim \mathcal{C}_i}[c_i = c]$ and similarly for $p'_i(c)$. Let $c^* = m_i/2 - 2$. We require only that $\mathcal{C}_i$ and $\mathcal{C}'_i$ satisfy the following conditions.

1. $p_i(c) = p'_i(c)$ for all $c' \leq c^*$.

2. $\Sigma_{c \leq c^*} p_i(c) = \Sigma_{c \leq c^*} p'_i(c) = 1 - 2\alpha m_i$.

[1] We assume $\mathbf{v}$ sends at least one unit to group i, otherwise it would be easy to construct an example where the algorithm allocating according to $\mathbf{v}$ is unfair.

3. $p_i(c^*+1) = 2\alpha m_i$ and $p_i(c) = 0$ for all $c \in \{c^*+1, \ldots, m_i\}$.

4. $p'_i(c) = 0$ for all $c \in \{c^*+1, \ldots, m_i - 1\}$ and $p'_i(m_i) = 2\alpha m_i$.

In other words, any two distributions that are the same up to c^*, have a CDF value of $2\alpha m_i$ at c^*, and differ in where in the tail they assign the remaining mass, will serve our purposes.

Let $f_i(v_i)$ and $f'_i(v_i)$ denote the discovery probability given allocation $\mathbf{v}$ which assigns v_i units to group i for candidate distributions $\mathcal{C}_i$ and $\mathcal{C}'_i$, respectively. Then

$$|f_i(v_i) - f'_i(v_i)| = 2\alpha m_i \left| \frac{v_i}{c^*+1} - \frac{v_i}{m_i} \right|.$$

This difference is minimized at $v_i = 1$, in which case

$$\left| \frac{v_i}{c^*+1} - \frac{v_i}{m_i} \right| > \frac{2}{m_i} - \frac{1}{m_i} = \frac{1}{m_i}.$$

Hence, for any allocation $\mathbf{v}$, $|f_i(v_i) - f'_i(v_i)| > 2\alpha$.

Finally, because $\mathcal{C}_i$ and $\mathcal{C}'_i$ do not differ on any potential observation less than c^*, distinguishing between the two candidate distributions requires observing at least one uncensored observation of c^* or higher. Under the precision model, this requires sending at least $m_i/2$ units to group i. However, conditioning on sending at least $m_i/2$ units, the probability of observing an uncensored observation is at most $2\alpha m_i$. Hence, to distinguish between $\mathcal{C}_i$ and $\mathcal{C}'_i$ (and thus to guarantee that an allocation $\mathbf{v}$ is α-fair) with probability of at least $1-\delta$, a learning algorithm must allocate $m_i/2$ units for $k\ln(1/\delta)/(\alpha m_i)$ rounds to group i. □

When m^*, the size of the largest group, is larger than $2\mathcal{V}$, then Theorem 3.6 implies that no algorithm can guarantee α-fairness for sufficiently small α. Moreover, even when $m^* \leq 2\mathcal{V}$, Theorem 3.6 shows that in general, if we want algorithms that have provable guarantees for *arbitrary* candidate distributions, it is impossible to avoid something akin to brute-force search, following a trivial algorithm of the kind that simply allocates *all* resources to each group in turn, for sufficiently many rounds to approximately learn the CDF of the candidate distribution, and then solves the offline problem. In the next section, we circumvent this by giving an algorithm with provable guarantees, assuming that the candidate distributions have a known parametric form.

3.3.4 Poisson Distributions and Convergence of the MLE

In this section, we assume that all the candidate distributions have a particular and known *parametric form* but that the parameters of the these distributions are not known to the allocator. Concretely, we assume that the candidate distribution for each group is Poisson[2] (denoted by $\mathcal{C}(\lambda)$) and write $\boldsymbol{\lambda}^* = (\lambda_1^*, \ldots, \lambda_G^*)$ for the true underlying parameters of the marginal candidate distributions. This assumption allows an algorithm to learn the tails of these distributions without needing to rely on brute-force search, thus circumventing the limitation given in Theorem 3.6. Indeed, we show that a small variant of the natural greedy algorithm incorporating these distributional assumptions converges to an optimal fair allocation.

At a high level, in each round, our algorithm uses Algorithm 3.2 to calculate an optimal fair allocation with respect to the current maximum likelihood estimates of the group distributions; then, it uses the new observations it obtains from this allocation to refine these estimates for the next round. This is summarized in Algorithm 3.3. Much of the work in this section is towards showing that by using this algorithm, the allocator's sequence of estimates of $\boldsymbol{\lambda}^*$ converge to the truth over time.

The algorithm differs from this purely greedy strategy in one respect, to overcome the following subtlety: there is a possibility that Algorithm 3.2, when operating on a preliminary estimate for the candidate distributions, will suggest sending zero units to some group, even when the optimal allocation for the true distributions sends some units to every group. Such a deployment would result in the algorithm receiving no feedback for the zero-allocated group that round. If this suggestion is followed and a lack of feedback is allowed to persist indefinitely, the algorithm's parameter estimate for the zero-allocated group will also stop updating — potentially at an incorrect value. In order

[2] To match our model, we would technically need to assume a *truncated* Poisson distribution to satisfy the bounded support condition. However, the distinction will not be important for the analysis, and so to minimize technical overhead, we perform the analysis assuming an untruncated Poisson.

to avoid this problem and continue making progress in learning, our algorithm chooses another allocation in this case. As we show, any allocation that allocates positive resources to all groups will suffice; in particular, our algorithm makes the natural choice of simply repeating the allocation from the previous round.

Algorithm 3.3 Learning an optimal fair allocation

```
Input: $\alpha$, $\mathcal{V}$ and $T$ (total number of rounds).
Output: An allocation $\mathbf{v}^{T+1}$ and estimates to parameters $\{\lambda_i^T\}$.
  $\mathbf{v}^1 \leftarrow (\lfloor(\mathcal{V}/\mathcal{G})\rfloor, \dots, \lfloor(\mathcal{V}/\mathcal{G})\rfloor)$.   ▷ Allocate uniformly.
  for rounds $t = 1, \dots, T$ do
    if $\exists i$ such that $v_i^t == 0$ then   ▷ Check whether every group is allocated a resource.
      $\mathbf{v}^t \leftarrow \mathbf{v}^{t-1}$.
    Observe $o_i^t = \min\{v_i^t, c_i^t\}$ for each group.
    for $i = 1, \dots, \mathcal{G}$ do
      Update history $\mathbf{h}_i^{t+1}$ with $o_i^t$ and $v_i^t$.
      $\hat{\lambda}_i^t \leftarrow \text{argmax}_{\lambda \in [\lambda_{\min}, \lambda_{\max}]} \hat{\mathcal{L}}(\mathbf{h}_i^{t+1}, \lambda)$.   ▷ Solve the maximum likelihood estimation problem.
    $\mathbf{v}^{t+1} \leftarrow$ Algorithm 3.2$(\alpha, \{\mathcal{C}(\hat{\lambda}_i^t)\}, \mathcal{V})$.   ▷ Compute an allocation to be deployed in the next round.
  return $\mathbf{v}^{T+1}$ and $\{\lambda_i^T\}$.
```

Notice that Algorithm 3.3 chooses an allocation at every round which is fair with respect to its estimates of the parameters of the candidate distributions; hence, asymptotic convergence of its output to an *optimal α-fair* allocation follows directly from the convergence of the estimates to true parameters. However, we seek a stronger, *finite sample* guarantee, as stated in Theorem 3.7.

Theorem 3.7. *Let $\varepsilon, \delta > 0$ and let $D = \max_{i \in [\mathcal{G}]} D_{TV}(\mathcal{C}(\lambda_i^*), \mathcal{C}(\hat{\lambda}_i))$ where D_{TV} denotes the total variation distance between two distributions. Suppose that the candidate distributions are Poisson distributions with unknown parameters in the vector $\boldsymbol{\lambda}^*$, where $\boldsymbol{\lambda}^*$ lies in the known interval $[\lambda_{\min}, \lambda_{\max}]^{\mathcal{G}}$. Suppose we run Algorithm 3.3 for $t > \tilde{\mathcal{O}}(\ln(\mathcal{G}/\delta)/(\eta(\varepsilon))^2) \triangleq T_{\max}$ rounds, where $\eta(\cdot)$ is some distribution-specific function[3] to get an allocation $\hat{\boldsymbol{v}}$ and estimated parameters $\hat{\lambda}_i$ for all groups i. Then for all i in $[\mathcal{G}]$, $|\hat{\lambda}_i - \lambda_i^*| \le \varepsilon$, with probability at least $1 - \delta$, $\hat{\boldsymbol{v}}$ is $(\alpha + 4D)$-fair and has utility at most $4D\mathcal{G}\mathcal{V}$ smaller than the utility of an optimal $(\alpha - 4D)$-fair allocation. That is, $\chi(\hat{\boldsymbol{v}}) \ge \chi(\boldsymbol{w}^{\alpha - 4D}) - 4D\mathcal{G}\mathcal{V}$.*

Remark 3.8. *Theorem 3.7 implies that in the limit, the allocation from Algorithm 3.3 will converge*

[3] See Corollary 3.19 for the relationship between η and ε. Also $\tilde{\mathcal{O}}$ hides poly-logarithmic terms in $1/\eta(\varepsilon)$.

to an optimal α-fair allocation. As $t \to \infty$, $\hat{\lambda}_i \xrightarrow{p} \lambda_i^*$ for all i, meaning $D \to 0$ and more importantly, $\hat{\boldsymbol{v}}$ will be α-fair and optimal.

The rest of this section is dedicated to the proof of Theorem 3.7. First, we introduce notation. Since we assume the candidate distribution for each group is Poisson, the *probability mass function* (PMF) and the CDF of the candidate distribution $\mathcal{C}_i$ for group i can be written as

$$\Pr_{c_i \sim \mathcal{C}(\lambda_i^*)}[c_i = c; \lambda_i^*] = \frac{{\lambda_i^*}^c e^{-\lambda_i^*}}{c!} \text{ and } F(c; \lambda_i^*) = \Pr_{c_i \sim \mathcal{C}(\lambda_i^*)}[c_i \leq c; \lambda_i^*] = e^{-\lambda_i^*} \sum_{x=0}^{c} \frac{{\lambda_i^*}^x}{x!}.$$

Given an allocation of v_i units of the resource to group i we use o_i to denote the (possibly censored) observation received by Algorithm 3.3. So while the candidates in group i are generated according to $\mathcal{C}(\lambda_i^*)$, the observations of Algorithm 3.3 follow a censored Poisson distribution which we abbreviate by $\mathcal{C}_o(\lambda_i^*, v_i)$. We can write the PMF of this distribution as

$$\Pr_{o_i \sim \mathcal{C}_o(\lambda_i^*, v_i)}[o_i = o; \lambda_i^*, v_i] = \begin{cases} \frac{{\lambda_i^*}^o e^{-\lambda_i^*}}{o!}, & o < v_i, \\ 1 - F(v_i - 1; \lambda_i^*), & o = v_i, \end{cases}$$

where $\mathcal{F}(v_i - 1; \lambda_i^*)$ is the CDF value of $\mathcal{C}(\lambda_i^*)$ at $v_i - 1$.

Since Algorithm 3.3 operates in rounds, we use the superscript t throughout to index the time period. For each round t, denote the history of the units allocated to group i and observations received (candidates discovered) in rounds up to t by $h_i^t = (v_i^1, o_i^1, \ldots, v_i^{t-1}, o_i^{t-1})$. We use $\mathbf{h}^t = (h_1^t, \ldots, h_{\mathcal{G}}^t)$ to denote the history for all groups. All the probabilities and expectations in this section are over the randomness of the observations drawn from the censored Poisson distributions unless otherwise noted; we suppress related notation for brevity and clarity. Finally, an allocation function $\mathcal{A}$ in round t is a mapping from the history of all groups $\mathbf{h}^t$ to the number of units to be allocated to each group i.e. $\mathcal{A} : \mathbf{h}^t \to \mathbf{v}^t$. We use $\mathcal{A}(\mathbf{h}^t)_i$ to denote the allocation v_i^t at round t. We now define the likelihood functions.

Definition 3.9 (Likelihood functions)**.** Let $p(v_i^t, o_i^t; \lambda) := \Pr[o_i^t; \lambda, v_i^t]$ denote the (censored) likelihood of discovering o_i^t candidates given an allocation v_i^t to group i assuming the candidate distribution follows $\mathcal{C}(\lambda)$. We write $\ell(v_i^t, o_i^t; \lambda)$ as $\log p(v_i^t, o_i^t; \lambda)$. So, given any history $\mathbf{h}^t$, the *empirical* log-likelihood function for group i is

$$\hat{\mathcal{L}}_i\left(\mathbf{h}^t,\lambda\right)=\frac{1}{t}\sum_{s=1}^{t}\ell\left(\mathcal{A}(\mathbf{h}^s)_i,o_i^s;\lambda\right).$$

The expected log-likelihood function given the history of allocations but over the randomness of the candidacy distribution can be written as

$$\mathcal{L}_i^*\left(\mathbf{h}^t,\lambda\right)=\frac{1}{t}\sum_{s=1}^{t}\mathbb{E}\left[\ell\left(\mathcal{A}(\mathbf{h}^s)_i,o_i^s;\lambda\right)\right],$$

where the expectation is over the randomness of o_i^s drawn from $\mathcal{C}_o(\lambda^*,\mathcal{A}(\mathbf{h}^s)_i)$.

Proof of Theorem 3.7

To prove Theorem 3.7, we first show that *any* sequence of allocations selected by Algorithm 3.3 will eventually recover the true parameters. There are two conceptual difficulties here: the first is that standard convergence results typically leverage the assumption of *independence*, which does not hold in this case as Algorithm 3.3 computes *adaptive* allocations which depend on the allocations and feedback from previous rounds, the second is the potential censoring of these observations. Despite these difficulties, we give quantifiable rates with which the estimates converge to the true parameters. Next, we show that computing an optimal α-fair allocation using the estimated parameters will result in an allocation that is $(\alpha+4D)$-fair with respect to the true candidate distributions where D denotes the maximum total variation distance between the true and estimated Poisson distributions across all groups. Finally, we show that this allocation also achieves a utility that is comparable to the utility of an optimal $(\alpha-4D)$-fair allocation. We note that while Theorem 3.7 is only stated for Poisson distributions, our results can be generalized to any unimodal, single parameter, Lipschitz-continuous family of distributions (see Remark 3.23).

3.3.5 Closeness of the Estimated Parameters

Our argument can be stated at a high level as follows: for any group i and any history $\mathbf{h}^t$, the empirical log-likelihood converges to the expected log-likelihood for any sequence of allocations made by Algorithm 3.3 as formalized in Lemma 3.16. We then show in Lemma 3.17 that the closeness of the empirical and expected log-likelihoods implies that the maximizers of these quantities (corresponding to the estimated and true parameters) will also become close. Since in our analysis we consider the groups separately, we fix a group i throughout the rest of this section and drop the subscript i for convenience.

Our first lemma shows that the true underlying parameter λ^* uniquely maximizes $\mathbb{E}[\ell]$ for any allocation. Since $\mathcal{L}^*$ is just a sum of $\mathbb{E}[\ell]$ terms, it follows as a corollary that $\mathcal{L}^*$ is uniquely maximized at λ^* for any sequence of allocations. This is stated as Lemma 3.10 and Corollary 3.11.

Lemma 3.10. *For any v,* $\operatorname*{argmax}_{\lambda} \mathbb{E}_o[\ell(v, o; \lambda)] = \{\lambda^*\}$.

Proof. Notice that since the expected log-likelihood function is the average over time periods of individual $\ell(v_i^t, c_i^t, \lambda)$ terms, λ^* being the unique maximizer of each term individually will imply that it is the unique maximizer of the the expected log-likelihood function. Thus we aim to show that

$$\mathbb{E}\left[\ell\left(v^t, o^t, \lambda^*\right)\right] \geq \mathbb{E}\left[\ell\left(v^t, o^t, \lambda\right)\right].$$

Notice that this is true if and only if

$$\mathbb{E}\left[-\log\left(\frac{p(v^t, o^t, \lambda)}{p(v^t, o^t, \lambda^*)}\right)\right] \geq 0. \tag{3.2}$$

The left hand side of Inequality (3.2) is the KL-divergence between two distributions, which is always non-negative, and thus the inequality holds. Since the KL-divergence is zero if and only if $\lambda = \lambda^*$, we have that λ^* is the unique maximizer of the expected log-likelihood. □

Corollary 3.11. *For any $\boldsymbol{h}^t$,* $\operatorname*{argmax}_{\lambda} \mathcal{L}^*(\lambda, \boldsymbol{h}^t) = \{\lambda^*\}$.

Lemma 3.12. $\left|\ell\left(v^t, o^t; \lambda\right)\right| \leq \max\left(\left|\ell\left(\mathcal{V}, \mathcal{V}; \lambda_{\min}\right)\right|, \left|\ell\left(\mathcal{V}-1, \mathcal{V}; \lambda_{\min}\right)\right|, \left|\ell\left(1, 0; \lambda_{\max}\right)\right|\right).$

Proof. The Poisson's PMF is unimodal and achieves its maximum at λ, where $p(v^t, o^t; \lambda)$ will be at most one, meaning $\ell(v^t, o^t; \lambda) \leq 0$. In order to bound the absolute value, we will bound how small ℓ can be.

For uncensored observations, the minimum log-likelihood is achieved at either 0 or at $\mathcal{V}-1$ due to unimodality. In this case, the choice of λ that can result in the minimum value is at $\lambda_{\max}$ or

$\lambda_{\min}$, respectively. In the case of a censored observation, $\ell(v^t, o^t; \lambda) = \log(1 - F(v^t - 1; \lambda))$. So the minimum will be achieved at $\ell(\mathcal{V}, \mathcal{V}; \lambda_{\min})$. □

Next we show that for any fixed λ, with high probability over the randomness of $\{o^t\}$, $\hat{\mathcal{L}}$ converges to $\mathcal{L}^*$ for any sequence of allocations $\{v^t\}$ that Algorithm 3.3 could have chosen.

Lemma 3.13. *For any $\lambda \in [\lambda_{min}, \lambda_{max}]$ and any $\boldsymbol{h}^t$*

$$\Pr\left[\left|\hat{\mathcal{L}}(\boldsymbol{h}^t, \lambda) - \mathcal{L}^*(\boldsymbol{h}^t, \lambda)\right| > \varepsilon\right] \leq 2e^{-\frac{t\varepsilon^2}{2C^2}},$$

where C is a constant, and in the case of Poisson distribution

$$C = \frac{1}{2}\max\Big(\left|\ell\left(\mathcal{V}, \mathcal{V}; \lambda_{min}\right)\right|, \left|\ell\left(\mathcal{V} - 1, \mathcal{V}; \lambda_{min}\right)\right|, \left|\ell\left(1, 0; \lambda_{\max}\right)\right|\Big).$$

Proof. Let $\mathcal{A}(h^s)$ denote the allocation to the group we are considering. We define Q^t as follows.

$$Q^t := t\left(\hat{\mathcal{L}}\left(\mathbf{h}^t, \lambda\right) - \mathcal{L}^*\left(\mathbf{h}^t, \lambda\right)\right) = \sum_{s=1}^{t} \ell\left(\mathcal{A}(h^s), o^s; \lambda\right) - \sum_{s=1}^{t} \mathbb{E}\left[\ell\left(\mathcal{A}(h^s), o^s; \lambda\right)\right].$$

Q^t is the sum of the difference between each period's observed and expected conditional log-likelihood function and is a martingale, as $\mathbb{E}[Q^{t+1}|Q^t] = Q^t$. Moreover, its terms form a bounded difference sequence since $\ell(\mathcal{A}(h^s), o^s; \lambda)$ is continuous in o_i^s with $o_i^s \in [0, \mathcal{V}]$ and $\lambda \in [\lambda_{\min}, \lambda_{\max}]$. In particular, we show in Lemma 3.12 that $\ell(v^t, o^t; \lambda)| \leq 2C$.

Since $\{Q^t\}$ is a bounded martingale difference sequence, we can apply Azuma's inequality to get

$$\Pr\left[\left|Q^t - Q^0\right| \geq t\varepsilon\right] \leq 2e^{-\frac{t\varepsilon^2}{2C^2}}.$$

Rearranging gives the claim. □

For k values of λ, taking the union bound and setting $\varepsilon = \sqrt{2C^2 \ln(2k\mathcal{G}/\delta)/t}$ provides the following corollary.

Corollary 3.14. *Let Λ be a set of k values in $[\lambda_{\min}, \lambda_{\max}]$. Then with probability at least $1 - \delta/\mathcal{G}$*

$$\max_{\lambda \in \Lambda}\left|\hat{\mathcal{L}}(\boldsymbol{h}^t, \lambda) - \mathcal{L}^*(\boldsymbol{h}^t, \lambda)\right| \leq \sqrt{\frac{2C^2 \ln(\frac{2k\mathcal{G}}{\delta})}{t}},$$

where C is as in Lemma 3.13.

We now need to show that the likelihood functions are Lipschitz-continuous:

Lemma 3.15. *For any λ, $\lambda' \in [\lambda_{\min}, \lambda_{\max}]$ such that $|\lambda - \lambda'| < \varepsilon$, we have that $|\ell(v_i^t, o_i^t, \lambda) - \ell(v_i^t, o_i^t, \lambda')| \leq b\varepsilon$ for some constant b.*

Proof. A differentiable function is Lipschitz-continuous if and only if its derivative is bounded, so, by definition,

$$\ell\left(v^t o^t; \lambda\right) := \begin{cases} \log\left(\frac{e^{-\lambda}\lambda^{o^t}}{o^t!}\right), & o^t < v^t, \\ \log\left(1 - F\left(v^t - 1; \lambda\right)\right), & \text{otherwise.} \end{cases}$$

In the uncensored case ($o^t < v^t$), we have that

$$\ell\left(v^t, o^t, \lambda\right) = -\lambda + o^t \log\lambda - \log o^t! \implies \frac{\partial \ell}{\partial \lambda} = -1 + \frac{o^t}{\lambda}.$$

For $\lambda \in [\lambda_{\min}, \lambda_{\max}]$ with $\lambda_{\min} > 0$, this function is continuous and its domain is bounded, so its image is bounded.

In the censored case ($o^t = v^t$), we can write

$$\ell\left(v^t, o^t; \lambda\right) = \log\left(1 - F\left(v^t - 1; \lambda\right)\right) = \log\left(\sum_{k=v^t}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!}\right) = -\lambda + \log\left(\sum_{k=v^t}^{\infty} \frac{\lambda^k}{k!}\right).$$

Again taking the derivative, we get

$$\begin{aligned} \frac{\partial \ell}{\partial \lambda} &= -1 + \frac{\frac{\partial}{\partial \lambda} \sum\limits_{k=v^t}^{\infty} \frac{\lambda^k}{k!}}{\sum\limits_{k=v^t} \frac{\lambda^k}{k!}} \\ &= -1 + \frac{\frac{\partial}{\partial \lambda}\left[e^{\lambda} - \sum\limits_{k=0}^{v^t-1} \frac{\lambda^k}{k!}\right]}{\sum\limits_{k=v^t}^{\infty} \frac{\lambda^k}{k!}} \\ &= -1 + \frac{\lambda - \sum\limits_{k=1}^{v^t-1} \frac{\lambda^{k-1}}{(k-1)!}}{\frac{\lambda^k}{k!}}. \end{aligned}$$

The fraction is the quotient of two continuous functions and the denominator is non-zero, so $\partial\ell/\partial\lambda$ is continuous in λ. Since the image of a continuous function on a compact set remains compact, $\partial\ell/\partial\lambda$ is bounded for all t. Thus $\ell(v^t, o^t; \lambda)$ is Lipschitz-continuous in this case as well. □

We next consider the rate of convergence of the empirical log-likelihood to the expected log-likelihood.

Lemma 3.16. *With probability at least $1-\delta/\mathcal{G}$, for any t and any $\boldsymbol{h}^t$ observed by Algorithm 3.3,*

$$\sup_{\lambda\in[\lambda_{min},\lambda_{max}]}\left|\hat{\mathcal{L}}\left(\boldsymbol{h}^t,\lambda\right)-\mathcal{L}^*\left(\boldsymbol{h}^t,\lambda\right)\right|\leq\mathcal{O}\left(\sqrt{\frac{\ln(t\mathcal{G}/\delta)}{t}}\right).$$

Proof. Define an ε-net $N_\varepsilon=\{\lambda_{\min},\lambda_{\min}+\varepsilon,\lambda_{\min}+2\varepsilon,\dots,\lambda_{\max}\}$ and let $k=|N_\varepsilon|$ denote its cardinality, so $k=\lceil\lambda_{\max}-\lambda_{\min}\rceil/\varepsilon$. Note that for any $\lambda\in[\lambda_{\min},\lambda_{\max}]$, there exists $\lambda'\in N_\varepsilon$ such that $|\lambda-\lambda'|\leq\varepsilon$.

By Corollary 3.14, for $\Lambda=\{\lambda_1,\lambda_2,\dots,\lambda_k\}$, with probability $1-\delta$,

$$\max_{\lambda\in\Lambda}\left|\hat{\mathcal{L}}\left(\mathbf{h}^t,\lambda\right)-\mathcal{L}^*\left(\mathbf{h}^t,\lambda\right)\right|\leq\sqrt{\frac{2C^2}{t}\ln\left(\frac{2k\mathcal{G}}{\delta}\right)}. \tag{3.3}$$

Now, for any $\lambda\in[\lambda_{\min},\lambda_{\max}]$ by the triangle inequality we have that

$$\begin{aligned}\left|\hat{\mathcal{L}}\left(\mathbf{h}^t,\lambda\right)-\mathcal{L}^*\left(\mathbf{h}^t,\lambda\right)\right|&\leq\left|\hat{\mathcal{L}}\left(\mathbf{h}^t,\lambda\right)-\hat{\mathcal{L}}\left(\mathbf{h}^t,\lambda'\right)\right|\\&\quad+\left|\hat{\mathcal{L}}\left(\mathbf{h}^t,\lambda'\right)-\mathcal{L}^*\left(\mathbf{h}^t,\lambda'\right)\right|\\&\quad+\left|\mathcal{L}^*\left(\mathbf{h}^t,\lambda'\right)-\mathcal{L}^*\left(\mathbf{h}^t,\lambda\right)\right|.\end{aligned}$$

By Lemma 3.15, the first and third term are at most εb where b is the Lipschitz constant in Lemma 3.15. Applying this to the closest $\lambda_k\in N_\varepsilon$ and noting that the inequality in Inequality (3.3) binds on the middle term with C as in Lemma 3.13, we have

$$\begin{aligned}\left|\hat{\mathcal{L}}\left(\mathbf{h}^t,\lambda\right)-\mathcal{L}^*\left(\mathbf{h}^t,\lambda\right)\right|&\leq\varepsilon b+\sqrt{\frac{2C^2\ln\left(\frac{2k\mathcal{G}}{\delta}\right)}{t}}+\varepsilon b\\&\leq 2\varepsilon b+\sqrt{\frac{2C^2\ln\left(\frac{2k\mathcal{G}}{\delta}\right)}{t}}\\&\leq 2\varepsilon b+\sqrt{\frac{2C^2\ln\left(\frac{2\mathcal{G}\lceil\lambda_{\max}-\lambda_{\min}\rceil}{\varepsilon\delta}\right)}{t}}.\end{aligned}$$

Setting $\varepsilon=1/t$ yields the claim. As $t\to\infty$, the difference approaches zero. □

The true and estimated parameters for each group correspond to the maximizers of the expected and empirical log-likelihoods, respectively (see Corollary 3.11). We next show that closeness of the empirical and expected log-likelihoods implies that the true and estimated parameters are also close.

Lemma 3.17. *Let $\hat{\lambda}$ denote the estimate of the Algorithm 3.3 after $T_{\max}$ rounds. Then with probability at least $1-\delta/\mathcal{G}$, $|\hat{\lambda}-\lambda^*|<\varepsilon$.*

Proof. Since Corollary 3.11 gives that $\mathcal{L}(\mathbf{h}^t,\lambda)$ has a unique maximizer at λ^* and Corollary 3.19 gives that there exists some $\eta(\varepsilon)$ such that for any λ' satisfying $|\mathcal{L}^*(\mathbf{h}^t,\lambda')-\mathcal{L}^*(\mathbf{h}^t,\lambda^*)|<\eta(\varepsilon)$, we must have that $|\lambda'-\lambda^*|<\varepsilon$. We denote $\eta(\varepsilon)$ by η for brevity and define the empirical maximizer $\hat{\lambda}$ to be the maximizer of $\hat{\mathcal{L}}(\mathbf{h}^t,\lambda)$. That is,

$$\hat{\lambda}\in\underset{\lambda\in[\lambda_{\min},\lambda_{\max}]}{\operatorname{argmax}}\ \hat{\mathcal{L}}\left(\mathbf{h}^t,\lambda\right). \tag{3.4}$$

Applying Lemma 3.16 implies that for any $\mathbf{h}^t$ with $t>T_{\max}$ and $\lambda'\in[\lambda_{\min},\lambda_{\max}]$, with probability at least $1-\delta/\mathcal{G}$,

$$\left|\hat{\mathcal{L}}(\mathbf{h}^t,\lambda')-\mathcal{L}^*(\mathbf{h}^t,\lambda')\right|<\frac{\eta}{2}.$$

In particular, we must have that

$$\left|\hat{\mathcal{L}}\left(\mathbf{h}^t,\lambda^*\right)-\mathcal{L}^*\left(\mathbf{h}^t,\lambda^*\right)\right|<\frac{\eta}{2}\qquad\text{and}\qquad\left|\hat{\mathcal{L}}\left(\mathbf{h}^t,\hat{\lambda}\right)-\mathcal{L}^*\left(\mathbf{h}^t,\hat{\lambda}\right)\right|<\frac{\eta}{2}. \tag{3.5}$$

Since $\hat{\lambda}$ is a maximizer of Inequality (3.4), we have that

$$\hat{\mathcal{L}}\left(\mathbf{h}^t,\hat{\lambda}\right)\geq\hat{\mathcal{L}}\left(\mathbf{h}^t,\lambda^*\right)\geq\mathcal{L}^*\left(\mathbf{h}^t,\lambda^*\right)-\frac{\eta}{2},$$

where the last inequality is by Inequality (3.5). This implies that

$$\begin{aligned}\mathcal{L}^*\left(\mathbf{h}^t,\hat{\lambda}\right)&>\hat{\mathcal{L}}\left(\mathbf{h}^t,\hat{\lambda}\right)-\frac{\eta}{2}\\&>\mathcal{L}^*\left(\mathbf{h}^t,\lambda^*\right)-\frac{\eta}{2}-\frac{\eta}{2}\\&=\mathcal{L}^*\left(\mathbf{h}^t,\lambda^*\right)-\eta,\end{aligned}$$

where the last inequality is by Inequality (3.5). Therefore, $\mathcal{L}^*(\mathbf{h}^t,\hat{\lambda})>\mathcal{L}^*(\mathbf{h}^t,\lambda^*)-\eta$, and thus Corollary 3.19 gives that $|\hat{\lambda}-\lambda^*|<\varepsilon$. □

Combining Lemma 3.17 with a union bound over all groups show that, with probability $1-\delta$, if Algorithm 3.3 is run for $T_{\max}$ rounds, then $|\hat{\lambda}_i-\lambda_i^*|<\varepsilon$, for all i. Note that as $t\to\infty$, the maximum total variation distance D between the estimated and the true distribution will converge in probability to 0.

3.3.6 Fairness of the Allocation

Now that we have shown convergence, we next need to demonstrate that the fairness violation (i.e. the maximum difference in discovery probabilities over all pairs of groups) is linear in terms of D. Therefore, as the running time of the Algorithm 3.3 increases and hence, $D \to 0$, the fairness violation of $\hat{\mathbf{v}}$ approaches α.

Lemma 3.18. *Suppose that a continuous function $g(x) : [a, b] \mapsto \mathbb{R}$ has a unique maximizer x^*. Then for every $\varepsilon > 0$ there exists a value $\eta > 0$ such that $g(x^*) - g(x) < \eta$ implies $|x - x^*| < \varepsilon$. In particular, this η can be written as*

$$\eta = g(x^*) - \max_{x \in [a,b] \setminus X_\varepsilon} g(x),$$

where $\mathcal{B}_\varepsilon(x^*)$ denotes the open ε-ball centered at x^*. When g is concave and differentiable, η can be found by evaluating g at a constant number of points.

Proof. Let X_ε be the ε-radius open ball centered at x^*, and let Θ be $[a, b] \setminus X_\varepsilon$. Since X_ε is open, Θ is closed and bounded, and therefore compact. Since g is continuous, the restriction of g to Θ has some maximum $g(\hat{x})$ for some (not necessarily unique) $\hat{x} \in \Theta$.

Observe that, if for any $x \in [a, b]$, we have that $g(x) > g(\hat{x})$, then x must be in X_ε. Otherwise $\hat{x}$ would not be a maximizer of the restriction of f to Θ. Choose $\eta = g(x^*) - g(\hat{x})$. Then, because $g(x) > g(\hat{x})$, we have that $g(x^*) - g(x) < g(x^*) - g(\hat{x}) = \eta$. Therefore, $|g(x^*) - g(x)| < \eta$ implies $|x^* - x| < \varepsilon$, completing the proof of existence.

The dependence of η on ε is function-dependent, but by construction, $\eta(\varepsilon)$ can be computed by taking the maximum of g on $\Theta \setminus X_\varepsilon$ and subtracting it from $g(x^*)$. Notice that in the case of concavity and differentiability, this maximization problem is easy to calculate. If g is concave and differentiable over $[a, b]$, its restrictions to $[a, x^* - \varepsilon]$ and $[x^* + \varepsilon, b]$ are as well. A differentiable, concave function on an interval can only be maximized at an interior critical point or at one of the two end points. Hence if x^* is interior, it must be a critical point, and by concavity it is the unique critical point on $[a, b]$, so g can have no critical points on $\Theta \setminus X_\varepsilon$. Thus g restricted to $[a, x^* - \varepsilon]$ is maximized at either a or $x^* - \varepsilon$; similarly for g restricted to $[x^* + \varepsilon, b]$. On the other hand, if x^* is either a or b, there is just one interval in $\Theta \setminus X_\varepsilon$ to check, and checking the endpoints of that interval plus exhaust the possible maximizers. In either case, no more than 4 points need be checked; in contrast, without concavity or differentiability, finding the maximum on g on $\Theta \setminus X_\varepsilon$ could require more involved optimization techniques. □

Corollary 3.19. *For any fixed $\boldsymbol{h}^t$ and $\lambda \in [\lambda_{min}, \lambda_{max}]$, the following must hold true for $\mathcal{L}^*(\boldsymbol{h}^t, \lambda)$ whose unique maximizer is λ^*. For every ε, there exists a value $\eta > 0$ such that $\mathcal{L}^*(\boldsymbol{h}^t, \lambda^*) - \mathcal{L}^*(\boldsymbol{h}^t, \lambda) < \eta$ implies $|\lambda - \lambda^*| < \varepsilon$, where η is*

$$\mathcal{L}^*(\boldsymbol{h}^t, \lambda^*) - \max_{[\lambda_{min}, \lambda_{max}] \setminus X_\varepsilon} \mathcal{L}^*(\boldsymbol{h}^t, \lambda).$$

Fixing any group, we show that for any fixed allocation v, the difference between the discovery probabilities with respect to the true and estimated candidate distributions is proportional to the total variation distance between the true and estimated distributions.

Lemma 3.20. *Let v be any fixed allocation to the group. Then*

$$\left| f(v, \mathcal{C}(\lambda^*)) - f(v, \mathcal{C}(\hat{\lambda})) \right| \leq 2 D_{TV}(\mathcal{C}(\lambda^*), \mathcal{C}(\hat{\lambda})).$$

Proof. We have

$$\begin{aligned}
\left| f(v, \mathcal{C}(\lambda^*)) - f(v, \mathcal{C}(\hat{\lambda})) \right| &= \left| \mathop{\mathbb{E}}_{c \sim \mathcal{C}(\lambda^*)} \left[\frac{\min(v, c)}{c} \right] - \mathop{\mathbb{E}}_{\mathbf{c} \sim \mathcal{C}(\hat{\lambda})} \left[\frac{\min(v, c)}{c} \right] \right| \\
&= \left| \sum_{c=0}^{\infty} \frac{\min(v, c)}{c} \left(\Pr[c; \lambda^*] - \Pr[c; \hat{\lambda}] \right) \right| \\
&\leq \sum_{c=0}^{\infty} \frac{\min(v, c)}{c} \left| \Pr[c; \lambda^*] - \Pr[c; \hat{\lambda}] \right| \\
&\leq \sum_{c=0}^{\infty} \left| \Pr[c; \lambda^*] - \Pr[c; \hat{\lambda}] \right| \\
&\leq 2 D_{TV}(\mathcal{C}(\lambda^*), \mathcal{C}(\hat{\lambda})).
\end{aligned}$$

□

Formally, the fairness violation approaches α in the following sense.

Lemma 3.21. *Let $\hat{\boldsymbol{v}}$ denote the allocation returned by Algorithm 3.3 after $T_{\max}$ rounds. Then with probability at least $1-\delta$, $|f_i(\hat{v}_i) - f_j(\hat{v}_j)| \leq \alpha + 4D, \forall i, j \in [\mathcal{G}]$.*

Proof. For any $i, j \in [\mathcal{G}]$ we have that

$$\begin{aligned}
|f_i(\hat{v}_i) - f_j(\hat{v}_j)| &= \left|f_i(\hat{v}_i, C(\lambda_i^*)) - f_j(\hat{v}_j, C(\lambda_j^*))\right| \\
&\leq \left|f_i(\hat{v}_i, \mathcal{C}(\lambda_i^*)) - f_i(\hat{v}_i, \mathcal{C}(\hat{\lambda}_i))\right| \\
&\quad + \left|f_i(\hat{v}_i, C(\hat{\lambda}_i)) - f_j(\hat{v}_j, \mathcal{C}(\hat{\lambda}_j))\right| \\
&\quad + \left|f_j(\hat{v}_j, \mathcal{C}(\lambda_j^*)) - f_j(\hat{v}_j, \mathcal{C}(\hat{\lambda}_j))\right| \\
&\leq 2D_{TV}(\mathcal{C}(\lambda_i^*), \mathcal{C}(\hat{\lambda}_i)) + \alpha + 2D_{TV}(\mathcal{C}(\lambda_j^*), \mathcal{C}(\hat{\lambda}_j)) \\
&= \alpha + 4D.
\end{aligned}$$

The first inequality follows from the triangle inequality. In the second inequality, the second term can be bounded because Algorithm 3.2 returns an α-fair allocation with respect to its input distribution. The first and third term in the second inequality can be bounded by Lemma 3.20. Lemma 3.20 shows that for any fixed allocation the difference between the discovery probability with respect to the true and estimated candidate distributions in group i is proportional to the total variation distance between the true and estimated distributions. □

3.3.7 Utility of the Allocation

The final component of proving theorem 3.7 is to analyze the utility of the allocation returned by Algorithm 3.3. We know that in the limit as the number of rounds Algorithm 3.3 is run grows, the algorithm learns an optimal α-fair allocation.

Lemma 3.22. *Let $\hat{\boldsymbol{v}}$ denote the allocation returned by Algorithm 3.3 after $T_{\max}$ rounds. Then with probability at least $1-\delta$, $\chi(\hat{\boldsymbol{v}}) > \chi(\boldsymbol{w}^{\alpha-4D}) - 4D\mathcal{G}\mathcal{V}$.*

Proof. Consider the following optimization problem, $\mathcal{P}(\alpha, \{\lambda_i\}, \{\bar{\lambda}_i\}, \mathcal{V})$.

$$
\begin{array}{ll}
\max\limits_{\mathbf{v}} & \chi\left(\mathbf{v}, \{\mathcal{C}(\lambda_i)\}\right), \\
\text{subject to} & \left|f_i\left(v_i, \mathcal{C}(\bar{\lambda}_i)\right) - f_j\left(v_j, \mathcal{C}(\bar{\lambda}_i)\right)\right| \leq \alpha, \forall i \text{ and } j, \\
& \sum\limits_{i \in [\mathcal{G}]} v_i \leq \mathcal{V}, \\
& v_i \in \mathcal{N}.
\end{array}
$$

We can think of the above optimization problem as the case where the underlying candidate distributions used for the objective value and the fairness constraints are different. Let us write $\mathcal{A}(\alpha, \{\lambda_i\}, \{\bar{\lambda}_i\}, \mathcal{V})$ to denote an optimal allocation in the above optimization problem, $\mathcal{P}(\alpha, \{\lambda_i\}, \{\bar{\lambda}_i\}, \mathcal{V})$. So an optimal fair allocation and the allocation returned by Algorithm 3.3 can be written as $\mathcal{A}(\alpha, \{\lambda_i^*\}, \{\lambda_i^*\}, \mathcal{V})$ and $\mathcal{A}(\alpha, \{\hat{\lambda}_i\}, \{\hat{\lambda}_i\}, \mathcal{V})$, respectively.

Note that for any fixed allocation $\mathbf{v}$

$$
\left|\chi(\mathbf{v}, \{\mathcal{C}(\lambda_i^*)\}) - \chi(\mathbf{v}, \{\mathcal{C}(\hat{\lambda}_i)\})\right| = \left|\sum_{i \in \mathcal{G}} \sum_{\mathbf{c}=0}^{v_i} \mathcal{T}_i^*(c) - \sum_{i \in \mathcal{G}} \sum_{\mathbf{c}=0}^{v_i} \hat{\mathcal{T}}_i(c)\right| \leq 2\mathcal{G}\mathcal{V}D, \tag{3.6}
$$

where $\hat{\mathcal{T}}_i$ is the tail probability of $\mathcal{C}(\hat{\lambda}_i)$. This is because $|\mathcal{T}_i^*(c) - \hat{\mathcal{T}}_i(c)|$ is bounded above by $2D_{TV}(\mathcal{C}(\lambda_i), \mathcal{C}(\hat{\lambda}_i))$. In other words, even when the underlying candidate distribution changes for the objective value, an allocation value can change by at most $2\mathcal{G}\mathcal{V}D$.

Now observe that

$$
\begin{aligned}
\chi(\hat{\mathbf{v}}, \{\mathcal{C}(\lambda_i^*)\}) &\geq \chi(\hat{\mathbf{v}}, \{\mathcal{C}(\hat{\lambda}_i)\}) - 2\mathcal{G}\mathcal{V}D \\
&= \chi\left(\mathcal{A}\left(\mathcal{V}, \{\mathcal{C}(\hat{\lambda}_i)\}, \{\mathcal{C}(\hat{\lambda}_i)\}, \alpha\right), \{\mathcal{C}(\hat{\lambda}_i)\}\right) - 2\mathcal{G}\mathcal{V}D \\
&\geq \chi\left(\mathcal{A}\left(\mathcal{V}, \{\mathcal{C}(\hat{\lambda}_i)\}, \{\mathcal{C}(\lambda_i^*)\}, \alpha - 4D\right), \{\mathcal{C}(\hat{\lambda}_i)\}\right) - 2\mathcal{G}\mathcal{V}D \\
&\geq \chi\left(\mathcal{A}\left(\mathcal{V}, \{\mathcal{C}(\lambda_i^*)\}, \{\mathcal{C}(\lambda_i^*)\}, \alpha - 4D\right), \{\mathcal{C}(\lambda_i^*)\}\right) - 4\mathcal{G}\mathcal{V}D \\
&= \chi(\mathbf{w}^{\alpha - 4D}) - 4D\mathcal{G}\mathcal{V}.
\end{aligned}
$$

The inequalities in the first and fourth lines are by Equation (3.6), which shows how the utility deteriorates when the underlying distribution for the objective function changes. The inequality in the third line follows from Lemma 3.21, as any $(\alpha - 4D)$ fair allocation is a feasible allocation to $\mathcal{P}(\mathcal{V}, \{\mathcal{C}(\hat{\lambda}_i)\}, \{\mathcal{C}(\hat{\lambda}_i)\}, \alpha)$, and $\mathcal{A}(\alpha, \{\lambda_i\}, \{\bar{\lambda}_i\}, \mathcal{V})$ is an optimal solution to this problem. □

Remark 3.23. *Although we assumed Poisson distributions in this section, all our results hold for any unimodal, single-parameter, Lipschitz-continuous distribution whose parameter is drawn from* *a compact set. However, the convergence rate of Theorem 3.7 depends on the quantity $\eta(\varepsilon)$ which* depends *on the family of distributions used to model the candidate distributions.*

3.4 The Random Discovery Model

Finally, we consider the *random model* of discovery. In the random model, when v_i units are allocated to a group with c_i candidates, the number of discovered candidates is a random variable corresponding to the number of candidates that appear in a uniformly random sample of v_i individuals from a group of size m_i. Equivalently, when v_i units are allocated to a group of size m_i with c_i candidates, the number of candidates discovered by $\text{disc}(\cdot)$ is a random variable $\text{disc}(v_i, c_i) \triangleq o_i$, where o_i is drawn from the hypergeometric distribution with parameters m_i, c_i and v_i. Furthermore, the expected number of candidates discovered when allocating v_i units to group i is $\mathbb{E}[\text{disc}(v_i, c_i)] = v_i\,\mathbb{E}[c_i]/m_i$.

For simplicity, at first we assume $m_i \geq \mathcal{V}$ for all i, and we completely relax this assumption towards the end of the section. Moreover, let $\mu_i = \mathbb{E}[c_i]/m_i$ denote the expected fraction of candidates in group i and without loss of generality that $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_{\mathcal{G}}$.

3.4.1 Optimal Allocation

In this section, we characterize optimal allocations. Note that the expected number of candidates discovered by the allocation choice $v_i \leq m_i$ in group i is simply $v_i \mu_i$. This suggests a simple algorithm to compute $\mathbf{w}^*$: allocating every unit of the resource to group 1. More generally, let $\mathcal{G}^* = \{i \mid \mu_i = \mu_1\}$ denote the subset of groups with the highest expected number of candidates. An allocation is optimal if and only if it only allocates *all* resources to groups in $\mathcal{G}^*$.

3.4.2 Properties of Fair Allocations

We next discuss the properties of fair allocations in the random discovery model. First, we point out that the discovery probability can be simplified as

$$f_i(v_i) = \mathop{\mathbb{E}}_{c_i \sim \mathcal{C}_i}\left[\frac{c_i v_i / m_i}{c_i}\right] = \frac{v_i}{m_i}.$$

So an allocation is α-fair in the random model if $|v_i/m_i - v_j/m_j| \leq \alpha$ for all groups i and j. Therefore, fair allocations (roughly) distribute resources in proportion to the size of the groups, essentially ignoring the candidate distributions within each group.

3.4.3 Price of Fairness

The *price of fairness* measures the discrepancy in the utility to the allocator between the fair and the optimal allocations. Because the optimal and fair allocations in this setting are both straightforward to describe and potentially dramatically different, we can give some bounds on this price in the random discovery model.

Definition 3.24 (Worst-case price of fairness)**.** Consider the random discovery model and let $\alpha \in [0, 1]$. We define the *price of fairness* as

$$\mathrm{PoF}(\alpha) = \max_{\mathcal{C}} \frac{\chi(\mathbf{w}^*, \mathcal{C})}{\chi(\mathbf{w}^\alpha, \mathcal{C})}.$$

where $\mathcal{C}$ ranges over all possible candidate distributions.

We can fully characterize this worst-case price of fairness in the random discovery model.

Theorem 3.25 (Bounds on worst-case price of fairness)**.** *The price of fairness in the random discovery model is*

$$PoF(\alpha) = \begin{cases} 1, & \frac{\mathcal{V}}{m_1} \leq \alpha, \\ \frac{M}{m_1 + \alpha(M - m_1)}, & \frac{\mathcal{V}}{m_1} > \alpha. \end{cases}$$

The price of fairness in the random model can be as high as M/m_1 in the worst case. If all groups are identically sized, this grows linearly with the number of groups.

Proof. First, we write down the integer program to compute an optimal α-fair allocation in the random model as

$$\begin{aligned} \max_{\mathbf{v}=\{v_1,\ldots,v_{\mathcal{G}}\}} \quad & \sum_{i=1}^{\mathcal{G}} \frac{v_i\,\mathbb{E}[c_i]}{m_i}, \\ \text{subject to} \quad & \left|\frac{v_i}{m_i} - \frac{v_j}{m_j}\right| \le \alpha, \forall i \text{ and } j, \\ & \sum_{i=1}^{\mathcal{G}} v_i \le \mathcal{V}, \\ & v_i \in \mathbb{N}, \forall i. \end{aligned}$$

Observe that when $\mathcal{V}/m_1 \le \alpha$ the allocation that sends all of the units of the resource to group 1 is both optimal and α-fair. For the case that $\mathcal{V}/m_1 > \alpha$, we will first provide an upper bound on the price of fairness by demonstrating an allocation $\mathbf{v}$ which is is α-fair and use it to show that imposing fairness does not decrease the total number of candidates discovered by $\mathbf{v}$ as compared to an optimal α-fair allocation. Then, we will construct specific candidate distributions $\mathcal{C}'$ and compute the price of fairness to show that the upper bound is tight.

Consider the following allocation $\mathbf{v}$.

$$v_i = \begin{cases} \left(\frac{\mathcal{V}+\alpha(M-m_1)}{M}\right) m_1, & i = 1, \\ \left(\frac{\mathcal{V}-\alpha m_1}{M}\right) m_i, & \text{otherwise.} \end{cases} \tag{3.7}$$

We show that $\mathbf{v}$ is a feasible α-fair allocation. To show feasibility, observe that

$$\sum_{i\in[\mathcal{G}]} v_i = \frac{\mathcal{V}-\alpha m_1}{M} \sum_{i\in[\mathcal{G}]} m_i + \alpha m_1 = \mathcal{V}.$$

To show that $\mathbf{v}$ is α-fair observe that

$$|f_1(v_1) - f_i(v_i)| = \left|\frac{v_1}{m_1} - \frac{v_i}{m_i}\right| = \left|\frac{\mathcal{V}+\alpha(M-m_1)}{M} - \frac{\mathcal{V}-\alpha m_1}{M}\right| = \alpha$$

$$|f_i(v_i) - f_j(v_j)| = 0, \text{for all } i, j \neq 1.$$

Since $\mathbf{v}$ is a feasible α-fair allocation, for any $\mathcal{C}$, we have that

$$\chi(\mathbf{w}^{\alpha}, \mathcal{C}) \ge \chi(\mathbf{v}, \mathcal{C}) \ge \mu_1 v_1,$$

where the last inequality is derived by counting only the candidates that allocation $\mathbf{v}$ discovers in group 1 according to the random model and ignoring all the discoveries in other groups. Moreover, for any $\mathcal{C}$, $\chi(\mathbf{w}^*, \mathcal{C}) = \mu_1 \mathcal{V}$ by the argument in Section 3.4.1. Therefore,

$$\text{PoF} = \max_{\mathcal{C}} \frac{\chi(\mathbf{w}^*, \mathcal{C})}{\chi(\mathbf{w}^\alpha, \mathcal{C})} \leq \frac{\mu_1 \mathcal{V}}{\mu_1 v_1} = \frac{\mathcal{V} M}{m_1 \left(\mathcal{V} + \alpha(M - m_1)\right)} \leq \frac{M}{m_1 + \alpha(M - m_1)},$$

where the last inequality uses the assumption that $\mathcal{V} \leq m_1$.

To derive the lower bound on the price of fairness, we construct a candidate distribution $\mathcal{C}'$ and perform the computation. We then show that the lower bound matches the upper bound, so the analysis is tight.

To construct $\mathcal{C}'$ assume all groups have size $\mathcal{V}$ i.e. $m_i = \mathcal{V}$ for all i. Furthermore, assume group 1 has $\mathcal{V}$ candidates and the rest of the groups have 0 candidates deterministically. Then the optimal allocation $\mathbf{w}^*$ is to send all the $\mathcal{V}$ units of resource to group 1, and doing so will discover $\mathcal{V}$ candidates (since $\mu_1 = 1$). As for the optimal α-fair allocation, we show that $\mathbf{w}^\alpha = \mathbf{v}$ where $\mathbf{v}$ is the same allocation as the allocation used in the proof for the upper bound.

For $\mathcal{C}'$, because all groups have the same size, $v_i = v_j$ for all $i, j \neq 1$. Since $\Sigma_{i \in [\mathcal{G}]} v_i = \mathcal{V}$, $v_i = (\mathcal{V} - v_1)/(\mathcal{G} - 1)$ for all $i \neq 1$. Now, any feasible α-fair allocation $\mathbf{v}'$ must have $v_1' \leq v_1$. Assume for the sake of contradiction that $v_1' > v_1$. Then, after assigning v_1' units of resource to group 1, the remaining units that will be strictly less than $\mathcal{V} - v_1$ will be distributed among the remaining $\mathcal{G} - 1$ groups. By the pigeonhole principle, there must exist at least one group j such that

$$v_j' < \frac{\mathcal{V} - v_1}{\mathcal{G} - 1} = v_j = \left(\frac{\mathcal{V} - \alpha m_1}{M}\right) m_j.$$

Now observe that

$$\left|f_1(v_1') - f_j(v_j')\right| = \left|\frac{v_1'}{m_1} - \frac{v_j'}{m_j}\right| > \left|\frac{\mathcal{V} + \alpha(M - m_1)}{M} - \frac{\mathcal{V} - \alpha m_1}{M}\right| = \alpha.$$

So $\mathbf{v}'$ cannot be α-fair. Therefore, $\mathbf{v}$ must be an optimal α-fair allocation since in $\mathbf{v}$ the maximum number of units of resources are allocated to group 1 which is the only group that contains candidates.

Note that the number of candidates discovered by $\mathbf{w}^\alpha$ is exactly v_1 since $\mu_1 = 1$. So given the

$\mathbf{w}^\alpha$ for $\mathcal{C}'$ we can compute the price of fairness as follows.

$$\begin{aligned}
\text{PoF} &= \frac{\chi(\mathbf{w}^*, \mathcal{C}')}{\chi(\mathbf{w}^\alpha, \mathcal{C}')} \\
&= \frac{\mathcal{V}}{v_1} \\
&= \frac{\mathcal{V} M}{(\mathcal{V} + \alpha(M - m_1))\, m_1} \\
&= \frac{m_1 M}{(m_1 + \alpha(M - m_1))\, m_1} = \frac{M}{m_1 + \alpha(M - m_1)}.
\end{aligned}$$

This lower bound matches the upper bound. □

3.4.4 Relaxing the Assumption of $\mathcal{V} \leq m_i$.

In this section we relax the assumption that $\mathcal{V} \leq m_i$ for all groups i. We first show how an optimal allocation can be computed using a greedy algorithm. Recall that we have assumed $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_{\mathcal{G}}$. The algorithm allocates $v_1 = \min(\mathcal{V}, m_1)$ units of resource to group 1. And recurse with the remaining $\mathcal{V} - v_1$ resources on the rest of the groups. If this algorithm allocates resources to groups 1 through k then the expected utility of the algorithm can be written as

$$\sum_{i=1}^{k-1} \mu_i m_i + \mu_k v_k = \sum_{i=1}^{k-1} \mu_i m_i + \mu_k(\mathcal{V} - \sum_{i=1}^{k-1} m_i).$$

In the case that $\mathcal{V} \leq m_i$, the algorithm allocates all the resources to group 1 without any leftover resources for other groups. We note that an optimal α-fair allocation can still be computed with the same integer program. Furthermore, the lower bound on the price of fairness continues to hold even when we relax this assumption.

However, our upper bound analysis on the price of fairness breaks when we relax the assumption. A similar upper bound can be derived in this case, but the analysis requires considering a large number of cases, depending on which groups receive as many units of the resource as they have people. We do not investigate this direction as the lower bound on the price of fairness shows that in the random model it can still be quite high even without this assumption on the group sizes.

3.5 Conclusion and Future Directions

Our presentation of allocative fairness provides a family of fairness definitions, modularly parameterized by *discovery models*. What counts as 'fair' or 'unfair' depends a great deal on the choice of discovery model, which makes explicit what would otherwise be unstated assumptions about the process of tasks like distributing resources across communities. The random and precision models of discovery studied in this paper represent two extreme points of a spectrum. The precision model assumes that resources can be perfectly targeted to candidates and the random model assumes the allocator has no power to direct the resource beyond the initial distribution. An interesting direction for future work is to study discovery models that lie in between these two.

We have also made a number of simplifying assumptions that could be relaxed. For example, we assumed the candidate distributions are *stationary* — fixed independently of the actions of the algorithm. Of course, the deployment of police officers can *change* crime distributions. Modeling this kind of dynamics, and designing learning algorithms that perform well in such dynamic settings would be interesting. Finally, we have assumed that the same discovery model applies to all groups. One friction to fairness that one might reasonably conjecture is that the discovery model may differ between groups — being closer to the precision model for one group, and closer to the random model for another. We leave the study of these extensions to future work.

Bibliography

A. Agarwal, P. Bartlett, and M. Dama. Optimal allocation strategies for the dark pool problem. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, pages 9–16, 2010.

H. Bastani, M. Bayati, and K. Khosravi. Exploiting the natural exploration in contextual bandits. *CoRR*, abs/1704.09011, 2017.

R. Berk, H. Heidari, S. Jabbari, M. Joseph, M. Kearns, J. Morgenstern, S. Neel, and A. Roth. A convex framework for fair regression. *CoRR*, abs/1706.02409, 2017.

R. Berk, H. Heidari, S. Jabbari, M. Kearns, and A. Roth. Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, 2018.

S. Bird, S. Barocas, K. Crawford, F. Diaz, and H. Wallach. Exploring or exploiting? social and ethical implications of autonomous experimentation in AI. 2016.

T. Calders, A. Karim, F. Kamiran, W. Ali, and X. Zhang. Controlling attribute effect in linear regression. In *Proceedings of 13th International Conference on Data Mining*, pages 71–80, 2013.

F. Chierichetti, R. Kumar, S. Lattanzi, and S. Vassilvitskii. Fair clustering through fairlets. In *Proceedings of the 31th Annual Conference on Neural Information Processing Systems*, pages 5029–5037, 2017.

S. Corbett-Davies, E. Pierson, A. Feller, S. Goel, and A. Huq. Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 797–806, 2017.

C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science*, pages 214–226, 2012.

D. Ensign, S. Friedler, S. Neville, C. Scheidegger, and S. Venkatasubramanian. Runaway feedback loops in predictive policing. In *Conference on Fairness, Accountability and Transparency*, pages 160–171, 2018a.

D. Ensign, S. Frielder, S. Neville, C. Scheidegger, and S. Venkatasubramanian. Decision making with limited feedback. In *Proceedings of the 29th Conference on Algorithmic Learning Theory*, pages 359–367, 2018b.

K. Ganchev, M. Kearns, Y. Nevmyvaka, and J. W. Vaughan. Censored exploration and the dark pool problem. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence*, pages 185–194, 2009.

M. Hardt, E. Price, and N. Srebro. Equality of opportunity in supervised learning. In *Proceedings of the 30th Annual Conference on Neural Information Processing Systems*, pages 3315–3323, 2016.

S. Jabbari, M. Joseph, M. Kearns, J. Morgenstern, and A. Roth. Fairness in reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning*, pages 1617–1626, 2017.

M. Joseph, M. Kearns, J. Morgenstern, and A. Roth. Fairness in learning: classic and contextual bandits. In *Proceedings of the 30th Annual Conference on Neural Information Processing Systems*, pages 325–333, 2016.

S. Kannan, J. Morgenstern, A. Roth, B. Waggoner, and Z. S. Wu. A smoothed analysis of the greedy algorithm for the linear contextual bandit problem. In *Proceedings of the 32nd Annual Conference on Neural Information Processing Systems*, 2018.

J. Kleinberg, S. Mullainathan, and M. Raghavan. Inherent trade-offs in the fair determination of risk scores. In *Proceedings of the 8th Conference on Innovations in Theoretical Computer Science*, pages 43:1–43:23, 2017.

L. Liu, S. Dean, E. Rolf, M. Simchowitz, and M. Hardt. Delayed impact of fair machine learning. In *Proceedings of the 35th International Conference on Machine Learning*, pages 3156–3164, 2018.

K. Lum and W. Isaac. To predict and serve? *Significance*, pages 14–18, October 2016.

A. Procaccia. Cake cutting: Not just child's play. *Communications of the ACM*, 56(7):78–87, 2013.

M. Raghavan, A. Slivkins, J. W. Vaughan, and Z. S. Wu. The externalities of exploration and how data diversity helps exploitation. In *Proceedings of the 31st Conference On Learning Theory*, pages 1724–1738, 2018.

B. Woodworth, S. Gunasekar, M. Ohannessian, and N. Srebro. Learning non-discriminatory predictors. In *Proceedings of the 30th Conference on Learning Theory*, pages 1920–1953, 2017.

M. B. Zafar, I. Valera, M. Gomez-Rodriguez, and K. Gummadi. Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1171–1180, 2017.

R. Zemel, Y. Wu, K. Swersky, T. Pitassi, and C. Dwork. Learning fair representations. In *Proceedings of the 30th International Conference on Machine Learning*, pages 325–333, 2013.

Chapter 4

The Price of Privacy in the Keynesian Beauty Contest

4.1 Introduction

In recent years, the mathematical study of privacy has become a major subject of inquiry. Much of the impetus for this work has been a series of data breaches and deanonymization of seemingly safe private information, perhaps most famously in the use of IMDb reviews to attack Netflix's data set over a decade ago (Narayanan and Shmatikov, 2008). Models such as differential privacy alleviate this problem by providing formal guarantees about how much about any individual an adversary can learn from the release of some statistic computed from a dataset. However, such techniques are generally predicated on the presence of a trusted central agent which applies the differentially private mechanism to the data it collects from individuals. Alternatively, in the case of the *local* differential privacy model, the agents are typically instructed in how to privatize their data.

In this work, we endogenize a notion of privacy in the absence of a trusted party to coordinate the mechanism. We analyze a formalization of a classical game called the *Keynesian Beauty Con-*

test, which has been used to study strategic interaction involving information acquisition and the coordination of collective action. In particular, we show how the traditional formulation of this game neglects privacy concerns in its equilibrium predictions, and we then provide a framework for extending the game to incorporate a flexible notion of privacy into the utility of the agents. Using this, we can characterize a *price of privacy*, somewhat akin to quantities such as the price of anarchy, which measures the loss of social welfare in a population of agents who act selfishly to guarantee their own privacy.

Abstractly, we think about players in a game being perfectly rational Bayesian agents who observe some information, perform a utility-maximizing computation, and play an action. However, by observing some player i's chosen action, player j may be able to learn something about player i's private information. For this reason, if players fear that their public actions may reveal private information, they may be incentivized to deviate from the strategy which maximizes utility in the absence of privacy concerns. If all players share such concerns, or anticipate others harboring them, equilibrium behavior may be significantly different than standard predictions.

One setting where there may be tension between privacy and coordination with other players and some underlying ground truth is intra-organizational information aggregation. Suppose that a firm wishes to poll a group of employees as to a particular decision, such as a prediction about the success of a particular product, an evaluation of a colleague, or a new procedure for evaluating and settling claims. In order to make the best decision possible, it is valuable to aggregate information, opinions, or signals received by individuals; however, individuals may be reluctant to fully share their opinion lest it be held against them or they prove to be ‘wrong’. In this situation, simple anonymization procedures may not be trustworthy or effective, and any such procedure requires trusting an internal mechanism to which the employee does not have access. If the respondent is concerned about other agents deanonymizing her survey, then it should be expected that she does not respond honestly.

4.1.1 Our results

We consider a formal model of the Keynesian Beauty Contest, a game in which each agent observes some information and submits an estimate about the true 'state of the world', then earns utility based on both how close her action is to the truth as well as how close it is to the *average action* over all agents. We describe the game, the information structure, and the strategy space and show the existence of a *symmetric linear Nash equilibrium*, where agents' actions are a convex combination of the public and private signals they observe, extending the results in Morris and Shin (2002).

We then turn to the privacy-augmented version of the game, where agents face the same utility function but also suffer a loss of utility based on the ability of other players to infer their private information. We show that this new game also has an equilibrium in strategies that can be written as strategies in the original game with added random noise. This leads to two different values which can be thought of as a 'price of privacy'. In the first, we consider the perspective of the agents and quantify the total loss of the players' utilities as a result of incorporating this concern for privacy. In the second, we think about an 'untrusted aggregator' who wants to compute some statistic using the agents' private beliefs but cannot convince them to participate in a privacy-protecting mechanism. Here, we compute the decrease in the quality of the aggregation due to the players' addition of noise to their previously optimal actions.

At a high level, we a consider setting in which there is no centralized planning mechanism, either to perform the differentially private computation or to instruct the agents who own the data to add a particular amount of noise to their information in order to perform a locally differentially private aggregation. Rather, we assume that agents are rational and derive some utility from both the aggregate-level computation being accurate as well as from the privacy gained by obscuring the information she releases. A major departure from the local differential privacy framework is in our treatment of outliers. In local differential privacy, an individual whose data is very different from the norm may have to add a considerable amount of noise before providing her information to the aggregator, since we need to worry that her unmodified data might shift the distributions of outcomes farther than we would like. In this work, we consider a different flavor of individual privacy, where an agent with outlying information does not necessarily care about being revealed as an outlier; she only cares how accurately an observer can guess her private information.

4.1.2 Related Work

Economics

Concretely, our work builds primarily on the results in Morris and Shin (2002), which formalizes the modern version of a Keynesian Beauty Contest in order to study the tendency for individuals to over-weight public information. We similarly explore linear equilibria in a Keynesian Beauty Contest game with public and private signals; the structure of our model is substantially very similar, and we recover their results as a special case when there is no concern for privacy. Viewing our model as a generalization of theirs, we expand their result on the existence of such equilibria to a privacy-aware setting. The results in Hellwig and Veldkamp (2009) build on those in Morris and Shin (2002), exploring a setting in which agents optimize a selection of information sources with different costs matching their qualities. While Hellwig and Veldkamp (2009) is quite different mechanically and in spirit, there is a certain sense in which our paper can be interpreted along similar lines: the equilibrium noisy action can be viewed as similar to choosing a private signal of differing variances, and the privacy cost can be thought of as a cost to more precise information. The Keynesian Beauty Contest is well-studied in the broader macroeconomics literature (see Nagel et al. (2017) for a recent survey) and is frequently used to study settings where agents derive some value for correctness and for coordination, such as in financial markets or strategic voting. These settings are natural cases for the use of formal privacy methods, such as financial analysts wanting to conceal the model they use to predict asset prices or voters being concerned about their preferences being held against them. Gradwohl and Smorodinsky (2017) study a setting in which strategic agents are concerned that their actions in a game will reveal sensitive private information, but their model and approach, which is based on the concept of *signaling games*, is considerably different from the setting we consider.

Differential Privacy

In the last decade, mathematical notions of privacy have been studied extensively in the computer science literature. The most influential is *differential privacy* (c.f. Dwork et al. (2006)) which states roughly that a statistical algorithm is differentially private if the probability that the algorithm gives any particular output does not change by much when we modify one row of the database the algorithm is run on. Standard differential privacy is principally a *central*, rather than *individual* notion of privacy, as the algorithms operate under the assumption that the adversary does not have access to the raw data provided to the analyst. The concept was introduced in Kasiviswanathan et al. (2011) and recent work has shown that this notion is a powerful framework for addressing privacy concerns (c.f. Kairouz et al. (2016); Joseph et al. (2018); Bassily (2018)). There is also a literature on *privacy in mechanism design.* The work in Chen et al. (2016) and Nissim et al. (2012) examines settings where agents participate in a data aggregation mechanism and earn utility which is increasing in the quality of the estimate but decreasing in the data leakage. The model in Chen et al. (2016) is one of *truthful voting.* In this work, we study a game-theoretic setting where privacy is a concern, rather than designing a privacy-preserving mechanism. Our work differs from the formal study of privacy in that we use a definition of which does not (necessarily) satisfy the strictness of the various versions of *differential* privacy. Rather, we consider the perspective of strategic agents who are concerned with other players' ability to learn 'too much' about their own private information. In this way, our work is related to the issue of *response bias* on surveys, where respondents do not answer questions honestly in order to avoid revealing sensitive information. Work examining the use of randomization to alleviate this effect goes back several decades (e.g. Warner (1965)).

Our extended game does bear a resemblance to the *local model* of differential privacy in that agents add noise to their actions, and one might intuitively map the case where agents add Gaussian noise to their actions to an instance of the local Gaussian mechanism (see, e.g. Dwork et al. (2006)). However, agents in our model are *not* using the Gaussian mechanism to achieve differential privacy, and such a guarantee cannot be recovered[1], as the fact that agents' data is unbounded means that this mechanism cannot give non-vacuous guarantees. To see this, recall that differential privacy requires that for any set of output of the mechanism must be approximately as likely when input is *neighboring*, which typically means differing in one record. In our case, this corresponds to having a different private signal. The problem is that for any choice of finite variance, adding Gaussian noise to private signals that are sufficiently distant will produce noisy actions that are far apart, and can thus be distinguished with high confidence.

More importantly, there is a significant conceptual difference between the models of differential privacy and our model. In differential privacy models, there is a strict separation of roles: there is an accuracy-concerned learner that wishes to perform and release the output of a query, and privacy-concerned agents that supply their data. The addition of noise either centrally or locally prevents any other party from learning the exact value of the private data with high confidence, and is assumed to be necessary to make agents willing to supply their data. In contrast, there is no external learner in our model; instead, agents attempt to coordinate their actions with other agents and the state of the world, but must add noise to their actions in order to prevent others from learning their private information. Noisier actions lead to greater variance around the state of the world and the average actions, and thus lower payoffs. Hence, the agents care about *both* privacy *and* accuracy, and this tradeoff determines how much noise they ultimately add.

4.2 Framework and Model

4.2.1 The Keynesian Beauty Contest

The Keynesian Beauty Contest dates back to John Maynard Keynes' 1936 *General Theory* in which he formulates a game by analogy to a newspaper beauty contest. In this game, players select the 'most beautiful' entrants from an array of photographs printed in the newspaper, and those players

[1] There are other, more complicated mechanisms that can achieve differential privacy with unbounded input data. We refer readers to Liu (2019) and Wang et al. (2019) for more discussion.

who choose the *most popular* faces are eligible for a prize. The salient idea is that a rational contestant must consider their own opinion about which entrants are the most beautiful as well as *beliefs* about the opinions of all other contestants, and perhaps beliefs about the beliefs of the other players' opinions of other players, and beliefs about the beliefs about the beliefs about the beliefs about the opinions of other players, and so on. Keynes originally proposed this as a model to explain the behavior of financial markets, where the value of an asset depends as much on its fundamental potential for returns as it does the *collective belief in its potential* for returns. Since its inception, this model has been used throughout the economics literature to describe the behavior of rational agents in strategic environments (c.f. Gao (2008); Allen et al. (2006); Bosch-Domenech et al. (2002); Cespa and Vives (2015); Nagel et al. (2017); Morris and Shin (2002)), particularly in macroeconomic settings.

In this paper, we work with a common abstraction of the Keynesian Beauty Contest (see e.g. Morris and Shin (2002); Hellwig and Veldkamp (2009)) in which there is some true state of the world s, and each agent submits a guess θ_i and earns utility equal to

$$u_i(\theta_i, \theta_{-i}) = -(1-\alpha)(\theta_i - \bar{\theta})^2 - \alpha(\theta_i - s)^2,$$

where $\bar{\theta}$ is the average choice of θ_i over all players and θ_{-i} denotes the action of all other players other than i. We refer to the $(\theta_i - \bar{\theta})^2$ term as the *coordination component*, which rewards how close player i's action is to that of the other players, and the $(\theta_i - s)^2$ term as the *guessing component*, which describes how close player i's action is to being the correct guess for the true state of the world s. The parameter $\alpha \in [0, 1]$ is common to all agents and describes the relative weighting of the values of coordination and guessing the true state of the world correctly.

We let n denote the number of players. In the economics literature, it is standard to consider a further abstraction of the game with a continuum of agents indexed by the unit interval $[0, 1]$. In computer science, particularly from an algorithmic perspective, we may be more interested in the game with finitely many players n, indexed by a finite set $[n] = \{1, 2, \ldots n\}$. We consider both perspectives and a contribution of this work is demonstrating that the results in the infinite setting can be realized as the limits of the corresponding results in the finite setting.

The fundamental difference between the two cases is in the definition of the average action $\bar{\theta}$. In the game with finitely many and infinitely many agents, these are

$$\bar{\theta} = \frac{1}{n}\sum_{j=1}^{n}\theta_j \qquad \text{and} \qquad \bar{\theta} = \int_0^1 \theta_j dj,$$

respectively. In the finite game, agent i's action θ_i has a measurable effect on the average action whereas in the infinite game, the impact is infinitesimal. Therefore, when there are finitely many players, agent i must consider her own action's effect on $\bar{\theta}$; in the infinite case, she can treat $\bar{\theta}$ as simply the average action of all players other than herself. Intuitively, as n grows, any individual's ability to unilaterally affect $\bar{\theta}$, so we should expect that the results in the infinite game emerge as the limits of the results in the finite game as n grows to infinity.

4.2.2 Information Structure

We now describe the information structure of the game. Each player has a common *improper uniform* prior distribution over the value of s, supported on the entire real line. There is a public signal y, drawn from a Gaussian distribution with mean s and variance σ_y^2. Additionally, each player i observes a private signal x_i independently drawn from a Gaussian distribution with mean s and variance σ_x^2. Each player observes y and her own x_i, but not any other player j's private signal x_j. Each player i's utility function $u_i(\theta_i, \theta_{-i})$ is identical as given above. The values of σ_y^2, σ_x^2, α, the utility functions, and the realization of y are public and common knowledge, and the form of the priors and structure of the game is also common knowledge. Furthermore, we assume that the agents are rational and Bayesian. That is, they seek to maximize their expected utility given their knowledge and their beliefs about other players.

We use the notion of an *improper uniform prior* belief to align with the economics literature; however, this concept is not necessarily standard across disciplines. One can think of a (proper) uniform prior with compact support as indicating that players believe any underlying parameter in the support is equally likely to be the truth. An 'improper uniform prior' expresses the same sentiment for an unbounded support. The term 'improper' arises because such an object is not a probability density function, but, when updated with the proper Gaussian signals in our setting, the posterior is a proper distribution. Berger et al. (2009) provide a useful discussion of settings in which the improper uniform prior is appropriate.

For readers uncomfortable with such a notion, we present an alternative framing of the information structure which is equivalent to the above. The private signals x_i, the utility functions, and the values of α and σ_x^2 are as before. Rather than y representing a public signal, each agent instead has a prior belief about the true state of the world s, which is a Gaussian distribution with mean y and variance σ_y^2. This prior is common to all agents and the parameters of the prior are common knowledge. This framing is equivalent to the previous one since an agent with an improper uniform prior, upon observing the realization of the public signal y, updates her posterior belief about s to be Gaussian with mean y and variance σ_y^2, which is what she would have believed if she instead began with the common Gaussian prior.

We write $\mathcal{I}_i$ for the information set of player i, which captures the structural information about the game as well as the values of x_i and y. Furthermore, we use $\mathbb{E}_i\left(\cdot|\mathcal{I}_i\right)$ to denote the *expectation of player i at $\mathcal{I}_i$*, which is player i's belief about some quantity given that she knows everything at $\mathcal{I}_i$. Where it is clear from context, we drop the $\mathcal{I}_i$ for notational clarity and simply write $\mathbb{E}_i(\cdot)$.

We now consider an equilibrium concept for this game.

Definition 4.1 (Symmetric linear Nash equilibrium)**.** A *symmetric linear Nash equilibrium* of this game is an action θ_i for each player i which can be written as a linear combination of the private and public signals x_i and y such that no player can profitably deviate unilaterally and every player i chooses the same weight to put on x_i and y. That is, a Nash equilibrium in which $\theta_i = \kappa x_i + (1-\kappa)y$ for all players and κ is identical for all players.

Morris and Shin (2002) describe the *unique* symmetric linear Nash equilibrium of a slightly different version of the game which has the same first order condition; our proof follows a very similar structure.

Lemma 4.2 (First order condition)**.** *In equilibrium, an agent's optimal choice of θ_i^* must satisfy*

$$\theta_i^* = \frac{\alpha n^2 \, \mathbb{E}_i[s]}{\alpha(2n-1) + (n-1)^2} + \frac{(1-\alpha)(n-1)\, \mathbb{E}_i \left[\sum_{j \neq i} \theta_j \right]}{\alpha(2n-1) + (n-1)^2}$$

Proof. Fix the actions of all other players $j \neq i$ and consider the utility of player i, recalling that in the finite setting,

$$\bar{\theta} = \frac{1}{n}\theta_i^* + \frac{1}{n}\sum_{j \neq i} \theta_j.$$

We have

$$\begin{aligned} u_i(\theta_i^*, \theta_{-i}) &= -\alpha(\theta_i^* - s)^2 - (1-\alpha)(\theta_i^* - \bar{\theta})^2 \\ &= -\alpha(\theta_i^* - s)^2 - (1-\alpha)\left(\theta_i^* - \left(\frac{1}{n}\theta_i^* + \frac{1}{n}\sum_{j \neq i}\theta_j\right)\right)^2 \\ &= -\alpha(\theta_i^* - s)^2 - (1-\alpha)\left(\frac{n-1}{n}\theta_i^* - \left(\frac{1}{n}\sum_{j \neq i}\theta_j\right)\right)^2 \\ &= -\alpha(\theta_i^* - s)^2 - (1-\alpha)\frac{1}{n^2}\left((n-1)\theta_i^* - \left(\sum_{j \neq i}\theta_j\right)\right)^2 \end{aligned}$$

Player i's first order condition will be to maximize this in expectation. Taking an expectation, we have

$$\begin{aligned} \mathop{\mathbb{E}}_i[u_i(\theta_i, \theta_{-i})] &= \mathop{\mathbb{E}}_i\left[-\alpha(\theta_i^* - s)^2 - (1-\alpha)\frac{1}{n^2}\left((n-1)\theta_i^* - \left(\sum_{j \neq i}\theta_j\right)\right)^2\right] \\ &= -\alpha \mathop{\mathbb{E}}_i[(\theta_i^* - s)^2] - (1-\alpha)\frac{1}{n^2}\mathop{\mathbb{E}}_i\left[\left((n-1)\theta_i^* - \left(\sum_{j \neq i}\theta_j\right)\right)^2\right] \\ &= -\alpha \mathop{\mathbb{E}}_i[(\theta_i^*)^2 - 2s\theta_i^* + s^2] \\ &\quad - (1-\alpha)\frac{1}{n^2}\mathop{\mathbb{E}}_i\left[(n-1)^2(\theta_i^*)^2 - 2(n-1)\theta_i^*\left(\sum_{j \neq i}\theta_j\right) + \left(\sum_{j \neq i}\theta_j\right)^2\right] \end{aligned}$$

Differentiating with respect to agent i's choice of θ_i gives us that, in equilibrium, θ_i^* must satisfy

$$\begin{aligned}\frac{\partial}{\partial\theta_i^*} = 0 &= -\alpha(2\theta_i^* - 2\underset{i}{\mathbb{E}}[s]) - (1-\alpha)\frac{1}{n^2}\left(2(n-1)^2\theta_i^* - 2(n-1)\underset{i}{\mathbb{E}}\left[\sum_{j\neq i}\theta_j\right]\right)\\ &= \alpha(\theta_i^* - \underset{i}{\mathbb{E}}[s]) + \frac{1-\alpha}{n^2}\left((n-1)^2\theta_i^* - (n-1)\underset{i}{\mathbb{E}}\left[\sum_{j\neq i}\theta_j\right]\right)\\ &= \theta_i^*\left(\alpha + \frac{(1-\alpha)(n-1)^2}{n^2}\right) - \left(\alpha\underset{i}{\mathbb{E}}[s] + \frac{(1-\alpha)(n-1)}{n^2}\underset{i}{\mathbb{E}}\left[\sum_{j\neq i}\theta_j\right]\right)\end{aligned}$$

Where we have used the fact that $\mathbb{E}_i[\theta_i^*] = \theta_i^*$, since it is agent i's choice and not a random variable.

Rearranging yields the claim.

□

Proposition 4.3. *In the game with infinitely many agents, in equilibrium, an agent's optimal choice of θ_i^* must satisfy*

$$\theta_i^* = \alpha\underset{i}{\mathbb{E}}[s] + (1-\alpha)\underset{i}{\mathbb{E}}[\bar{\theta}]$$

Proof. Since agent i chooses his action after observing signals x_i and y, they optimize their action given those signals. We can write:

$$\theta_i^* \in \operatorname{argmax}\underset{i}{\mathbb{E}}[u_i(\theta_i)]$$

Expanding, we have:

$$\underset{i}{\mathbb{E}}[u_i(\theta_i,\theta_{-i})] = -\alpha(\underset{i}{\mathbb{E}}[\theta_i^2 - 2\theta_i s + s^2]) - (1-\alpha)(\underset{i}{\mathbb{E}}[\theta_i^2 - 2\theta_i\bar{\theta} + \bar{\theta}^2]).$$

We can differentiate with respect to θ_i^* and set the derivative equal to zero to find

$$0 = \frac{\partial}{\partial\theta_i}\underset{i}{\mathbb{E}}[u_i(\theta_i,\theta_{-i})] = -\alpha(2\theta_i - 2\underset{i}{\mathbb{E}}[s]) - (1-\alpha)[2\theta_i - \bar{\theta}].$$

Where we have used the fact that $\mathbb{E}_i[\theta_i^*] = \theta_i^*$, since it is agent i's choice and not a random variable, as well as the fact that $\frac{\partial\bar{\theta}}{\partial\theta_i^*} = 0$ because there are infinitely many agents, so agent i's action cannot affect the average of all of the actions.

Rearranging yields the claim. □

These results say that the optimal action is a convex combination of agents' expectations about the state and the average actions; this follows directly from the structure of the payoff function.

Now, if other agents' strategies are convex combinations of signals, then the unique best response is also a convex combination of signals. We can match coefficients to solve for this symmetric linear Nash equilibrium.

Lemma 4.4 (Nash equilibrium and Value of κ)**.** *The symmetric linear Nash equilibrium is given by*

$$\theta_i = \kappa x_i + (1-\kappa)y, \textit{ where}$$
$$\kappa = \frac{\alpha n^2 \tau_x}{\alpha n^2 \tau_x + \left((n-1)^2 + \alpha(2n-1)\right)\tau_y}, \quad \tau_x = \frac{1}{\sigma_x^2}, \textit{ and } \tau_y = \frac{1}{\sigma_y^2}$$

for all players i.

Proof. Suppose that all agents play $\theta_i = \kappa_i + (1-\kappa)y$, and consider a representative agent i.

Since agent i is Bayesian, she aggregates the public and private signals according to their precisions to compute

$$\mathop{\mathbb{E}}_i[s] = \frac{\tau_x x_i + \tau_y y}{\tau_x + \tau_y}$$

and her belief about any other agent j's private signal x_j is that

$$\mathop{\mathbb{E}}_i[x_j] = \mathop{\mathbb{E}}_i[s].$$

Supposing that each agent acts optimally and chooses θ_i^* according to Lemma 4.2, agent i will write the following. We let $\mathcal{C} = \frac{1}{\alpha(2n-1)+(n-1)^2}$ for clarity of notation.

$$
\begin{aligned}
\theta_i^* &= \frac{\alpha n^2\,\mathbb{E}_i[s]}{\alpha(2n-1)+(n-1)^2} + \frac{(1-\alpha)(n-1)\,\mathbb{E}_i\left[\sum\limits_{j\neq i}\theta_j\right]}{\alpha(2n-1)+(n-1)^2} \\
&= \mathcal{C}\left(\alpha n^2 \mathop{\mathbb{E}}_i[s] + (1-\alpha)(n-1)\mathop{\mathbb{E}}_i\left[\sum_{j\neq i}(\kappa x_j + (1-\kappa)y)\right]\right) \\
&= \mathcal{C}\left(\alpha n^2 \mathop{\mathbb{E}}_i[s] + (1-\alpha)(n-1)\left[\sum_{j\neq i}\left(\kappa \mathop{\mathbb{E}}_i[s] + (1-\kappa)y\right)\right]\right) \\
&= \mathcal{C}\left(\alpha n^2 \mathop{\mathbb{E}}_i[s] + (1-\alpha)(n-1)^2\left(\kappa \mathop{\mathbb{E}}_i[s] + (1-\kappa)y\right)\right) \\
&= \mathcal{C}\left(\alpha n^2 \frac{\tau_x x_i + \tau_y y}{\tau_x+\tau_y} + (1-\alpha)(n-1)^2\left(\kappa\frac{\tau_x x_i + \tau_y y}{\tau_x+\tau_y} + (1-\kappa)y\right)\right) \\
&= \mathcal{C}\left(\alpha n^2 \frac{\tau_x}{\tau_x+\tau_y} + (1-\alpha)(n-1)^2\kappa\frac{\tau_x}{\tau_x+\tau_y}\right)x_i + \\
&\quad + \mathcal{C}\left(\alpha n^2 \frac{\tau_y}{\tau_x+\tau_y} + (1-\alpha)(n-1)^2\left(\kappa\frac{\tau_y}{\tau_x+\tau_y} + (1-\kappa)\right)\right)y
\end{aligned}
$$

We can then equate κ and the coefficient on x_i or the coefficient on y to $(1-\kappa)$. By construction, it's easy to see that the sum of these two coefficients will be one, so it suffices to perform the computation for κ.

$$
\begin{aligned}
\kappa &= \mathcal{C}\left(\alpha n^2 \frac{\tau_x}{\tau_x+\tau_y} + (1-\alpha)(n-1)^2\kappa\frac{\tau_x}{\tau_x+\tau_y}\right) \\
&= \mathcal{C}\alpha n^2 \frac{\tau_x}{\tau_x+\tau_y} + \kappa\mathcal{C}(1-\alpha)(n-1)^2\frac{\tau_x}{\tau_x+\tau_y} \\
&= \frac{\mathcal{C}\alpha n^2 \frac{\tau_x}{\tau_x+\tau_y}}{1-\mathcal{C}(1-\alpha)(n-1)^2\frac{\tau_x}{\tau_x+\tau_y}}
\end{aligned}
$$

Plugging the value for $\mathcal{C}$ back in and simplifying, we get

$$
\kappa = \frac{\alpha n^2 \tau_x}{\alpha n^2\tau_x + \left((n-1)^2+\alpha(2n-1)\right)\tau_y}
$$

□

Proposition 4.5. *The symmetric linear Nash equilibrium in the game with infinitely many players is given by*

$$
\begin{aligned}
&\theta_i = \kappa x_i + (1-\kappa)y, \textit{ where} \\
&\kappa = \frac{\tau_x\alpha}{\tau_x\alpha+\tau_y},\ \tau_x = \frac{1}{\sigma_x^2},\ \textit{and } \tau_y = \frac{1}{\sigma_y^2}
\end{aligned}
$$

for all players i. This recovers the analogous result in Morris and Shin (2002).

Proof. Proposition 4.3 shows that agent i's optimal action is a convex combination of his belief about the true state s and his belief about the average action of all players $\bar{\theta}$. $\mathbb{E}_i[s]$ is independent of the equilibrium profile, and given by

$$\mathop{\mathbb{E}}_i[s] = \frac{\tau_x x_i + \tau_y y}{\tau_x + \tau_y},$$

which is the standard prior-free aggregation of independent Gaussian signals.

On the other hand, $\mathbb{E}_i[\bar{\theta}]$ does depend on the equilibrium profile. In a SLNE, all other players $j \neq i$ play

$$\theta_j = \kappa x_j + (1-\kappa) y.$$

Since agent i is Bayesian:

$$\mathop{\mathbb{E}}_i[x_j] = \mathop{\mathbb{E}}_i[s]$$

and

$$\mathop{\mathbb{E}}_i[\theta_j] = \mathop{\mathbb{E}}_i[\kappa x_j + (1-\kappa) y] = \kappa \mathop{\mathbb{E}}_i[x_j] + (1-\kappa) y.$$

Notice that the expectation of agent i about the belief of any representative player $j \neq i$ is the same, since his information is symmetric. Thus, we write $\mathbb{E}_i[\theta_{-i}]$ to emphasize that this is the belief of player i about any other player. Thus:

$$\mathop{\mathbb{E}}_i[\bar{\theta}] = \mathop{\mathbb{E}}_i \int_0^1 \theta_j dj = \int_0^1 \mathop{\mathbb{E}}_i[\theta_{-i}] = \mathop{\mathbb{E}}_i[\theta_{-i}] = \kappa \mathop{\mathbb{E}}_i[x_{-i}] + (1-\kappa) y$$

where we have used the Fubini-Tonelli theorem to exchange the integral with the expectation.

But now note that

$$\begin{aligned}\theta_i^* &= \alpha \mathop{\mathbb{E}}_i[s] + (1-\alpha)[\kappa \mathop{\mathbb{E}}_i[s] + (1-\kappa) y] \\ &= \alpha \frac{\tau_x x_i + \tau_y y}{\tau_x + \tau_y} + (1-\alpha)(\kappa \frac{\tau_x x_i + \tau_y y}{\tau_x + \tau_y}) + (1-\alpha)(1-\kappa) y\end{aligned}$$

We can rearrange this as:

$$\begin{aligned}\theta_i^* &= \left[\frac{\alpha \tau_x}{\tau_x + \tau_y} + (1-\alpha)\frac{\kappa \tau_x}{\tau_x + \tau_y}\right] x_i + \left[\frac{\tau_y \alpha}{\tau_x + \tau_y} + (1-\alpha)\frac{\kappa \tau_y}{\tau_x + \tau_y} + (1-\alpha)(1-\kappa)\right] y \\ &= \left[\frac{\tau_x(\alpha + (1-\alpha)\kappa)}{\tau_x + \tau_y}\right] x_i + \left[(1-\alpha)(1-\kappa) + \frac{\alpha \tau_y + (1-\alpha)\kappa \tau_y}{\tau_x + \tau_y}\right] y\end{aligned}$$

Matching coefficients and solving:

$$\left[(\alpha + (1-\alpha)\kappa)\frac{\tau_x}{\tau_x + \tau_y}\right] = \kappa \implies \kappa = \frac{\alpha\tau_x}{\alpha\tau_x + \tau_y}$$

as desired. □

Notice that if $\alpha = 1$, the agent is only rewarded based on his closeness to the state, and θ_i^* then dictates that agent i play his best guess of the state. As α falls, the weight Agent i places on his private signal diminishes, until at $\alpha = 0$, the agent chooses θ_i^* to be equal to y, the public signal.

We can now write the expected utility to each agent in this equilibrium:

Lemma 4.6 (Expected utility). *The expected utility of agents for playing $\theta_i = \kappa x_i + (1-\kappa)y$, given s is:*

$$\mathop{\mathbb{E}}_i[u_i|s] = -\alpha(\kappa^2\sigma_x^2 + (1-\kappa)^2\sigma_y^2) - \frac{(1-\alpha)\kappa^2(n-1)}{n}\,\sigma_x^2$$

Proof. We can write the expected utility of agents as

$$\mathop{\mathbb{E}}_i[u_i|s] = -\alpha\mathop{\mathbb{E}}_i[(\theta_i - s)^2|s] - (1-\alpha)\mathop{\mathbb{E}}_i\left[\left(\theta_i - \bar{\theta}\right)^2\middle| s\right]$$

We can plug in the equilibrium strategies and recall the definition of $\bar{\theta}$ to write

$$\mathop{\mathbb{E}}_i[u_i|s] = -\alpha\mathop{\mathbb{E}}_i[(\theta_i - s)^2|s] - (1-\alpha)\mathop{\mathbb{E}}_i\left[\left(\theta_i - \frac{1}{n}\sum_{j=1}^{n}\theta_j\right)^2\middle| s\right]$$

We can write each $\theta_j = \kappa x_j + (1-\kappa)y$ and since we are taking expectations conditional on knowing s, we can write the signals x_j and y as the state s plus mean-zero Gaussian noise with variance σ_x^2 and σ_y^2, respectively. We write these as $s + \varepsilon_{x_j}$ and $s + \varepsilon_y$, so we can write $\theta_j = s + \kappa\varepsilon_{x_j} + (1-\kappa)\varepsilon_y$.

$$\begin{aligned}
\mathop{\mathbb{E}}_i[u_i|s] &= -\alpha \mathop{\mathbb{E}}_i[(s+\kappa\varepsilon_{x_i}+(1-\kappa)\varepsilon_y-s)^2|s] \\
&\qquad -(1-\alpha)\mathop{\mathbb{E}}_i\left[\left(s+\kappa\varepsilon_{x_i}+(1-\kappa)\varepsilon_y-\frac{1}{n}\sum_{j=1}^{n} s+\kappa\varepsilon_{x_j}+(1-\kappa)\varepsilon_y\right)^2\middle| s\right] \\
&= -\alpha \mathop{\mathbb{E}}_i[(\kappa\varepsilon_{x_i}+(1-\kappa)\varepsilon_y)^2]-(1-\alpha)\mathop{\mathbb{E}}_i\left[\left(\kappa\varepsilon_{x_i}+-\frac{1}{n}\sum_{j=1}^{n}\kappa\varepsilon_{x_j}\right)^2\right] \\
&= -\alpha \mathop{\mathbb{E}}_i[(\kappa\varepsilon_{x_i}+(1-\kappa)\varepsilon_y)^2]-(1-\alpha)\kappa^2\mathop{\mathbb{E}}_i\left[\left(\varepsilon_{x_i}+-\frac{1}{n}\sum_{j=1}^{n}\varepsilon_{x_j}\right)^2\right] \\
&= -\alpha \mathop{\mathbb{E}}_i[(\kappa\varepsilon_{x_i}+(1-\kappa)\varepsilon_y)^2]-(1-\alpha)\kappa^2\mathop{\mathbb{E}}_i\left[\left(\frac{n-1}{n}\varepsilon_{x_i}+-\frac{1}{n}\sum_{j\neq i}^{n}\varepsilon_{x_j}\right)^2\right] \\
&= -\alpha \mathop{\mathbb{E}}_i[(\kappa\varepsilon_{x_i}+(1-\kappa)\varepsilon_y)^2]-\frac{(1-\alpha)\kappa^2}{n^2}\mathop{\mathbb{E}}_i\left[\left((n-1)\varepsilon_{x_i}+-\sum_{j\neq i}^{n}\varepsilon_{x_j}\right)^2\right]
\end{aligned}$$

Because all of the ε_{x_j} and ε_y are independent with mean zero, we can write

$$\begin{aligned}
\mathop{\mathbb{E}}_i[u_i|s] &= -\alpha(\kappa^2\sigma_x^2+(1-\kappa)^2\sigma_y^2)-\frac{(1-\alpha)\kappa^2}{n^2}\left((n-1)^2\sigma_x^2+(n-1)\sigma_x^2\right)^2 \\
&= -\alpha(\kappa^2\sigma_x^2+(1-\kappa)^2\sigma_y^2)-\frac{(1-\alpha)\kappa^2(n-1)}{n}\sigma_x^2
\end{aligned}$$

□

Proposition 4.7. *The expected utility of agents for playing $\theta_i = \kappa x_i + (1-\kappa)y$, conditional on s in the game with infinitely many players, is:*

$$\mathop{\mathbb{E}}_i[u_i|s] = -\alpha(1-\kappa)^2\sigma_y^2 - \kappa^2\sigma_x^2$$

Proof. We can write the expected utility of agents as

$$\mathop{\mathbb{E}}_i[u_i|s] = -\alpha\mathop{\mathbb{E}}_i[(\theta_i-s)^2|s]-(1-\alpha)\mathop{\mathbb{E}}_i\left[\left(\theta_i-\bar{\theta}\right)^2\middle| s\right]$$

We can plug in the equilibrium strategies and recall the definition of $\bar{\theta}$ to write

$$\mathop{\mathbb{E}}_i[u_i|s] = -\alpha\mathop{\mathbb{E}}_i[(\theta_i-s)^2|s]-(1-\alpha)\mathop{\mathbb{E}}_i\left[\left(\theta_i-\int_0^1\theta_j dj\right)^2\middle| s\right]$$

We can write each $\theta_j = \kappa x_j + (1-\kappa)y$ and since we are taking expectations conditional on knowing s, we can write the signals x_j and y as the state s plus mean-zero Gaussian noise with variance σ_x^2 and σ_y^2, respectively. We write these as $s+\varepsilon_{x_j}$ and $s+\varepsilon_y$, so we can write $\theta_j = s + \kappa\varepsilon_{x_j} + (1-\kappa)\varepsilon_y$.

$$\begin{aligned}\mathop{\mathbb{E}}_i[u_i|s] &= -\alpha \mathop{\mathbb{E}}_i[(s+\kappa\varepsilon_{x_i} + (1-\kappa)\varepsilon_y - s)^2|s] \\ &\quad - (1-\alpha)\mathop{\mathbb{E}}_i\left[\left(s+\kappa\varepsilon_{x_i}+(1-\kappa)\varepsilon_y - \int_0^1 \left(s+\kappa\varepsilon_{x_j}+(1-\kappa)\varepsilon_y dj\right)\right)^2 \middle| s\right] \\ &= -\alpha\mathop{\mathbb{E}}_i[(\kappa\varepsilon_{x_i}+(1-\kappa)\varepsilon_y)^2] - (1-\alpha)\mathop{\mathbb{E}}_i\left[\left(\kappa\varepsilon_{x_i} + \int_0^1 \left(\kappa\varepsilon_{x_j} dj\right)\right)^2\right]\end{aligned}$$

Because $\mathbb{E}[\varepsilon_{x_j}] = 0$, $\int_0^1 \left(\kappa\varepsilon_{x_j} dj\right) = 0$. Furthermore, since all of the ε_{x_j} and ε_y are independent we have

$$\begin{aligned}\mathop{\mathbb{E}}_i[u_i|s] &= -\alpha\mathop{\mathbb{E}}_i[(\kappa\varepsilon_{x_i}+(1-\kappa)\varepsilon_y)^2] - (1-\alpha)\mathop{\mathbb{E}}_i\left[(\kappa\varepsilon_{x_i})^2\right] \\ &= -\alpha(\kappa^2\sigma_x^2 + (1-\kappa)^2\sigma_y^2) - (1-\alpha)\kappa^2\sigma_x^2 \\ &= -\alpha(1-\kappa)^2\sigma_y^2 - \kappa^2\sigma_x^2\end{aligned}$$

□

Using the value of κ from the infinite setting for simplicity[2] we can now examine how utility changes as the variances of the signals and number of agents do. We briefly analyze this next.

Corollary 4.8 (Comparative statics)**.** *We have*

$$\begin{aligned}\frac{\partial\,\mathbb{E}_i[u_i|s]}{\partial\sigma_x^2} &= -\frac{(\alpha\sigma_y^2)^2}{(\alpha\sigma_y^2+\sigma_x^2)^3}\left((2-\alpha)\alpha^2\sigma_x^2\sigma_y^2 + \sigma_y^2 - \frac{n-1}{n}(1-\alpha)(\alpha\sigma_y^2 - \sigma_x^2)\right), \\ \frac{\partial\,\mathbb{E}_i[u_i|s]}{\partial\sigma_y^2} &= -\frac{\alpha(\sigma_x^2)^2}{(\alpha\sigma_y^2+\sigma_x^2)^3}\left(2\alpha^2\sigma_y^2 - \alpha\sigma_y^2 + \sigma_x^2 + \frac{n-1}{n}2\alpha(1-\alpha)\sigma_y^2\right),\end{aligned}$$

and

$$\frac{\partial\,\mathbb{E}_i[u_i|s]}{\partial n} = -\frac{(1-\alpha)\alpha^2\sigma_x^2(\sigma_y^2)^2}{n^2(\alpha\sigma_y^2+\sigma_x^2)^2}.$$

[2] For a fixed n, the infinite version of κ differs from the finite version by a multiplicative factor of $1-\frac{1}{n^2}$ while the utilities differ by a factor of $1-\frac{1}{n}$. Therefore, the value of κ in the infinite case is very close to that in the finite case for modest values of n, and any countervailing effect of changing the parameters through κ will be dominated by the direct effects on the utility.

All else fixed, decreasing σ_x^2 unambiguously increases utility. The fractional term in $\frac{\partial \mathbb{E}_i[u_i|s]}{\partial \sigma_x^2}$ is always positive and the term in the parentheses is also always positive, so $\frac{\partial \mathbb{E}_i[u_i|s]}{\partial \sigma_x^2}$ is strictly decreasing as σ_x^2 increases. This aligns with the intuition that higher quality information means that agents will be able to both guess the true state of the world more precisely as well as coordinate better with one another. A similar intuition holds for the effect of decreasing σ_y^2, since the term in the parentheses is always positive when $n \geq 2$. However, the *rate* at which utility increases when the signal variances decrease is not the same. Supposing $\sigma_x^2 = \sigma_y^2$, agents gain more value from decreasing σ_y^2 than σ_x^2 for two reasons. The first is the over-weighting of y. Since agents more heavily weight the public signal y compared to their private information x_i, a small improvement in the precision of y will help agents choose an action θ_i closer to s than an identical improvement in the precision of the x_i signals. The second reason can be thought of as a second-order effect of this over-weighting. Since agents over-weight it, improving the quality of the public signal y even further increases the weight agents will place on it. Therefore, not only will agents actions be closer (in expectation) to s, but they will be closer to *each other*, thus also improving the utility in the coordination component.

As n increases, utility decreases. There is an $\frac{n-1}{n}$ coefficient on the second term in the utility function, and as n grows, this grows towards 1. We can think of this term as measuring the amount of non-impact any one agent can have on the average action $\bar{\theta}$. When the number of agents is very few, agent i's action θ_i can't be too far from the average since the construction of $\bar{\theta}$ will have a high weight on θ_i, and this weight decreases as n grows. For very large values of n, agent i has a small impact on the average action, so the risk of being far away is greater.

4.2.3 Privacy-Awareness

Now we come to the heart of the privacy issue. Suppose that at the end of the game, each player's action θ_i is made public. Then, upon seeing θ_i, player j can simply write

$$x_i = \frac{\theta_i - (1-\kappa)y}{\kappa},$$

and since everything on the right hand side is known to player j, she can learn player i's private signal x_i with perfect precision. This suggests that if the players care at all about preserving the privacy of their private signals, they ought not to play exactly the equilibrium prediction θ_i.

As written, privacy is not in the players' utility functions, and it is too much to ask for a prediction that incorporates privacy without actually incorporating privacy into the utility function. To overcome this, we let $\rho(\tilde{\theta}_i)$ denote, abstractly, a measure which describes how 'private' the (possibly randomized) action $\tilde{\theta}_i$ is. For example, ρ could represent the *maximum precision (i.e. the reciprocal of the variance)*[3] *to which some player j can infer the private signal x_i of player i* after observing his action and given her information set $\mathcal{I}_j$ and knowledge of the equilibrium strategies. Here, we consider both this precision measure of privacy as well as an *entropy* measure, where $\rho(\tilde{\theta}_i)$ is the (information-theoretic) entropy of this inference, rather than the variance.

It is clear that any equilibrium which prescribes a deterministic mapping from the available information to an action θ_i^* cannot offer any measure of privacy to the players. To correct for this, we enhance players' utility functions to incorporate the value they gain from obscuring their signals and extend the equilibrium concept to a *noisy* one, where players select not only a guess about the state, but also a *noise-generating distribution.* Formally, we extend the utility function to

$$v_i(\tilde{\theta}_i, \tilde{\theta}_{-i}) = (1-\beta)u_i(\tilde{\theta}_i, \tilde{\theta}_{-i}) + \beta\rho(\tilde{\theta}_i),$$

where u_i is as before and $\beta \in [0,1]$ denotes the agents' *relative value for obfuscation.*

[3] Taken as a worst-case over all $j \neq i$.

We can observe that if $\beta \neq 0$, the equilibrium actions in the original game do not support an equilibrium in this game, since upon observing θ_i^*, any player j can exactly recover the private signal x_i, and player j's 'distribution' over the possible values of x_i is degenerate with variance or entropy equal to zero. Concretely, in the precision setting, $\beta\rho(\tilde{\theta}_i)$ is $-\infty$ and $v_i(\theta_i^*, \tilde{\theta}_{-i}) = -\infty$. Conversely, if player i were to choose $\kappa' = 0$ (and therefore $\tilde{\theta}_i = y$), her utility would be finite, a profitable deviation.

Thus, any deterministic equilibrium cannot be supported if privacy is incorporated into the utility function, and some degree of randomization is necessary; however, the next section will show that the prediction of the deterministic optimum will serve as the core deterministic component to an optimal randomized strategy.

We remark here that if agents do not care about others learning their signal at all (i.e. $\beta = 0$), the utility function degenerates to that of the original game. If $\beta = 1$, then players *only* care about protecting their private signal, and a dominant strategy is to take the action which maximizes the obfuscation of the private signal.

4.3 The Extended Game

We now formalize this modified game and show that it still has a symmetric linear Nash equilibrium when we consider *randomized* strategies. Here, optimal actions are of the form $\tilde{\theta}_i = \theta_i + \eta_i$, where θ_i is the same linear aggregation of the public and private signals as in the original game and η_i is independent noise drawn from a distribution whose form depends on how we measure the obfuscation component $\rho(\tilde{\theta}_i)$. That is to say, there is a symmetric linear Nash equilibrium in which players' optimal action is a noisy modification of their optimal action in the original game. Note that randomized strategies are *not* mixed strategies in the classical sense. Each player's choice of a noise-generating distribution can be thought of as committing to a collection of parameters which describe that distribution. In this sense, our equilibrium concept is a 'pure strategies' one, since each player will commit to a single collection of parameters, not a distribution over such collections.

4.3.1 Defining the Extended Game

In the extended game, s, x_i, σ_x^2, y, σ_y^2, and $\bar{\theta}$ are as before. Each player chooses an action $\tilde{\theta}_i$, and receives utility

$$v_i(\tilde{\theta}_i) = (1-\beta)u_i(\tilde{\theta}_i) + \beta\rho(\tilde{\theta}_i),$$

where $\beta \in [0,1]$, and $u_i(\tilde{\theta}_i, \tilde{\theta}_{-i}) = -(1-\alpha)(\tilde{\theta}_i - \bar{\tilde{\theta}})^2 - \alpha(\tilde{\theta}_i - s)^2$ as before. We write $\bar{\tilde{\theta}}$ to represent the average (noisy) action of the players. After each player announces $\tilde{\theta}_i$, players also learn the function each player used to select $\tilde{\theta}_i$ as a function of $\mathcal{I}_i$. For example, if a player's strategy is to play x_i plus some random noise, player j learns the distribution from which this noise was drawn, though (crucially) not the *realization* of this draw. We denote the *privacy loss* of each player after actions are revealed $\rho(\tilde{\theta}_i)$. The two possibilities we consider for ρ are both functions of the belief distribution that a Bayesian agent, given their information set, observation of $\tilde{\theta}_i$, and knowledge of equilibrium strategies and the randomization mechanism used by player i, assigns to x_i; one is the *precision* of these beliefs, and the other is the *entropy*. Formally,

Definition 4.9 (Precision privacy loss in equilibrium)**.** For a given equilibrium profile, Player i's *precision* privacy loss is the expected precision with which a Bayesian opponent, knowing the equilibrium profile and observing their own signal, the public signal, and Player i's action $\tilde{\theta}_i$, can estimate Player i's signal x_i. That is, if γ represents the belief distribution about x_i, we define:

$$\rho_{\text{prec}}(\tilde{\theta}_i) = \frac{1}{\operatorname{Var}\gamma(x_i | s, \tilde{\theta}_i, \mathcal{I}_j, H)}$$

Definition 4.10 (Entropy privacy loss in equilibrium)**.** For a given equilibrium profile, Player i's *entropy* privacy loss s the expected precision with which a Bayesian opponent, knowing the equilibrium profile and observing their own signal, the public signal, and Player i's action $\tilde{\theta}_i$, can estimate Player i's signal x_i. That is, if γ represents the belief distribution about x_i, we define:

$$\rho_{\text{ent}}(\tilde{\theta}_i) = -\int \gamma(x_i | s, \tilde{\theta}_i, \mathcal{I}_j, H) \log \gamma(x_i | s, \tilde{\theta}_i, \mathcal{I}_j, H) d\tilde{\theta}_i$$

Before we move on, it is worth considering kinds of distributions these choices consider to be 'private'. For example, considering the following two belief distributions:

$$\gamma_1 \ : \ x_j = U([-\varepsilon, \varepsilon]) \qquad \gamma_2 \ : \ \begin{cases} M & \text{with probability } \delta \\ -\varepsilon & \text{otherwise} \end{cases}$$

By choosing M large and ε small, the appropriate choice of δ can be made to force γ_2 to have arbitrarily large variance (small precision), but extremely low entropy. On the other hand, a uniform distribution γ_1 has relatively high entropy as compared to its variance. If agents measure their privacy by precision, they will be indifferent between adding noise from a fairly narrow uniform distribution and a discrete distribution which almost always adds a very small amount of noise but rarely adds an enormous amount. However, if they measure their privacy with respect to entropy, the uniform option offers a much greater amount of privacy, and since (as we will show in Corollary 4.16) agents pay a price in their utility equal to the variance of the noise-generating distribution, for a fixed variance the higher entropy uniform distribution will be strictly preferred to the discrete one.

We aim to show the following theorem:

Theorem 4.11. *Consider the modified game as defined above. Suppose players' actions $\tilde{\theta}_i$ and the (possibly randomized) mapping $(y, x_i) \mapsto \tilde{\theta}_i$ are revealed at the end of the game, and players' loss of privacy $\rho(\tilde{\theta}_i)$ is measured as either the reciprocal of the variance or the entropy of a representative player j's posterior belief about x_i at the end of the game. Then, this game has a symmetric linear Nash equilibrium where each player chooses $\tilde{\theta}_i = \kappa x_i + (1-\kappa)y + \eta_i$ where κ is as it is in the original game and η_i is a random variable drawn from a distribution whose form depends on whether we use reciprocal variance or entropy to measure privacy.*

Proof sketch. We prove this theorem over the course of several steps throughout the remainder of this section, which contains the formal statements of the following:

1. First, we show (Claim 4.13) that if players do indeed choose their action by adding noise to their privacy-unaware equilibrium action, then the distribution which generates the noise must

have mean zero.

2. Using this, we prove (Lemma 4.15) that the utility function *separates* into the sum of three parts: the utility in the original game, a penalty paid in the variance of the noise distribution, and the privacy component.

3. Then, since the utility function separates additively, we can use the first-order conditions (Lemma 4.2) to solve for the optimal choice of distribution from which to draw the noise (Corollary 4.19), which depends on whether the loss of privacy is measured by reciprocal variance or negative entropy.

4. Finally, we demonstrate that this strategy profile supports a Nash equilibrium by arguing that there is no profitable deviation for any player in the game (Lemmas 4.22 and 4.23). We call this a *symmetric noisy linear Nash equilibrium.*

□

We first define a Noisy Strategy:

Definition 4.12 ((Linear) H-noisy strategy)**.** An H-noisy strategy has the form

$$\tilde{\theta}(x, y, \nu) = \theta_i(x, y) + \eta_i$$

where η_i is a random variable drawn from a distribution H and $\theta_i(x, y)$ is a deterministic function of signals x and y. We say that $\tilde{\theta}$ is a *normal* noisy strategy if H is a Gaussian distribution. If, moreover, $\theta_i(x, y)$ is *linear* and can be written as $\theta_i(x, y) = \kappa x + (1 - \kappa)y$, we say that $\tilde{\theta}$ is a *linear* noisy strategy.

Note that *any* randomized strategy that can be decomposed into a deterministic component and a random component can be described as a noisy strategy. This means that Corollary 4.16 applies whether or not the underlying deterministic component is linear.

Claim 4.13. *If there exists an equilibrium in noisy strategies, where player i chooses $\tilde{\theta}_i = \theta_i + \eta_i$ and each player's η_i is drawn independently from a distribution H_i, then there exists such an equilibrium strategy profile in which the mean of each H_i is zero.*

Proof. Since the distributions which generate the η_i are revealed after each player announces $\tilde{\theta}_i$, choosing to draw noise from a distribution with non-zero mean cannot improve the privacy that player i achieves, since a representative player j can simply subtract this mean when constructing her posterior distribution over x_i. Additionally, choosing a mean other than zero makes the utility from the guessing component strictly worse. Finally, if we assume that all players other than i choose to add noise drawn from a distribution with the same mean, then in the coordination portion of the utility function, player i choosing a noise distribution mean other than the common one is dominated by choosing the common one. In particular, if everyone else chooses mean-zero noise, player i should as well.

□

If the parameter α is greater than $\frac{1}{2}$, then *every* such equilibrium requires players to choose a mean-zero noise-generating distribution. To see this, suppose all players $j \neq i$ choose a mean $\mu > 0$. Then for a small enough value of ε, player i choosing a noise-generating distribution with mean $\mu - \varepsilon$ is a profitable deviation, since $\alpha > \frac{1}{2}$ means that her gain from moving $\tilde{\theta}_i$ closer to s more than offsets the loss from moving further from $\bar{\tilde{\theta}}$. Going forward, we assume without loss of generality that any noisy equilibrium is one in which the added noise comes from a mean-zero distribution.

We now define our equilibrium concept, which is analogous to that of the original game.

Definition 4.14 (Symmetric noisy linear Nash equilibrium)**.** A *symmetric noisy linear Nash equilibrium* of this game is a strategy profile $\tilde{\theta}$ where each player i chooses

$$\tilde{\theta}_i = \theta_i + \eta_i$$

and $\theta_i = \kappa x_i + (1-\kappa) y$, κ is the same for all agents, and each player's noise η_i is drawn independently from the same distribution H.

One important point is that $\rho(\tilde{\theta}_i)$ is a function of the *optimal Bayesian posterior* distribution of x_i given $\tilde{\theta}_i$ and y, so an agent must consider his choice of action both in relation to the change in coordination in addition to others' beliefs about him; however, these beliefs are only what an optimal Bayes estimator, knowing the equilibrium profile, could *infer.* In particular, agents are concerned only by the knowledge of agents that are *correct.* It is conceivable that agents could worry about opponents *incorrectly* inferring their signals, but that falls outside the scope of this model.

4.3.2 Solving the Extended Game

Having defined the privacy-extended game, we can solve for the parameter values which support a symmetric noisy linear Nash equilibrium. Much of the analysis follows either directly from or along similar arguments as in the original game. We proceed as follows. First, we demonstrate in Corollary 4.16 that the utility function separates additively into three components: the utility in the original game, a penalty in this utility due to the added noise, and the privacy term. We then derive the first order condition of a representative agent in Proposition 4.18 and use this to find the optimal parameter values in Corollary 4.20, depending on whether we consider the precision-based or entropy-based measure of obfuscation. Finally, in Lemmas 4.22 and 4.23, we show that these values support an equilibrium. This completes the proof of Theorem 4.11.

We begin with the separability of the utility function.

Lemma 4.15 (Separability)**.** *In the game with finitely many players, the players' utility functions in the privacy-aware game separate additively into the utility in the privacy unaware game, a penalty in ν_i, and a privacy term as*

$$\begin{aligned} v_i(\tilde{\theta}_i, \tilde{\theta}_{-i}) &= (1-\beta)\left(-\alpha(\tilde{\theta}_i - s)^2 - (1-\alpha)(\tilde{\theta}_i - \bar{\tilde{\theta}})^2\right) + \beta\rho(\tilde{\theta}_i) \\ &= (1-\beta)u_i(\theta_i, \tilde{\theta}_{-i}) + (1-\beta)\left(\alpha + \left(1 - \frac{1}{n}\right)^2(1-\alpha)\right)\nu_i + \beta\rho(\tilde{\theta}_i), \end{aligned}$$

where ν_i denotes the variance of the noise-generating distribution H_i of player i and u_i is the utility function in the privacy-unaware game.

Proof. The proof is nearly identical to that of Corollary 4.16. Writing $\tilde{\theta}_i$ as $\theta_i + \eta_i$, i.e. a deterministic component plus random noise, we can decompose the various pieces of the utility function as follows. The first part is

$$-\alpha \mathop{\mathbb{E}}_i[(\theta_i + \eta_i - s)^2] = -\alpha \mathop{\mathbb{E}}_i \left[(\theta_i - s)^2 + \eta_i(\theta_i - s) + \eta_i^2\right] = -\alpha \mathop{\mathbb{E}}_i[(\theta_i - s)^2] - \alpha\nu_i,$$

where we have again used the independence of η_i to conclude that $E_i[\eta_i(\theta_i - s)] = 0$.

The second term is

$$-(1-\alpha) \mathop{\mathbb{E}}_i \left[\left(\theta_i - \eta_i - \frac{1}{n}\sum_{j=1}^{n} \tilde{\theta}_j\right)^2\right],$$

which can be rewritten as

$$-(1-\alpha) \mathop{\mathbb{E}}_i \left[\left(\theta_i \left(1 - \frac{1}{n}\right) + \eta_i \left(1 - \frac{1}{n}\right) - \frac{1}{n}\sum_{j \neq i} \tilde{\theta}_j\right)^2\right].$$

This can then be simplified to

$$(1-\alpha) \mathop{\mathbb{E}}_i \left[\eta_i \left(1 - \frac{1}{n}\right)^2 \frac{1}{n}\theta_i \sum_{j \neq i} \tilde{\theta}_j\right],$$

this term is equal to zero because η_i is independent of all the other parameters of the game as well as the other η_j.

□

Corollary 4.16. *Suppose that all agents play a noisy strategy $\tilde{\theta}_i = \theta_i + \eta_i$, with η_i being a random variable and $\mathbb{E}(\eta_i) = 0$. Then an agent's utility can be decomposed into*

$$\mathop{\mathbb{E}}_i[v_i(\tilde{\theta}, \tilde{\theta}_{-i})] = (1-\beta) \mathop{\mathbb{E}}_i[u(\theta_i, \tilde{\theta}_{-i})] + (1-\beta)\nu_i + \beta\rho(\tilde{\theta}_i)$$

where ν_i denotes the variance of the noise-generating distribution H_i of player i and u_i is the utility function in the privacy-unaware game.

Proof. By definition,

$$\mathop{\mathbb{E}}_i[v_i(\tilde{\theta}, \tilde{\theta}_{-i})] = (1-\beta) \mathop{\mathbb{E}}_i[u_i(\tilde{\theta}_i, \tilde{\theta}_{-i})] + \beta\rho(\tilde{\theta}_i),$$

so if we show that $u_i(\tilde{\theta}_i, \tilde{\theta}_{-i}) = u_i(\theta_i, \tilde{\theta}_{-i}) - \nu_i$ we will be done. We can write

$$\begin{aligned}\mathop{\mathbb{E}}_i[u_i(\tilde{\theta}_i, \tilde{\theta}_{-i})] &= -\alpha \mathop{\mathbb{E}}_i[(\tilde{\theta}_i - s)^2] - (1-\alpha) \mathop{\mathbb{E}}_i[(\tilde{\theta}_i - \bar{\tilde{\theta}})^2] \\ &= -\alpha \mathop{\mathbb{E}}_i[(\theta_i + \eta_i - s)^2] - (1-\alpha) \mathop{\mathbb{E}}_i[(\theta_i + \eta_i - \bar{\tilde{\theta}})^2] \\ &= -\alpha \mathop{\mathbb{E}}_i[(\theta_i - s + \eta_i)^2] - (1-\alpha) \mathop{\mathbb{E}}_i[(\theta_i - \bar{\tilde{\theta}} + \eta_i)^2]\end{aligned}$$

Expanding these terms, we have

$$\begin{aligned}\mathop{\mathbb{E}}_i[u_i(\tilde{\theta}_i, \tilde{\theta}_{-i})] = -\alpha &\left(\mathop{\mathbb{E}}_i[(\theta_i - s)^2] + 2\mathop{\mathbb{E}}_i[\eta_i(\theta_i - s)] + \mathop{\mathbb{E}}_i[(\eta_i^2)]\right) \\ &- (1-\alpha)\left(\mathop{\mathbb{E}}_i[(\theta_i - \bar{\tilde{\theta}})^2] + 2\mathop{\mathbb{E}}_i[(\eta_i)(\theta_i - \bar{\tilde{\theta}})] + \mathop{\mathbb{E}}_i[\eta_i^2]\right)\end{aligned}$$

Now the first terms of each line sum to exactly $u_i(\theta_i, \tilde{\theta}_{-i})$. On the other hand, the sum of the last two terms is $-\mathbb{E}_i[\eta_i^2] = -\nu_i$. To complete the proof, we show that these middle to terms are, in fact, zero. To see this, notice that at $\mathcal{I}_i$, η_i is yet unrealized with $\mathbb{E}_i[\eta_i] = 0$, but is independent of s and θ_i and thus of $\bar{\theta}$. Hence,

$$\mathop{\mathbb{E}}_i[\eta_i(\theta_i - s)] = \mathop{\mathbb{E}}_i[\eta_i]\mathop{\mathbb{E}}_i[\theta_i - s] = 0,$$

Moreover, η_i is independent of each $\tilde{\theta}_{-i}$, and agent i's action cannot unilaterally change $\bar{\tilde{\theta}}$, so

$$\mathop{\mathbb{E}}_i[\eta_i(\theta_i - \bar{\tilde{\theta}})] = \mathop{\mathbb{E}}_i[\eta_i(\theta_i - \bar{\tilde{\theta}}_{-i})] = \mathop{\mathbb{E}}_i[\eta_i]\mathop{\mathbb{E}}_i[\theta_i - \bar{\tilde{\theta}}_{-i}] = 0$$

□

This shows that agents can evaluate what their optimal action would be in the original game and then decide the optimal amount of noise to add, choosing their action as the realized value of the noise plus their optimal action.

We next derive the first order conditions for the privacy-extended game. By Lemma 4.15 and Corollary 4.16, the first order conditions can be disentangled into a first order condition on the deterministic component, which must be as in Proposition 4.3 and Lemma 4.2, and a separate first order condition on the variance of the random component. Because the added noise is mean-zero, expectations about average action and state are as before, and the κ in the optimal deterministic action is the same as before.

Lemma 4.17 (Privacy-aware first order conditions)**.** *In an equilibrium of the game with finitely many players where the optimal action is $\tilde{\theta}_i^* = \theta_i^* + \eta_i$, the optimal choice of θ_i^* and the variance ν^**

for the noise-generating distribution H_i from which η_i is drawn must satisfy

$$\theta_i^* = \frac{\alpha n^2 \mathbb{E}_i[s]}{\alpha(2n-1)+(n-1)^2} + \frac{(1-\alpha)(n-1)\mathbb{E}_i\left[\sum\limits_{j\neq i}\theta_j\right]}{\alpha(2n-1)+(n-1)^2},$$

$$\frac{\partial\rho}{\partial\nu^*} = -\frac{-(1-\beta)\left(\alpha+\left(1-\frac{1}{n}\right)^2(1-\alpha)\right)}{\beta}.$$

Proof. Using Lemma 4.15, we can decompose the expected utility of agent i as

$$\mathop{\mathbb{E}}_i[v_i(\tilde{\theta}_i,\tilde{\theta}_{-i}] = (1-\beta)\mathop{\mathbb{E}}_i[u_i(\theta_i,\tilde{\theta}_{-i})] - (1-\beta)\left(\alpha+\left(1-\frac{1}{n}\right)^2(1-\alpha)\right)\nu_i + \beta\rho(\tilde{\theta}_i),$$

which is the sum of a piece that depends on θ_i and a piece that depends on ν_i.

The agent can therefore optimize each piece separately with her choice of θ_i and ν_i. Lemma 4.2 gives the first order condition on θ_i^*.

To find the first order condition on ν^*, we can write

$$0 = \frac{\partial v_i}{\partial\nu^*} = -(1-\beta)\left(\alpha+\left(1-\frac{1}{n}\right)^2(1-\alpha)\right) + \beta\frac{\partial\rho}{\partial\nu^*}$$

and solve for $\frac{\partial\rho}{\partial\nu^*}$ to get the result.

□

Proposition 4.18. *In the game with infinitely many players, in an equilibrium where the optimal action is $\tilde{\theta}_i^* = \theta_i^* + \eta_i$, the optimal choice of θ_i^* and the variance ν^* for the noise-generating distribution H_i from which η_i is drawn must satisfy*

$$\theta_i^* = \alpha\mathop{\mathbb{E}}_i[s] + (1-\alpha)\mathop{\mathbb{E}}_i[\bar{\tilde{\theta}}] \qquad \frac{\partial\rho}{\partial\nu^*} = -\frac{1-\beta}{\beta}.$$

Proof. Using Corollary 4.16, we can decompose the utility of agent i as

$$\mathop{\mathbb{E}}_i[v_i(\tilde{\theta},\tilde{\theta}_{-i})] = (1-\beta)\mathop{\mathbb{E}}_i[u(\theta_i,\tilde{\theta}_{-i})] - (1-\beta)\nu_i - \beta\rho(\tilde{\theta}_i),$$

which is the sum of a piece that depends on θ_i and a piece that depends on ν_i.

The agent can therefore optimize each piece separately with her choice of θ_i and ν_i. Proposition 4.3 gives the first order condition on θ_i^*.

To find the first order condition on ν^*, we can write

$$0 = \frac{\partial v_i}{\partial \nu^*} = -(1-\beta) - \beta \frac{\partial \rho}{\partial \nu^*}$$

and solve for $\frac{\partial \rho}{\partial \nu^*}$ to get the result.

□

Corollary 4.19 (Finite game privacy parameters)**.** *The optimal deterministic component θ_i is, as before,*

$$\theta_i^* = \kappa x_i + (1-\kappa) y$$

and the optimal choice of variance for the noise distribution is

$$\nu_{i,prec}^* = \sqrt{\frac{\beta}{1-\beta}\left(\alpha + (1-\alpha)\left(1-\frac{1}{n}\right)^2\right)^{-1}}, \quad or$$
$$\nu_{i,ent}^* = \frac{\beta}{1-\beta}\left(\alpha + (1-\alpha)\left(1-\frac{1}{n}\right)^2\right)^{-1}$$

where $\nu_{i,prec}^*$ and $\nu_{i,ent}^*$ are the optimal variances under ρ being the precision and entropy privacy measures, respectively.

Corollary 4.20 (Infinite game privacy parameters)**.** *The optimal deterministic θ_i is, as before,*

$$\theta_i^* = \kappa x_i + (1-\kappa) y$$

and the optimal choice of variance for the noise distribution is

$$\nu_{i,prec}^* = \sqrt{\frac{\beta}{1-\beta}} \qquad \nu_{i,ent}^* = \frac{\beta}{1-\beta}$$

where $\nu_{i,prec}^*$ and $\nu_{i,ent}^*$ are the optimal choices of variance when ρ is the precision-based or entropy-based privacy measure, respectively.

Before proving this, we state a fact about the entropy of Gaussian distributions.

Fact 4.21. *Among all distributions supported on the entire real line with a fixed mean μ and variance σ^2, the Gaussian $\mathcal{N}(\mu, \sigma^2)$ achieves the maximum entropy, and its entropy is given by $\frac{1}{2}\log(2\pi e \sigma^2)$.*

Since agents pay a penalty in the variance of their noise distribution, agents who measure their privacy loss with entropy pay the same penalty for any distribution with a fixed variance, and their gain from privacy is maximized by picking the distribution with maximum entropy. Therefore, the choice of a mean-zero Gaussian with appropriate variance is the dominant strategy for such agents. A proof of this fact can be found in Chapter 9 of Cover and Thomas (2012).

Proof of Corollaries 4.19 and 4.20. The fact that the first order condition on θ_i^* is the same as that of θ_i^* in the original game, and that agents add mean-zero noise (implying $\mathbb{E}_i[\bar{\tilde{\theta}}] = \mathbb{E}_i[\bar{\theta}]$) implies the optimal choice of θ_i^*. To find the optimal ν^*, we note that the reciprocal variance and entropy penalties (respectively) are

$$\rho_{prec}(\nu) = \frac{-1}{\nu} \qquad \rho_{ent}(\nu) = \frac{1}{2}\log(2e\pi\nu).$$

Differentiating these with respect to ν and rearranging gives the results. □

We can observe that the optimal choice of ν_i^* is similar for both of these functions, even if their constructions are not. Each is increasing at a decreasing rate as ν grows, since their derivatives are of the form $\frac{1}{\nu^c}$ for $c \geq 1$. Since players gain a diminishing marginal benefit for adding additional noise as they increase ν but pay a penalty linear in ν, we should expect this trade-off to point to an equilibrium. This will hold for any penalty function whose derivative is of this type, although the interpretation of such a function may not be as natural of a property of a distribution as precision or entropy.

Using these, a symmetric linear noisy Nash equilibrium of this game follows:

Lemma 4.22 (Equilibrium – precision loss)**.** *There exists a symmetric linear noisy Nash equilibrium in which all agents use κ as defined in Lemma 4.4 and draw independent noise from a distribution H_i with variance ν^*, where H_i can be any distribution with mean zero and variance $\nu_{i,prec}^* = \sqrt{\frac{\beta}{1-\beta}}$. Here, agents may choose* any *distribution with these properties.*

Lemma 4.23 (Equilibrium – entropy loss)**.** *There exists a symmetric linear noisy Nash equilibrium in which all agents use κ as defined in Lemma 4.4 and draw independent noise from the Gaussian distribution $N(0, \nu_{ent}^*)$.*

Since the Gaussian is the maximum entropy distribution with fixed variance ν^*, all agents draw noise (independently) from the same distribution.

This completes the proof of Theorem 4.11, since we have found a symmetric linear noisy action for each player which satisfies the first order conditions of the privacy aware game. We next analyze and discuss the *price of privacy* in the new game.

4.4 Price of Privacy

Informally, the price of privacy describes the loss in quality of some measure as a result of introducing privacy-awareness into the game. We describe two different, but related, quantities which represent this effect and discuss some settings where one may be interested in each.

4.4.1 The Agents' Cost

Our first notion of the price of privacy can be viewed as the utility the *agents* pay in the original game in order to express their value for obfuscation. Formally, recall that $u_i(\theta_i, \theta_{-i}) = -(1-\alpha)(\theta_i - \bar{\theta})^2 - \alpha(\theta_i - s)^2$ is the utility function in the unmodified game. In the symmetric linear Nash equilibrium, each player chooses the prescribed optimal action θ_i^* and earns utility $u_i(\theta_i^*, \theta_{-i}^*)$. Similarly, in the modified game, in the symmetric linear Nash equilibrium chooses the optimal noisy action $\tilde{\theta}_i^*$. We can now ask how much worse-off playing the actions $\tilde{\theta}^*$ in the original game makes a representative player as compared to playing the actions θ^*. Formally, we can quantify the price of privacy by taking the ratio of these utilities.

Definition 4.24 (Agents' price of privacy)**.** The *price of privacy* given an information structure is

$$\mathrm{PoP}(\tau_x, \tau_y, \beta) = \frac{\mathbb{E}_i[u_i(\tilde{\theta}_i^*, \tilde{\theta}_{-i}^*)]}{\mathbb{E}_i[u_i(\theta_i^*, \theta_{-i}^*)]}$$

where the expected utility is with respect to the game with signal variances σ_x^2 and σ_y^2.

Lemma 4.25 (Agents' price of privacy – form)**.** *The price of privacy in the game where agents play a linear strategy has the form*

$$\mathrm{PoP}(\tau_x, \tau_y, \beta) = 1 + \frac{\nu_i^*}{\mathbb{E}_i[u_i(\theta_i^*, \theta_{-i}^*)]}$$

where the expected utility is with respect to the game with signal variances σ_x^2 *and* σ_y^2*.*

Proof. First note that

$$\mathop{\mathbb{E}}_i[u_i(\theta_i, \tilde{\theta}_{-i})] = \mathop{\mathbb{E}}_i[u_i(\theta_i, \theta_{-i})]$$

because the noise added to θ_i has a mean of zero. Now, Corollary 4.16 lets us write

$$\mathrm{PoP}(\tau_x, \tau_y, \beta) = \frac{\mathbb{E}_i[u_i(\tilde{\theta}_i^*, \tilde{\theta}_{-i}^*)]}{\mathbb{E}_i[u_i(\theta_i^*, \theta_{-i}^*)]} = \frac{\mathbb{E}_i[u_i(\tilde{\theta}_i^*, \tilde{\theta}_{-i}^*)] + \nu_i^*}{\mathbb{E}_i[u_i(\theta_i^*, \theta_{-i}^*)]}$$

where we have factored out all of the negative signs. Combining with the previous part, we have that

$$\mathrm{PoP}(\tau_x, \tau_y, \beta) = \frac{\mathbb{E}_i[u_i(\tilde{\theta}_i^*, \tilde{\theta}_{-i}^*)] + \nu_i^*}{\mathbb{E}_i[u_i(\theta_i^*, \theta_{-i}^*)]} = \frac{\mathbb{E}_i[u_i(\tilde{\theta}_i^*, \theta{-}i^*)] + \nu_i^*}{\mathbb{E}_i[u_i(\theta_i^*, \theta_{-i}^*)]} = 1 + \frac{\nu_i^*}{\mathbb{E}_i[u_i(\theta_i^*, \theta_{-i}^*)]},$$

as desired. □

Theorem 4.26 (Price of privacy – value)**.** *In the game with finitely many agents, the price of privacy is:*

$$\mathrm{PoP}(\tau_x, \tau_y, \beta) = 1 + \frac{\nu_i^*}{\alpha(\kappa^2\sigma_x^2 + (1-\kappa)^2\sigma_y^2) + \frac{(1-\alpha)\kappa^2(n-1)}{n}\sigma_x^2},$$

for

$$\nu_{i,prec}^* = \sqrt{\frac{\beta}{1-\beta}\left(\alpha + (1-\alpha)\left(1 - \frac{1}{n}\right)^2\right)^{-1}}, \quad or$$

$$\nu_{i,ent}^* = \frac{\beta}{1-\beta}\left(\alpha + (1-\alpha)\left(1 - \frac{1}{n}\right)^2\right)^{-1}$$

if we measure the privacy loss using precision or entropy, respectively, and

$$\kappa = \frac{\alpha n^2 \tau_x}{\alpha n^2 \tau_x + ((n-1)^2 + \alpha(2n-1))\,\tau_y}.$$

Proposition 4.27. *In the game with infinitely many agents, the price of privacy is:*

$$\mathrm{PoP}(\tau_x, \tau_y, \beta) = 1 + \frac{\nu_i^*}{(\alpha(1-\kappa)^2\sigma_y^2 + \kappa^2\sigma_x^2)},$$

where ν_i^* *is* $\sqrt{\frac{\beta}{1-\beta}}$ *if privacy is measured by precision and* $\frac{\beta}{1-\beta}$ *if privacy is measured by entropy and recalling that* $\kappa = \frac{\tau_x\alpha}{\tau_x\alpha + \tau_y}$*.*

Proof. This follows directly from the expected utility computation in Proposition 4.7 and the functional form of the price of privacy in Lemma 4.25. □

We can observe several features of the price of privacy. First, it can be *arbitrarily large*, depending on the value of the parameters. If β is close to 1, then agents have a relatively high value for obfuscation and will add large amounts of noise to their actions in order to protect their private signals. As β becomes closer to zero, agents do not care too much about privacy and their actions become more and more similar to those in the original game, so the price of privacy is minimized when $\beta = 0$. Second, if we fix a value of β, then the price of privacy decreases as the variance of the public and private signals *increase*. This is because the value of β determines the noise that players will add to their signals, and therefore fixes the numerator of the fractional part of the price of privacy. The price of privacy is *decreasing* in any factor that improves the expected utility, such as decreasing the signal variances or n (as in Corollary 4.8). The effect is ambiguous when changing σ_y^2, due to the over-weighting phenomenon.

This ratio measures the degree to which agents are worse-off in the coordination and guessing components by playing the privacy-aware equilibrium noisy actions as compared to the privacy-unaware equilibrium actions for a fixed α, σ_x^2, and σ_y^2 and the realizations of x_i and y. This can be thought of as describing the increased risk of making the 'wrong' decision in an opinion-aggregation setting or of misvaluing an asset in a financial markets setting as a result of agents adding noise to their actions.

4.4.2 The Aggregator's Cost

Suppose instead we are viewing this game from the position of a data aggregator. The aggregator can observe the actions of some finite number of agents n and takes the average of these to estimate the true state of the world s by taking a simple average. The aggregator does not know the realizations of any of the x_i or y, but they do know the signal variances and the value of κ; they also know that $\mathbb{E}[\theta_i] = \mathbb{E}[\tilde{\theta}_i] = s$ for all agents i, so the sample average will provide an unbiased estimate of s. If we define the aggregator's 'utility' to be the variance of its sample about s, then we can make the following observation:

Lemma 4.28 (Aggregator's utility)**.** *Consider an instance of the privacy-aware game where an aggregator observes the actions of n agents (either all all of the agents in the finite case or some uniformly random sample in the finite or infinite case), the signal variances are σ_x^2 and σ_y^2, and players choose to add mean-zero noise with variance ν_i^*. Then the utility of the aggregator, as measured by the variance of the sample average about the true state s is given by*

$$\begin{aligned}\mathcal{U}_{agg}(\sigma_x^2, \sigma^2, \nu_i^*, n) &= \mathbb{E}\left[\left(\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\theta}_i\right) - s\right)^2 \middle| s\right] \\ &= \frac{\kappa^2}{n}\sigma_x^2 + \frac{\nu_i^*}{n} + (1-\kappa)^2\sigma_y^2.\end{aligned}$$

We can find the aggregator's utility in the privacy-unaware game by letting $\nu_i^* = 0$, which recovers a result in Morris and Shin (2002).

Proof. By definition,

$$\begin{aligned}\mathcal{U}_{agg}(\sigma_x^2, \nu_i^*, \sigma^2, n) &= \mathbb{E}\left[\left(\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\theta}_i\right) - s\right)^2 \middle| s\right] \\ &= \mathbb{E}\left[\left(\left(\frac{1}{n}\sum_{i=1}^{n}\kappa x_i + (1-\kappa)y + \eta_i\right) - s\right)^2 \middle| s\right] \\ &= \mathbb{E}\left[\left(\left(\frac{1}{n}\sum_{i=1}^{n}\kappa x_i + \eta_i\right) + (1-\kappa)y - s\right)^2 \middle| s\right] \\ &= \mathbb{E}\left[\left(\left(\frac{1}{n}\sum_{i=1}^{n}\kappa(s + \varepsilon_{x_i}) + \eta_i\right) + (1-\kappa)(s + \varepsilon_y) - s\right)^2 \middle| s\right] \\ &= \mathbb{E}\left[\left(\left(\frac{1}{n}\sum_{i=1}^{n}\kappa\varepsilon_{x_i} + \eta_i\right) + (1-\kappa)\varepsilon_y\right)^2 \middle| s\right] \\ &= \mathbb{E}\left[\left(\frac{\kappa}{n}\sum_{i=1}^{n}\varepsilon_{x_i} + \frac{1}{n}\sum_{i=1}^{n}\eta_i + (1-\kappa)\varepsilon_{)}\right)^2\right] \\ &= \frac{\kappa^2}{n}\sigma_x^2 + \frac{\nu_i^*}{n} + (1-\kappa)^2\sigma_y^2\end{aligned}$$

where we have decomposed x_i and y into s plus mean-zero Gaussian noise ε_{x_i} and ε_y. □

We can again take the ratio of the aggregator's utility in the privacy-aware game to that in the privacy-unaware game to quantify the extent to which privacy-awareness degrades the quality of the aggregator's sample mean as an estimate of s.

Lemma 4.29 (The Aggregator's price of privacy)**.** *The price of privacy for the aggregator in equilibrium is given by*

$$\mathrm{PoP}_{agg}(\sigma_x^2, \sigma_y^2, n) = \frac{\mathbb{E}[\mathcal{U}_{agg}(\sigma_x^2, \sigma_y^2, \nu_i^*, n)]}{\mathbb{E}[\mathcal{U}_{agg}(\sigma_x^2, \sigma_y^2, 0, n)]} = 1 + \frac{\nu_i^*}{\kappa^2\sigma_x^2 + n(1-\kappa)^2\sigma_y^2}$$

where ν_i^* is $\sqrt{\frac{\beta}{1-\beta}}$ if privacy is measured by precision and $\frac{\beta}{1-\beta}$ if privacy is measured by entropy and recalling that $\kappa = \frac{\tau_x\alpha}{\tau_x\alpha+\tau_y}$ in the game with infinitely many agents and $\kappa = \frac{\alpha n^2\tau_x}{\alpha n^2\tau_x+((n-1)^2+\alpha(2n-1))\tau_y}$ in the game with finitely many agents.

Proof. The result follows from plugging in the forms for the expected utilities from Lemma 4.28 and simplifying. □

We can again observe that, for the same reason as the agents' price of privacy, that this quantity can be arbitrarily large as β approaches one and that it is decreasing as the signal variances grow. We can also observe that this price decreases towards one as n grows, all else fixed. This is consistent with the intuition that an aggregator can offset the cost of agents adding more noise to their actions by making more observations.

The agents' and aggregator's prices of privacy are somewhat similar both in functional form and in interpretation and this results from their values almost aligning. The agents' cost is measured by a weighted deviation from both the true state and the average while the aggregator's is measured by a deviation from the true state alone. Any configuration of the game parameters which cause $\kappa = 1$ (i.e. agents ignore the public signal) causes them to be exactly equal. This again highlights the problem of over-weighting the public signal, and this issue is exacerbated for the aggregator, who, in effect, observes n 'copies' of y rolled into the actions θ_i and $\tilde{\theta}_i$. If the aggregator knows the realization of y, they can, similarly to the agents, use that information to find (an estimate of) the private signals x_i from the observations of θ_i or $\tilde{\theta}_i$. This motivates the same privacy concern as in the case where other agents were trying to infer agent i's private signal, this time in the context of a distrusted central aggregator.

4.5 Discussion and Future Directions

In this work, we used a game-theoretic model to examine a game with agents acting to obscure their private information. We began with the Keynesian Beauty Contest and modified the agents' utility functions so as to endogenize a notion of privacy. Using this, we can quantify the social costs of this 'selfish' privacy protection as a 'price of privacy', both with respect to the agents as well as an (untrusted) central aggregator.

A clear next-step would be to repeat this analysis for other stylized models of strategic interaction, such as bargaining games or resource allocation problems, where individuals seek to balance expressing their preferences well enough to achieve a high payoff but choose to deviate slightly from this in order to obscure their true valuations.

Finally, the high-level motivation in this work is to examine how strategic agents behave when they are concerned with privacy, so another interesting direction would be to integrate this approach with existing formal notions of privacy. For example, examining what agents' utility functions and action spaces must look like in order for there to exist an equilibrium which implements a locally differentially private mechanism could be a fruitful avenue of research.

Bibliography

F. Allen, S. Morris, and H. S. Shin. Beauty contests and iterated expectations in asset markets. *The Review of Financial Studies*, 19(3):719–752, 2006.

R. Bassily. Linear queries estimation with local differential privacy. *arXiv preprint arXiv:1810.02810*, 2018.

J. O. Berger, J. M. Bernardo, D. Sun, et al. The formal definition of reference priors. *The Annals of Statistics*, 37(2):905–938, 2009.

A. Bosch-Domenech, J. G. Montalvo, R. Nagel, and A. Satorra. One, two,(three), infinity,...: Newspaper and lab beauty-contest experiments. *American Economic Review*, 92(5):1687–1701, 2002.

G. Cespa and X. Vives. The beauty contest and short-term trading. *The Journal of Finance*, 70(5): 2099–2154, 2015.

Y. Chen, S. Chong, I. A. Kash, T. Moran, and S. Vadhan. Truthful mechanisms for agents that value privacy. *ACM Transactions on Economics and Computation (TEAC)*, 4(3):13, 2016.

T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In S. Halevi and T. Rabin, editors, *Theory of Cryptography*, pages 265–284, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg. ISBN 978-3-540-32732-5.

P. Gao. Keynesian beauty contest, accounting disclosure, and market efficiency. *Journal of Accounting Research*, 46(4):785–807, 2008.

R. Gradwohl and R. Smorodinsky. Perception games and privacy. *Games and Economic Behavior*, 104:293–308, 2017.

C. Hellwig and L. Veldkamp. Knowing what others know: Coordination motives in information acquisition. *The Review of Economic Studies*, 76(1):223–251, 2009. doi: 10.1111/j.1467-937X.2008.00515.x.

M. Joseph, J. Kulkarni, J. Mao, and Z. S. Wu. Locally private gaussian estimation. *arXiv preprint arXiv:1811.08382*, 2018.

P. Kairouz, S. Oh, and P. Viswanath. Extremal mechanisms for local differential privacy. *J. Mach.*

Learn. Res., 17(1):492–542, Jan. 2016. ISSN 1532-4435. URL http://dl.acm.org/citation.cfm?id=2946645.2946662.

S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011.

J. M. Keynes. *The general theory of employment, interest, and money*. Macmillan, 1936.

F. Liu. Generalized gaussian mechanism for differential privacy. *IEEE Transactions on Knowledge and Data Engineering*, 31(4):747–756, 2019.

S. Morris and H. S. Shin. Social value of public information. *American Economic Review*, 92(5): 1521–1534, December 2002. doi: 10.1257/000282802762024610. URL http://www.aeaweb.org/articles?id=10.1257/000282802762024610.

R. Nagel, C. Bühren, and B. Frank. Inspired and inspiring: Hervé moulin and the discovery of the beauty contest game. *Mathematical Social Sciences*, 90:191–207, 2017.

A. Narayanan and V. Shmatikov. Robust de-anonymization of large sparse datasets. In *Security and Privacy, 2008. SP 2008. IEEE Symposium on*, pages 111–125. IEEE, 2008.

K. Nissim, C. Orlandi, and R. Smorodinsky. Privacy-aware mechanism design. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 774–789. ACM, 2012.

N. Wang, X. Xiao, Y. Yang, J. Zhao, S. C. Hui, H. Shin, J. Shin, and G. Yu. Collecting and analyzing multidimensional data with local differential privacy. In *Proceedings of IEEE ICDE*, 2019.

S. L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965. ISSN 01621459. URL http://www.jstor.org/stable/2283137.

Chapter 5

Algorithms and Redistricting

5.1 Introduction to district compactness

Striving for the *geographic compactness* of electoral districts is a traditional principle of redistricting Altman (1998), and, to that end, many jurisdictions have included the criterion of compactness in their legal code for drawing districts. Some of these include Iowa's measuring the perimeter of districts Iowa Code §42.4(4), Maine's minimizing travel time within a district Maine Statute §1206-A, and Idaho's avoiding drawing districts which are 'oddly shaped' Idaho Statute 72-1506(4). Such measures can vary widely in their precision, both mathematical and otherwise. Computing the perimeter of districts is a very clear definition, minimizing travel time is less so, and what makes a district oddly shaped or not seems rather challenging to consider from a rigorous standpoint.

While a strict definition of when a district is or is not 'compact' is quite elusive, the purpose of such a criterion is much easier to articulate. Simply put, a district which is bizarrely shaped, such as one with small tendrils grabbing many distant chunks of territory, probably wasn't drawn like that by accident. Such a shape need not be drawn for nefarious purposes, but its unusual nature should trigger closer scrutiny. Measures to compute the geographic compactness of districts are intended to formalize this quality of 'bizarreness' mathematically. We briefly note here that the term *compactness* is somewhat overloaded, and that we exclusively use the term to refer to the shape of geographic regions and not to the topological definition of the word.

People have formally studied geographic compactness for nearly two hundred years, and, over that period, scientists and legal scholars have developed many formulas to assign a numerical measure of 'compactness' to a region such as an electoral district Young (1988). Three of the most commonly discussed formulations are the *Polsby-Popper score*, which measures the normalized ratio of a district's area to the square of its perimeter, the *convex hull score*, which measures the ratio of the area of a district to the smallest convex region containing it, and the *Reock score*, which measures the ratio of the area of a district to the area of the smallest circular disc containing it. Each of these measures is appealing at an intuitive level, since they assign to a district a single scalar value between zero and one, which presents a simple method to compare the relative compactness of two or more districts. Additionally, the mathematics underpinning each is widely understandable by the relevant parties, including lawmakers, judges, advocacy groups, and the general public.

However, none of these measures truly discerns which districts are 'compact' and which are not. For each score, we can construct a mathematical counterexample for which a human's intuition and the score's evaluation of a shape's compactness differ. A region which is roughly circular but has a jagged boundary may appear compact to a human's eye, but such a shape has a very poor Polsby-Popper score. Similarly, a very long, thin rectangle appears non-compact to a person, but has a perfect convex hull score. Additionally, these scores often do not agree. The long, thin rectangle has a very poor Polsby-Popper score, and the ragged circle has an excellent convex hull score. These issues are well-studied by political scientists and mathematicians alike Polsby and Popper (1991); Frolov (1975); Maceachren (1985); Barnes and Solomon (Forthcoming).

In this paper, we propose a further critique of these measures, namely *sensitivity under the choice of map projection.* Each of the compactness scores named above is defined as a tool to evaluate geometric shapes in the plane, but in reality we are interested in analyzing shapes which sit on the surface of the planet Earth, which is (roughly) spherical. When we analyze the geometric properties of a geographic region, we work with a *projection* of the Earth onto a flat plane, such as a piece of paper or the screen of a computer. Therefore, when a shape is assigned a compactness score, it is implicitly done with respect to some choice of map projection. We prove that this may have serious consequences for the comparison of districts by these scores. Because there is no projection from the sphere to the plane which preserves 'too many' metric properties and most compactness scores synthesize several of these properties, it is unreasonable to expect any projection to preserve the numerical values of these scores for all regions. However, since there are projections which preserve *some* geometric properties, such as those which preserve the area of all regions or conformal projections which preserve the angle of intersection of all line segments, we might ask a weaker question and consider whether there is a projection which can preserve the *induced ordering* of a compactness score over all regions.

In particular, we consider the Polsby-Popper, convex hull, and Reock scores on the sphere, and demonstrate that for any choice of map projection, there are two regions, A and B, such that A is more compact than B on the sphere but B is more compact than A when projected to the plane. We prove our results in a theoretical context before evaluating the extent of this phenomenon empirically. We find that with real-world examples of Congressional districts, the effect of the commonly-used *Plate carée* projection, which treats latitude-longitude coordinates as Cartesian coordinate pairs, on the convex hull and Polsby-Popper scores is relatively minor, but the impact on Reock scores is quite dramatic, which may have serious implications for the use of this measure as a tool to evaluate geographic compactness.

5.1.1 Organization

For each of the compactness scores we analyze, our proof that no map projection can preserve their order follows a similar recipe. We first use the fact that any map projection which preserves an ordering must preserve the *maximizers* in that ordering. In other words, if there is some shape which a score says is "the most compact" on the sphere but the projection sends this to a shape in the plane which is “not the most compact”, then whatever shape *does* get sent to the most compact shape in the plane leapfrogs the first shape in the induced ordering. For all three of the scores we study, such a maximizer exists.

Using this observation, we can restrict our attention to those map projections which preserve the maximizers in the induced ordering, then argue that any projection in this restricted set must permute the order of scores of some pair of regions.

Preliminaries We first introduce some definitions and results which we will use to prove our three main theorems. Since spherical geometry differs from the more familiar planar geometry, we carefully describe a few properties of spherical lines and triangles to build some intuition in this domain.

Convex Hull For the convex hull score, we first show that any projection which preserves the maximizers of the convex hull score ordering must maintain certain geometric properties of shapes and line segments between the sphere and the plane. Using this, we demonstrate that no map projection from the sphere to the plane can preserve these properties, and therefore no such convex hull score order preserving projection exists.

Reock For the Reock score, we follow a similar tack, first showing that any order-preserving map projection must also preserve some geometric properties and then demonstrating that such a map projection cannot exist.

Polsby-Popper To demonstrate that there is no projection which maintains the score ordering induced by the Polsby-Popper score, we leverage the difference between the *isoperimetric inequalities* on the sphere and in the plane, in that the inequality for the plane is scale invariant in that setting but not on the sphere, in order to find a pair of regions in the sphere, one more compact than the other, such that the less compact one is sent to a circle under the map projection.

Empirical Results We finally examine the impact of the Cartesian latitude-longitude map projection on the convex hull, Reock, and Polsby-Popper scores and the ordering of regions under these scores. While the impacts of the projection on the convex hull and Polsby-Popper scores and their orderings are not severe, the Reock score and the Reock score ordering both change dramatically under the map projection.

5.2 Preliminaries

We begin by introducing some necessary observations, definitions, and terminology which will be of use later.

5.2.1 Spherical Geometry

In this section, we present some basic results about spherical geometry with the goal of proving Girard's Theorem, which states that the area of a triangle on the unit sphere is the sum of its interior angles minus π. Readers familiar with this result should feel free to skip ahead.

We use $\mathbb{R}^2$ to denote the Euclidean plane with the usual way of measuring distances,

$$d(x, y) = \sqrt{(x-y)^2};$$

similarly, $\mathbb{R}^3$ denotes Euclidean 3-space. We use $\mathbb{S}^2$ to denote the *unit 2-sphere*, which can be thought of as the set of points in $\mathbb{R}^3$ at Euclidean distance one from the origin.

In this paper, we only consider the sphere and the plane, and leave the consideration of other surfaces, measures, and metrics to future work.

Definition 5.1. On the sphere, a **great circle** is the intersection of the sphere with a plane passing through the origin. These are the circles on the sphere with radius equal to that of the sphere. See Figure 5.1 for an illustration.

Definition 5.2. Lines in the plane and great circles on the sphere are called **geodesics**. A **geodesic segment** is a line segment in the plane and an arc of a great circle on the sphere.

The idea of geodesics generalizes the notion of 'straight lines' in the plane to other settings. One critical difference is that in the plane, there is a unique line passing through any two distinct points and a unique line segment joining them. On the sphere, there will typically be a unique great circle and *two* geodesic segments through a pair of points, with the exception of one case.

Definition 5.3. A **triangle** in the plane or the sphere is defined by three distinct points and the shortest geodesics connecting each pair of points.

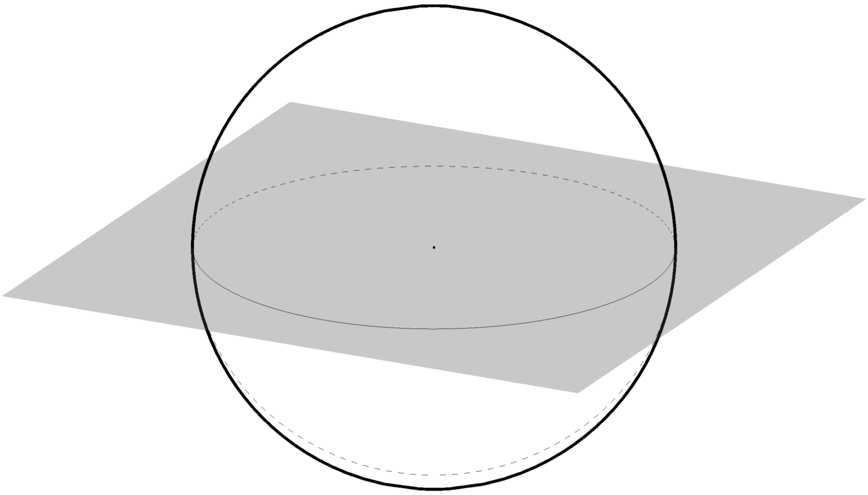


Figure 5.1: A great circle on the sphere with its identifying plane.

Observation 5.4. *Given any two points p and q on the sphere which are not antipodal, there is a unique great circle through p and q and therefore two geodesic segments joining them.*

If p and q *are* antipodal, then any great circle containing one must contain the other as well, so there are infinitely many such great circles. For any two non-antipodal points on the sphere, one of the geodesic segments will be shorter than the other. This shorter geodesic segment is the shortest path between the points and its length is the metric distance between p and q.

We now have enough terminology to show a very important fact about spherical geometry.

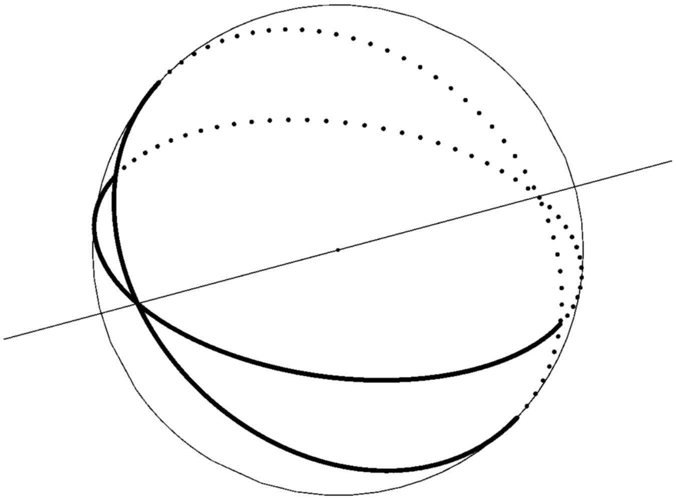


Figure 5.2: Two great circles meet at antipodal points.

This observation is one of the salient features which distinguishes it from the more familiar planar geometry.

Observation 5.5. *Any pair of distinct great circles on the sphere intersect exactly twice, and the points of intersection are antipodes.*

In the plane, it is always the case that any pair of distinct lines intersects exactly once or never, in which case we call them *parallel*. Since distinct great circles on the sphere intersect exactly twice, there is no such thing as 'parallel lines' on the sphere, and we have to be careful about discussing 'the' intersection of two great circles since they do not meet at a unique point. Another difference between spherical and planar geometry appears when computing the angles of triangles. In the planar setting, the sum of the interior angles of a triangle is always π, regardless of its area. However, in the spherical case we can construct a triangle with three right angles. The north pole and two points on the equator, one a quarter of the way around the sphere from the other, form such a triangle. Its area is one eighth of the whole sphere, or $\frac{\pi}{2}$, which is, not coincidentally, equal to $\frac{\pi}{2} + \frac{\pi}{2} + \frac{\pi}{2} - \pi$. Girard's theorem, which we will prove below, connects the total angle to the area of a spherical triangle.

In order to show Girard's Theorem, we need some way to translate between *angles* and *area*. To do that, we'll use a shape which doesn't even exist in the plane: the *diangle* or *lune*. We know that two great circles intersect at two antipodal points, and we can also see that they cut the surface of the sphere into four regions. Consider one of these regions. Its boundary is a pair of great circle segments which connect antipodal points and meet at some angle $\theta \leq \pi$ at both of these points.

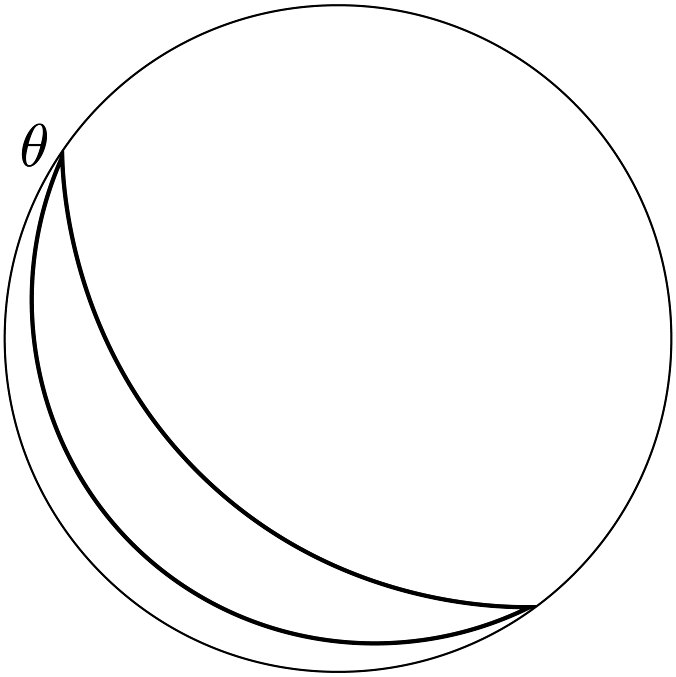


Figure 5.3: A lune corresponding to an angle θ.

Using that the surface area of a unit sphere is 4π, computing the area of a lune with angle θ is straightforward.

Claim 5.6. *Consider a lune whose boundary segments meet at angle θ. Then the area of this lune is 2θ.*

Now that we have a tool that lets us relate angles and areas, we can prove Girard's Theorem.

Lemma 5.7. *(Girard's Theorem)*

The sum of the interior angles of a spherical triangle is strictly greater than π. More specifically, the sum of the interior angles is equal to π plus the area of the triangle.

Proof. Consider a triangle T on the sphere with angles θ_1, θ_2, and θ_3. Let area(T) denote the area of this triangle. If we extend the sides of the triangle to their entire great circles, each pair intersects at the vertices of T as well as the three points antipodal to the vertices of T, and at the same angles at antipodal points. This second triangle is congruent to T, so its area is also area(T). Each pair of great circles cuts the sphere into four lunes, one which contains T, one which contains the antipodal triangle, and two which do not contain either triangle. We are interested in the three pairs of lunes which do contain the triangles. We will label these lunes by their angles, so we have a lune $L(\theta_1)$ and its antipodal lune $L'(\theta_1)$, and we can similarly define $L(\theta_2)$, $L'(\theta_2)$, $L(\theta_3)$, and $L'(\theta_3)$.

We have six lunes. In total, they cover the sphere, but share some overlap. If we remove T from two of the three which contain it and the antipodal triangle from two of the three which contain it, then we have six non-overlapping regions which cover the sphere, so the area of the sphere must be equal to the sum of the areas of these six regions.

By the earlier claim, we know that the areas of the lunes are twice their angles, so we can write this as

$$4\pi = 2\theta_1 + 2\theta_1 + (2\theta_2 - \text{area}(T)) + (2\theta_2 - \text{area}(T)) + (2\theta_3 - \text{area}(T)) + (2\theta_3 - \text{area}(T))$$

and rearrange to get

$$\theta_1 + \theta_2 + \theta_3 = \pi + \text{area}(T),$$

which is exactly the statement we wanted to show. □

We will need one more fact about spherical triangles before we conclude this section.

Fact 5.8. *An equilateral triangle is equiangular, and vice versa, where equilateral means that the three sides have equal length and equiangular means that the three angles all have the same measure.*

This result is also true of planar triangles, and the planar version follows from Propositions I.6 and I.8 in Euclid's *Elements* (Byrne, 1847; Crowell, 2016). Since Euclid's proof doesn't rely on the existence of parallel lines, this fact can be shown using his argument.

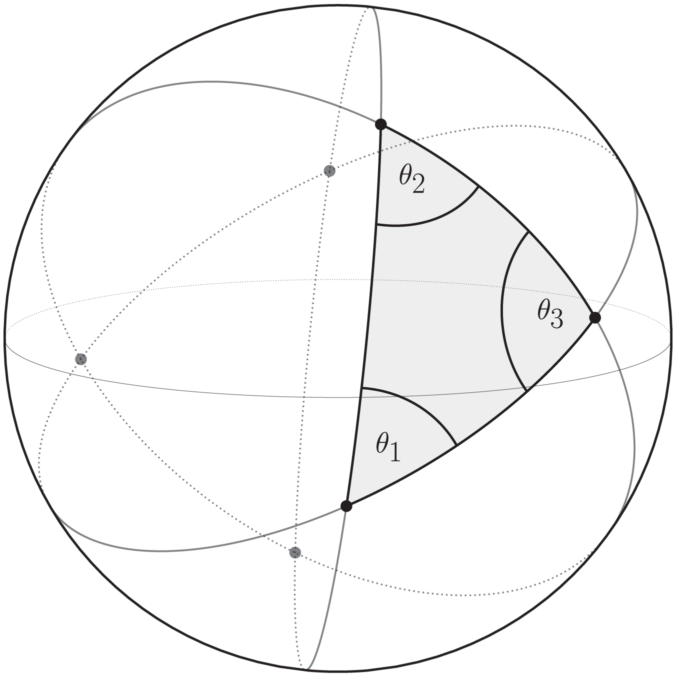


Figure 5.4: A spherical triangle and the antipodal triangle define six lunes.

5.2.2 Some Definitions

Now that we have the necessary tools of spherical geometry, we will wrap up this section with a battery of definitions. We carefully lay these out so as to align with an intuitive understanding of the concepts and to appease the astute reader who may be concerned with edge cases, geometric weirdness, and nonmeasurability. Throughout, we implicitly consider all figures on the sphere to be strictly contained in a hemisphere.

Definition 5.9. A **region** is a non-empty, open subset Ω of $\mathbb{S}^2$ or $\mathbb{R}^2$ such that Ω is bounded and its boundary is piecewise smooth.

We choose this definition to ensure that the *area* and *perimeter* of the region are well-defined concepts. This eliminates pathological examples of open sets whose boundaries have non-zero area or edge cases like considering the whole plane a 'region'.

Definition 5.10. A **compactness score function** $\mathcal{C}$ is a function from the set of all regions to the non-negative real numbers or infinity. We can compare the scores of any two regions, and we adopt the convention that *more compact* regions have *higher* scores. That is, region A is at least as compact as region B if and only if $\mathcal{C}(A) \geq \mathcal{C}(B)$.

The final major definition we need is that of a *map projection*. In reality, the regions we are interested in comparing sit on the surface of the Earth (i.e. a sphere), but these regions are often examined after being projected onto a flat sheet of paper or computer screen, and so have been subject to such a projection.

Definition 5.11. A **map projection** φ is a diffeomorphism from a region on the sphere to a region of the plane.

We choose this definition, and particularly the term *diffeomorphism*, to ensure that φ is smooth, its inverse φ^{-1} exists and is smooth, and both φ and φ^{-1} send regions in their domain to regions in their codomain. Throughout, we use φ to denote such a function from a region of the sphere to a region of the plane and φ^{-1}, to denote the inverse which is a function from a region of the plane back to a region of the sphere.

Since the image of a region under a map projection φ is also a region, we can examine the compactness score of that region both before and after applying φ, and this is the heart of the problem we address in this paper. We demonstrate, for several examples of compactness scores $\mathcal{C}$, that the order induced by $\mathcal{C}$ is different than the order induced by $\mathcal{C} \circ \varphi$ for *any* choice of map projection φ.

Definition 5.12. We say that a map projection φ **preserves the compactness score ordering** of a score $\mathcal{C}$ if for any regions Ω, Ω' in the domain of φ, $\mathcal{C}(\Omega) \geq \mathcal{C}(\Omega')$ if and only if $\mathcal{C}(\varphi(\Omega)) \geq \mathcal{C}(\varphi(\Omega'))$ in the plane.

This is a weaker condition than simply preserving the raw compactness scores. If there is some map projection which results in adding .1 to the score of each region, the raw scores are certainly not preserved, but the ordering of regions by their scores is. Additionally, φ preserves a compactness score ordering if and only if φ^{-1} does.

Definition 5.13. A **cap** on the sphere $\mathbb{S}^2$ is a region on the sphere which can be described as all of the points on the sphere to one side of some plane in $\mathbb{R}^3$. A cap has a *height*, which is the largest distance between this cutting plane and the cap, and a *radius*, which is the radius of the circle formed by the intersection of the plane and the sphere. See Figure 5.5 for an illustration.

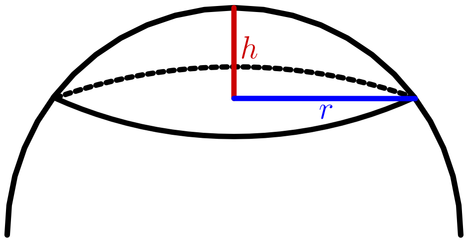


Figure 5.5: The height h and radius r of a spherical cap.

5.3 Convex Hull

We first consider the *convex hull score*. We briefly recall the definition of a convex set and then define this score function.

Definition 5.14. A set in $\mathbb{R}^2$ or $\mathbb{S}^2$ is **convex** if every shortest geodesic segment between any two points in the set is entirely contained within that set.

Definition 5.15. Let $\text{conv}(\Omega)$ denote the *convex hull* of a region Ω in either the sphere or the plane, which is the smallest convex region containing Ω. Then we define the *convex hull score* of Ω as

$$\text{CH}(\Omega) = \frac{\text{area}(\Omega)}{\text{area}(\text{conv}(\Omega)).}$$

Since the intersection of convex sets is a convex set, there is a unique smallest (by containment) convex hull for any region Ω.

Suppose that our map projection φ does preserve the ordering of regions induced by the convex hull score. We begin by observing that such a projection must preserve certain geometric properties

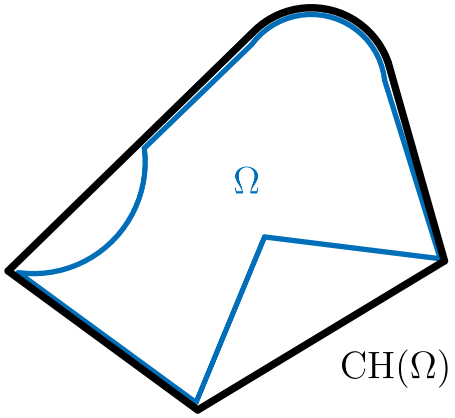


Figure 5.6: A region Ω and its convex hull.

of regions within its domain.

Lemma 5.16. *Let φ be a map projection from some region of the sphere to a region of the plane. If φ preserves the convex hull compactness score ordering, then the following must hold:*

1. *φ and φ^{-1} send convex regions in their domains to convex regions in their codomains.*

2. *φ sends every segment of a great circle in its domain to a line segment in its codomain. That is, it preserves geodesics.*[1]

3. *There exists a region U in the domain of φ such that for any regions $A, B \subset U$, if A and B have equal area on the sphere, then $\varphi(A)$ and $\varphi(B)$ have equal area in the plane. The same holds for φ^{-1} for all pairs of regions inside of $\varphi(U)$.*

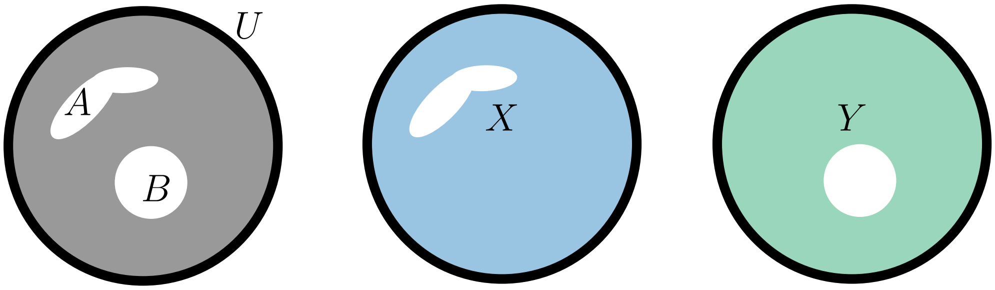


Figure 5.7: Two equal area regions A and B removed from U to form the regions X and Y.

[1] Such a projection is sometimes called a *geodesic map*.

Proof. The proof of (1) follows from the idea that any projection which preserves the convex hull score ordering of regions must preserve the maximizers in that ordering. Here, the maximizers are convex sets.

To show (2) we suppose, for the sake of contradiction, that there is some geodesic segment s in U such that $\varphi(s)$ is not a line segment. Construct two convex spherical polygons L and M inside of U which both have s as a side.

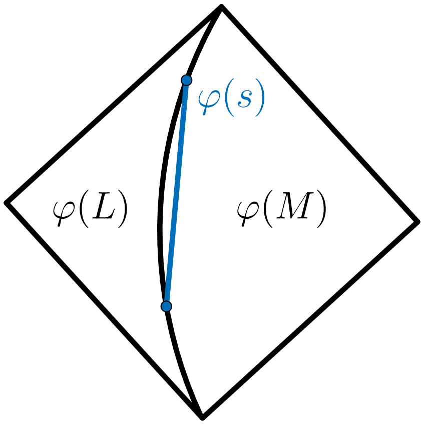


Figure 5.8: If $\varphi(s)$ is not a line segment, then one of $\varphi(M)$ or $\varphi(L)$ is not convex.

By (1), φ must send both of these polygons to convex regions in the plane, but this is not the case. All of the points along $\varphi(s)$ belong to both $\varphi(L)$ and $\varphi(M)$, but since $\varphi(s)$ is not a line segment, we can find two points along it which are joined by some line segment which contains points which only belong to $\varphi(L)$ or $\varphi(M)$, which means that at least one of these convex spherical polygons is sent to something non-convex in the plane, which contradicts our assumption. See Figure 5.8 for an illustration.

That φ^{-1} sends line segments in the plane to great circle segments on the sphere is shown analogously. This completes the proof of (2).

To show (3), let U be some convex region in the domain of φ. Take A, B to be regions of equal area such that A and B are properly contained in the interior of U, as in Figure 5.7. Define two new regions $X = U \setminus A$ and $Y = U \setminus B$, i.e. these regions are equal to U with A or B deleted, respectively.

The cap U is itself the convex hull of both X and Y, and since A and B have equal area, we have that $\mathrm{CH}(X) = \mathrm{CH}(Y)$. Since U is a cap, it is convex, so by (1), $\varphi(U)$ is also convex. Since φ preserves the ordering of convex hull scores and X and Y had equal scores on the sphere, φ must send X and Y to regions in the plane which also have the same convex hull score as each other. Furthermore, the convex hulls of $\varphi(X)$ and $\varphi(Y)$ are $\varphi(U)$.

By definition, we have $\mathrm{CH}(X) = \mathrm{CH}(Y)$ and by the construction of X and Y, we have

$$\frac{\operatorname{area}(\varphi(U)) - \operatorname{area}(\varphi(A))}{\operatorname{area}(\varphi(U))} = \frac{\operatorname{area}(\varphi(U)) - \operatorname{area}(\varphi(B))}{\operatorname{area}(\varphi(U))}$$

and so

$$\operatorname{area}(\varphi(A)) = \operatorname{area}(\varphi(B)),$$

which is what we wanted to show. The proof that φ^{-1} also has this property is analogous.

□

We can now show that no map projection can preserve the convex hull score ordering of regions by demonstrating that there is no projection from a patch on the sphere to the plane which has all three of the properties described in Lemma 5.16.

Theorem 5.17. *There does not exist a map projection with the three properties in Lemma 5.16*

Proof. Assume that such a map, φ, exists, and restrict it to U as above. Let $T \subset U$ be a small equilateral spherical triangle centered at the center of U. Let T_1 and T_2 be two congruent triangles meeting at a point and each sharing a face with T, as in Figure 5.9.

We first argue that the images of $T \cup T_1$ and $T \cup T_2$ are parallelograms.

Without loss of generality, consider $T \cup T_1$. By construction, it is a convex spherical quadrilateral. By symmetry, its geodesic diagonals on the sphere divide it into four triangles of equal area. To see this, consider the geodesic segment which passes through the vertex of T opposite the side shared with T_1 which divides T into two smaller triangles of equal area. Since T is equilateral, this segment meets the shared side at a right angle at the midpoint, and the same is true for the area bisector

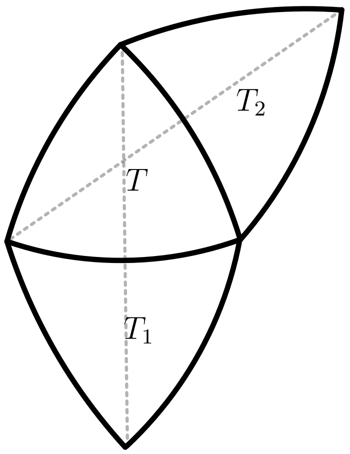


Figure 5.9: The spherical regions T, T_1, T_2.

of T_1. Since both of these bisectors meet the shared side at a right angle and at the same point, together they form a single geodesic segment, the diagonal of the quadrilateral. Since the diagonal cuts each of T and T_1 in half, and T and T_1 have the same area, the four triangles formed in this construction have the same area.

Since φ sends spherical geodesics to line segments in the plane, it must send $T \cup T_1$ to a Euclidean quadrilateral Q whose diagonals are the images of the diagonals of the spherical quadrilateral $T \cup T_1$.

Since φ sends equal area regions to equal area regions, it follows that the diagonals of Q split it into four equal area triangles.

We now argue that this implies that Q is a Euclidean parallelogram by showing that its diagonals bisect each other. Since the four triangles formed by the diagonals of Q are all the same area, we can pick two of these triangles which share a side and consider the larger triangle formed by their union. One side of this triangle is a diagonal d_1 of Q and its area is bisected by the other diagonal d_2, which passes through d_1 and its opposite vertex. The area bisector from a vertex, called the *median*, passes through the midpoint of the side d_1, meaning that the diagonal d_2 bisects the diagonal d_1. Since this holds for any choice of two adjacent triangles in Q, the diagonals must bisect each other, so Q is a parallelogram.

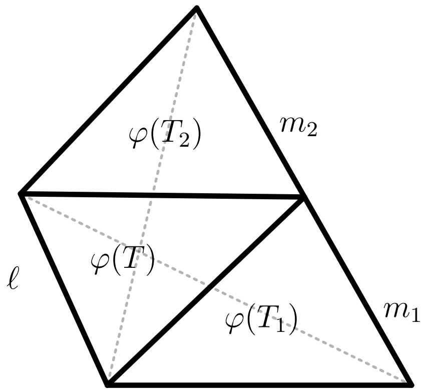


Figure 5.10: The image under φ of T, T_1, T_2 which form the quadrilateral in the plane.

Since $T \cup T_1$ and $T \cup T_2$ are both spherical quadrilaterals which overlap on the spherical triangle T, the images of $T \cup T_1$ and $T \cup T_2$ are Euclidean parallelograms of equal area which overlap on a shared triangle $\varphi(T)$. See Figure 5.10 for an illustration.

Because the segment ℓ is parallel to m_1 and m_2, m_1 and m_2 are parallel to each other, and because they meet at the point shared by all three triangles, m_1 and m_2 together form a single segment parallel to ℓ. Therefore, the image of the three triangles forms a quadrilateral in the plane. Therefore, the image of $T \cup T_1 \cup T_2$ has a boundary consisting of four line segments.

To find the contradiction, consider the point on the sphere shared by T, T_1, and T_2. Since these triangles are all equilateral spherical triangles, the three angles at this point are each strictly greater than $\frac{\pi}{3}$ radians, because the sum of interior angles on a triangle is strictly greater than π. so, the total measure of the three angles at this point is greater than π, Therefore, the two geodesic segments which are part of the boundaries of T_1 and T_2 meet at this point at an angle of measure strictly greater than π. Therefore, together they do not form a single geodesic. On the sphere, the region $T \cup T_1 \cup T_2$ has a boundary consisting of five geodesic segments whereas its image has a boundary consisting of four, which contradicts the assumption that φ and φ^{-1} preserve geodesics. □

This implies that no map projection can preserve the ordering of regions by their convex hull scores, which is what we aimed to show.

5.4 Reock

Let circ(Ω) denote the *smallest bounding circle* (smallest bounding *cap* on the sphere) of a region Ω. Then the *Reock score* of Ω is

$$\operatorname{Reock}(\Omega) = \frac{\operatorname{area}(\Omega)}{\operatorname{area}(\operatorname{circ}(\Omega))}.$$

We again consider what properties a map projection φ must have in order to preserve the ordering of regions by their Reock scores.

Lemma 5.18. *If φ preserves the ordering of regions induced by their Reock scores, then the following must hold:*

1. *φ sends spherical caps in its domain to Euclidean circles in the plane, and φ^{-1} does the opposite.*

2. *There exists a region U in the domain of φ such that for any regions $A, B \subset U$, if A and B have equal area on the sphere, then $\varphi(A)$ and $\varphi(B)$ have equal area in the plane. The same holds for φ^{-1} for all pairs of regions inside of $\varphi(U)$.*

Proof. Similarly to the convex hull setting, the proof of (1) follows from the requirement that φ preserves the maximizers in the compactness score ordering. In the case of the Reock score, the maximizers are caps in the sphere and circles in the plane.

To show (2), let κ be a cap in the domain of φ, and let $A, B \subset \kappa$ be two regions of equal area properly contained in the interior of κ. Then, define two new regions $X = \kappa \setminus A$ and $Y = \kappa \setminus B$, which can be thought of as κ with A and B deleted, respectively.

Since κ is the smallest bounding cap of X and Y and since A and B have equal areas, $\operatorname{Reock}(X) = \operatorname{Reock}(Y)$. Furthermore, by (1), φ must send κ to some circle in the plane, which is the smallest bounding circle of $\varphi(X)$ and $\varphi(Y)$. Since φ preserves the ordering of Reock scores, it must be that $\varphi(X)$ and $\varphi(Y)$ have identical Reock scores in the plane.

By definition, we can write

$$\text{Reock}(X) = \text{Reock}(Y)$$

and so

$$\frac{\text{area}(\varphi(X))}{\text{area}(\varphi(\kappa))} = \frac{\text{area}(\varphi(Y))}{\text{area}(\varphi(\kappa))},$$

and by the construction of X and Y, we have

$$\frac{\text{area}(\varphi(\kappa)) - \text{area}(\varphi(A))}{\text{area}(\varphi(\kappa))} = \frac{\text{area}(\varphi(\kappa)) - \text{area}(\varphi(B))}{\text{area}(\varphi(\kappa))}$$

meaning that $\text{area}(\varphi(A)) = \text{area}(\varphi(B))$. Thus, for all pairs of regions of the same area inside of κ, the images under φ of those regions will have the same area as well.

The same construction works in reverse, which demonstrates that φ^{-1} also sends regions of equal area in some circle in the plane to regions of equal area in the sphere. □

We can now show that no such φ exists. Rather than constructing a figure on the sphere and examining its image under φ, it will be more convenient to construct a figure in the plane and reason about φ^{-1}.

Theorem 5.19. *There does not exist a map projection with the two properties in Lemma 5.18.*

Proof. Assume that such a φ does exist and restrict its domain to a cap κ as above. This corresponds to a restriction of the domain of φ^{-1} to a circle in the plane. Inside of this circle, draw seven smaller circles of equal area tangent to each other as in Figure 5.11.

Under φ^{-1}, they must be sent to a similar configuration of equal-area caps on the sphere .

However, the radius of a of a spherical cap is determined by its area, so since the areas of these caps are all the same, their radii must be as well. Thus, the midpoints of these caps form six equilateral triangles on the sphere which meet at a point. However, this is impossible, as the three angles of an equilateral triangle on the sphere must all be greater than $\frac{\pi}{3}$, but the total measure of all the angles at a point must be equal to 2π, which contradicts the assumption that such a φ exists. □

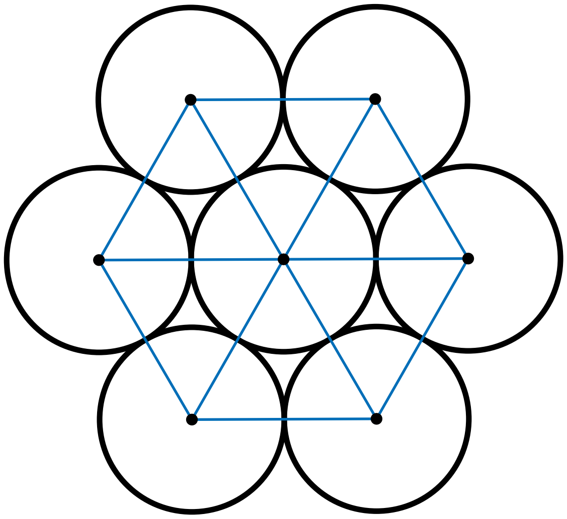


Figure 5.11: Seven circles arranged as in the construction for Theorem 5.19.

This shows that no map projection exists which preserves the ordering of regions by their Reock scores.

5.5 Polsby-Popper

The final compactness score we analyze is the *Polsby-Popper score*, which takes the form of an *isoperimetric quotient*, meaning it measures how much area a region's perimeter encloses, relative to all other regions with the same perimeter.

Definition 5.20. The Polsby-Popper score of a region Ω is defined to be

$$\mathrm{PP}(\Omega) = \frac{4\pi \cdot \mathrm{area}(\Omega)}{\mathrm{perim}(\Omega)^2}$$

in either the sphere or the plane, and area and perim are the area and perimeter of Ω, respectively.

The ancient Greeks were first to observe that if Ω is a region in the plane, then $4\pi \cdot \mathrm{area}(\Omega) \leq \mathrm{perim}(\Omega)^2$, with equality if and only if Ω is a circle. This became known as the *isoperimetric inequality* in the plane. The Polsby-Popper score measures the normalized ratio of a district's area to the square of its perimeter and takes the form of an an *isoperimetric quotient*. With respect to this measure, a circle is the most compact shape with a Polsby-Popper score of one, and deviations from this ideal decrease the score towards zero. In other words, in the plane, $0 \leq \mathrm{PP}(\Omega) \leq 1$, and the Polsby-Popper score is equal to 1 only in the case of a circle. We can observe that the Polsby-Popper score is scale-invariant in the plane.

An isoperimetric inequality for the sphere exists, and we state it as the following fact. For a more detailed treatment of isoperimetry in general, see Osserman (1979), and for a proof of this inequality for the sphere, see Rado (1935). For a modern perspective on isoperimetry in the context of redistricting, see DeFord et al. (2019b).

Fact 5.21. *If Ω is a region on the sphere with area A and perimeter P, then $P^2 \geq 4\pi A - A^2$ with equality if and only if Ω is a spherical cap.*

A consequence of this is that among all regions on the sphere with a fixed area A, a spherical cap with area A has the shortest perimeter. However, the key difference between the Polsby-Popper score in the plane and on the sphere is that on the sphere, there is no scale invariance; two spherical caps of different sizes will have different scores.

Lemma 5.22. *Let S be the unit sphere, and let $\kappa(h)$ be a cap of height h. Then $\mathrm{PP}(\kappa(h))$ is a monotonically increasing function of h.*

Proof. Let $r(h)$ be the radius of the circle bounding $\kappa(h)$. We compute:

$$\begin{aligned} 1 &= r(h)^2 + (1-h)^2, \text{ by right triangle trigonometry} \\ &= r(h)^2 + 1 - 2h + h^2 \end{aligned}$$

Rearranging, we get that $r(h)^2 = 2h - h^2$, which we can plug in to the standard formula for perimeter:

$$\mathrm{perim}_S(\kappa(h)) = 2\pi r(h) = 2\pi\sqrt{2h - h^2}$$

We can now use the Archimedian equal-area projection defined by

$$(x, y, z) \to \left(\frac{x}{\sqrt{x^2 + y^2}}, \frac{y}{\sqrt{x^2 + y^2}}, z \right)$$

to compute $\mathrm{area}_S(\kappa(h)) = 2\pi h$ and plug it in to get:

$$\mathrm{PP}_S(\kappa(h)) = \frac{4\pi(2\pi h)}{4\pi^2(2h - h^2)} = \frac{2}{2-h}$$

Which is a monotonically increasing function of h. □

Corollary 5.23. *On the sphere, Polsby-Popper scores of caps are monotonically increasing with area.*

Using this, we can show the main theorem of this section, that no map projection from a region on the sphere to the plane can preserve the ordering of Polsby-Popper scores for all regions.

Theorem 5.24. *If $\varphi : U \to V$ is a map projection from the sphere to the plane, then there exist two regions $A, B \subset U$ such that the Polsby-Popper score of B is greater than that of A in the sphere, but the Polsby-Popper score of $\varphi(A)$ is greater than that of $\varphi(B)$ in the plane.*

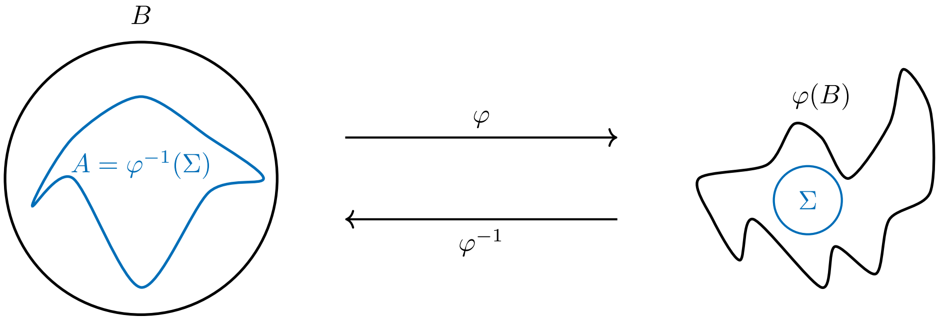


Figure 5.12: The construction of regions A and B in the proof of Theorem 5.24.

Proof. Let φ be a map projection, and let $\kappa \subset U$ be some cap. We will take our regions A and B to lie in κ. Set B to be a cap contained in κ. Let Σ be a circle in the plane such that $\Sigma \subsetneq \varphi(B)$ and let $A = \varphi^{-1}(\Sigma)$. See Figure 5.12 for an illustration.

We now use the isoperimetric inequality for the sphere and Corollary 5.23 to claim that A does not maximize the Polsby-Popper score in the sphere.

To see this, take $\hat{A}$ to be a cap in the sphere with area equal to that of A. Note that since the area of $\hat{A}$ is less than the area of the cap B, it follows that we can choose $\hat{A} \subset B$.

By the isoperimetric inequality of the sphere, $\mathrm{PP}_S(\hat{A}) \geq \mathrm{PP}_S(A)$. Since map projections preserve containment, $\Sigma \subsetneq \varphi(B)$ implies that $A \subsetneq B$, meaning that $\mathrm{area}(\hat{A}) = \mathrm{area}(A) \lneq \mathrm{area}(B)$. By Corollary 5.23, we know that $\mathrm{PP_S}(\hat{A}) < \mathrm{PP_S}(B)$, and combining this with the earlier inequality, we get

$$\mathrm{PP_S}(A) \leq \mathrm{PP_S}(\hat{A}) < \mathrm{PP_S}(B)$$

Since $\Sigma = \varphi(A)$ maximizes the Polsby-Popper score in the plane, but A does not do so in the sphere, we have shown that φ does not preserve the maximal elements in the score ordering, and therefore it cannot preserve the ordering itself. □

The reason why every map projection fails to preserve the ordering of Polsby-Popper scores is because the score itself is constructed from the *planar* notion of isoperimetry, and there is no reason to expect this formula to move nicely back and forth between the sphere and the plane. This proof crucially exploits a scale invariance present in the plane but not the sphere. If we consider any circle in the plane, its Polsby-Popper score is equal to one, but that is not true of every cap in the sphere.

5.6 Empirical Evaluation

In the previous sections we showed that no projection from the sphere to the plane can preserve various compactness orderings. These theorems suggest that in general maps that distort shape cannot preserve compactness orderings. In this section we investigate empirically the consequences of calculating compactness in different map projections, demonstrating the practical relevance of our investigation and providing evidence for possible generalizations.

5.6.1 Commonly-Used Projections

We briefly identify four commonly-used projections in the redistricting domain, which we will use in the next section to compare the empirical effects of the choice of map projection on the compactness orderings.

Plate Carrée The *plate carrée* projection, sometimes called an *equirectangular* projection interprets latitude-longitude coordinates on the Earth as planar x-y coordinates. This map projection does not accurately reflect most geographic figures and is therefore inappropriate for most applications. The U.S. Census Bureau distributes its shapefiles in this format, trusting the user to reproject the data into a format suited for the relevant application. However, because the data is distributed in this format, redistricting analysts and stakeholders (e.g. Chikina et al. (2017); League of Women Voters of Pennsylvania, *et al.* (2018); Chen (2017)) often do not perform this reprojection step, and this has led to the *plate carrée* projection becoming a *de facto* standard in this domain.

Mercator The *Web Mercator* projection is a cylindrical projection which is popular in Web mapping applications. As a result, this is the projection used in several online redistricting software tools available to the public, including DistrictBuilder (Public Mapping Project, 2018), Dave's Redistricting App (Bradlee et al., 2019), and Districtr (Metric Geometry and Gerrymandering Group, 2019a).

Lambert The *Lambert conic* projection is a conformal projection, which means that it preserves the angles of intersection of all segments. This is colloquially interpreted as 'preserving shape at a small scale'. This projection is used in some portions of the U.S. State Plate Coordinate System, and is therefore used in an official capacity for some states.

Albers The *Albers* projection is an *equal area* conic projection, meaning it preserves the areas of all figures. The U.S. Atlas projection for the conterminous 48 states is an Albers projection and is the default in the Maptitude for Redistricting software, which is widely used by redistricting professionals, including legislators, consultants, and advocacy groups.

5.6.2 Results

While we have shown mathematically that the ordering of compactness scores is necessarily permuted by any map projection, we now consider the possibility of this effect occurring in reality. If it is the case that congressional districts all have scores far enough apart that the distortion introduced by the choice of projection is not sufficient to swap the ordering of this score, then the results above are merely mathematical curiosities. Precisely stated, we ask whether reprojection affects compactness score rankings of real districts in the context of commonly-used map projections. In previous both the scientific literature and the legal landscape, this question was either unaddressed, or asserted to be answered in the negative (c.f. Chen (2017); League of Women Voters of Pennsylvania, *et al.* (2018); Chikina et al. (2017)).

In this section, we demonstrate that for the commonly-used map projections listed above and the three compactness scores we examine in the previous sections, that this permutation effect does occur in practice, using the congressional districts from the 115th Congress.[2] We extract the boundaries of the districts from a U.S. Census Bureau shapefile, using the highest resolution available, drawn at a scale of 1:500,000. We then compute the convex hull, Reock, and Polsby-Popper scores with respect to common map projections[3] and examine the ordering of the districts with respect to both. While this is slightly different from the mathematical framework where we compare an abstract map projection to the computation on the surface of the sphere, computing the spherical values of these scores is not a simple task, even in modern geographic information systems (GIS) software.[4] Rather, we can observe that the numerical values of all three scores on all districts are very similar with respect to the Lambert and Albers projections. These projections preserve local shape and area, respectively, and so we can imagine the ground-truth spherical value to also be concordant with

[2] Used for the 2016 congressional elections.

[3] The code to compute the various compactness scores is based on Lee Hachadoorian's *compactr* project (Hachadoorian, 2018).

[4] Provided that the region is contained in an open hemisphere, and that the earth is assumed to be a perfect sphere sitting in $\mathbb{R}^3$, a simple algorithm to calculate a minimum bounding cap is as follows: find the minimum bounding 3-ball of the region and intersect that ball with the Earth. Efficient algorithms exist for computing the minimum bounding ball of a collection of points (Ritter, 1990). However, since computing the minimum bounding sphere of a region is not a typical problem in GIS, the data necessary to run the algorithm is not readily available.

these measures.

With four different map projections and three different compactness scores, we explore several instances in which the choice of map projection distorts the compactness score ranking of districts.

We first consider the 36 congressional districts in Texas. In Section 5.6.2, we plot the Polsby-Popper score ordering of these districts, comparing several pairs of projections. A perfect preservation of the order would result in these points all falling on the diagonal. However, what we see in practice is that most points do lie on the diagonal but several are not, indicating a disagreement between the ordering between the two projections, although the score orders totally agree in the Mercator and Albers projections. The distortion is clearly present, although fairly mild, with the only swaps occurring being between pairs nearby in the orderings.

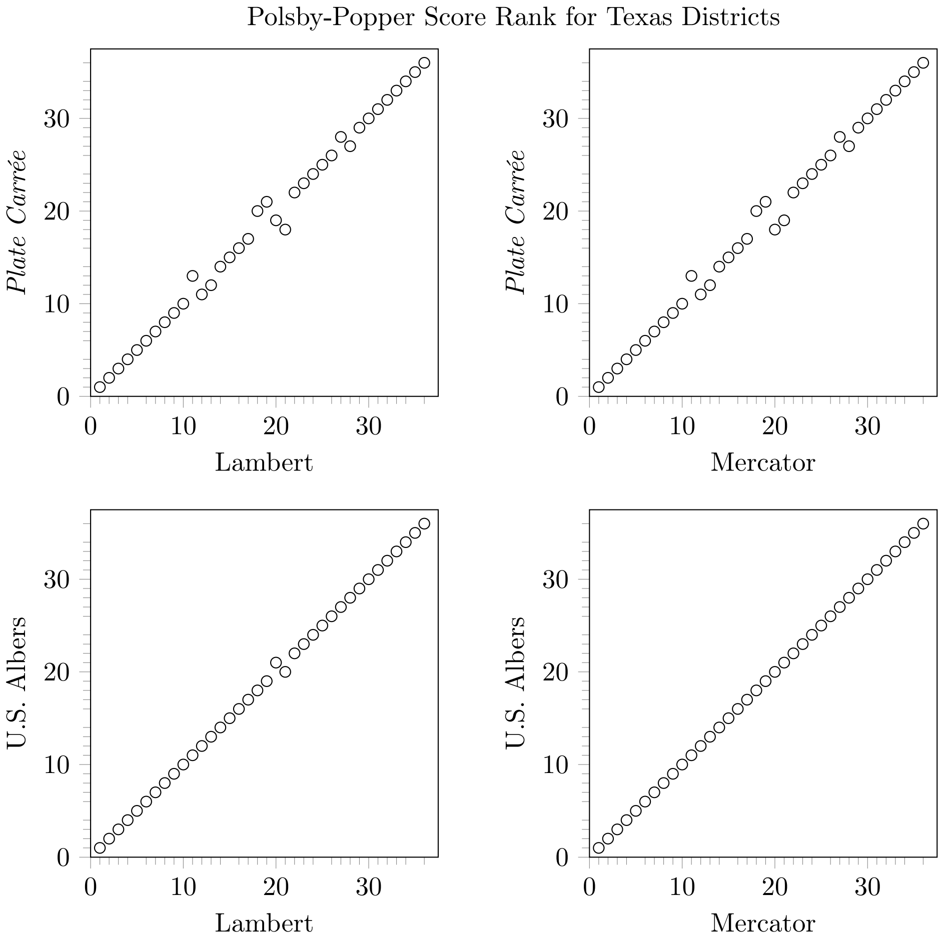


Figure 5.13: Permutation of districts' Polsby-Popper scores under different projections.

We observe a similarly mild, though still present, perturbation in the convex hull score orderings, shown in Figure 5.14. In this setting, however, the score ordering is identical between the Lambert and Albers projections. A similar observation holds at the national level, considering all 433 districts in the coterminous United States. Thus, we empirically observe that the Polsby-Popper and convex hull score orders are fairly robust to the choice of projection, although not entirely.

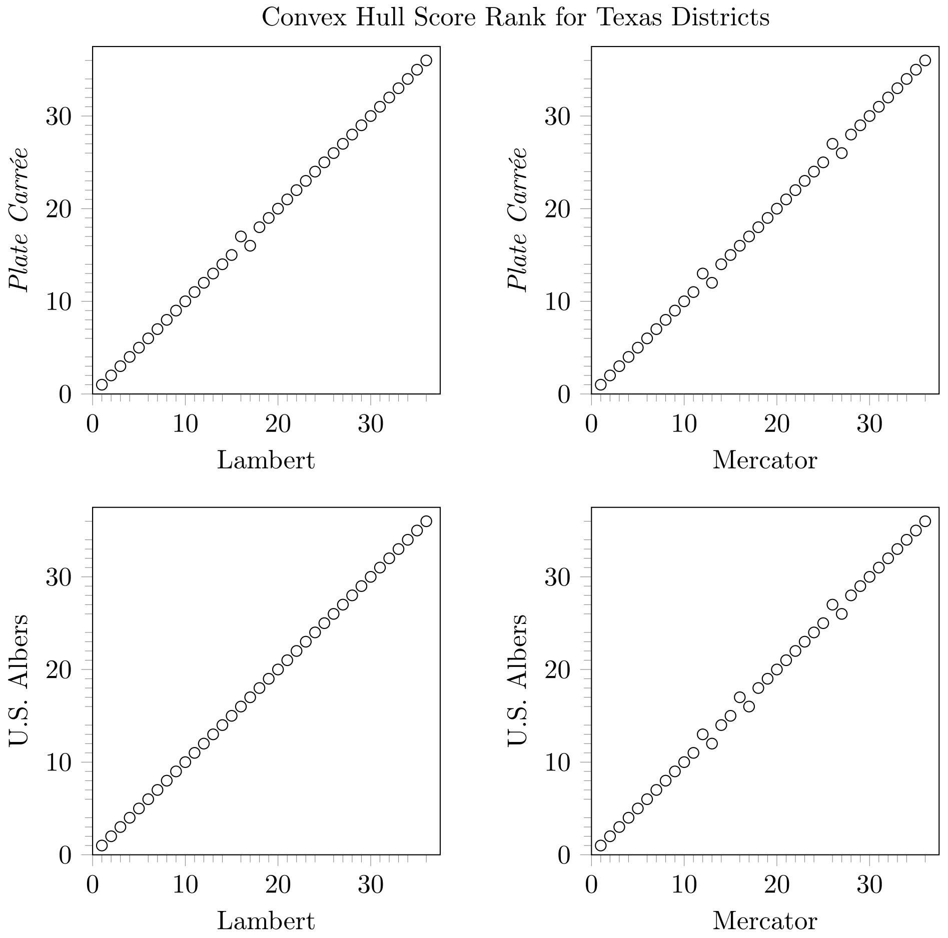


Figure 5.14: Permutation of districts' convex hull scores under different projections.

However, some compactness score orderings are more sensitive than others. In Figure 5.15, we examine the Reock score ordering for the same pairs of projections. While the permutation between the Albers and the Lambert or Mercator projections is still relatively mild, although more complex than for the Polsby-Popper score, the distortion between *Plate Carrée* and these two projections is

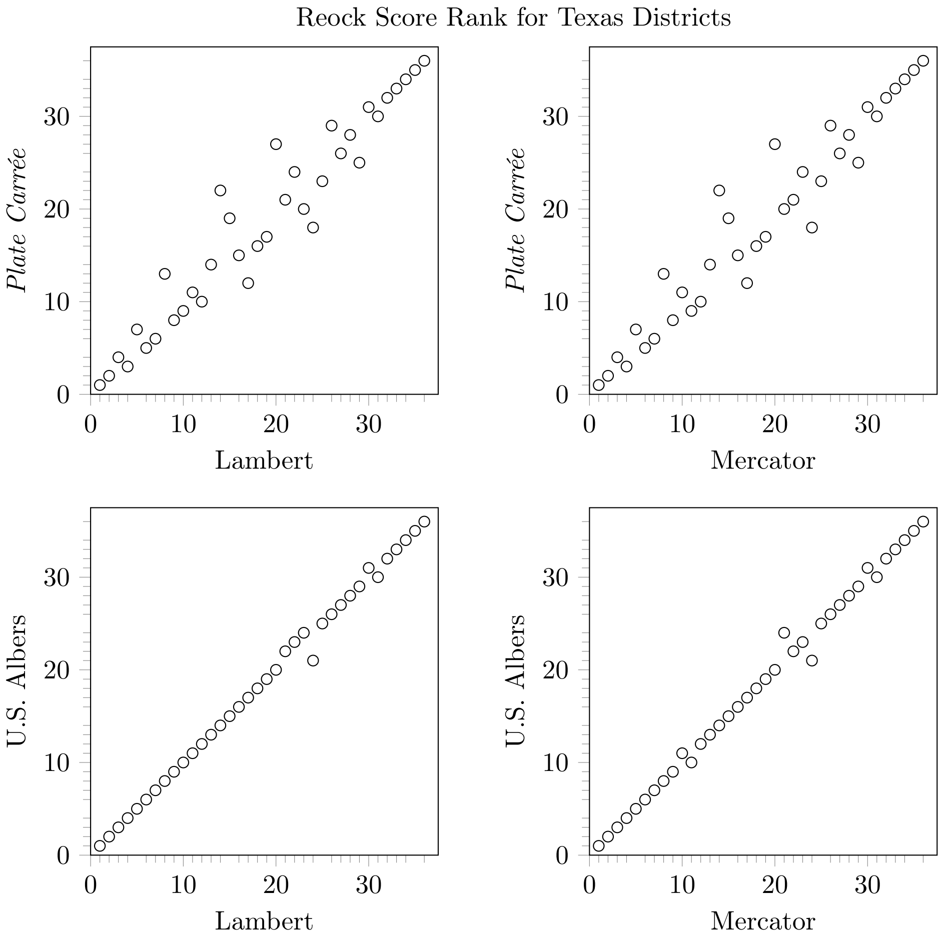


Figure 5.15: Permutation of districts' Reock scores under different projections.

quite dramatic. The districts at the extreme ends of the ordering are relatively undisturbed, but the districts in the middle portion get shuffled around significantly. We observe that this effect is not a result of some idiosyncracy of Texas' districts, since a similar effect persists when we consider all of the districts in the coterminous United States, shown in Figure 5.16.

The Polsby-Popper score is calculated by considering a portion of the map that contains only the district itself. Since some of the projections we consider are locally very similar, and the districts themselves are very small, this gives an explanation for the robustness of the compactness orderings for that score. On the other hand, the more extreme reprojection order reversal we see in Reock scores results from the fact that its computation depends on the potentially large smallest bounding

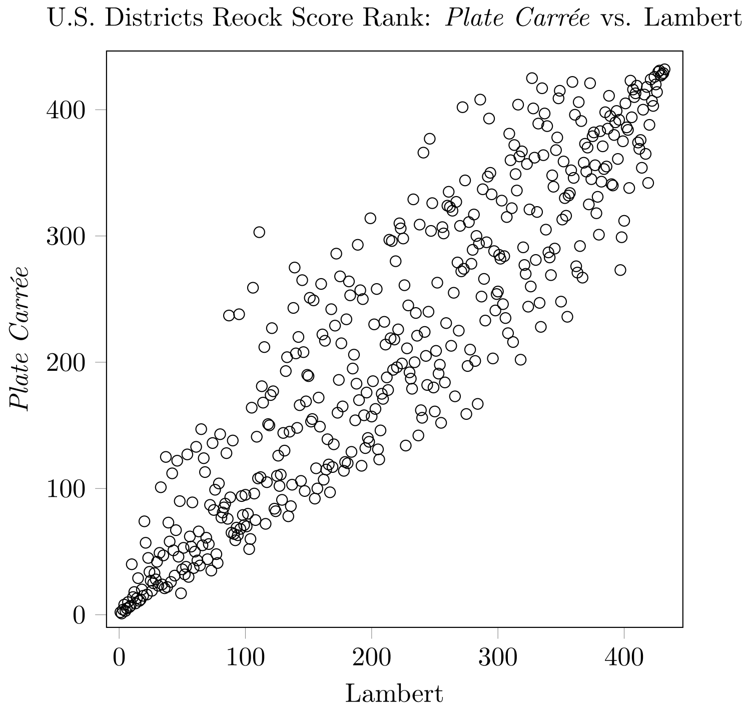


Figure 5.16: Permutation of all U.S. Congressional districts' Reock scores between the *Plate Carrée* and Lambert projections.

circle around the district. This circle will always be larger than the region relevant for the calculation of the convex hull score, since the convex hull of a district is always contained in any bounding circle, and all the map projections we consider distort larger shapes more severely than smaller ones. Thus, we should expect the distortion from reprojections to affect the Reock score more significantly than the convex hull score.

While the results outlined here are by no means comprehensive, they are a representative sample of the prevalence of the order-reversal phenomenon in practice. In all cases, extreme shapes remain extreme under reprojection, but the rankings of the middle-ranked districts are distorted. While the actual numerical discrepancies between the scores computed under the different projections is small, that this permutation can even occur when using 'nice' projections like Albers and Lambert muddies the water in discussing compactness. If value of using mathematics to describe the shape of districts is to provide a small objective frame of reference in a setting where subjective political factors play a large role, then the inconsistency even in the ordering of the districts under the scores works counter to this purpose.

Furthermore, compactness scores are used directly and indirectly in the rapidly growing area of statistical analysis of gerrymandering using *ensembles* of districting plans (Herschlag et al., 2018; Chikina et al., 2017; Chen and Rodden, 2015; Liu et al., 2015), where many possible maps are generated by a computer and used to contextualize properties of a proposed plan. In that context, compactness scores are often aggregated into a score for a districting plan, which is then used to constrain the universe of plans the algorithm generates. For example, we might insist that the average Polsby-Popper score of the generated plans not be larger than our plan of interest or assert a lower threshold for the scores of the districts individually. One underexplored question is the extent to which the dependence on the map projection affects the resulting statistical analysis of these ensembles. We emphasize that it is possible that changes to the compactness scores of the "middle of the pack" districting plans can affect the distribution from which the algorithm draws samples; for instance, under the cut-off approach the universe of allowable plans itself could change significantly if the choice of map projection shuffles which kinds of shapes have scores above and below the threshold. Najt et al. (2019) refer to this line of questioning as 'the extreme outlier hypothesis' and discuss the implications of it in greater detail.

5.7 Weaknesses of compactness scores

We have identified a major *mathematical* weakness in the commonly discussed compactness scores in that no map projection can preserve the ordering over regions induced by these scores. This leads to several important considerations in the mathematical and popular examinations of the detection of gerrymandering.

From the mathematical perspective, rigorous definitions of compactness require more nuance than the simple score functions which assign a single real-number value to each district. *Multiscale* methods, such as those proposed by DeFord et al. (2019b), assign a vector of numbers or a function to a region, rather than a single number. The richer information contained in such constructions is less susceptible to perturbations of map projections. Alternatively, we can look to capturing the geometric information of a district without having to work with respect to a particular spherical or planar representation. So-called *discrete compactness* methods, such as those proposed in Duchin and Tenner (2018), extract a graph structure from the geography and are therefore unaffected by the choice of map projection, and our results suggest that this is an important advantage of these kinds of scores over traditional ones. Finally, recent work has used lab experiments to discern what qualities of a region humans use to determine whether they believe a region is compact or not (Kaufman et al., 2019). Incorporating more qualitative techniques is important, especially in this setting where the social impacts of a particular districting plan may be hard to quantify. To further complicate matters, as highlighted by Barnes and Solomon (Forthcoming), the *resolution* of the shapefile influence the computation of compactness scores, particularly the Polsby-Popper score where the detail of features like coastlines can have a massive impact on the measured perimeter of a region. For this reason, repeating the experiment in Section 5.6.2 for different choices of resolution results in quantitiatively different (although qualitatively similar) results.

We proved our non-preservation results for three particular compactness scores which appear frequently in the context of electoral redistricting. There are countless other scores offered in legal codes and academic writing, such as definitions analogous to the Reock and convex hull scores which use different kinds of bounding regions, scores which measure the ratio of the area of the largest inscribed shape of some kind to the area of the district, and versions of these scores which replace the notion of 'area' with the population of that landmass. Many of these and others suffer from similar flaws as the three scores we examined in this work.

While compactness scores are not used critically in a *legal* context, they appear frequently in the popular discourse about redistricting issues and frame the perception of the 'fairness' of a plan. An Internet search for a term like 'most gerrymandered districts' will invariably return results naming-and-shaming the districts with the most convoluted shapes rather than highlighting where more pleasant looking shapes resulted in unfair electoral outcomes.

Similarly, a sizable amount of work towards remedying such abuses focuses primarily on the geometry rather than the politics of the problem. Popular press pieces (e.g. Ingraham (2014)) and academic research alike (e.g. Cohen-Addad et al. (2017); Svec et al. (2007); Levin and Friedler (2019)) describe algorithmic approaches to redistricting which use geometric methods to generate districts with appealing shapes. However, these approaches ignore all of the social and political information which are critical to the analysis of whether a districting plan treats some group of people unfairly in some way. A purely geometric approach to drawing districts implicitly supposes that the mathematics used to evaluate the geometric features of districts are unbiased and unmanipulable and therefore can provide true insight into the fairness of electoral districts. We proved here that the use of geographic compactness as a proxy for fairness is much less clear and rigid than some might expect.

In the next part of the chapter, we dive into this issue of fairness more deeply and show that attempting to remedy a perceived unfairness withy respect to a particular definition and framework might exacerbate the perceived unfairness with respect to a different one.

Future Work

This work opens several promising avenues for further investigation. We prove strong results for the most common compactness scores, but the question remains what the most general mathematical results in this domain might be, such as giving a set of necessary and sufficient conditions for a map projection to preserve the compactness ordering with respect to a particular score, and which kinds of surfaces do and do not admit such an order-preserving diffeomorphism or describing the permutation of scores as a function of the change in curvature between the two spaces of interest. Our work demonstrates a potential issue arising from the lack of standardization in the use of map projections in redistricting applications, for instance in the statistical analysis of gerrymandering, as discussed at the end of Section 5.6. Gaining a better understanding of these effects is crucial as these statistical methods gain both academic and legal traction. From a cartographic standpoint, understanding other redistricting-related topics beyond compactness scores where the choice of map projection might have a significant effect on the outcome is important, particularly as access to mapmaking tools and data become more widely available to the general public.

5.8 Introduction to Fair Redistricting

Gerrymandering, the careful crafting of electoral districts to favor or disfavor a particular outcome, is a hot topic in contemporary political discourse. In advance of the 2020 U.S. Census and subsequent redistricting processes, several high-profile court cases, reform initiatives, and new lines of academic research have ignited discussions about what kinds of processes and outcomes lead to the 'fairest' districts. However, fairness in this setting is loosely defined. Since the early days of the republic, politicians have used the power of the pen to draw districts which help their political allies and harm their rivals. The term *gerrymander* itself comes from a portmanteau used in an 1812 political cartoon lampooning Massachusetts governor Elbridge *Gerry* and a sala*mander*-shaped state senate district which was part of a plan advantaging the governor's Democratic-Republican party. Since then, districts and districting plans have been identified as unfair for various reasons, but a singular framework for determining when a districting plan is *fair* remains elusive.

Since the early 1960s, advocates for fair districts and districting procedures have proposed using algorithmic techniques to remove the human element, and therefore potential for human bias, from the system. In a letter, economist William Vickrey proposed an algorithmic framework with a large amount of randomness to even further separate human decisions from the eventual output Vickrey (1961). Over the last sixty years, the growth in computational power and availability of data brings us to a point where Vickrey's dream of an autonomous redistricting machine could be realized Altman and McDonald (2010). However, the use of an algorithm does not imply that the internal process of drawing the lines or the districting plan it outputs is unbiased or fair. Given the renewed interest in the redistricting problem, the emergence of computational districting methods in legal settings, and the availability of the necessary data resources to properly implement an algorithmic redistricter, it is important to understand how differing views of fairness may or may not be compatible with each other in such a system.

Our Work We begin by highlighting some recent algorithmic approaches to drawing districts. We then discuss two conceptualizations of fairness in this domain: drawing districts by a *neutral process* and drawing districts to achieve a particular *outcome* which aligns with certain values. We consider these two approaches in an empirical domain using computer-generated districting plans for North Carolina and Pennsylvania and construct Pareto frontiers to examine the trade-off between optimizing for the *compactness* of the districts and the *partisan symmetry* of the contests in those districts. Finally, we discuss some future directions for inquiry and research in the domain of automated and algorithmic redistricting.

5.8.1 Automated Redistricting

Several works have proposed purely algorithmic approaches to constructing electoral districts, and the prototypical formulation is to minimize a functional evaluating a geometric property of the districting plan, subject to a few standard constraints including population balance and connectedness. Vickrey's proposal as well as the algorithms of Levin and Friedler (2019), Chen and Rodden (2013, 2015), Hess et al. (1965), and Cohen-Addad et al. (2018) involve selecting a random location as a 'seed' for each district and then assigning territory to each of those seeds based on proximity in a particular way, such as with Voronoi diagrams or an iterative flood fill procedure. The shortest splitline algorithm Kalcsics et al. (2005); Smith (2011) and the diminishing halves algorithm Spann et al. (2007) choose to iteratively cut the state along the shortest line meeting a particular criterion. Another family of computational redistricting proposals include the Markov chain Monte Carlo approach as in Bangia et al. (2017); Chikina et al. (2017); DeFord and Duchin (2019); DeFord et al.

(2019a); Carter et al. (2019), simulated annealing Browdy (1990), and genetic algorithms Liu et al. (2015), which are all algorithms which involve making random perturbations to a districting plan or a collection of districting plans to improve its score according to some measure. These can be used to search for a maximally compact plan as the other algorithmic approaches do, but can also be instantiated with other objective functions, and so have a more versatile functionality but are less clear in how they arrive at a particular 'final' plan.

Another family of algorithmic redistricting methods is the game-theoretic ones, which involve framing the process of drawing districts as a game between two opposing players (usually viewed as representatives of the two major political parties) and then designing rules for this game such that when the players act in an optimal way, induces a 'fair' outcome. For example, the framework of Mixon and Villar (2018) promises that if the two parties have similar numbers of voters, they will win in a similar number of districts. The game of Pegden et al. (2017) similarly makes guarantees about the ability of one player to arbitrarily limit the number of seats their opponent can win.

5.9 Procedural Neutrality

The first conceptualization of fairness we discuss is that of a *neutral process*; districts ought to be drawn without considering of any of the potentially sensitive attributes of the underlying population such as racial or partisan information. The most common proposal under this framework is to draw districts which maximize a particular notion of *compactness* subject to the basic constraints of connectedness and equal population; indeed this is the underlying objective of several of the algorithms outlined in the previous section. From the legal side, many jurisdictions specify that districts should split political subunits, such as counties or municipalities, as little as possible. The degree to which the preservation of political subunits binds the process in practice varies widely. It is treated very seriously, for example, in Iowa and West Virginia where the congressional districts enacted after the 2010 Census do not split any counties. In other states, it is treated more as a guiding principle.

Taking neutrality as a definition of fairness has several advantages. First, many of the *redistricting principles* Altman (1998); National Conference of State Legislatures (2019), including contiguity, compactness, avoidance of partisan data, and preserving political subunits, fall under the framework of neutrality. Additionally, these neutral criteria are typically easy to operationalize and quantify. For example, we can count the number of municipalities two districting plans split and objectively observe that one splits fewer than the other. Such an analysis is not as straightforward for the outcome-centered criteria described in the next section. Stern (1974) argues that by rigorously adhering to a standard of compactness, districts may contain fragments of many different communities, encouraging the formation of coalitions, which then has a positive impact on the democratic process. For these reasons, a clear set of neutral criteria for drawing districts has been a common approach for redistricting reform since the 1960s Dixon (1981).

On the other hand, districts being composed of fragmented communities can impede the ability of minority groups to achieve representation in the legislative body. The 1982 amendments to the Voting Rights Act and subsequent court opinions specifically prohibit this kind of fragmentation, colloquially referred to as the 'cracking' of voters. Through the history of the United States, political mechanisms including the drawing of district lines have been used to intentionally limit the political power and access of minority groups. Arguments along these lines admit that adhering solely to neutral criteria can perpetuate these kinds of inequities, and this undermines the use of neutrality as the standard of fairness in this context.

5.10 Fairness of Outcome

At the other end of the spectrum is that we should only consider the outcomes of the elections in the districts, irrespective of the procedure used to actually generate those districts. Arguments with respect to this viewpoint underpinned several high-profile court cases in recent years, including cases where Democrats earned over half of the statewide vote but a minority of seats. Roughly half of the vote translated into winning 36 of 99 seats in Wisconsin's General Assembly and three of thirteen of North Carolina's congressional districts, for example. In Maryland, the Democratic legislature redrew the state's congressional districts to tilt the partisan balance in one district so as to force a long-time Republican incumbent to narrowly lose to a Democratic challenger.

If we demand that districts lead to a fair outcome, the question which must be addressed is how to define a fair outcome? Even restricting to a solely partisan perspective, where we simply ask for the outcomes to be fair with respect to the voters' partisan identities, this question is very hard to answer. The idea of *proportionality*, that each party should win a fraction of the districts (roughly) equal to its statewide vote share, seems appealing for its simplicity, but it is often not possible to achieve. In Massachusetts, for example, Republicans typically win approximately 35 percent of the statewide vote share in Senatorial and Presidential elections, and a demand for proportionality would demand they win approximately three of the nine congressional seats in the state. However, because Republican voters are distributed roughly evenly around the state, it is very difficult to draw even a single district with a reliable Republican majority, let alone three districts Duchin et al. (2018).

A similar issue of geographic concentration appears when designing districts which satisfy a notion of proportionality with respect to providing minority groups the ability to elect a candidate of choice. This is further complicated by the observation that, while electing a Republican candidate requires a majority (or at least a plurality) of voters in a district to favor the Republican, electing a minority group's candidate of choice does not require drawing a district in which that group constitutes a majority if there are other voters who will reliably support that group's favored candidate.

Using outcomes as the baseline for fairness is sensible for many reasons. First, if the purpose of representative government is to represent the populace, then any evaluation of fairness should be with respect to the groups and viewpoints elected from the districts to the legislature rather than the process by which the districts themselves come about. Providing communities-of-interest and historically marginalized groups access to representation requires drawing the districts in a way that facilitates these desired outcomes because a neutral process risks fragmenting these communities. Additionally, there are other redistricting principles which require considering the outcomes of potential elections, such as avoiding pitting two incumbents against one another. On the other hand, many seemingly desirable ‘fair’ outcomes are mutually exclusive. Even narrowly focusing on partisan measures, one person may believe that districts ought to facilitate as near a proportional outcome as is possible while another may believe that they should be drawn such that the individual district-level elections are as competitive as possible. These are, of course, largely incompatible ideas, since if the elections are competitive, then a small surge in support for one party will tip several of the seats, resulting in a highly disproportionate outcome.

5.11 Empirical Evaluation

To empirically evaluate a quantitative trade-off between adhering to different conceptualizations of fairness, we need to first pin down a metric by which to evaluate a districting plan along each of these dimensions. Here, we select two measures from the literature and use them to evaluate a computer-generated sample of districting plans. We demonstrate that this setting has a clear trade-off between using the two different notions of fairness described previously.

For procedural neutrality, we take the maxim that ‘districts ought to be drawn to be as compact as possible’. We measure the compactness of the districts with the Polsby-Popper score, discussed more precisely earlier in this chapter. This is the most common such measure in the literature and discourse. Recall that the score of a region Ω is computed as

$$\mathrm{PP}(\Omega) = \frac{4\pi \cdot \mathrm{Area}(\Omega)}{\mathrm{Perim}(\Omega)^2},$$

a region attains a Polsby-Popper score of 1 if and only if it is a circle, the measure is scale-invariant, and it degrades toward zero as the boundary becomes more and more contorted.

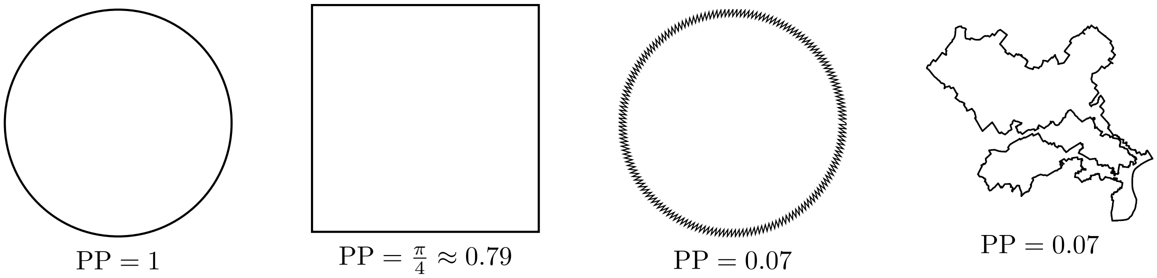


Figure 5.17: Polsby-Popper scores for various figures.
Polsby-Popper scores of four example regions: a perfect circle, a square, a circle with a ragged boundary, and a district from the Pennsylvania plan shown in Figure 5.27.

This score is not without its flaws, in particular it is highly sensitive to minor perturbations of the boundary, which may penalize features like coastlines in an undesireable way. We discussed another major weakness, the potential sensitivity to the choice of map projection, earlier in the chapter. We compute a few basic examples of this score in Figure 5.17.

As our measure of fairness-of-outcome, we use a measure of *partisan symmetry*, evaluating the extent to which Democratic and Republican voters are treated equally under a districting plan. To make this more concrete, we briefly introduce the *seats-votes curve*, which uses the results of an election to extrapolate the necessary statewide vote share for a party to win a particular number of seats. We describe a simple example here, illustrated in Figure 5.18. Suppose in our fictional election, the Republican party earned 45 percent of the statewide vote share across five individual district contests. In these races, they won 20 percent, 25 percent, 55 percent, 60 percent, and 65 percent of the vote, respectively, and therefore winning three of the five seats. The point (45,3) is therefore on the seats-votes curve for this election. We can also see that if the Republican vote share increased or decreased a little bit, the number of resulting seats would not change, so points such as (48,3) and (42,3) are also on our seats-votes curve. However, if the Republicans' statewide vote share dropped by seven percent or increased by 27 percent, the number of seats they win would change, so points like (38,2) and (72,4) are also on the seats-votes curve. Performing this exercise for all potential vote shares yields the final curve. The modelling assumption that the percentage point change in vote share is equal across all districts is called the *uniform partisan swing* assumption and

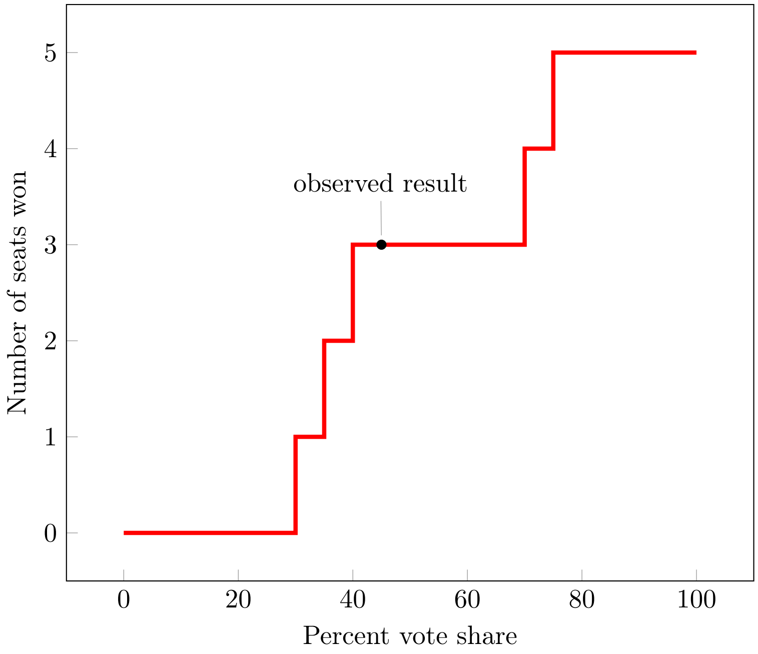


Figure 5.18: An example of a seats-votes curve.

is discussed thoroughly by Katz et al. (Forthcoming).

The seats-votes curve is a simple but powerful picture which captures many standard notions of partisan asymmetry including the *mean-median* score, which measures how far a party's statewide vote share is from its vote share in the median district, and the *efficiency gap* which measures how many votes one party wastes relative to the other. Here, we choose a measure designed to capture asymmetry at all points in the picture: we compute the area between the seats-votes curve as described above and its *inversion* about the midpoint of the figure Nagle (2015). This synthesizes, over all vote shares x, how different the number of seats the Democrats would win with x percent of the vote versus the number of seats the Republicans would win with x percent of the vote. In other words, there is an asymmetry if Republicans win y seats with $50 + x$ percent of the vote but do not lose y seats with $50 - x$ percent of the vote. The area between the seats-votes curve and its inversion about the midpoint is the sum over all possible vote shares of the amount of asymmetry for all of the $50 + x$ and $50 - x$ percent pairs. By dividing this area by the total number of seats and subtracting from one, we obtain a score between zero and one, where a score of one means that

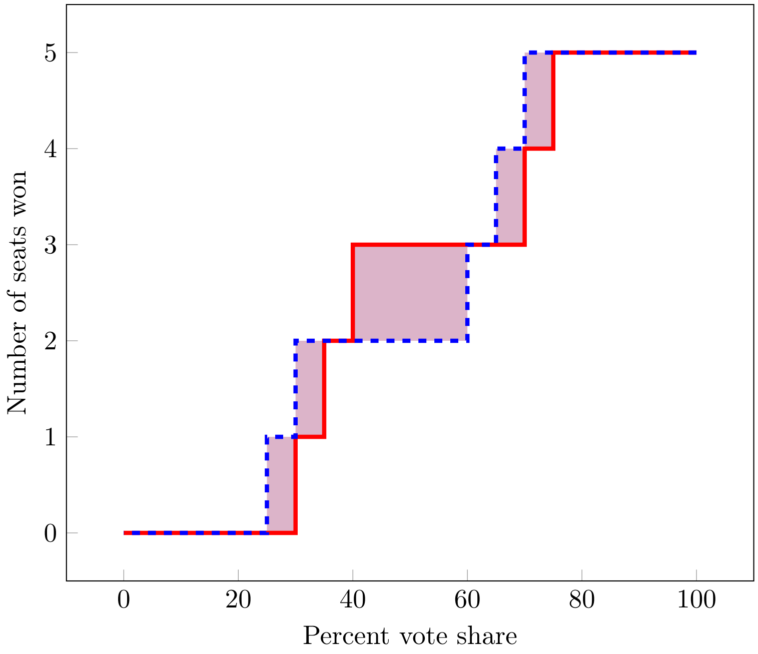


Figure 5.19: The partisan symmetry score from the example seats-votes curve in Figure 5.18, drawn with a solid red line and its inversion about the midpoint, $(50, 2.5)$, drawn with a dashed blue line. The shaded area corresponds to the amount of asymmetry, and this plan achieves a score of 0.92.

the plan is perfectly symmetric with respect to both parties, and the score declines towards zero as one party is better able to translate votes to seats, relative to the other.

5.11.1 Generating Plans

We use the GerryChain Python package Metric Geometry and Gerrymandering Group (2019b) to examine hypothetical districting plans for two states: North Carolina's 13 congressional districts and Pennsylvania's 18 congressional districts. Both of these states are reasonably close to having an equal number of Democrats and Republicans and both have had high-profile court cases challenging their congressional districts in recent years. The data for both states comes from the mggg-states repository on GitHub Metric Geometry and Gerrymandering Group and Buck (2019). We evaluate the partisan symmetry score using a statewide U.S. senatorial race for both states, the 2014 election in North Carolina and the 2016 election in Pennsylvania. In both contests, the Republican candidate narrowly won the election.

We are interested in finding plans at the *Pareto frontier* of compactness and partisan symmetry; districting plans for which there is no other plan which is both more compact and has a higher degree of partisan symmetry. We call a plan on the Pareto frontier *Pareto-optimal* and one that is not we call *Pareto-dominated*. Because the collection of all districting plans which meet the basic criteria of connectedness and population equality is unfathomably large, directly constructing plans of interest is extremely challenging. Instead, GerryChain allows us to use a Markov chain Monte Carlo procedure to generate a large number of plans and extract the Pareto-optimal subset as an approximation to the true Pareto frontier. In brief, our algorithm first generates a random plan then attempts to make small random modifications which improve either its compactness, its partisan symmetry, or both, thereby performing a guided random walk through the space of districting plans. After repeating this for a large number of random seeds and a large number of steps for each walk, we can extract all of the Pareto-optimal plans and use these to draw the empirical Pareto frontier.

The data is in the form of a graph with a vertex for each *voting tabulation district* (VTD), which is the smallest geographic units at which election results are aggregated. Two vertices are joined by an edge if their corresponding VTDs are geographically adjacent. All modifications to plans are made at this level, that is, our problem can be viewed as a *graph partitioning problem* where each VTD must be assigned to exactly one district. For this reason, the universe of possible plans this procedure can generate is more restricted than when working with smaller units such as U.S. Census blocks or drawing free-hand contours through the state. The only constraints we use are connectedness and population equality, which for the sake of tractability is taken to mean that a districting plan is valid if the deviation from the ideal of the population of any district is no more than 2.5 percent.

5.11.2 Results

In Figures 5.20 and 5.21, we plot the compactness and partisan symmetry of our samples of plans, highlighting the empirical Pareto frontier. We can observe several similarities and differences between

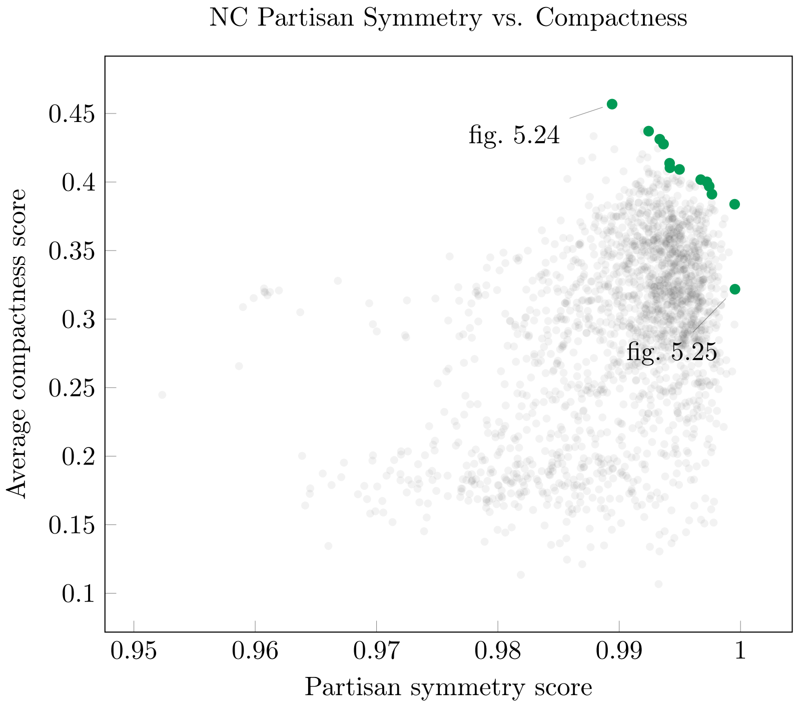


Figure 5.20: Comparison of partisan symmetry and compactness for North Carolina. Each point corresponds to one plan in the sample. Green points are Pareto-optimal, grey points are Pareto-dominated.

these two figures. First, the general shapes of the observed Pareto frontiers are the same. For large values of the partisan symmetry score, one can dramatically increase the achieveable compactness score by relaxing the demand for a high partisan symmetry score a little bit. We do not see a similar effect for large values of the compactness score, where the trade-off between compactness and partisan symmetry appears roughly linear everywhere except at the extreme end of the partisan symmetry score. In both states, we see that it is possible to find plans with nearly perfect partisan symmetry scores. We show the plan with the highest compactness score and highest partisan symmetry score in Figures 5.24 to 5.27 and the full set of Pareto-optimal plans are available online, along with replication code.[5]

While the shapes of the plots are similar, the numerical values of the scores associated to points at the Pareto frontier are very different in the two figures, which we highlight in Figure 5.22. In

[5] `https://zachschutzman.com/tradeoffs-fair-dist`

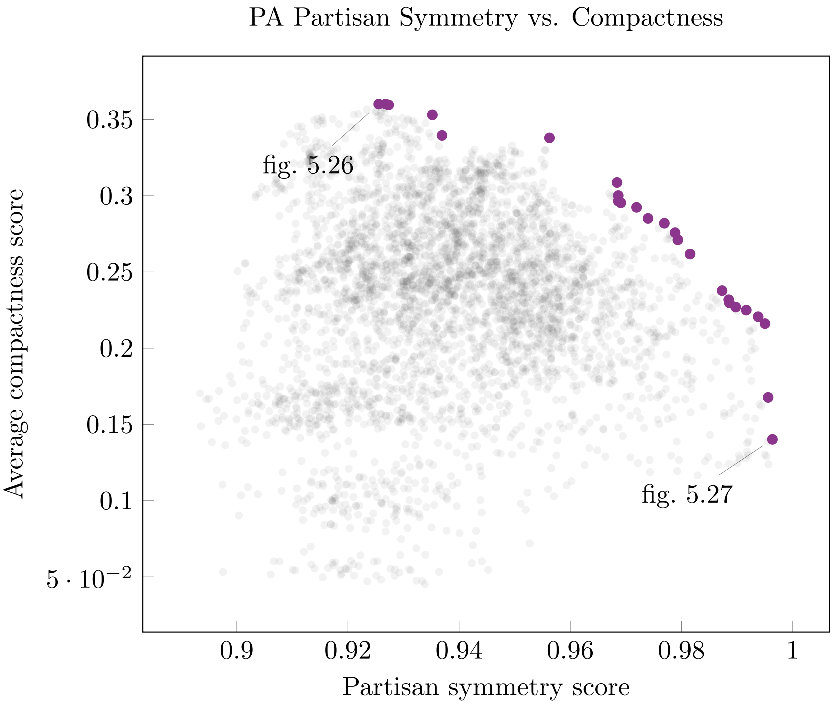


Figure 5.21: Comparison of partisan symmetry and compactness for Pennsylvania. Each point corresponds to one plan in the sample. Purple points are Pareto-optimal, grey points are Pareto-dominated.

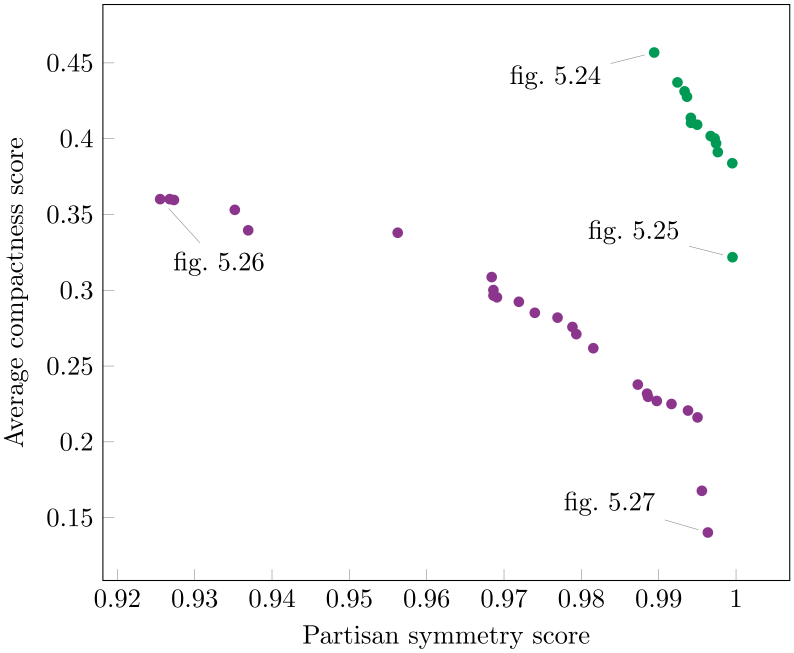


Figure 5.22: The Pareto-optimal points from Figures 5.20 and 5.21.

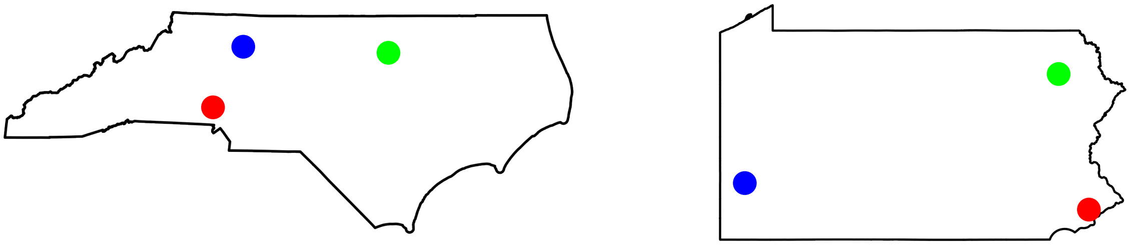


Figure 5.23: The approximate locations of urban areas in North Carolina and Pennsylvania

North Carolina, the most compact plans we found have a partisan symmetry of roughly 0.99, whereas in Pennsylvania, the most compact plans have a partisan symmetry score around 0.92. As a point of reference, the congressional districts enacted in North Carolina in 2016 and those enacted in Pennsylvania in 2011 were found in court to be egregious Republican-favoring gerrymanders and have partisan symmetry scores around 0.9 with respect to the election data used here, so a score of 0.92 suggests that this plan does indeed have a significant partisan tilt.

One explanation for this is the differing *political geography* of the two states Chen and Rodden (2013). Pennsylvania has high concentrations of Democrats in the densely populated corners of the state: the Philadelphia, Pittsburgh, and Scranton–Wilkes-Barre areas. On the other hand, the vast middle of the state has a much more sparse population and is largely Republican-favoring, although the Democratic lean of the cities is somewhat stronger than the Republican tilt of the rest of the state. This means that, even though the balance of Democrats and Republicans is roughly equal, the urban districts will 'use up' more of the Democratic vote than the rural districts do of the Republican vote. Because the Democratic centers are geographically distant from each other, it is difficult to draw districts to balance this effect while remaining highly compact. On the other hand, North Carolina's population is much less concentrated. The largest county in North Carolina has about two-thirds the population of the largest county in Pennsylvania. Furthermore, there are a number of metropolitan regions with high concentrations of Democrats in the middle of the state, including the Raleigh–Durham and the Greensboro areas. The city of Charlotte is also somewhat centrally located in the state. For this reason, districts can remain highly compact and also include portions of these urban regions and rural regions, which helps to balance the asymmetry that arises from large numbers of Democrats living in more dense areas, as in Pennsylvania. The difference in achievable compactness scores may be attributable to the shapes of the precincts themselves, rather than any deeper political reason.

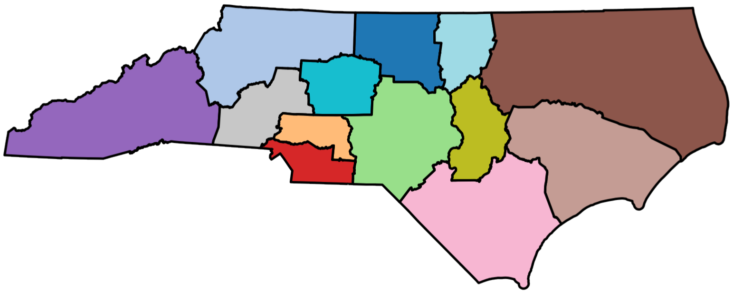


Figure 5.24: The most compact Pareto-optimal plan for North Carolina.

This analysis demonstrates that the 'cost' of partisan symmetry in terms of compactness (and vice versa) is different in the two states. In North Carolina, adhering to a neutral criterion of compactness gives us a high degree of partisan symmetry almost for free. We can see in Figure 5.25 that, with the exception of the two in the eastern portion of the state, the districts are relatively nicely shaped with much of the noncompactness coming from the jagged boundary, in contrast with

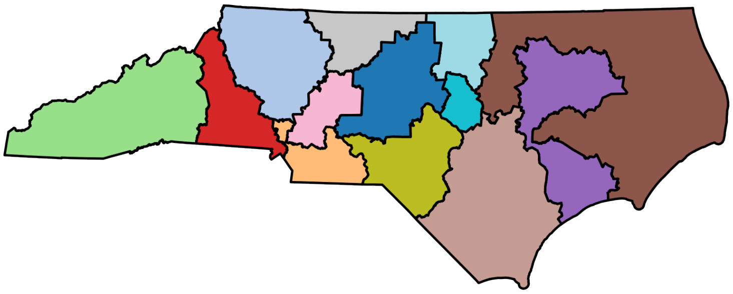


Figure 5.25: The Pareto-optimal plan with the highest degree of partisan symmetry for North Carolina.

the contorted shapes in Figure 5.27. This suggests that the converse is true as well: in North Carolina, aiming for districts which treat the two parties symmetrically doesn't require a severe deviation from nicely shaped districts.

On the other hand, in Pennsylvania, seeking a high degree of partisan symmetry comes at a high cost in terms of compactness. In Figure 5.27, we can see that in order to achieve partisan symmetry, the districts must contort around the Democratic strongholds to properly distribute votes among the less dense, Republican-leaning rural areas. In the southeast, we see five districts extending little tendrils into the Philadelphia area, in the southwest we see the Pittsburgh area divided among four districts. A closer look at the districts in the southeast is shown in Section 5.11.2. The large purple district across the northern part of the state balances a chunk of the Scranton–Wilkes-Barre area with a massive swath of low population rural regions along the New York border. Where there is a strongly Democratic district in the Raleigh–Durham area and a strongly Republican district in the northwestern part of the state and the remaining 11 districts balance mostly rural Republican populations with urban Democratic ones, but splitting up these urban areas does not require drawing the same kinds of contorted shapes as are necessary in Pennsylvania. In contrast, the districts in Figure 5.26 are much more regularly shaped, but achieve a very low degree of partisan symmetry. The four districts nestled in the southeast portion of the state as well as the teardrop shaped one in the southwest encompassing much of Pittsburgh are very strongly Democratic while most of the others have a solid, but relatively weaker, Republican tilt.

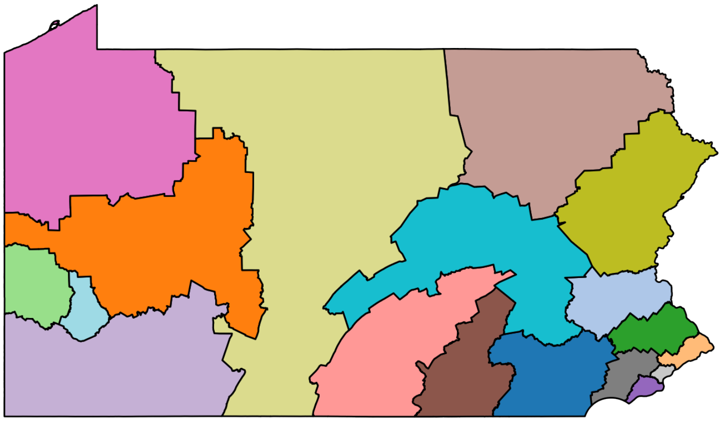


Figure 5.26: The most compact Pareto-optimal plan for Pennsylvania.

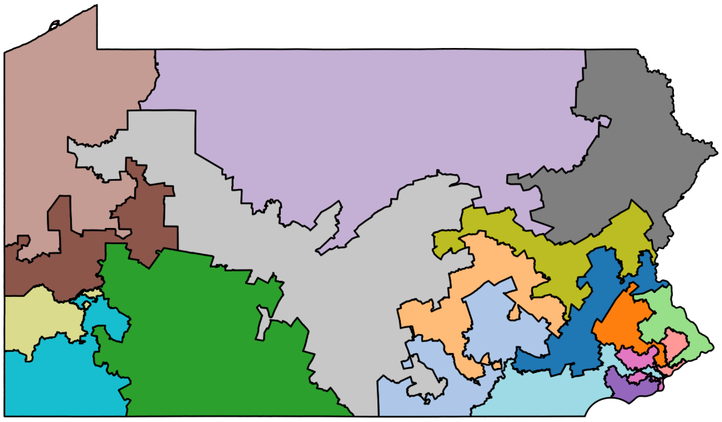


Figure 5.27: The Pareto-optimal plan with the highest degree of partisan symmetry for Pennsylvania.

5.12 Discussion and Future Work

This work points toward several avenues for future research, and we highlight a few here. First, we only considered two instantiations of two particular notions of fairness in this domain. Repeating this analysis for other partisan measures, such as the competitiveness of districts, or other neutral procedures, such as avoiding the splitting of municipalities or counties, would shed more light on what the space of possible districting plans looks like. Additionally, we demonstrate our analysis on Pennsylvania and North Carolina, and the results are considerably different. We posit that this is due to the political geographies of the two states, and examining this effect is an important thread for understanding what kinds of reforms might or might not be effective in various jurisdictions.

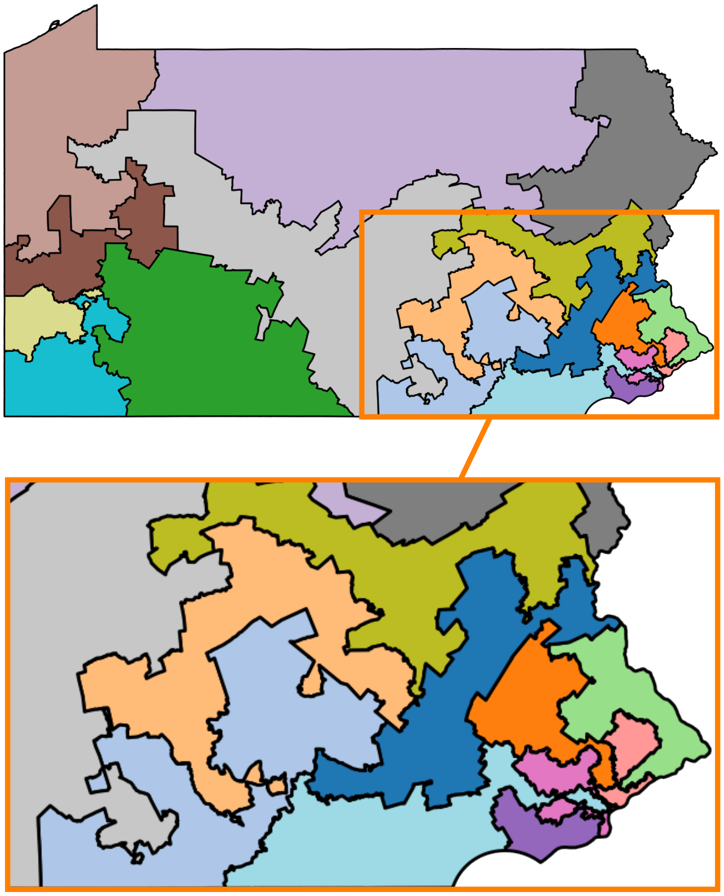


Figure 5.28: A closer look at southeastern Pennsylvania in the plan shown in Figure 5.27.

Future work could use more sophisticated mathematical and statistical techniques to describe a relationship between political geography and the trade-offs we consider here. Our analysis suggests that a one-size-fits-all approach to drawing 'fair' districts is inappropriate and that individual states and localities should carefully consider the relevant trade-offs when redistricting or implementing redistricting reform initiatives. One factor ignored in this analysis, which is critical to the process of drawing districts, is *respecting communities-of-interest*. Even defining and locating geographically such communities is a very difficult problem, let alone the determination of whether or not to preserve that group in a single district. We therefore propose our analysis as a framework for discussion about trade-offs in redistricting rather than as a policy recommendation.

In this work, we have demonstrated with a simple model that demanding districts be drawn to be as compact as possible and demanding that they satisfy a notion of partisan symmetry are incompatible, but to different degrees depending on the particular features of the geographic region in question. Since existing proposals and methodologies for automated and algorithmic redistricting involve finding an approximate solution to an optimization problem, it is important to understand how changing the objective function of these procedures can affect the outcome. As more jurisdictions consider redistricting reforms, they should be cautious about abdicating the line drawing process to algorithms which encode values different from those of the voters who use the districts to elect their representatives.

Bibliography

M. Altman. Traditional districting principles: Judicial myths vs. reality. *Social Science History*, 22 (2):159–200, 1998.

M. Altman and M. McDonald. The promise and perils of computers in redistricting how will computers be used in the next round of redistributing. *Duke Journal of Constitutional Law & Public Policy*, 5:69, 2010.

S. Bangia, C. V. Graves, G. Herschlag, H. S. Kang, J. Luo, J. C. Mattingly, and R. Ravier. Redistricting: Drawing the line. *arXiv:1704.03360*, 2017.

R. Barnes and J. Solomon. Gerrymandering and compactness: Implementation flexibility and abuse. *Political Analysis*, Forthcoming.

D. Bradlee, T. Crowley, M. Matheiu, and A. Ramesey. Dave's redistricting app, 2019. URL `http://gardow.com/davebradlee/redistricting/`.

M. H. Browdy. Simulated annealing: an improved computer model for political redistricting. *Yale Law & Policy Review*, 8(1):163–179, 1990.

O. Byrne. *The first six books of the Elements of Euclid: in which coloured diagrams and symbols are used instead of letters for the greater ease of learners*. William Pickering, 1847.

D. Carter, G. Herschlag, Z. Hunter, and J. Mattingly. A merge-split proposal for reversible monte carlo markov chain sampling of redistricting plans. *arXiv preprint arXiv:1911.01503*, 2019.

J. Chen. Common Cause, et al., vs. Robert A. Rucho, League Of Women Voters of North Carolina vs. Robert A. Rucho, deposition of Jowei Chen. *In The United States District Court For The Middle District Of North Carolina*, page 167, 2017.

J. Chen and J. Rodden. Unintentional gerrymandering: Political geography and electoral bias in legislatures. *Quarterly Journal of Political Science*, 8(3):239–269, 2013.

J. Chen and J. Rodden. Cutting through the thicket: Redistricting simulations and the detection of partisan gerrymanders. *Election Law Journal*, 14(4):331–345, 2015.

M. Chikina, A. Frieze, and W. Pegden. Assessing significance in a Markov chain without mixing. *Proceedings of the National Academy of Sciences*, 114(11):2860–2864, 2017.

V. Cohen-Addad, P. N. Klein, and N. E. Young. Balanced power diagrams for redistricting. *CoRR*, abs/1710.03358, 2017. URL `http://arxiv.org/abs/1710.03358`.

V. Cohen-Addad, P. N. Klein, and N. E. Young. Balanced centroidal power diagrams for redistricting. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 389–396. ACM, 2018.

B. Crowell. Is an equilateral triangle the same as an equiangular triangle, in any geometry? Mathematics Stack Exchange, 2016. URL `https://math.stackexchange.com/q/95080`. URL:https://math.stackexchange.com/q/95080 (version: 2016-11-10).

D. DeFord and M. Duchin. Redistricting reform in Virginia: Districting criteria in context. *Virginia Policy Review*, 2019.

D. DeFord, M. Duchin, and J. Solomon. Recombination: A family of markov chains for redistricting. *arXiv preprint arXiv:1911.05725*, 2019a.

D. DeFord, H. Lavenant, Z. Schutzman, and J. Solomon. Total variation isoperimetric profiles. *SIAM Aplied Algebra and Geometry*, 2019b.

R. Dixon. Fair criteria and procedures for establishing legislative districts. *Policy Studies Journal*, 9(6):839, Spring 1981. URL https://proxy.library.upenn.edu/login?url=https://search.proquest.com/docview/1300129298?accountid=14707. Last updated - 2013-02-23.

M. Duchin and B. E. Tenner. Discrete geometry for electoral geography. *arXiv:1808.05860*, 2018.

M. Duchin, T. Gladkova, E. Henninger-Voss, B. Klingensmith, H. Newman, and H. Wheelen. Locating the representational baseline: Republicans in Massachusetts. *arXiv preprint arXiv:1810.09051*, 2018.

Y. S. Frolov. Measuring the shape of geographical phenomena: A history of the issue. *Soviet Geography*, 16(10):676–687, 1975.

L. Hachadoorian. Compactr. https://github.com/gerrymandr/compactr, 2018.

G. Herschlag, H. S. Kang, J. Luo, C. V. Graves, S. Bangia, R. Ravier, and J. C. Mattingly. Quantifying gerrymandering in north carolina. *arXiv preprint arXiv:1801.03783*, 2018.

S. W. Hess, J. Weaver, H. Siegfeldt, J. Whelan, and P. Zitlau. Nonpartisan political redistricting by computer. *Operations Research*, 13(6):998–1006, 1965.

Idaho Statute 72-1506(4).

C. Ingraham. This computer programmer solved gerrymandering in his spare time, Jun 2014. URL https://www.washingtonpost.com/news/wonk/wp/2014/06/03/this-computer-programmer-solved-gerrymandering-in-his-spare-time/?noredirect=on&utm_term=.47fccb34f63d.

Iowa Code §42.4(4).

J. Kalcsics, S. Nickel, and M. Schröder. Towards a unified territorial design approach—applications, algorithms and gis integration. *Top*, 13(1):1–56, 2005.

J. N. Katz, G. King, and E. Rosenblatt. Theoretical foundations and empirical evaluations of partisan fairness in district-based democracies. *American Political Science Review*, Forthcoming.

A. Kaufman, G. King, and M. Komisarchik. How to measure legislative district compactness if you only know it when you see it. *American Journal of Political Science*, 2019.

League of Women Voters of Pennsylvania, *et al.* Petitioners' brief in support of proposed remedial plans, 2 2018.

H. Levin and S. Friedler. Automated congressional redistricting. *ACM Journal of Experimental Algorithms*, 2019.

Y. Y. Liu, W. K. T. Cho, and S. Wang. A scalable computational approach to political redistricting optimization. In *Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure*, page 6. ACM, 2015.

A. M. Maceachren. Compactness of geographic shape: Comparison and evaluation of measures. *Geografiska Annaler. Series B, Human Geography*, 67(1):53–67, 1985. ISSN 04353684, 14680467. URL `http://www.jstor.org/stable/490799`.

Maine Statute §1206-A.

Metric Geometry and Gerrymandering Group. Districtr, 2019a. URL `https://districtr.org/`.

Metric Geometry and Gerrymandering Group. mggg/gerrychain: v0.2.12, July 2019b. URL `https://github.com/mggg/gerrychain`.

Metric Geometry and Gerrymandering Group and R. Buck. mggg-states, September 2019. URL `https://github.com/mggg-states`.

D. G. Mixon and S. Villar. Utility ghost: Gamified redistricting with partisan symmetry. *arXiv preprint arXiv:1812.07377*, 2018.

J. F. Nagle. Measures of partisan bias for legislating fair elections. *Election Law Journal*, 14(4): 346–360, 2015.

L. Najt, D. DeFord, and J. Solomon. Complexity and geometry of sampling connected graph partitions. *arXiv:1908.08881*, 2019.

National Conference of State Legislatures. Redistricting criteria, Apr 2019. URL `http://www.ncsl.org/research/redistricting/redistricting-criteria.aspx`.

R. Osserman. Bonnesen-style isoperimetric inequalities. *The American Mathematical Monthly*, 86 (1):1–29, 1979.

W. Pegden, A. D. Procaccia, and D. Yu. A partisan districting protocol with provably nonpartisan outcomes. *arXiv preprint arXiv:1710.08781*, 2017.

D. D. Polsby and R. D. Popper. The third criterion: Compactness as a procedural safeguard against partisan gerrymandering. *Yale Law & Policy Review*, 9(2):301–353, 1991.

Public Mapping Project. Publicmapping/districtbuilder, Dec 2018. URL `https://github.com/PublicMapping/DistrictBuilder`.

T. Rado. The isoperimetric inequality on the sphere. *American Journal of Mathematics*, 57(4): 765–770, 1935. ISSN 00029327, 10806377. URL `http://www.jstor.org/stable/2371011`.

J. Ritter. An efficient bounding sphere. *Graphics gems*, 1:301–303, 1990.

W. D. Smith. Rangevoting.org - gerrymandering and a cure - shortest splitline algorithm, 2011. URL `https://rangevoting.org/GerryExamples.html`.

A. Spann, D. Kane, and D. Gulotta. Electoral redistricting with moment of inertia and diminishing halves models. *The UMAP Journal*, 28(3):281–299, 2007.

R. S. Stern. Political gerrymandering: A statutory compactness standard as an antidote for judicial impotence comment. *University of Chicago Law Review*, 41:398, 1974.

L. Svec, S. Burden, and A. Dilley. Applying voronoi diagrams to the redistricting problem. *The UMAP Journal*, 28(3):313–329, 2007.

W. Vickrey. On the prevention of gerrymandering. *Political Science Quarterly*, 76(1):105–110, 1961.

H. P. Young. Measuring the compactness of legislative districts. *Legislative Studies Quarterly*, 13 (1):105–115, 1988. URL `http://www.jstor.org/stable/439947`.

Chapter 6

Conclusion

Endless societal concerns and research questions arise when the algorithmic principles of computational efficiency and optimization come into conflict with human values like fairness, equity, and privacy. We have touched on a few in this dissertation, highlighting a broad range of applications, algorithmic techniques, and social harms.

Each chapter contains specific discussion points and ideas for future research extending that work. More broadly, the problems and solutions described point towards a need for a better understanding of existing systems which are not strictly 'computerized', 'automated', or 'mechanical' as *algorithms*. For example, interpreting an individual allocating resources as a 'black box' algorithm or understanding government data-sharing processes and legal institutions like redlining and segregation as rule-based algorithmic procedures. By taking this viewpoint, we can potentially gain new insights into how they impact our world today, both from a socioeconomic perspective as well as how they color our observations and data collection in developing modern algorithmic solutions to societal problems. Furthermore, by understanding these existing and historical systems as algorithms, we can more directly work on developing new systems and algorithms which aren't subject to the same flaws or inflict the same harms.

The work here makes some steps towards this mode of thinking. By reasoning about allocation problems as learning algorithms, portfolio advising as an optimization problem, privacy-aware individuals as strategic agents, and redistricting as a composite of many algorithmic techniques from computational geometry and graph partitioning, we can begin to understand how modern computational tools can play a role in these crucial social settings and, just as importantly, articulate the limitations of these algorithmic methodologies.