Social Dynamics of Machine Learning for Decision Making

Lydia Liu

Electrical Engineering and Computer Sciences
University of California, Berkeley

Technical Report No. UCB/EECS-2022-41
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2022/EECS-2022-41.html

May 8, 2022

Acknowledgement

My time at Berkeley has been more than I could have imagined, and for that I am deeply grateful to the many people who have not only supported my journey in research, but touched my life.

I am fortunate to have not one but two very supportive PhD advisors, Moritz Hardt and Michael Jordan, who are both relentlessly forward thinkers. [...]

Social Dynamics of Machine Learning for Decision Making

by

Lydia Tingruo Liu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Associate Professor Moritz Hardt, Co-chair
Professor Michael I. Jordan, Co-chair
Professor Chris Shannon

Spring 2022

Social Dynamics of Machine Learning for Decision Making

Abstract

Social Dynamics of Machine Learning for Decision Making

by

Lydia Tingruo Liu

Doctor of Philosophy in Electrical Engineering and Computer Sciences

University of California, Berkeley

Associate Professor Moritz Hardt, Co-chair

Professor Michael I. Jordan, Co-chair

From education to hiring, consequential decisions in society increasingly rely on data-driven algorithms. Yet the long-term impact of algorithmic decision making is largely ill-understood, and there exist serious challenges to ensuring equitable benefits, in theory and practice. In this thesis, I examine the social dynamics of machine learning algorithms from two angles: (i) *long-term fairness of algorithmic decisions*, and (ii) *long-term stability in matching markets*.

In computer science, the subject of algorithmic fairness has received much attention, yet the understanding that algorithms can have disparate impact on populations through various dynamic mechanisms is recent. We contribute to this evolving understanding by presenting two different models of the dynamic interactions of machine learning algorithms and populations of interest. First, we introduce the notion of *delayed impact*—the welfare impact of decision-making algorithms on populations after decision outcomes are observed, motivated, for example, by the change in average credit scores after a new loan approval algorithm is applied. We demonstrate that several statistical criteria for fair machine learning proposed by the research community, if applied as a constraint to decision-making, can result in harm to the welfare of a disadvantaged population. Next, we consider a dynamic setting where individuals invest in a positive outcome based on their expected reward from an algorithmic decision rule. We show that undesirable long-term outcomes arise due to heterogeneity across groups and the lack of realizability, and study the effectiveness of interventions such as 'decoupling' the decision rule by group and providing subsidies.

Adjacent to the question of long-term fairness, another challenge faced in the utilization of machine learning for societal benefit is that of social choice. In markets, individual learning objectives—typically conceived—may conflict with the long-term social objective of reaching an efficient market outcome. Motivated by repeated matching problems in online marketplaces and platforms, we study two-sided matching markets where participants match

repeatedly and gain imperfect information about their preferences through matching. Due to *competition*, one participant's attempt to learn their preferences may affect the utility of other participants. We design a machine learning algorithm for a market platform that enables the market as a whole to learn their preferences efficiently enough to quickly attain a notion of market fairness known as *stability*. Further, we study a decentralized version of the aforementioned problem, and design learning algorithms for participants to strategically avoid competition given past data, thus removing the need for a central platform. We also investigate whether strategic participants with the temptation to act independently should still follow the algorithm's recommendations, showing several positive results on the algorithms' incentive compatibility.

To my parents.

Contents

List of Figures

List of Tables

Acknowledgments

My time at Berkeley has been more than I could have imagined, and for that I am deeply grateful to the many people who have not only supported my journey in research, but touched my life.

I am fortunate to have not one but two very supportive PhD advisors, Moritz Hardt and Michael Jordan, who are both relentlessly forward thinkers. Moritz introduced me to the topic of algorithmic fairness and societal impact, and always encouraged me to work on questions that are interesting to me. I admire Moritz's knack for giving memorable advice—often concise—that crystallizes years of thought. Our discussions have been invaluable whenever I've had a conundrum about academic life. It's been wonderful getting to know Moritz's family over Thanksgiving dinners and barbecues: Petra, Mila, Leonora, and Quentin. I'm grateful for Petra's kind readership and friendship.

Since convincing me to come to Berkeley, Mike has been a constant source of encouragement and optimism. I cannot say how much Mike's exuberance has helped in moments of challenge and doubt. Mike's vision in research has inspired me to formulate new problems at the intersection of research communities. I also admire Mike's dedication towards building an academic community. The weekly SAIL group meeting, rain or shine, was a hallmark of my graduate school experience. My thanks to Mike and Barbara for opening their home to the group and hosting many great pizza parties.

I am grateful to Chris Shannon, for her open and enthusiastic engagement with my work across disciplinary borders, as the outside member of my thesis committee, and her invaluable feedback on my quals.

I have been lucky to have tremendous mentors in teaching and research; to them I owe my heartfelt thanks: Jennifer Chayes and Christian Borgs hosted me during my internship at Microsoft Research New England, and—serendipitously—then joined Berkeley, where we continued to collaborate and exchange ideas. I was fortunate to work with Josh Blumenstock at the I School early in my PhD, and it was our project that first sparked my interest in interdisciplinary work. Ben Recht and his group have been a welcoming and stimulating presence throughout my time at Berkeley, and I'm grateful to have been able to instruct in his graduate optimization class. Rediet Abebe and Tolani Britton have been a privilege to work with on our interdisciplinary project on machine learning in education with Serena Wang, and Rediet also chaired my quals committee.

I have been fortunate to have incredible co-authors who taught me a lot about doing research: Sarah Dean, Esther Rolf, Max Simchowitz, Dan Bjorkegren, Ashia Wilson, Nika Haghtalab, Adam Kalai, Horia Mania, Feng Ruan, Chi Jin, Rong Ge, and Nikhil Garg. Special thanks to: Chi who mentored me through one of my first research projects at Berkeley; Sarah, Esther, Max for being a fantastic team; Horia for being my serial collaborator and friend; Ashia for being a kindred spirit and giving great advice.

I am grateful to mentors from my undergraduate days who inspired me to go to graduate school and continued to be a source of friendship and support thereafter, especially: Amit Singer, Peter Ramadge, Ramon van Handel, and Barbara Engelhardt at Princeton, and

Katja Hofmann at Microsoft Research. Ramon gave me what is still one of the best pieces of advice—"Follow your curiosity". It was a pleasure to work with Edgar Dobriban and Jane Zhizhen Zhao on projects started at Princeton.

Berkeley has been a most vibrant intellectual community, and I've learned much from interesting conversations in the office and beyond. I am happy to count among my research group peers: John Miller, Tijana Zrnic, Smitha Milli, Yu Sun, Melih Eliboh, and Karl Krauth. Many thanks to Ludwig Schmidt, Francis Ding, Celestine Mendler-Dünner, Yixin Wang, Nicolas Flammarion, Lihua Lei, Clara Wong-Fannjiang, Mariel Werner, Manolis Zampetakis, Eric Mazumdar, Lillian Ratliff, Wenshuo Guo, and others for exploring various topics and problems together. It would be remiss not to mention my appreciation of Juanky Perdomo, Stephen Bates, Anastasios Angelopoulos, Neha Wadia and Paula Gradu for enriching our office days with social planning, and Akosua Busia, Nilesh Tripuraneni, Becca Roelofs, Aditi Raghunathan, Angela Zhou, Deb Raji for hanging out outside of work.

My graduate school experience has been enriched by various academic visits and exchanges. I am grateful to my hosts: Ofer Dekel at Microsoft Research Redmond, Krishna Gummadi at MPI SWS, Bernhard Schölkopf at MPI for Intelligent Systems, Jon Kleinberg and Karen Levy at Cornell. I thank Niki Kilbertus for inviting me to co-organize a NeurIPS workshop, and my illustrious co-organizers: Niki, Angela Zhou, Ashia Wilson, John Miller, Lily Hu, Nathan Kallus, and Shira Mitchell.

I would not have completed my degree half as smoothly if not for the amazing administrative staff at Berkeley EECS. A million thanks to Shirley Salanio, Naomi Yamasaki, Susan Kauer, Kattt Atchley, Boban Zarkovich, Angie Abbatecola, and Ami Katagiri.

I am grateful to Veronique Munoz-Darde and Josh Cohen for organizing the Fall 2021 Kadish Workshop in Law, Philosophy, and Political Theory, and for always being happy to discuss Rawls and beyond.

Poetry has been a vital counterpoint to my academic life. I thank my Berkeley teachers, importantly Robert Hass and C.S. Giscombe. I'm also grateful to: Jess Laser and Leo Dunsker for organizing the last five years of CPPWG, Mary Mussman and Erika Higbee for introducing me at the Holloway Reading Series, Hannah Kirwan for being a wonderful reader and friend, and the "Virtual Valley Poets Monthly Meet-ups on last Sundays" (Karen, Yeva, Amy, Angela, Rachel, Therí...) for being a true artistic home.

My friends during my time at Berkeley have been the great joys of my life. I remember Matt Brennan (1994-2021) for the brightness of his light. I thank Hallie, Rachel Lawrence, Rachel Chen, Deb, Evelyn, Clara, Yifan, Irene, Dheeraj, Chitra, Justin, Shao Wei, Deepa, and Gloria for cherished memories, and all there is to come. My thanks to Serena for being my ride or die (on wheels and off!); Anwar for being my wisest older sister.

Finally, I would not be anywhere if not for the love and support of my parents. Thank you Mom and Dad for giving me the world.

Chapter 1

Introduction

From admissions to hiring, consequential decisions increasingly rely on data-driven algorithms. Yet the widespread use of machine learning (ML) techniques in decision-making remains controversial, due to striking examples of racial and gender bias, a lack of transparency in data collection and algorithm design, as well as harmful outcomes for vulnerable populations. The long-term impact of algorithmic decision making is largely ill-understood, and there exist serious challenges to ensuring equitable benefits, in theory and in practice.

When an algorithm makes decisions in applications such as hiring, lending and admissions, these decisions often impact multiple stakeholders and may have longer-term consequences in society at large. Existing computational and statistical tools for building ML algorithms do not usually account for broader social dynamics.

In this thesis, we examine algorithms within their societal context, focusing on the long-term distributive impact of ML algorithms on populations. We develop dynamic models of how algorithmic systems distribute value and opportunity among diverse stakeholders, and apply ensuing insights to design interventions that bring machine learning technology into alignment with societal values—fairness and long-term welfare.

The following chapters address two settings, broadly:

1. *Long-term fairness of algorithmic decisions:* In Chapters 2 and 3, we study the dynamic interactions of machine learning algorithms and populations, for the purpose of mitigating disparate impact in lending and hiring. By presenting two different mathematical models where the dynamic interactions of machine learning algorithms and populations render standard algorithmic fairness interventions ineffective, even harmful, we demonstrate that a view toward long-term outcomes in the discussion of "fair" machine learning is necessary.

2. *Long-term stability in matching markets:* In Chapters 4 and 5, we articulate new design challenges for data-driven decision systems in matching markets that arise from the interplay between uncertainty and competition. Motivated by repeated matching problems in online marketplaces and platforms, we examine dynamic two-sided matching markets where participants learn about their preferences over time, and propose

new learning algorithms for market participants to ensure long-term preference learning and stability, as well as incentive compatibility. A centralized platform is considered in Chapter 4, whereas Chapter 5 studies a decentralized matching market.

The material presented in this thesis is based on previously published work co-authored with Sarah Dean, Esther Rolf, Max Simchowitz, Moritz Hardt [Liu et al., 2018], Ashia Wilson, Nika Haghtalab, Adam Kalai, Christian Borgs, and Jennifer Chayes [Liu et al., 2020b], Horia Mania, Feng Ruan, and Michael I. Jordan [Liu et al., 2020a, 2021a].

1.1 Long-term fairness of algorithmic decisions

Existing scholarship on fairness in automated decision-making criticizes unconstrained machine learning for its potential to harm historically underrepresented or disadvantaged groups in the population [Executive Office of the President, 2016, Barocas and Selbst, 2016]. Consequently, a variety of algorithmic *fairness criteria* have been proposed as constraints on standard machine learning objectives [Calders et al., 2009, Dwork et al., 2012, Zafar et al., 2017, Hardt et al., 2016c]. Even though, in each case, these constraints are clearly intended to protect the disadvantaged group by an appeal to intuition, the concern has been raised that narrow technical definitions of fairness will continue to reproduce discriminatory outcomes [Benjamin, 2019]. Meanwhile, recent work has begun to demonstrate that algorithms can have disparate impact on populations through various dynamic mechanisms [e.g. Ensign et al., 2018, Fuster et al., 2022]

In Chapter 2, we introduce the notion of *delayed impact*—the welfare impact of decision-making algorithms on populations *after* decision outcomes are observed—motivated, for example, by the change in average credit scores after a new loan approval algorithm is applied. We demonstrate that two statistical criteria previously proposed for fair machine learning—*demographic parity* and *equality of opportunity*—if applied as a constraint to decision-making, will not in general promote the welfare improvement of a disadvantaged group and could even result in harm, under reasonable circumstances.

In Chapter 3, we consider a different dynamic setting motivated by applications such as hiring and school admissions, where individuals invest in a positive outcome based on their expected reward from an algorithmic decision rule, thus generalizing Coate and Loury [1993]'s model of labor markets to more flexible feature distributions. We show that undesirable equilibria, in terms of a low long-term rate of skill investment, arise due to heterogeneity across groups and the lack of realizability. We also study the effectiveness of fairness-promoting interventions such as "decoupling" the decision rule by group [Dwork et al., 2018] and providing subsidies, finding that decoupling can be beneficial in the short term, but harmful under longer-term dynamics.

1.2 Long-term stability in matching markets

How should rational learning agents act when they have to *compete* in the same uncertain social environment, such as a market? While there has been a long line of work on learning in games [Fudenberg and Levine, 1998, Hu et al., 1998, Littman, 1994], recent developments in statistical learning theory and online learning have opened the door to a new line of work that aims to quantify precisely the amount of data players require to achieve good performance in games with stochasticity. The problems studied are motivated by a range of modern applications, from modeling competition among firms [Mansour et al., 2018, Aridor et al., 2019] to implementing protocols for wireless networks [Liu and Zhao, 2010, Cesa-Bianchi et al., 2016, Shahrampour et al., 2017]. A particularly salient application is the online marketplace[1], where two sides of a market need to be matched and market participants have uncertainty about their preferences, leading to a need for exploration and statistical learning.

Even in the more complex setting involving multiple players participating in a two-sided matching market, the multi-armed bandit problem can be extended to model how players simultaneously learn and acquire information about their preferences, while satisfying economic constraints. Such a blend of bandit learning with two-sided matching markets was proposed by Das and Kamenica [2005], who formulated a problem in which the players and the arms form the two sides of the market, and each side has preferences over the other side. In contrast to the classical formulation of matching markets [Gale and Shapley, 1962], the preferences of the players are assumed to be unknown a priori and must be learned from the rewards that are received when arms are pulled successfully.

In Chapter 4, we introduce the first bandit algorithm for two-sided markets with theoretical guarantees, focusing on a *centralized* setting in which the players are able to communicate with a central platform that computes matchings for the entire market. We define a notion of regret called *stable regret*, which is the average reward a player obtains less the rewards achieved under a *stable matching* with respect to the true preferences of the market. We show that an algorithm that combines the upper confidence bound principle from the bandit literature [Lai and Robbins, 1985b] with the Gale-Shapley algorithm from the matching market literature [Gale and Shapley, 1962] can achieve low stable regret. In other words, the algorithm enables the market as a whole to learn their preferences efficiently enough to approximately attain a stable market outcome, at the optimal rate. The learning rate provided quantifies exactly how much the sample complexity of preference learning increases when there is an added social objective of reaching a stable matching.

In Chapter 5, we study a decentralized version of the aforementioned problem, and design learning algorithms for participants to strategically avoid competition given past data, thus removing the need for a central platform. Indeed, most online marketplaces have varying degrees of decentralization, that is, there is no central clearinghouse and players are unable to coordinate their actions with each other directly. However, players may observe limited

[1]Examples include online labor markets (Upwork, TaskRabbit, Handy), online crowdsourcing platforms (Amazon Mechanical Turk), online dating services (Match.com) and peer-to-peer sharing platforms (Airbnb).

information about past matchings, hence motivating the setting we consider. New theoretical challenges also arise in the decentralized setting, in both the design and the analysis of algorithms.

In both chapters, we ask whether strategic participants with the temptation to act independently at any time should still follow the algorithm's recommendations. We show several positive results on the algorithms' long-term incentive compatibility, as well as a negative result for the decentralized setting, where long-term incentive compatibility may be more challenging.

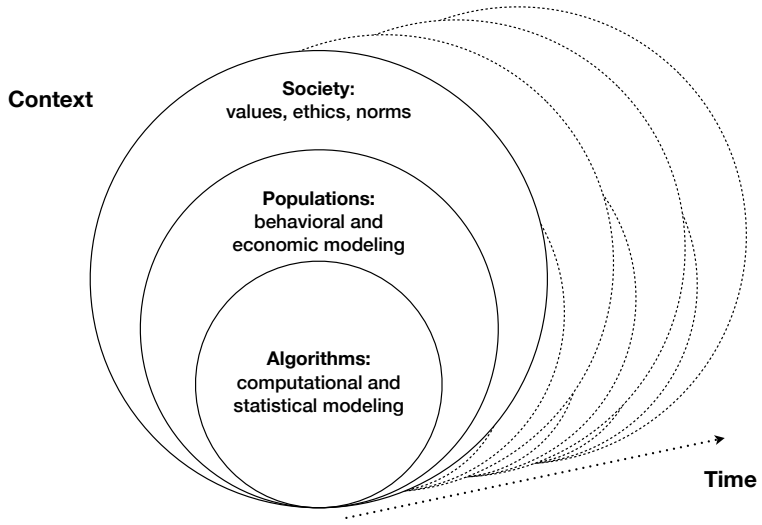


Figure 1.1: Nested model of contexts in algorithmic decision making

In this thesis, we seek to examine algorithms within their broader societal context. We do so by considering their interactions with stakeholders at both the population and the individual level. As seen in Figure 1.1, this goes beyond existing computational and statistical tools and involves behavioral and economic modeling in tandem. Time (e.g., short-term v.s. long-term) is also an important context for algorithmic decision making, as the interactions between algorithms, populations, and society at large are dynamic.

Understanding and improving the impact of machine decision-making ultimately transcends computational research [Abebe et al., 2020]. The final layer of the nested model considers the values, ethics and norms that are promoted by algorithmic systems. Questions such as fairness that involve moral judgment within context cannot be resolved by computational expertise alone, or even primarily. In high-stakes domains such as education [Madaio et al., 2021, Liu et al., 2021b] and credit [Heaven, 2022], where machine learning tools have become prevalent, issues of disparate impact and long-term implications of algorithmic tools are increasingly urgent and pertinent. Further interdisciplinary work is needed to bridge the gap between the prevailing technical understanding of machine learning and the perspective of domain practitioners and policy makers on fairness and social good.

Chapter 2

Delayed Impact of Fair Machine Learning

2.1 Introduction

Machine learning commonly considers static objectives defined on a snapshot ofthe population at one instant in time; consequential decisions, in contrast, reshape the population over time. Lending practices, for example, can shift the distribution of debt and wealth in the population. In this chapter, we formally examine under what circumstances fairness criteria do indeed promote the long-term well-being of disadvantaged groups measured in terms of a temporal variable of interest. Going beyond the standard classification setting, we introduce a one-step feedback model of decision-making that exposes how decisions change the underlying population over time.

Our running example is a hypothetical lending scenario. There are two groups in the population with features described by a summary statistic, such as a *credit score*, whose distribution differs between the two groups. The bank can choose thresholds for each group at which loans are offered. While group-dependent thresholds may face legal challenges [Ross and Yinger, 2006], they are generally inevitable for some of the criteria we examine. The impact of a lending decision has multiple facets. A default event not only diminishes profit for the bank, it also worsens the financial situation of the borrower as reflected in a subsequent decline in credit score. A successful lending outcome leads to profit for the bank and also to an increase in credit score for the borrower.

When thinking of one of the two groups as disadvantaged, it makes sense to ask what lending policies (choices of thresholds) lead to an expected improvement in the score distribution within that group. An unconstrained bank would maximize profit, choosing thresholds that meet a break-even point above which it is profitable to give out loans. One frequently proposed fairness criterion, sometimes called demographic parity, requires the bank to lend to both groups at an equal rate. Subject to this requirement the bank would continue to maximize profit to the extent possible. Another criterion, originally called equality of oppor-

tunity, equalizes the *true positive rates* between the two groups, thus requiring the bank to lend in both groups at an equal rate among individuals who repay their loan. Other criteria are natural, but for clarity we restrict our attention to these three.

Do these fairness criteria benefit the disadvantaged group? When do they show a clear advantage over unconstrained classification? Under what circumstances does profit maximization work in the interest of the individual? These are important questions that we begin to address in this work.

Contributions

We introduce a one-step feedback model that allows us to quantify the long-term impact of classification on different groups in the population. We represent each of the two groups A and B by a *score* distribution $\boldsymbol{\pi}_{\mathsf{A}}$ and $\boldsymbol{\pi}_{\mathsf{B}}$, respectively. The support of these distributions is a finite set $\mathcal{X}$ corresponding to the possible values that the score can assume. We think of the score as highlighting one variable of interest in a specific domain such that higher score values correspond to a higher probability of a positive outcome. An *institution* chooses selection policies $\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}} \colon \mathcal{X} \to [0, 1]$ that assign to each value in $\mathcal{X}$ a number representing the rate of selection for that value. In our example, these policies specify the lending rate at a given credit score within a given group. The institution will always maximize their utility (defined formally later) subject to either (a) no constraint, or (b) equality of selection rates, or (c) equality of true positive rates.

We assume the availability of a function $\boldsymbol{\Delta} \colon \mathcal{X} \to \mathbb{R}$ such that $\boldsymbol{\Delta}(x)$ provides the expected change in score for a selected individual at score x. The central quantity we study is the expected difference in the mean score in group $j \in \{\mathsf{A}, \mathsf{B}\}$ that results from an institutions policy, $\Delta\boldsymbol{\mu}_{\mathsf{j}}$ defined formally in Equation (2.2). When modeling the problem, the expected mean difference can also absorb external factors such as "reversion to the mean" so long as they are mean-preserving. Qualitatively, we distinguish between *long-term improvement* ($\Delta\boldsymbol{\mu}_{\mathsf{j}} > 0$), *stagnation* ($\Delta\boldsymbol{\mu}_{\mathsf{j}} = 0$), and *decline* ($\Delta\boldsymbol{\mu}_{\mathsf{j}} < 0$). Our findings can be summarized as follows:

1. Both fairness criteria (equal selection rates, equal true positive rates) can lead to all possible outcomes (improvement, stagnation, and decline) in natural parameter regimes. We provide a complete characterization of when each criterion leads to each outcome in Section 2.3.

    - There are a class of settings where equal selection rates cause decline, whereas equal true positive rates do not (Corollary 2.3.5),
    - Under a mild assumption, the institution's optimal unconstrained selection policy can never lead to decline (Proposition 2.3.1).

2. We introduce the notion of an *outcome curve* (Figure 2.1) which succinctly describes the different regimes in which one criterion is preferable over the others.

3. We perform experiments on FICO credit score data from 2003 and show that under various models of bank utility and score change, the outcomes of applying fairness criteria are in line with our theoretical predictions.

4. We discuss how certain types of measurement error (e.g., the bank underestimating the repayment ability of the disadvantaged group) affect our comparison. We find that measurement error narrows the regime in which fairness criteria cause decline, suggesting that measurement should be a factor when motivating these criteria.

5. We consider alternatives to hard fairness constraints.

    - We evaluate the optimization problem where fairness criterion is a regularization term in the objective. Qualitatively, this leads to the same findings.
    - We discuss the possibility of optimizing for group score improvement $\Delta\boldsymbol{\mu}_{\mathrm{j}}$ directly subject to institution utility constraints. The resulting solution provides an interesting possible alternative to existing fairness criteria.

We focus on the impact of a selection policy over a single epoch. The motivation is that the designer of a system usually has an understanding of the time horizon after which the system is evaluated and possibly redesigned. Formally, nothing prevents us from repeatedly applying our model and tracing changes over multiple epochs. In reality, however, it is plausible that over greater time periods, economic background variables might dominate the effect of selection.

Reflecting on our findings, we argue that careful temporal modeling is necessary in order to accurately evaluate the impact of different fairness criteria on the population. Moreover, an understanding of measurement error is important in assessing the advantages of fairness criteria relative to unconstrained selection. Finally, the nuances of our characterization underline how intuition may be a poor guide in judging the long-term impact of fairness constraints.

Related work

Recent work by Hu and Chen [2018b] considers a model for long-term outcomes and fairness in the labor market. They propose imposing the demographic parity constraint in a *temporary* labor market in order to provably achieve an equitable long-term equilibrium in the *permanent* labor market, reminiscent of economic arguments for affirmative action [Foster and Vohra, 1992]. The equilibrium analysis of the labor market dynamics model allows for specific conclusions relating fairness criteria to long term outcomes. Our general framework is complementary to this type of domain specific approach.

[Fuster et al., 2022] consider the problem of fairness in credit markets from a different perspective. Their goal is to study the effect of machine learning on interest rates in different groups at an equilibrium, under a static model without feedback.

Ensign et al. [2018] consider feedback loops in predictive policing, where the police more heavily monitor high crime neighborhoods, thus further increasing the measured number of crimes in those neighborhoods. While the work addresses an important temporal phenomenon using the theory of urns, it is rather different from our one-step feedback model both conceptually and technically.

Demographic parity and its related formulations have been considered in numerous papers [e.g. Calders et al., 2009, Zafar et al., 2017]. Hardt et al. [2016c] introduced the equality of opportunity constraint that we consider and demonstrated limitations of a broad class of criteria. Kleinberg et al. and Chouldechova [2017] point out the tension between "calibration by group" and equal true/false positive rates. These trade-offs carry over to some extent to the case where we only equalize true positive rates [Pleiss et al., 2017].

A growing literature on fairness in the "bandits" setting of learning [see Joseph et al., 2016, *et sequelae*] deals with online decision making that ought not to be confused with our one-step feedback setting. Finally, there has been much work in the social sciences on analyzing the effect of affirmative action [see e.g., Keith et al., 1985, Kalev et al., 2006].

2.2 Problem Setting

We consider two *groups* A and B, which comprise a g_{A} and $g_{\mathsf{B}} = 1 - g_{\mathsf{A}}$ fraction of the total population, and an *institution* which makes a binary decision for each individual in each group, called *selection*. Individuals in each group are assigned *scores* in $\mathcal{X} := [C]$, and the scores for group $\mathsf{j} \in \{\mathsf{A}, \mathsf{B}\}$ are distributed according $\boldsymbol{\pi}_{\mathsf{j}} \in \text{Simplex}^{C-1}$. The institution selects a *policy* $\boldsymbol{\tau} := (\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}}) \in [0,1]^{2C}$, where $\boldsymbol{\tau}_{\mathsf{j}}(x)$ corresponds to the probability the institution selects an individual in group j with score x. One should think of a score as an abstract quantity which summarizes how well an individual is suited to being selected; examples are provided at the end of this section.

We assume that the institution is utility-maximizing, but may impose certain constraints to ensure that the policy $\boldsymbol{\tau}$ is *fair*, in a sense described in Section 2.2. We assume that there exists a function $\boldsymbol{u} : C \to \mathbb{R}$, such that the institution's expected utility for a policy $\boldsymbol{\tau}$ is given by

$$\mathcal{U}(\boldsymbol{\tau}) = \textstyle\sum_{\mathsf{j} \in \{\mathsf{A},\mathsf{B}\}} g_{\mathsf{j}} \sum_{x \in \mathcal{X}} \boldsymbol{\tau}_{\mathsf{j}}(x) \boldsymbol{\pi}_{\mathsf{j}}(x) \boldsymbol{u}(x). \tag{2.1}$$

Novel to this work, we focus on the effect of the selection policy $\boldsymbol{\tau}$ on the groups A and B. We quantify these *outcomes* in terms of an average effect that a policy $\boldsymbol{\tau}_{\mathsf{j}}$ has on group j. Formally, for a function $\boldsymbol{\Delta}(x) : \mathcal{X} \to \mathbb{R}$, we define the average change of the mean score $\boldsymbol{\mu}_{\mathsf{j}}$ for group j

$$\Delta\boldsymbol{\mu}_{\mathsf{j}}(\boldsymbol{\tau}) := \textstyle\sum_{x \in \mathcal{X}} \boldsymbol{\pi}_{\mathsf{j}}(x) \boldsymbol{\tau}_{\mathsf{j}}(x) \boldsymbol{\Delta}(x)\,. \tag{2.2}$$

We remark that many of our results also go through if $\Delta\boldsymbol{\mu}_{\mathsf{j}}(\boldsymbol{\tau})$ simply refers to an abstract change in well-being, not necessarily a change in the mean score. Furthermore, it is possible

to modify the definition of $\Delta\boldsymbol{\mu}_{\mathrm{j}}(\boldsymbol{\tau})$ such that it directly considers outcomes of those who are not selected.[1] Lastly, we assume that the *success* of an individual is independent of their group given the score; that is, the score summarizes all relevant information about the success event, so there exists a function $\boldsymbol{\rho} : \mathcal{X} \to [0, 1]$ such that individuals of score x succeed with probability $\boldsymbol{\rho}(x)$.

We now introduce the specific domain of credit scores as a running example in the rest of the paper, after which we present two more examples showing the general applicability of our formulation to many domains.

Example 2.2.1 (Credit scores)**.** *In the setting of loans, scores $x \in [C]$ represent credit scores, and the bank serves as the institution. The bank chooses to grant or refuse loans to individuals according to a policy $\boldsymbol{\tau}$. Both bank and personal utilities are given as functions of loan repayment, and therefore depend on the success probabilities $\boldsymbol{\rho}(x)$, representing the probability that any individual with credit score x can repay a loan within a fixed time frame. The expected utility to the bank is given by the expected return from a loan, which can be modeled as an affine function of $\boldsymbol{\rho}(x)$: $\boldsymbol{u}(x) = u_+\boldsymbol{\rho}(x) + u_-(1 - \boldsymbol{\rho}(x))$, where u_+ denotes the profit when loans are repaid and u_- the loss when they are defaulted on. Individual outcomes of being granted a loan are based on whether or not an individual repays the loan, and a simple model for $\boldsymbol{\Delta}(x)$ may also be affine in $\boldsymbol{\rho}(x)$: $\boldsymbol{\Delta}(x) = c_+\boldsymbol{\rho}(x) + c_-(1 - \boldsymbol{\rho}(x))$, modified accordingly at boundary states. The constant c_+ denotes the gain in credit score if loans are repaid and c_- is the score penalty in case of default.*

Example 2.2.2 (Advertising)**.** *A second illustrative example is given by the case of advertising agencies making decisions about which groups to target. An individual with product interest score x responds positively to an ad with probability $\boldsymbol{\rho}(x)$. The ad agency experiences utility $\boldsymbol{u}(x)$ related to click-through rates, which increases with $\boldsymbol{\rho}(x)$. Individuals who see the ad but are uninterested may react negatively (becoming less interested in the product), and $\boldsymbol{\Delta}(x)$ encodes the interest change. If the product is a positive good like education or employment opportunities, interest can correspond to well-being. Thus the advertising agency's incentives to only show ads to individuals with extremely high interest may leave behind groups whose interest is lower on average. A related historical example occurred in advertisements for computers in the 1980s, where male consumers were targeted over female consumers, arguably contributing to the current gender gap in computing.*

Example 2.2.3 (College Admissions)**.** *The scenario of college admissions or scholarship allotments can also be considered within our framework. Colleges may select certain applicants for acceptance according to a score x, which could be thought encode a "college preparedness"*

[1] If we consider functions $\boldsymbol{\Delta}_p(x) : \mathcal{X} \to \mathbb{R}$ and $\boldsymbol{\Delta}_n(x) : \mathcal{X} \to \mathbb{R}$ to represent the average effect of selection and non-selection respectively, then $\Delta\boldsymbol{\mu}_{\mathrm{j}}(\boldsymbol{\tau}) := \sum_{x\in\mathcal{X}} \boldsymbol{\pi}_{\mathrm{j}}(x)\,(\boldsymbol{\tau}_{\mathrm{j}}(x)\boldsymbol{\Delta}_p(x) + (1 - \boldsymbol{\tau}_{\mathrm{j}}(x))\boldsymbol{\Delta}_n(x))$. This model corresponds to replacing $\boldsymbol{\Delta}(x)$ in the original outcome definition with $\boldsymbol{\Delta}_p(x) - \boldsymbol{\Delta}_n(x)$, and adding a offset $\sum_{x\in\mathcal{X}} \boldsymbol{\pi}_{\mathrm{j}}(x)\boldsymbol{\Delta}_n(x)$. Under the assumption that $\boldsymbol{\Delta}_p(x) - \boldsymbol{\Delta}_n(x)$ increases in x, this model gives rise to outcomes curves resembling those in Figure 2.1 up to vertical translation. All presented results hold unchanged under the further assumption that $\Delta\boldsymbol{\mu}(\beta^{\texttt{MaxUtil}}) \geq 0$.

measure. The students who are admitted might "succeed" (this could be interpreted as graduating, graduating with honors, finding a job placement, etc.) with some probability $\boldsymbol{\rho}(x)$ depending on their preparedness. The college might experience a utility $\boldsymbol{u}(x)$ corresponding to alumni donations, or positive rating when a student succeeds; they might also show a drop in rating or a loss of invested scholarship money when a student is unsuccessful. The student's success in college will affect their later success, which could be modeled generally by $\boldsymbol{\Delta}(x)$. In this scenario, it is challenging to ensure that a single summary statistic x captures enough information about a student; it may be more appropriate to consider x as a vector as well as more complex forms of $\boldsymbol{\rho}(x)$.

While a variety of applications are modeled faithfully within our framework, there are limitations to the accuracy with which real-life phenomenon can be measured by strictly binary decisions and success probabilities. Such binary rules are necessary for the definition and execution of existing fairness criteria, (see Sec. 2.2) and as we will see, even modeling these facets of decision making as binary allows for complex and interesting behavior.

The Outcome Curve

We now introduce important outcome regimes, stated in terms of the change in average group score. A policy $(\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}})$ is said to cause *active harm* to group j if $\Delta\boldsymbol{\mu}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}) < 0$, *stagnation* if $\Delta\boldsymbol{\mu}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}) = 0$, and *improvement* if $\Delta\boldsymbol{\mu}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}) > 0$. Under our model, `MaxUtil` policies can be chosen in a standard fashion which applies the same threshold $\boldsymbol{\tau}^{\texttt{MaxUtil}}$ for both groups, and is agnostic to the distributions $\boldsymbol{\pi}_{\mathsf{A}}$ and $\boldsymbol{\pi}_{\mathsf{B}}$. Hence, if we define

$$\Delta\boldsymbol{\mu}_{\mathsf{j}}^{\texttt{MaxUtil}} := \Delta\boldsymbol{\mu}_{\mathsf{j}}(\boldsymbol{\tau}^{\texttt{MaxUtil}}) \tag{2.3}$$

we say that a policy causes *relative harm* to group j if $\Delta\boldsymbol{\mu}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}) < \Delta\boldsymbol{\mu}_{\mathsf{j}}^{\texttt{MaxUtil}}$, and *relative improvement* if $\Delta\boldsymbol{\mu}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}) > \Delta\boldsymbol{\mu}_{\mathsf{j}}^{\texttt{MaxUtil}}$. In particular, we focus on these outcomes for a disadvantaged group, and consider whether imposing a fairness constraint improves their outcomes relative to the `MaxUtil` strategy. From this point forward, we take A to be disadvantaged or protected group.

Figure 2.1 displays the important outcome regimes in terms of *selection rates* $\beta_{\mathsf{j}} := \sum_{x\in\mathcal{X}} \boldsymbol{\pi}_{\mathsf{j}}(x)\boldsymbol{\tau}_{\mathsf{j}}(x)$. This succinct characterization is possible when considering decision rules based on (possibly randomized) score thresholding, in which all individuals with scores above a threshold are selected. In Section 2.5, we justify the restriction to such *threshold policies* by showing it preserves optimality. In Section 2.5, we show that the outcome curve is concave, thus implying that it takes the shape depicted in Figure 2.1. To explicitly connect selection rates to decision policies, we define the rate function $r_{\boldsymbol{\pi}}(\boldsymbol{\tau}_{\mathsf{j}})$ which returns the proportion of group j selected by the policy. We show that this function is invertible for a suitable class of threshold policies, and in fact the outcome curve is precisely the graph of the map from selection rate to outcome $\beta \mapsto \Delta\boldsymbol{\mu}_{\mathsf{A}}(r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(\beta))$. Next, we define the values of β that mark boundaries of the outcome regions.

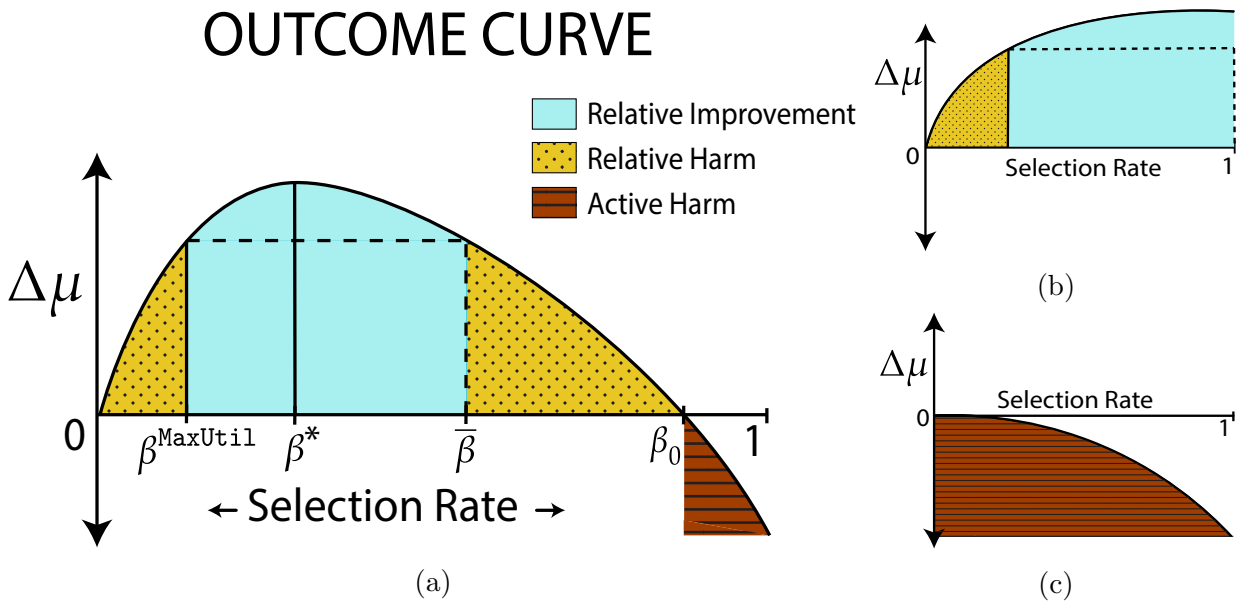


Figure 2.1: The above figure shows the *outcome curve*. The horizontal axis represents the selection rate for the population; the vertical axis represents the mean change in score. (a) depicts the full spectrum of outcome regimes, and colors indicate regions of active harm, relative harm, and no harm. In (b): a group that has much potential for gain, in (c): a group that has no potential for gain.

Definition 2.2.1 (Selection rates of interest)**.** *Given the protected group* A*, the following selection rates are of interest in distinguishing between qualitatively different classes of outcomes (Figure 2.1). We define* $\beta^{\texttt{MaxUtil}}$ *as the selection rate for* A *under* `MaxUtil`*;* β_0 *as the harm threshold, such that* $\Delta\boldsymbol{\mu}_{\mathsf{A}}(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta_0)) = 0$*;* β^* *as the selection rate such that* $\Delta\boldsymbol{\mu}_{\mathsf{A}}$ *is maximized;* $\overline{\beta}$ *as the outcome-complement of the* `MaxUtil` *selection rate,* $\Delta\boldsymbol{\mu}_{\mathsf{A}} r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\overline{\beta})) = \Delta\boldsymbol{\mu}_{\mathsf{A}}(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta^{\texttt{MaxUtil}}))$ *with* $\overline{\beta} > \beta^{\texttt{MaxUtil}}$*.*

Decision Rules and Fairness Criteria

We will consider policies that maximize the institution's total expected utility, potentially subject to a constraint: $\boldsymbol{\tau} \in \mathcal{C} \in [0,1]^{2C}$ which enforces some notion of "fairness". Formally, the institution selects $\tau_* \in \text{argmax } \mathcal{U}(\boldsymbol{\tau})$ s.t. $\boldsymbol{\tau} \in \mathcal{C}$. We consider the three following constraints:

Definition 2.2.2 (Fairness criteria)**.** *The* maximum utility *(*`MaxUtil`*) policy corresponds to the null-constraint* $\mathcal{C} = [0,1]^{2C}$*, so that the institution is free to focus solely on utility. The* demographic parity *(*`DemParity`*) policy results in equal selection rates between both groups. Formally, the constraint is* $\mathcal{C} = \left\{(\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}}) : \sum_{x\in\mathcal{X}} \boldsymbol{\pi}_{\mathsf{A}}(x)\boldsymbol{\tau}_{\mathsf{A}} = \sum_{x\in\mathcal{X}} \boldsymbol{\pi}_{\mathsf{B}}(x)\boldsymbol{\tau}_{\mathsf{B}}\right\}$*. The* equal op-

portunity *(*`EqOpt`*) policy results in equal true positive rates (TPR) between both group, where TPR is defined as* $\mathrm{TPR}_{\mathsf{j}}(\boldsymbol{\tau}) := \frac{\sum_{x \in \mathcal{X}} \boldsymbol{\pi}_{\mathsf{j}}(x)\boldsymbol{\rho}(x)\boldsymbol{\tau}(x)}{\sum_{x \in \mathcal{X}} \boldsymbol{\pi}_{\mathsf{j}}(x)\boldsymbol{\rho}(x)}$*.* `EqOpt` *ensures that the conditional probability of selection given that the individual will be successful is independent of the population, formally enforced by the constraint* $\mathcal{C} = \{(\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}}) : \mathrm{TPR}_{\mathsf{A}}(\boldsymbol{\tau}_{\mathsf{A}}) = \mathrm{TPR}_{\mathsf{B}}(\boldsymbol{\tau}_{\mathsf{B}})\}$.

Just as the expected outcome $\Delta\boldsymbol{\mu}$ can be expressed in terms of selection rate for threshold policies, so can the total utility $\mathcal{U}$. In the unconstrained cause, $\mathcal{U}$ varies independently over the selection rates for group A and B; however, in the presence of fairness constraints the selection rate for one group determines the allowable selection rate for the other. The selection rates must be equal for `DemParity`, but for `EqOpt` we can define a *transfer function*, $G^{(\mathsf{A} \to \mathsf{B})}$, which for every loan rate β in group A gives the loan rate in group B that has the same true positive rate. Therefore, when considering threshold policies, decision rules amount to maximizing functions of single parameters. This idea is expressed in Figure 2.2, and underpins the results to follow.

2.3 Results

In order to clearly characterize the outcome of applying fairness constraints, we make the following assumption.

Assumption 1 (Institution utilities)**.** *The institution's individual utility function is more stringent than the expected score changes,* $\boldsymbol{u}(x) > 0 \implies \boldsymbol{\Delta}(x) > 0$*. (For the linear form presented in Example 2.2.1,* $\frac{u_-}{u_+} < \frac{c_-}{c_+}$ *is necessary and sufficient.)*

This simplifying assumption quantifies the intuitive notion that institutions take a greater risk by accepting than the individual does by applying. For example, in the credit setting, a bank loses the amount loaned in the case of a default, but makes only interest in case of a payback. Using Assumption 1, we can restrict the position of `MaxUtil` on the outcome curve in the following sense.

Proposition 2.3.1 (`MaxUtil` does not cause active harm)**.** *Under Assumption 1,* $0 \leq \Delta\boldsymbol{\mu}^{\texttt{MaxUtil}} \leq \Delta\boldsymbol{\mu}^{*}$*.*

We direct the reader to Section 2.9 for the proof of the above proposition, and all subsequent results presented in this section. The results are corollaries to theorems presented in Section 2.6.

Prospects and Pitfalls of Fairness Criteria

We begin by characterizing general settings under which fairness criteria act to improve outcomes over unconstrained `MaxUtil` strategies. For this result, we will assume that group A is disadvantaged in the sense that the `MaxUtil` acceptance rate for B is large compared to relevant acceptance rates for A.

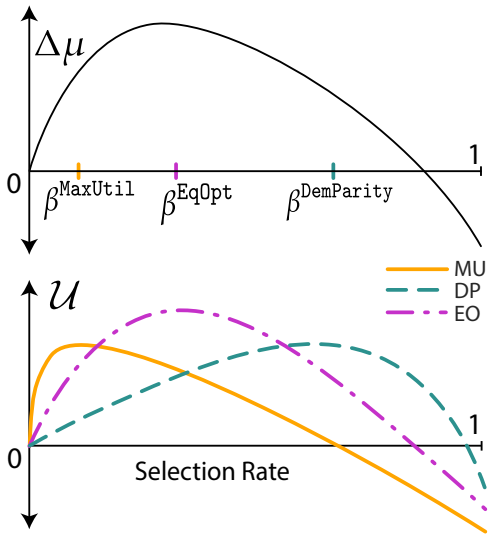


Figure 2.2: Both outcomes $\Delta\boldsymbol{\mu}$ and institution utilities $\mathcal{U}$ can be plotted as a function of selection rate for one group. The maxima of the utility curves determine the selection rates resulting from various decision rules.

Corollary 2.3.2 (Fairness Criteria can cause Relative Improvement)**.** *(a) Under the assumption that* $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \overline{\beta}$ *and* $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > \beta_{\mathsf{A}}^{\texttt{MaxUtil}}$*, there exist population proportions* $g_0 < g_1 < 1$ *such that, for all* $g_{\mathsf{A}} \in [g_0, g_1]$*,* $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta_{\mathsf{A}}^{\texttt{DemParity}} < \overline{\beta}$*. That is,* `DemParity` *causes relative improvement.*

(b) Under the assumption that there exist $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta < \beta' < \overline{\beta}$ *such that* $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > G^{(\mathsf{A}\to\mathsf{B})}(\beta), G^{(\mathsf{A}\to\mathsf{B})}(\beta')$*, there exist population proportions* $g_2 < g_3 < 1$ *such that, for all* $g_{\mathsf{A}} \in [g_2, g_3]$*,* $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta_{\mathsf{A}}^{\texttt{EqOpt}} < \overline{\beta}$*. That is,* `EqOpt` *causes relative improvement.*

This result gives the conditions under which we can guarantee the existence of settings in which fairness criteria cause improvement relative to `MaxUtil`. Relying on machinery proved in Section 2.6, the result follows from comparing the position of optima on the utility curve to the outcome curve. Figure 2.2 displays a illustrative example of both the outcome curve and the institutions' utility $\mathcal{U}$ as a function of the selection rates in group A. In the utility function (2.1), the contributions of each group are weighted by their population proportions g_{j}, and thus the resulting selection rates are sensitive to these proportions.

As we see in the remainder of this section, fairness criteria can achieve nearly any position along the outcome curve under the right conditions. This fact comes from the potential mismatch between the outcomes, controlled by $\boldsymbol{\Delta}$, and the institution's utility $\boldsymbol{u}$.

The next theorem implies that `DemParity` can be bad for long term well-being of the protected group by being over-generous, under the mild assumption that $\Delta\boldsymbol{\mu}_{\mathsf{A}}(\beta_{\mathsf{B}}^{\texttt{MaxUtil}}) < 0$:

Corollary 2.3.3 (DemParity can cause harm by being over-eager)**.** *Fix a selection rate β. Assume that $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > \beta > \beta_{\mathsf{A}}^{\texttt{MaxUtil}}$. Then, there exists a population proportion g_0 such that, for all $g_{\mathsf{A}} \in [0, g_0]$, $\beta_{\mathsf{A}}^{\texttt{DemParity}} > \beta$. In particular, when $\beta = \beta_0$,* DemParity *causes active harm, and when $\beta = \overline{\beta}$,* DemParity *causes relative harm.*

The assumption $\Delta\boldsymbol{\mu}_{\mathsf{A}}(\beta_{\mathsf{B}}^{\texttt{MaxUtil}}) < 0$ implies that a policy which selects individuals from group A at the selection rate that MaxUtil would have used for group B necessarily lowers average score in A. This is one natural notion of protected group A's 'disadvantage' relative to group B. In this case, DemParity penalizes the scores of group A even more than a naive MaxUtil policy, as long as group proportion g_{A} is small enough. Again, small g_{A} is another notion of group disadvantage.

Using credit scores as an example, Corollary 2.3.3 tells us that an overly aggressive fairness criterion will give too many loans to people in a protected group who cannot pay them back, hurting the group's credit scores on average. In the following theorem, we show that an analogous result holds for EqOpt.

Corollary 2.3.4 (EqOpt can cause harm by being over-eager)**.** *Suppose that $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > G^{(\mathsf{A}\to\mathsf{B})}(\beta)$ and $\beta > \beta_{\mathsf{A}}^{\texttt{MaxUtil}}$. Then, there exists a population proportion g_0 such that, for all $g_{\mathsf{A}} \in [0, g_0]$, $\beta_{\mathsf{A}}^{\texttt{EqOpt}} > \beta$. In particular, when $\beta = \beta_0$,* EqOpt *causes active harm, and when $\beta = \overline{\beta}$,* EqOpt *causes relative harm.*

We remark that in Corollary 2.3.4, we rely on the *transfer function*, $G^{(\mathsf{A}\to\mathsf{B})}$, which for every loan rate β in group A gives the loan rate in group B that has the same true positive rate. Notice that if $G^{(\mathsf{A}\to\mathsf{B})}$ were the identity function, Corollary 2.3.3 and Corollary 2.3.4 would be exactly the same. Indeed, our framework (detailed in Section 2.6 and Section 2.9) unifies the analyses for a large class of fairness constraints that includes DemParity and EqOpt as specific cases, and allows us to derive results about impact on $\Delta\boldsymbol{\mu}$ using general techniques. In the next section, we present further results that compare the fairness criteria, demonstrating the usefulness of our technical framework.

Comparing EqOpt and DemParity

Our analysis of the acceptance rates of EqOpt and DemParity in Section 2.6 suggests that it is difficult to compare DemParity and EqOpt without knowing the full distributions $\boldsymbol{\pi}_{\mathsf{A}}, \boldsymbol{\pi}_{\mathsf{B}}$, which is necessary to compute the transfer function $G^{(\mathsf{A}\to\mathsf{B})}$. In fact, we have found that settings exist both in which DemParity causes harm while EqOpt causes improvement and in which DemParity causes improvement while EqOpt causes harm. There cannot be one general rule as to which fairness criteria provides better outcomes in all settings. We now present simple sufficient conditions on the geometry of the distributions for which EqOpt is always better than DemParity in terms of $\Delta\boldsymbol{\mu}_{\mathsf{A}}$.

Corollary 2.3.5 (EqOpt may avoid active harm where DemParity fails)**.** *Fix a selection rate β. Suppose $\boldsymbol{\pi}_{\mathsf{A}}, \boldsymbol{\pi}_{\mathsf{B}}$ are identical up to a translation with $\boldsymbol{\mu}_{\mathsf{A}} < \boldsymbol{\mu}_{\mathsf{B}}$, i.e. $\boldsymbol{\pi}_{\mathsf{A}}(x) = \boldsymbol{\pi}_{\mathsf{B}}(x +$*

$(\boldsymbol{\mu}_{\mathsf{B}} - \boldsymbol{\mu}_{\mathsf{A}})$). *For simplicity, take* $\boldsymbol{\rho}(x)$ *to be linear in* x. *Suppose*

$$\beta > \sum_{x > \boldsymbol{\mu}_{\mathsf{A}}} \boldsymbol{\pi}_{\mathsf{A}}.$$

Then there exists an interval $[g_1, g_2] \subseteq [0, 1]$, *such that* $\forall g_{\mathsf{A}} > g_1$, $\beta^{\texttt{EqOpt}} < \beta$ *while* $\forall g_{\mathsf{A}} < g_2$, $\beta^{\texttt{DemParity}} > \beta$. *In particular, when* $\beta = \beta_0$, *this implies* `DemParity` *causes active harm but* `EqOpt` *causes improvement for* $g_{\mathsf{A}} \in [g_1, g_2]$, *but for any* g_{A} *such that* `DemParity` *causes improvement,* `EqOpt` *also causes improvement.*

To interpret the conditions under which Corollary 2.3.5 holds, consider when we might have $\beta_0 > \sum_{x > \boldsymbol{\mu}_{\mathsf{A}}} \boldsymbol{\pi}_{\mathsf{A}}$. This is precisely when $\Delta\boldsymbol{\mu}_{\mathsf{A}}(\sum_{x > \boldsymbol{\mu}_{\mathsf{A}}} \boldsymbol{\pi}_{\mathsf{A}}) > 0$, that is, $\Delta\boldsymbol{\mu}_{\mathsf{A}} > 0$ for a policy that selects every individual whose score is above the group A mean, which is reasonable in reality. Indeed, the converse would imply that group A has such low scores that even selecting all above average individuals in A would hurt the average score. In such a case, Corollary 2.3.5 suggests that `EqOpt` is better than `DemParity` at avoiding active harm, because it is more conservative. A natural question then is: can `EqOpt` cause relative harm by being *too* conservative?

Corollary 2.3.6 (`DemParity` never loans less than `MaxUtil`, but `EqOpt` might)**.** *Recall the definition of the TPR functions* $\mathrm{TPR}_{\mathsf{j}}$, *and suppose that the* `MaxUtil` *policy* $\boldsymbol{\tau}^{\texttt{MaxUtil}}$ *is such that*

$$\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta_{\mathsf{B}}^{\texttt{MaxUtil}} \text{ and } \mathrm{TPR}_{\mathsf{A}}(\boldsymbol{\tau}^{\texttt{MaxUtil}}) > \mathrm{TPR}_{\mathsf{B}}(\boldsymbol{\tau}^{\texttt{MaxUtil}}) \tag{2.4}$$

Then $\beta_{\mathsf{A}}^{\texttt{EqOpt}} < \beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta_{\mathsf{A}}^{\texttt{DemParity}}$. *That is,* `EqOpt` *causes relative harm by selecting at a rate lower than* `MaxUtil`.

The above theorem shows that `DemParity` is never stingier than `MaxUtil` to the protected group A, as long as a A is disadvantaged in the sense that `MaxUtil` selects a larger proportion of B than A. On the other hand, `EqOpt` can select less of group A than `MaxUtil`, and by definition, cause relative harm. This is a surprising result about `EqOpt`, and this phenomenon arises from high levels of in-group inequality for group A. Moreover, we show in Section 2.9 that there are parameter settings where the conditions in Corollary 2.3.6 are satisfied even under a stringent notion of disadvantage we call CDF domination, described therein.

2.4 Relaxations of Constrained Fairness

Fairness Under Measurement Error

Next, consider the implications of an institution with imperfect knowledge of scores. Under a simple model in which the estimate of an individual's score $X \sim \boldsymbol{\pi}$ is prone to errors $e(X)$ such that $X + e(X) := \widehat{X} \sim \widehat{\boldsymbol{\pi}}$. Constraining the error to be negative results in the setting

that scores are systematically *underestimated*. In this setting, it is equivalent to consider the CDF of underestimated distribution $\widehat{\boldsymbol{\pi}}$ to be *dominated* by the CDF true distribution $\boldsymbol{\pi}$, that is $\sum_{x \geq c} \widehat{\boldsymbol{\pi}}(x) \leq \sum_{x \geq c} \boldsymbol{\pi}(x)$ for all $c \in [C]$. Then we can compare the institution's behavior under this estimation to its behavior under the truth.

Proposition 2.4.1 (Underestimation causes underselection)**.** *Fix the distribution of* B *as* $\boldsymbol{\pi}_{\mathsf{B}}$ *and let* β *be the acceptance rate of* A *when the institution makes the decision using perfect knowledge of the distribution* $\boldsymbol{\pi}_{\mathsf{A}}$*. Denote* $\widehat{\beta}$ *as the acceptance rate when the group is instead taken as* $\widehat{\boldsymbol{\pi}}_{\mathsf{A}}$*. Then* $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} > \widehat{\beta}_{\mathsf{A}}^{\texttt{MaxUtil}}$ *and* $\beta_{\mathsf{A}}^{\texttt{DemParity}} > \widehat{\beta}_{\mathsf{A}}^{\texttt{DemParity}}$*. If the errors are further such that the true TPR dominates the estimated TPR, it is also true that* $\beta_{\mathsf{A}}^{\texttt{EqOpt}} > \widehat{\beta}_{\mathsf{A}}^{\texttt{EqOpt}}$*.*

Because fairness criteria encourage a higher selection rate for disadvantaged groups (Corollary 2.3.2), systematic underestimation widens the regime of their applicability. Furthermore, since the estimated `MaxUtil` policy underloans, the region for relative improvement in the outcome curve (Figure 2.1) is larger, corresponding to more regimes under which fairness criteria can yield favorable outcomes. Thus the potential for measurement error should be a factor when motivating these criteria.

Outcome-based alternative

As explained in the preceding sections, fairness criteria may actively harm disadvantaged groups. It is thus natural to consider a modified decision rule which involves the explicit maximization of $\Delta\boldsymbol{\mu}_{\mathsf{A}}$. In this case, imagine that the institution's primary goal is to aid the disadvantaged group, subject to a limited profit loss compared to the maximum possible expected profit $\mathcal{U}^{\texttt{MaxUtil}}$. The corresponding problem is as follows.

$$\max_{\boldsymbol{\tau}_{\mathsf{A}}} \Delta\boldsymbol{\mu}_{\mathsf{A}}(\boldsymbol{\tau}_{\mathsf{A}}) \;\; \text{s.t.} \;\; \mathcal{U}_{\mathsf{A}}^{\texttt{MaxUtil}} - \mathcal{U}(\boldsymbol{\tau}) < \delta \,. \tag{2.5}$$

Unlike the fairness constrained objective, this objective no longer depends on group B and instead depends on our model of the mean score change in group A, $\Delta\boldsymbol{\mu}_{\mathsf{A}}$.

Proposition 2.4.2 (Outcome-based solution)**.** *In the above setting, the optimal bank policy* $\boldsymbol{\tau}_{\mathsf{A}}$ *is a threshold policy with selection rate* $\beta = \min\{\beta^*, \beta^{\max}\}$*, where* β^* *is the outcome-optimal loan rate and* $\beta^{\max}$ *is the maximum loan rate under the bank's "budget".*

The above formulation's advantage over fairness constraints is that it directly optimizes the outcome of A and can be approximately implemented given reasonable ability to predict outcomes. Importantly, this objective shifts the focus to outcome modeling, highlighting the importance of domain specific knowledge. Future work can consider strategies that are robust to outcome model errors.

2.5 Optimality of Threshold Policies

Next, we move towards statements of the main theorems underlying the results presented in Section 2.3. We begin by establishing notation which we shall use throughout. Recall that $\circ$ denotes the Hadamard product between vectors. We identify functions mapping $\mathcal{X} \to \mathbb{R}$ with vectors in $\mathbb{R}^C$. We also define the group-wise utilities

$$\mathcal{U}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}) := \sum_{x \in \mathcal{X}} \boldsymbol{\pi}_{\mathsf{j}}(x) \boldsymbol{\tau}_{\mathsf{j}}(x) \boldsymbol{u}(x) \ , \tag{2.6}$$

so that for $\boldsymbol{\tau} = (\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}})$, $\mathcal{U}(\boldsymbol{\tau}) := g_{\mathsf{A}} \mathcal{U}_{\mathsf{A}}(\boldsymbol{\tau}_{\mathsf{A}}) + g_{\mathsf{B}} \mathcal{U}_{\mathsf{B}}(\boldsymbol{\tau}_{\mathsf{B}})$.

First, we formally describe threshold policies, and rigorously justify why we may always assume without loss of generality that the institution adopts policies of this form.

Definition 2.5.1 (Threshold selection policy)**.** *A single group selection policy $\boldsymbol{\tau} \in [0,1]^C$ is called a* threshold policy *if it has the form of a randomized threshold on score:*

$$\boldsymbol{\tau}_{c,\gamma} = \begin{cases} 1, & x > c \\ \gamma, & x = c \\ 0, & x < c \end{cases} \ , \ \textit{for some} \ \ c \in [C] \ \textit{and} \ \gamma \in (0,1] \, . \tag{2.7}$$

As a technicality, if no members of a population have a given score $x \in \mathcal{X}$, there may be multiple threshold policies which yield equivalent selection rates for a given population. To avoid redundancy, we introduce the notation $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} \boldsymbol{\tau}'_{\mathsf{j}}$ to mean that the set of scores on which $\boldsymbol{\tau}_{\mathsf{j}}$ and $\boldsymbol{\tau}'_{\mathsf{j}}$ differ has probability 0 under $\boldsymbol{\pi}_{\mathsf{j}}$; formally, $\sum_{x : \boldsymbol{\tau}_{\mathsf{j}}(x) \neq \boldsymbol{\tau}_{\mathsf{j}}(x)} \boldsymbol{\pi}_{\mathsf{j}}(x) = 0$. For any distribution $\boldsymbol{\pi}_{\mathsf{j}}$, $\cong_{\boldsymbol{\pi}_{\mathsf{j}}}$ is an equivalence relation. Moreover, we see that if $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} \boldsymbol{\tau}'_{\mathsf{j}}$, then $\boldsymbol{\tau}_{\mathsf{j}}$ and $\boldsymbol{\tau}'_{\mathsf{j}}$ both provide the same utility for the institution, induce the same outcomes for individuals in group j, and have the same selection and true positive rates. Hence, if $(\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}})$ is an optimal solution to any of `MaxUtil`, `EqOpt`, or `DemParity`, so is any $(\boldsymbol{\tau}'_{\mathsf{A}}, \boldsymbol{\tau}'_{\mathsf{B}})$ for which $\boldsymbol{\tau}_{\mathsf{A}} \cong_{\boldsymbol{\pi}_{\mathsf{A}}} \boldsymbol{\tau}'_{\mathsf{A}}$ and $\boldsymbol{\tau}_{\mathsf{B}} \cong_{\boldsymbol{\pi}_{\mathsf{B}}} \boldsymbol{\tau}'_{\mathsf{B}}$.

For threshold policies in particular, their equivalence class under $\cong_{\boldsymbol{\pi}_{\mathsf{j}}}$ is uniquely determined by the selection rate function,

$$r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}) := \sum_{x \in \mathcal{X}} \boldsymbol{\pi}_{\mathsf{j}}(x) \boldsymbol{\tau}_{\mathsf{j}}(x) \ , \tag{2.8}$$

which denotes the fraction of group j which is selected. Indeed, we have the following lemma (proved in Section 2.9):

Lemma 2.5.1. *Let $\boldsymbol{\tau}_{\mathsf{j}}$ and $\boldsymbol{\tau}'_{\mathsf{j}}$ be threshold policies. Then $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} \boldsymbol{\tau}'_{\mathsf{j}}$ if and only if $r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}) = r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}'_{\mathsf{j}})$. Further, $r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}})$ is a bijection from $\mathcal{T}_{\text{thresh}}(\boldsymbol{\pi}_{\mathsf{j}})$ to $[0,1]$, where $\mathcal{T}_{\text{thresh}}(\boldsymbol{\pi}_{\mathsf{j}})$ is the set of equivalence classes between threshold policies under $\cong_{\boldsymbol{\pi}_{\mathsf{j}}}$. Finally, $\boldsymbol{\pi}_{\mathsf{j}} \circ r^{-1}_{\boldsymbol{\pi}_{\mathsf{j}}}(\beta_{\mathsf{j}})$ is well defined.*

Remark that $r_{\boldsymbol{\pi}_\mathsf{j}}^{-1}(\beta_\mathsf{j})$ is an equivalence class rather than a single policy. However, $\boldsymbol{\pi}_\mathsf{j} \circ r_{\boldsymbol{\pi}_\mathsf{j}}^{-1}(\boldsymbol{\tau}_\mathsf{j})$ is well defined, meaning that $\boldsymbol{\pi}_\mathsf{j} \circ \boldsymbol{\tau}_\mathsf{j} = \boldsymbol{\pi}_\mathsf{j} \circ \boldsymbol{\tau}_\mathsf{j}'$ for any two policies in the same equivalence class. Since all quantities of interest will only depend on policies $\boldsymbol{\tau}_\mathsf{j}$ through $\boldsymbol{\pi}_\mathsf{j} \circ \boldsymbol{\tau}_\mathsf{j}$, it does not matter *which* representative of $r_{\boldsymbol{\pi}_\mathsf{j}}^{-1}(\beta_\mathsf{j})$ we pick. Hence, abusing notation slightly, we shall represent $\mathcal{T}_{\text{thresh}}(\boldsymbol{\pi}_\mathsf{j})$ by choosing one representative from each equivalence class under $\cong_{\boldsymbol{\pi}_\mathsf{j}}$[2].

It turns out the policies which arise in this away are always optimal in the sense that, for a given loan rate β_j, the threshold policy $r_{\boldsymbol{\pi}_\mathsf{j}}^{-1}(\beta_j)$ is the (essentially unique) policy which maximizes both the institution's utility and the utility of the group. Defining the group-wise utility,

$$\mathcal{U}_\mathsf{j}(\boldsymbol{\tau}_\mathsf{j}) := \sum_{x \in \mathcal{X}} \boldsymbol{u}(x)\boldsymbol{\pi}_\mathsf{j}(x)\boldsymbol{\tau}_\mathsf{j}(x) \ , \tag{2.9}$$

we have the following result:

Proposition 2.5.2 (Threshold policies are preferable)**.** *Suppose that $\boldsymbol{u}(x)$ and $\boldsymbol{\Delta}(x)$ are strictly increasing in x. Given any loaning policy $\boldsymbol{\tau}_\mathsf{j}$ for population with distribution $\boldsymbol{\pi}_j$, then the policy $\boldsymbol{\tau}_\mathsf{j}^{\text{thresh}} := r_{\boldsymbol{\pi}_\mathsf{j}}^{-1}(r_{\boldsymbol{\pi}_\mathsf{j}}(\boldsymbol{\tau}_\mathsf{j})) \in \mathcal{T}_{\text{thresh}}(\boldsymbol{\pi}_\mathsf{j})$ satisfies*

$$\Delta\boldsymbol{\mu}_\mathsf{j}(\boldsymbol{\tau}_\mathsf{j}^{\text{thresh}}) \geq \Delta\boldsymbol{\mu}_\mathsf{j}(\boldsymbol{\tau}_\mathsf{j}) \text{ and } \mathcal{U}_\mathsf{j}(\boldsymbol{\tau}_\mathsf{j}^{\text{thresh}}) \geq \mathcal{U}_\mathsf{j}(\boldsymbol{\tau}_\mathsf{j}) \ . \tag{2.10}$$

Moreover, both inequalities hold with equality if and only if $\boldsymbol{\tau}_\mathsf{j} \cong_{\boldsymbol{\pi}_\mathsf{j}} \boldsymbol{\tau}_\mathsf{j}^{\text{thresh}}$.

The map $\boldsymbol{\tau}_\mathsf{j} \mapsto r_{\boldsymbol{\pi}_\mathsf{j}}^{-1}(r_{\boldsymbol{\pi}_\mathsf{j}}(\boldsymbol{\tau}_\mathsf{j}))$ can be thought of transforming an arbitrary policy $\boldsymbol{\tau}_\mathsf{j}$ into a threshold policy with the same selection rate. In this language, the above proposition states that this map never reduces institution utility or individual outcomes. We can also show that optimal `MaxUtil` and `DemParity` policies are threshold policies, as well as all `EqOpt` policies under an additional assumption:

Proposition 2.5.3 (Existance of optimal threshold policies under fairness constraints)**.** *Suppose that $\boldsymbol{u}(x)$ is strictly increasing in x. Then all optimal* `MaxUtil` *policies $(\boldsymbol{\tau}_\mathsf{A}, \boldsymbol{\tau}_\mathsf{B})$ satisfy $\boldsymbol{\tau}_\mathsf{j} \cong_{\boldsymbol{\pi}_\mathsf{j}} r_{\boldsymbol{\pi}_\mathsf{j}}^{-1}\left(r_{\boldsymbol{\pi}_\mathsf{j}}(\boldsymbol{\tau}_\mathsf{j})\right)$ for $\mathsf{j} \in \{\mathsf{A}, \mathsf{B}\}$. The same holds for all optimal* `DemParity` *policies, and if in addition $\boldsymbol{u}(x)/\boldsymbol{\rho}(x)$ is increasing, the same is true for all optimal* `EqOpt` *policies.*

To prove proposition 2.5.2, we invoke the following general lemma which is proved using standard convex analysis arguments (in Section 2.9):

Lemma 2.5.4. *Let $\boldsymbol{v} \in \mathbb{R}^C$, and let $\boldsymbol{w} \in \mathbb{R}^C_{>0}$, and suppose either that $\boldsymbol{v}(x)$ is increasing in x, and $\boldsymbol{v}(x)/\boldsymbol{w}(x)$ is increasing or, $\forall x \in \mathcal{X}$, $\boldsymbol{w}(x) = 0$. Let $\boldsymbol{\pi} \in \text{Simplex}^{C-1}$ and fix $t \in [0, \sum_{x \in \mathcal{X}} \boldsymbol{\pi}(x) \cdot \boldsymbol{w}(x)]$. Then any*

$$\boldsymbol{\tau}^* \in \arg\max_{\boldsymbol{\tau} \in [0,1]^C} \langle \boldsymbol{v} \circ \boldsymbol{\pi}, \boldsymbol{\tau} \rangle \quad \text{s.t.} \quad \langle \boldsymbol{\pi} \circ \boldsymbol{w}, \boldsymbol{\tau} \rangle = t \tag{2.11}$$

satisfies $\boldsymbol{\tau}^* \cong_{\boldsymbol{\pi}} r_{\boldsymbol{\pi}}^{-1}(r_{\boldsymbol{\pi}}(\boldsymbol{\tau}^*))$. Moreover, at least one maximizer $\boldsymbol{\tau}^* \in \mathcal{T}_{\text{thresh}}(\boldsymbol{\pi})$ exists.

[2]One way to do this is to consider the set of all threshold policies $\boldsymbol{\tau}_{c,\gamma}$ such that, $\gamma = 1$ if $\boldsymbol{\pi}_\mathsf{j}(c) = 0$ and $\boldsymbol{\pi}_\mathsf{j}(c-1) > 0$ if $\gamma = 1$ and $c > 1$.

Proof of Proposition 2.5.2. We will first prove Proposition 2.5.2 for the function $\mathcal{U}_{\mathsf{j}}$. Given our nominal policy $\boldsymbol{\tau}_{\mathsf{j}}$, let $\beta_j = r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}})$. We now apply Lemma 2.5.4 with $\boldsymbol{v}(x) = \boldsymbol{u}(x)$ and $\boldsymbol{w}(x) = 1$. For this choice of $\boldsymbol{v}$ and $\boldsymbol{w}$, $\langle \boldsymbol{v}, \boldsymbol{\tau} \rangle = \mathcal{U}_{\mathsf{j}}(\boldsymbol{\tau})$ and that $\langle \boldsymbol{\pi}_j \circ \boldsymbol{w}, \boldsymbol{\tau} = r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau})$. Then, if $\boldsymbol{\tau}_{\mathsf{j}} \in \arg\max_{\boldsymbol{\tau}} \mathcal{U}_{\mathsf{j}}(\boldsymbol{\tau})$ s.t. $r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}) = \beta_{\mathsf{j}}$, Lemma 2.11 implies that $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} r^{-1}_{\boldsymbol{\pi}_{\mathsf{j}}}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}))$.

On the other hand, assume that $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} r^{-1}_{\boldsymbol{\pi}_{\mathsf{j}}}\left(r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}})\right)$. We show that $r^{-1}_{\boldsymbol{\pi}_{\mathsf{j}}}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}))$ is a maximizer; which will imply that $\boldsymbol{\tau}_{\mathsf{j}}$ is a maximizer since $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} r^{-1}_{\boldsymbol{\pi}_{\mathsf{j}}}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}))$ implies that $\mathcal{U}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}) = \boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} r^{-1}_{\boldsymbol{\pi}_{\mathsf{j}}}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}))$. By Lemma 2.5.4 there exists a maximizer $\boldsymbol{\tau}^*_{\mathsf{j}} \in \mathcal{T}_{\text{thresh}}(\boldsymbol{\pi})$, which means that $\boldsymbol{\tau}^*_{\mathsf{j}} = r^{-1}_{\boldsymbol{\pi}_{\mathsf{j}}}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}^*_{\mathsf{j}}))$. Since $\boldsymbol{\tau}^*_{\mathsf{j}}$ is feasible, we must have $r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}^*_{\mathsf{j}}) = r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}})$, and thus $\boldsymbol{\tau}^*_{\mathsf{j}} = r^{-1}_{\boldsymbol{\pi}_{\mathsf{j}}}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}))$, as needed. The same argument follows verbatim if we instead choose $\boldsymbol{v}(x) = \boldsymbol{\Delta}(x)$, and compute $\langle \boldsymbol{v}, \boldsymbol{\tau} \rangle = \Delta\boldsymbol{\mu}_{\mathsf{j}}(\boldsymbol{\tau})$. □

We now argue Proposition 2.5.3 for `MaxUtil`, as it is a straightforward application of Lemma 2.5.4. We will prove Proposition 2.5.3 for `DemParity` and `EqOpt` separately in Sections 2.6 and 2.6.

Proof of Proposition 2.5.3 for `MaxUtil`. `MaxUtil` follows from lemma 2.5.4 with $\boldsymbol{v}(x) = \boldsymbol{u}(x)$, and $t = 0$ and $\boldsymbol{w} = \boldsymbol{0}$.

□

Quantiles and Concavity of the Outcome Curve

To further our analysis, we now introduce left and right quantile functions, allowing us to specify thresholds in terms of both selection rate and score cutoffs.

Definition 2.5.2 (Upper quantile function)**.** *Define* Q *to be the upper quantile function corresponding to* $\boldsymbol{\pi}$*, i.e.*

$$\mathrm{Q}_{\mathsf{j}}(\beta) = \operatorname{argmax}\{c : \sum_{x=c}^{C} \boldsymbol{\pi}_{\mathsf{j}}(x) > \beta\} \quad \textit{and} \quad \mathrm{Q}^{+}_{\mathsf{j}}(\beta) := \operatorname{argmax}\{c : \sum_{x=c}^{C} \boldsymbol{\pi}_{\mathsf{j}}(x) \geq \beta\}\,. \tag{2.12}$$

Crucially $\mathrm{Q}(\beta)$ is continuous from the right, and $\mathrm{Q}^{+}(\beta)$ is continuous from the left. Further, $\mathrm{Q}(\cdot)$ and $\mathrm{Q}^{+}(\cdot)$ allow us to compute derivatives of key functions, like the mapping from selection rate β to the group outcome associated with a policy of that rate, $\Delta\boldsymbol{\mu}(r^{-1}_{\pi}(\beta))$. Because we take $\boldsymbol{\pi}$ to have discrete support, all functions in this work are *piecewise linear*, so we shall need to distinguish between the left and right derivatives, defined as follows

$$\partial_{-} f(x) := \lim_{t \to 0^{-}} \frac{f(x+t) - f(x)}{t} \quad \text{and} \quad \partial_{+} f(y) := \lim_{t \to 0^{+}} \frac{f(y+t) - f(y)}{t}\,. \tag{2.13}$$

For f supported on $[a, b]$, we say that f is left- (resp. right-) differentiable if $\partial_{-} f(x)$ exists for all $x \in (a, b]$ (resp. $\partial_{+} f(y)$ exists for all $y \in [a, b)$). We now state the fundamental derivative computation which underpins the results to follow:

Lemma 2.5.5. *Let $\boldsymbol{e}_x$ denote the vector such that $\boldsymbol{e}_x(x) = 1$, and $\boldsymbol{e}_x(x') = 0$ for $x' \neq x$. Then $\boldsymbol{\pi}_{\mathsf{j}} \circ r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(\beta) : [0,1] \to [0,1]^C$ is continuous, and has left and right derivatives*

$$\partial_+ \left(\boldsymbol{\pi}_{\mathsf{j}} \circ r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(\beta) \right) = \boldsymbol{e}_{\mathrm{Q}(\beta)} \quad \textit{and} \quad \partial_- \left(\boldsymbol{\pi}_{\mathsf{j}} \circ r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(\beta) \right) = \boldsymbol{e}_{\mathrm{Q}^+(\beta)} \,. \tag{2.14}$$

The above lemma is proved in Section 2.9. Moreover, Lemma 2.5.5 implies that the outcome curve is concave under the assumption that $\boldsymbol{\Delta}(x)$ is monotone:

Proposition 2.5.6. *Let $\boldsymbol{\pi}$ be a distribution over C states. Then $\beta \mapsto \Delta\boldsymbol{\mu}(r_{\boldsymbol{\pi}}^{-1}(\beta))$ is concave. In fact, if $\boldsymbol{w}(x)$ is any non-decreasing map from $\mathcal{X} \to \mathbb{R}$, $\beta \mapsto \langle \boldsymbol{w}, r_{\boldsymbol{\pi}}^{-1}(\beta) \rangle$ is concave.*

Proof. Recall that a univariate function f is concave (and finite) on $[a,b]$ if and only (a) f is left- and right-differentiable, (b) for all $x \in (a,b)$, $\partial_- f(x) \geq \partial_+ f(x)$ and (c) for any $x > y$, $\partial_- f(x) \leq \partial_+ f(y)$.

Observe that $\Delta\boldsymbol{\mu}(r_{\boldsymbol{\pi}}^{-1}(\beta)) = \langle \boldsymbol{\Delta}, \boldsymbol{\pi} \circ r_{\boldsymbol{\pi}}^{-1}(\beta) \rangle$. By Lemma 2.5.5, $\boldsymbol{\pi} \circ r_{\boldsymbol{\pi}}^{-1}(\beta)$ has right and left derivatives $\boldsymbol{e}_{\mathrm{Q}(\beta)}$ and $\boldsymbol{e}_{\mathrm{Q}^+(\beta)}$. Hence, we have that

$$\partial_+ \Delta\boldsymbol{\mu}(\beta_{\mathsf{B}}) = \boldsymbol{\Delta}(\mathrm{Q}(\beta_{\mathsf{B}})) \quad \text{and} \quad \partial_- \Delta\boldsymbol{\mu}(\beta_{\mathsf{B}}) = \boldsymbol{\Delta}(\mathrm{Q}^+(\beta_{\mathsf{B}})) \,. \tag{2.15}$$

Using the fact that $\boldsymbol{\Delta}(x)$ is monotone, and that $\mathrm{Q} \leq \mathrm{Q}^+$, we see that $\partial_+ \Delta\boldsymbol{\mu}(f_{\boldsymbol{\pi}}^{-1}(\beta_{\mathsf{B}})) \leq \partial_- \Delta\boldsymbol{\mu}(f_{\boldsymbol{\pi}}^{-1}(\beta_{\mathsf{B}}))$, and that $\partial \Delta\boldsymbol{\mu}(f_{\boldsymbol{\pi}}^{-1}(\beta_{\mathsf{B}}))$ and $\partial_+ \Delta\boldsymbol{\mu}(f_{\boldsymbol{\pi}}^{-1}(\beta_{\mathsf{B}}))$ are non-increasing, from which it follows that $\Delta\boldsymbol{\mu}(f_{\boldsymbol{\pi}}^{-1}(\beta_{\mathsf{B}}))$ is concave. The general concavity result holds by replacing $\boldsymbol{\Delta}(x)$ with $\boldsymbol{w}(x)$. □

2.6 Proofs of Main Theorems

We are now ready to present and prove theorems that characterize the selection rates under fairness constraints, namely `DemParity` and `EqOpt`. These characterizations are crucial for proving the results in Section 2.3. Our computations also generalize readily to other linear constraints, in a way that will become clear in Section 2.6.

A Characterization Theorem for `DemParity`

In this section, we provide a theorem that gives an explicit characterization for the range of selection rates β_{A} for A when the bank loans according to `DemParity`. Observe that the `DemParity` objective corresponds to solving the following linear program:

$$\max_{\boldsymbol{\tau} = (\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}}) \in [0,1]^{2C}} \mathcal{U}(\boldsymbol{\tau}) \quad \text{s.t.} \quad \langle \boldsymbol{\pi}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{A}} \rangle = \langle \boldsymbol{\pi}_{\mathsf{B}}, \boldsymbol{\tau}_{\mathsf{B}} \rangle \,.$$

Let us introduce the auxiliary variable $\beta := \langle \boldsymbol{\pi}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{A}} \rangle = \langle \boldsymbol{\pi}_{\mathsf{B}}, \boldsymbol{\tau}_{\mathsf{B}} \rangle$ corresponding to the selection rate which is held constant across groups, so that all feasible solutions lie on the green DP line in Figure 2.3. We can then express the following equivalent linear program:

$$\max_{\boldsymbol{\tau} = (\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}}) \in [0,1]^{2C}, \beta \in [0,1]} \mathcal{U}(\boldsymbol{\tau}) \quad \text{s.t.} \quad \beta = \langle \boldsymbol{\pi}_{\mathsf{j}}, \boldsymbol{\tau}_{\mathsf{j}} \rangle, \ \mathsf{j} \in \{\mathsf{A}, \mathsf{B}\} \,.$$

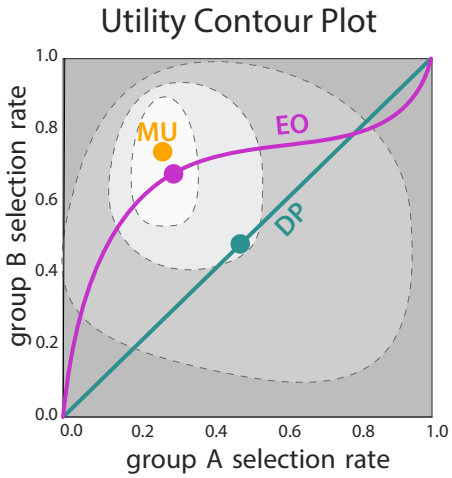


Figure 2.3: Considering the utility as a function of selection rates, fairness constraints correspond to restricting the optimization to one-dimensional curves. The `DemParity` (DP) constraint is a straight line with slope 1, while the `EqOpt` (EO) constraint is a curve given by the graph of $G^{(\mathsf{A}\to\mathsf{B})}$. The derivatives considered throughout Section 2.6 are taken with respect to the selection rate β_{A} (horizontal axis); projecting the EO and DP constraint curves to the horizontal axis recovers concave utility curves such as those shown in the lower panel of Figure 2.2 (where `MaxUtil` in is represented by a horizontal line through the MU optimal solution).

This is equivalent because, for a given β, Proposition 2.5.3 says that the utility maximizing policies are of the form $\boldsymbol{\tau}_{\mathsf{j}} = r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(\beta)$. We now prove this:

Proof of Proposition 2.5.3 for `DemParity`. Noting that $r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}) = \langle \boldsymbol{\pi}_{\mathsf{j}}, \boldsymbol{\tau}_{\mathsf{j}} \rangle$, we see that, by Lemma 2.5.4, under the special case where $\boldsymbol{v}(x) = \boldsymbol{u}(x)$ and $\boldsymbol{w}(x) = 1$, the optimal solution $(\boldsymbol{\tau}_{\mathsf{A}}^*(\beta), \boldsymbol{\tau}_{\mathsf{B}}^*(\beta))$ for fixed $r_{\boldsymbol{\pi}_{\mathsf{A}}}(\boldsymbol{\tau}_{\mathsf{A}}) = r_{\boldsymbol{\pi}_{\mathsf{B}}}(\boldsymbol{\tau}_{\mathsf{B}}) = \beta$ can be chosen to coincide with the threshold policies. Optimizing over β, the global optimal must coincide with thresholds.

□

Hence, any optimal policy is equivalent to the threshold policy $\boldsymbol{\tau} = (r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta), r_{\boldsymbol{\pi}_{\mathsf{B}}}^{-1}(\beta))$, where β solves the following optimization:

$$\max_{\beta \in [0,1]} \mathcal{U}\left(\left(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta), r_{\boldsymbol{\pi}_{\mathsf{B}}}^{-1}(\beta) \right) \right) . \tag{2.16}$$

We shall show that the above expression is in fact a *concave* function in β, and hence the set of optimal selection rates can be characterized by first order conditions. This is presented formally in the following theorem:

Theorem 2.6.1 (Selection rates for `DemParity`)**.** *The set of optimal selection rates β^* satisfying* (2.16) *forms a continuous interval $[\beta^-_{\texttt{DemParity}}, \beta^+_{\texttt{DemParity}}]$, such that for any $\beta \in [0,1]$, we have*

$$\begin{aligned} \beta < \beta^-_{\texttt{DemParity}} \quad &\text{if} \quad g_{\mathsf{A}}\boldsymbol{u}\left(\mathrm{Q}_{\mathsf{A}}(\beta)\right) + g_{\mathsf{B}}\boldsymbol{u}\left(\mathrm{Q}_{\mathsf{B}}(\beta)\right) > 0\,, \\ \beta > \beta^+_{\texttt{DemParity}} \quad &\text{if} \quad g_{\mathsf{A}}\boldsymbol{u}\left(\mathrm{Q}^+_{\mathsf{A}}(\beta)\right) + g_{\mathsf{B}}\boldsymbol{u}\left(\mathrm{Q}^+_{\mathsf{B}}(\beta)\right) < 0\,. \end{aligned}$$

Proof. Note that we can write

$$\mathcal{U}\left(\left(r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(\beta), r^{-1}_{\boldsymbol{\pi}_{\mathsf{B}}}(\beta)\right)\right) = g_{\mathsf{A}}\langle \boldsymbol{u}, \boldsymbol{\pi}_{\mathsf{A}} \circ r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(\beta)\rangle + g_{\mathsf{B}}\langle \boldsymbol{u}, \boldsymbol{\pi}_{\mathsf{B}} \circ r^{-1}_{\boldsymbol{\pi}_{\mathsf{B}}}(\beta)\rangle\,.$$

Since $\boldsymbol{u}(x)$ is non-decreasing in x, Proposition 2.5.6 implies that $\beta \mapsto \mathcal{U}\left(\left(r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(\beta), r^{-1}_{\boldsymbol{\pi}_{\mathsf{B}}}(\beta)\right)\right)$ is concave in β. Hence, all optimal selection rates β^* lie in an interval $[\beta^-, \beta^+]$. To further characterize this interval, let us us compute left- and right-derivatives.

$$\begin{aligned} \partial_+\mathcal{U}\left(\left(r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(\beta), r^{-1}_{\boldsymbol{\pi}_{\mathsf{B}}}(\beta)\right)\right) &= \partial_+ g_{\mathsf{A}}\langle \boldsymbol{u}, \boldsymbol{\pi}_{\mathsf{A}} \circ r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(\beta)\rangle + \partial_+ g_{\mathsf{B}}\langle \boldsymbol{u}, \boldsymbol{\pi}_{\mathsf{B}} \circ r^{-1}_{\boldsymbol{\pi}_{\mathsf{B}}}(\beta)\rangle \\ &= g_{\mathsf{A}}\langle \boldsymbol{u}, \partial_+\left(\boldsymbol{\pi}_{\mathsf{A}} \circ r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(\beta)\right)\rangle + g_{\mathsf{B}}\langle \boldsymbol{u}, \partial_+\left(\boldsymbol{\pi}_{\mathsf{B}} \circ r^{-1}_{\boldsymbol{\pi}_{\mathsf{B}}}(\beta)\right)\rangle \\ &\overset{\text{Lemma 2.5.5}}{=} g_{\mathsf{A}}\langle \boldsymbol{u}, \boldsymbol{e}_{\mathrm{Q}_{\mathsf{A}}(\beta)}\rangle + g_{\mathsf{B}}\langle \boldsymbol{u}, \boldsymbol{e}_{\mathrm{Q}_{\mathsf{B}}(\beta)}\rangle \\ &= g_{\mathsf{A}}\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta)) + g_{\mathsf{B}}\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta))\,. \end{aligned}$$

The same argument shows that

$$\partial_-\mathcal{U}((r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(\beta), r^{-1}_{\boldsymbol{\pi}_{\mathsf{B}}}(\beta))) = g_{\mathsf{A}}\boldsymbol{u}(\mathrm{Q}^+_{\mathsf{A}}(\beta)) + g_{\mathsf{B}}\boldsymbol{u}(\mathrm{Q}^+_{\mathsf{B}}(\beta)).$$

By concavity of $\mathcal{U}\left(\left(r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(\beta), r^{-1}_{\boldsymbol{\pi}_{\mathsf{B}}}(\beta)\right)\right)$, a positive right derivative at β implies that $\beta < \beta^*$ for all β^* satisfying (2.16), and similarly, a negative left derivative at β implies that $\beta > \beta^*$ for all β^* satisfying (2.16).

□

With a result of the above form, we can now easily prove statements such as that in Corollary 2.3.3 (see Section 2.9 for proofs), by fixing a selection rate of interest (e.g. β_0) and inverting the inequalities in Theorem 2.6.1 to find the exact population proportions under which, for example, `DemParity` results in a higher selection rate than β_0.

`EqOpt` and General Constraints

Next, we will provide a theorem that gives an explicit characterization for the range of selection rates β_{A} for A when the bank loans according to `EqOpt`. Observe that the `EqOpt` objective corresponds to solving the following linear program:

$$\max_{\boldsymbol{\tau}=(\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}}) \in [0,1]^{2C}} \mathcal{U}(\boldsymbol{\tau}) \quad \text{s.t.} \quad \langle \boldsymbol{w}_{\mathsf{A}} \circ \boldsymbol{\pi}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{A}}\rangle = \langle \boldsymbol{w}_{\mathsf{B}} \circ \boldsymbol{\pi}_{\mathsf{B}}, \boldsymbol{\tau}_{\mathsf{B}}\rangle\,, \tag{2.17}$$

where $\boldsymbol{w}_{\mathsf{j}} = \frac{\boldsymbol{\rho}}{\langle \boldsymbol{\rho}, \boldsymbol{\pi}_{\mathsf{j}} \rangle}$. This problem is similar to the demographic parity optimization in (2.16), except for the fact that the constraint includes the weights. Whereas we parameterized demographic parity solutions in terms of the acceptance rate β in equation (2.16), we will parameterize equation (2.17) in terms of the true positive rate (TPR), $t := \langle \boldsymbol{w}_{\mathsf{A}} \circ \boldsymbol{\pi}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{A}} \rangle$. Thus, (2.17) becomes

$$\max_{t \in [0, t_{\max}]} \max_{(\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}}) \in [0,1]^{2C}} \sum_{j \in \{\mathsf{A},\mathsf{B}\}} g_{\mathsf{j}} \mathcal{U}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}) \quad \text{s.t.} \quad \langle \boldsymbol{w}_{\mathsf{j}} \circ \boldsymbol{\pi}_{\mathsf{j}}, \boldsymbol{\tau}_{\mathsf{j}} \rangle = t, \ \mathsf{j} \in \{\mathsf{A}, \mathsf{B}\} \ , \tag{2.18}$$

where $t_{\max} = \min_{\mathsf{j} \in \{\mathsf{A},\mathsf{B}\}} \{ \langle \boldsymbol{\pi}_{\mathsf{j}}, \boldsymbol{w}_{\mathsf{j}} \rangle \}$ is the largest possible TPR. The magenta EO curve in Figure 2.3 illustrates that feasible solutions to this optimization problem lie on a curve parametrized by t. Note that the objective function decouples for $\mathsf{j} \in \{\mathsf{A}, \mathsf{B}\}$ for the inner optimization problem,

$$\max_{\boldsymbol{\tau}_{\mathsf{j}} \in [0,1]^C} \sum_{j \in \{\mathsf{A},\mathsf{B}\}} g_{\mathsf{j}} \mathcal{U}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}) \quad \text{s.t.} \quad \langle \boldsymbol{w}_{\mathsf{j}} \circ \boldsymbol{\pi}_{\mathsf{j}}, \boldsymbol{\tau}_{\mathsf{j}} \rangle = t \, . \tag{2.19}$$

We will now show that all optimal solutions for this inner optimization problem are $\boldsymbol{\pi}_{\mathsf{j}}$-a.e. equal to a policy in $\mathcal{T}_{\text{thresh}}(\boldsymbol{\pi}_{\mathsf{j}})$, and thus can be written as $r^{-1}_{\boldsymbol{\pi}_{\mathsf{j}}}(\beta_{\mathsf{j}})$, depending only on the resulting selection rate.

Proof of Proposition 2.5.3 for `EqOpt`. We apply Lemma 2.5.4 to the inner optimization in (2.19) with $\boldsymbol{v}(x) = \boldsymbol{u}(x)$ and $\boldsymbol{w}(x) = \frac{\boldsymbol{\rho}(x)}{\langle \boldsymbol{\rho}, \boldsymbol{\pi}_{\mathsf{j}} \rangle}$. The claim follows from the assumption that $\boldsymbol{u}(x)/\boldsymbol{\rho}(x)$ is increasing by optimizing over t. □

This selection rate β_{j} is uniquely determined by the TPR t (proof appears in Section 2.9): Suppose that $\boldsymbol{w}(x) > 0$ for all x. Then the function

$$T_{\mathsf{j}, \boldsymbol{w}_{\mathsf{j}}}(\beta) := \langle r^{-1}_{\boldsymbol{\pi}_{\mathsf{j}}}(\beta), \boldsymbol{\pi}_{\mathsf{j}} \circ \boldsymbol{w}_{\mathsf{j}} \rangle$$

is a bijection from $[0, 1]$ to $[0, \langle \boldsymbol{\pi}_{\mathsf{j}}, \boldsymbol{w} \rangle]$. Hence, for any $t \in [0, t_{\max}]$, the mapping from TPR to acceptance rate, $T^{-1}_{\mathsf{j}, \boldsymbol{w}_{\mathsf{j}}}(t)$, is well defined and any solution to (2.19) is $\boldsymbol{\pi}_{\mathsf{j}}$-a.e. equal to the policy $r^{-1}_{\boldsymbol{\pi}_{\mathsf{j}}}(T^{-1}_{\mathsf{j}, \boldsymbol{w}_{\mathsf{j}}}(t))$. Thus (2.18) reduces to

$$\max_{t \in [0, t_{\max}]} \sum_{j \in \{\mathsf{A},\mathsf{B}\}} g_{\mathsf{j}} \mathcal{U}_{\mathsf{j}} \left(r^{-1}_{\boldsymbol{\pi}_{\mathsf{j}}} \left(T^{-1}_{\mathsf{j}, \boldsymbol{w}_{\mathsf{j}}}(t) \right) \right) \, . \tag{2.20}$$

The above expression parametrizes the optimization problem in terms of a single variable. We shall show that the above expression is in fact a *concave* function in t, and hence the set of optimal selection rates can be characterized by first order conditions. This is presented formally in the following theorem:

Theorem 2.6.2 (Selection rates for `EqOpt`)**.** *The set of optimal selection rates β^* for group A satsifying* (2.18) *forms a continuous interval $[\beta^-_{\texttt{EqOpt}}, \beta^+_{\texttt{EqOpt}}]$, such that for any $\beta \in [0,1]$, we have*

$$\beta < \beta^-_{\texttt{EqOpt}} \quad \textit{if} \quad g_{\mathsf{A}} \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta))}{\boldsymbol{w}_{\mathsf{A}}(\mathrm{Q}_{\mathsf{A}}(\beta))} + g_{\mathsf{B}} \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G_{\boldsymbol{w}}^{(\mathsf{A}\to\mathsf{B})}(\beta)))}{\boldsymbol{w}_{\mathsf{B}}(\mathrm{Q}_{\mathsf{B}}(G_{\boldsymbol{w}}^{(\mathsf{A}\to\mathsf{B})}(\beta)))} > 0\,,$$

$$\beta > \beta^+_{\texttt{EqOpt}} \quad \textit{if} \quad g_{\mathsf{A}} \frac{\boldsymbol{u}(\mathrm{Q}^+_{\mathsf{A}}(\beta))}{\boldsymbol{w}_{\mathsf{A}}(\mathrm{Q}^+_{\mathsf{A}}(\beta))} + g_{\mathsf{B}} \frac{\boldsymbol{u}(\mathrm{Q}^+_{\mathsf{B}}(G_{\boldsymbol{w}}^{(\mathsf{A}\to\mathsf{B})}(\beta)))}{\boldsymbol{w}_{\mathsf{B}}(\mathrm{Q}^+_{\mathsf{B}}(G_{\boldsymbol{w}}^{(\mathsf{A}\to\mathsf{B})}(\beta)))} < 0\,.$$

Here, $G_{\boldsymbol{w}}^{(\mathsf{A}\to\mathsf{B})}(\beta) := T^{-1}_{\mathsf{B},\boldsymbol{w}_{\mathsf{B}}}(T^{-1}_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}(\beta))$ denotes the (well-defined) map from selection rates β_{A} for A to the selection rate β_{B} for B such that the policies $\boldsymbol{\tau}^*_{\mathsf{A}} := r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(\beta_{\mathsf{A}})$ and $\boldsymbol{\tau}^*_{\mathsf{B}} := r^{-1}_{\boldsymbol{\pi}_{\mathsf{B}}}(\beta_{\mathsf{B}})$ satisfy the constraint in (2.17)*.*

Proof. Starting with the equivalent problem in (2.20), we use the concavity result of Lemma 2.9.1. Because the objective function is the positive weighted sum of two concave functions, it is also concave. Hence, all optimal true positive rates t^* lie in an interval $[t^-, t^+]$. To further characterize $[t^-, t^+]$, we can compute left- and right-derivatives, again using the result of Lemma 2.9.1.

$$\begin{aligned}
\partial_+ \sum_{j\in\{\mathsf{A},\mathsf{B}\}} g_{\mathsf{j}} \mathcal{U}_{\mathsf{j}} \left(r^{-1}_{\boldsymbol{\pi}_{\mathsf{j}}}(T^{-1}_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}(t)) \right) &= g_{\mathsf{A}} \partial_+ \mathcal{U}_{\mathsf{A}} \left(r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(T^{-1}_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}(t)) \right) + g_{\mathsf{A}} \partial_+ \mathcal{U}_{\mathsf{A}} \left(r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(T^{-1}_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}(t)) \right) \\
&= g_{\mathsf{A}} \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(T^{-1}_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}(t)))}{\boldsymbol{w}_{\mathsf{A}}(\mathrm{Q}_{\mathsf{A}}(T^{-1}_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}(t)))} + g_{\mathsf{B}} \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(T^{-1}_{\mathsf{B},\boldsymbol{w}_{\mathsf{B}}}(t)))}{\boldsymbol{w}_{\mathsf{B}}(\mathrm{Q}_{\mathsf{B}}(T^{-1}_{\mathsf{B},\boldsymbol{w}_{\mathsf{B}}}(t)))}
\end{aligned}$$

The same argument shows that

$$\partial_- \sum_{j\in\{\mathsf{A},\mathsf{B}\}} g_{\mathsf{j}} \mathcal{U}_{\mathsf{j}} \left(r^{-1}_{\boldsymbol{\pi}_{\mathsf{j}}}(T^{-1}_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}(t)) \right) = g_{\mathsf{A}} \frac{\boldsymbol{u}(\mathrm{Q}^+_{\mathsf{A}}(T^{-1}_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}(t))}{\boldsymbol{w}_{\mathsf{A}}(\mathrm{Q}^+_{\mathsf{A}}(T^{-1}_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}(t)))} + g_{\mathsf{B}} \frac{\boldsymbol{u}(\mathrm{Q}^+_{\mathsf{B}}(T^{-1}_{\mathsf{B},\boldsymbol{w}_{\mathsf{B}}}(t)))}{\boldsymbol{w}_{\mathsf{B}}(\mathrm{Q}^+_{\mathsf{B}}(T^{-1}_{\mathsf{B},\boldsymbol{w}_{\mathsf{B}}}(t)))}.$$

By concavity, a positive right derivative at t implies that $t < t^*$ for all t^* satisfying (2.20), and similarly, a negative left derivative at t implies that $t > t^*$ for all t^* satisfying (2.20).

Finally, by Lemma 2.6, this interval in t uniquely characterizes an interval of acceptance rates. Thus we translate directly into a statement about the selection rates β for group A by seeing that $T^{-1}_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}(t) = \beta$ and $T^{-1}_{\mathsf{B},\boldsymbol{w}_{\mathsf{B}}}(t) = G_{\boldsymbol{w}}^{(\mathsf{A}\to\mathsf{B})}(\beta)$. □

Lastly, we remark that the results derived in this section go through verbatim for any linear constraint of the form $\langle \boldsymbol{w}, \boldsymbol{\pi}_{\mathsf{A}} \circ \boldsymbol{\tau}_{\mathsf{A}} \rangle = \langle \boldsymbol{w}, \boldsymbol{\pi}_{\mathsf{B}} \circ \boldsymbol{\tau}_{\mathsf{B}} \rangle$, as long as $\boldsymbol{u}(x)/\boldsymbol{w}(x)$ is increasing in x, and $\boldsymbol{w}(x) > 0$.

2.7 Simulations

We examine the outcomes induced by fairness constraints in the context of FICO scores for two race groups. FICO scores are a proprietary classifier widely used in the United States

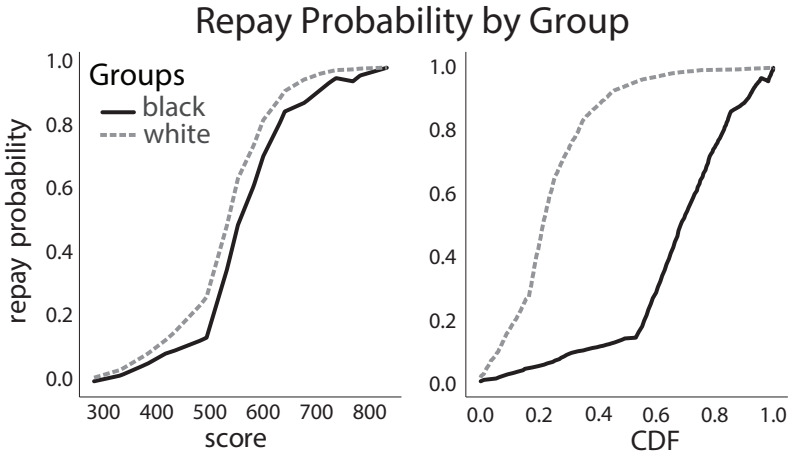


Figure 2.4: The empirical payback rates as a function of credit score and CDF for both groups from the TransUnion TransRisk dataset.

to predict credit worthiness. Our FICO data is based on a sample of 301,536 TransUnion TransRisk scores from 2003 [US Federal Reserve, 2007], preprocessed by Hardt et al. [2016c]. These scores, corresponding to x in our model, range from 300 to 850 and are meant to predict credit risk. Empirical data labeled by race allows us to estimate the distributions π_{j}, where j represents race, which is restricted to two values: white non-Hispanic (labeled "white" in figures), and black. Using national demographic data, we set the population proportions to be 18% and 82%.

Individuals were labeled as defaulted if they failed to pay a debt for at least 90 days on at least one account in the ensuing 18-24 month period; we use this data to estimate the success probability given score, $\boldsymbol{\rho}_{\mathsf{j}}(x)$, which we allow to vary by group to match the empirical data (see Figure 2.4). Our outcome curve framework allows for this relaxation; however, this discrepancy can also be attributed to group-dependent mismeasurement of score, and adjusting the scores accordingly would allow for a single $\boldsymbol{\rho}(x)$. We use the success probabilities to define the affine utility and score change functions defined in Example 2.2.1. We model individual penalties as a score drop of $c_- = -150$ in the case of a default, and in increase of $c_+ = 75$ in the case of successful repayment.

In Figure 2.5, we display the empirical CDFs along with selection rates resulting from different loaning strategies for two different settings of bank utilities. In the case that the bank experiences a loss/profit ratio of $\frac{u_-}{u_+} = -10$, no fairness criteria surpass the active harm rate β_0; however, in the case of $\frac{u_-}{u_+} = -4$, `DemParity` overloans, in line with the statement in Corollary 2.3.3.

These results are further examined in Figure 2.6, which displays the normalized outcome curves and the utility curves for both the white and the black group. To plot the `MaxUtil` utility curves, the group that is not on display has selection rate fixed at $\beta^{\texttt{MaxUtil}}$. In this

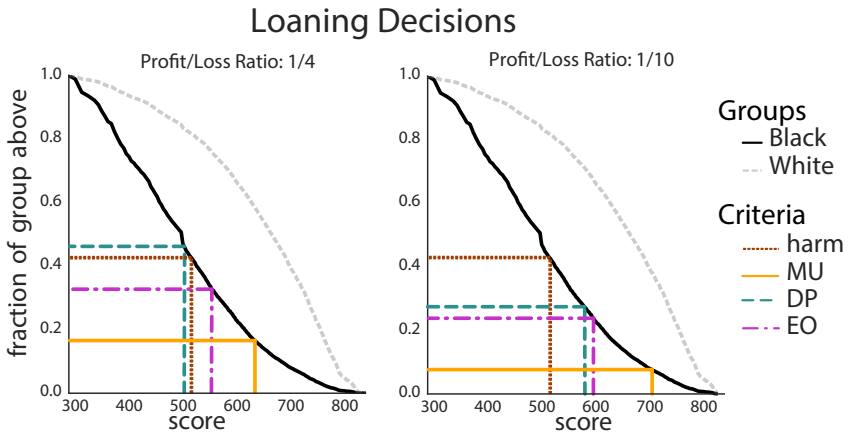


Figure 2.5: The empirical CDFs of both groups are plotted along with the decision thresholds resulting from `MaxUtil`, `DemParity`, and `EqOpt` for a model with bank utilities set to (a) $\frac{u_-}{u_+} = -4$ and (b) $\frac{u_-}{u_+} = -10$. The threshold for active harm is displayed; in (a) `DemParity` causes active harm while in (b) it does not. `EqOpt` and `MaxUtil` never cause active harm.

figure, the top panel corresponds to the average change in credit scores for each group under different loaning rates β; the bottom panels shows the corresponding *total* utility $\mathcal{U}$ (summed over both groups and weighted by group population sizes) for the bank.

Figure 2.6 highlights that the position of the utility optima in the lower panel determines the loan (selection) rates. In this specific instance, the utility and change ratios are fairly close, $\frac{u_-}{u_+} = -4$, and $\frac{c_-}{c_+} = -2$, meaning that the bank's profit motivations align with individual outcomes to some extent. Here, we can see that `EqOpt` loans much closer to optimal than `DemParity`, similar to the setting suggested by Corollary 2.3.2.

Although one might hope for decisions made under fairness constraints to positively affect the black group, we observe the opposite behavior. The `MaxUtil` policy (solid orange line) and the `EqOpt` policy result in similar expected credit score change for the black group. However, `DemParity` (dashed green line) causes a negative expected credit score change in the black group, corresponding to active harm. For the white group, the bank utility curve has almost the same shape under the fairness criteria as it does under `MaxUtil`, the main difference being that fairness criteria lowers the total expected profit from this group.

This behavior stems from a discrepancy in the outcome and profit curves for each population. While incentives for the bank and positive results for individuals are somewhat aligned for the majority group, under fairness constraints, they are more heavily misaligned in the minority group, as seen in graphs (left) in Figure 2.6. We remark that in other settings where the *unconstrained* profit maximization is misaligned with individual outcomes (e.g., when $\frac{u_-}{u_+} = -10$), fairness criteria may perform more favorably for the minority group by pulling the utility curve into a shape consistent with the outcome curve.

By analyzing the resulting affects of `MaxUtil`, `DemParity`, and `EqOpt` on actual credit

score lending data, we show the applicability of our model to real-world applications. In particular, some results shown in Section 2.3 hold empirically for the FICO TransUnion TransRisk scores.

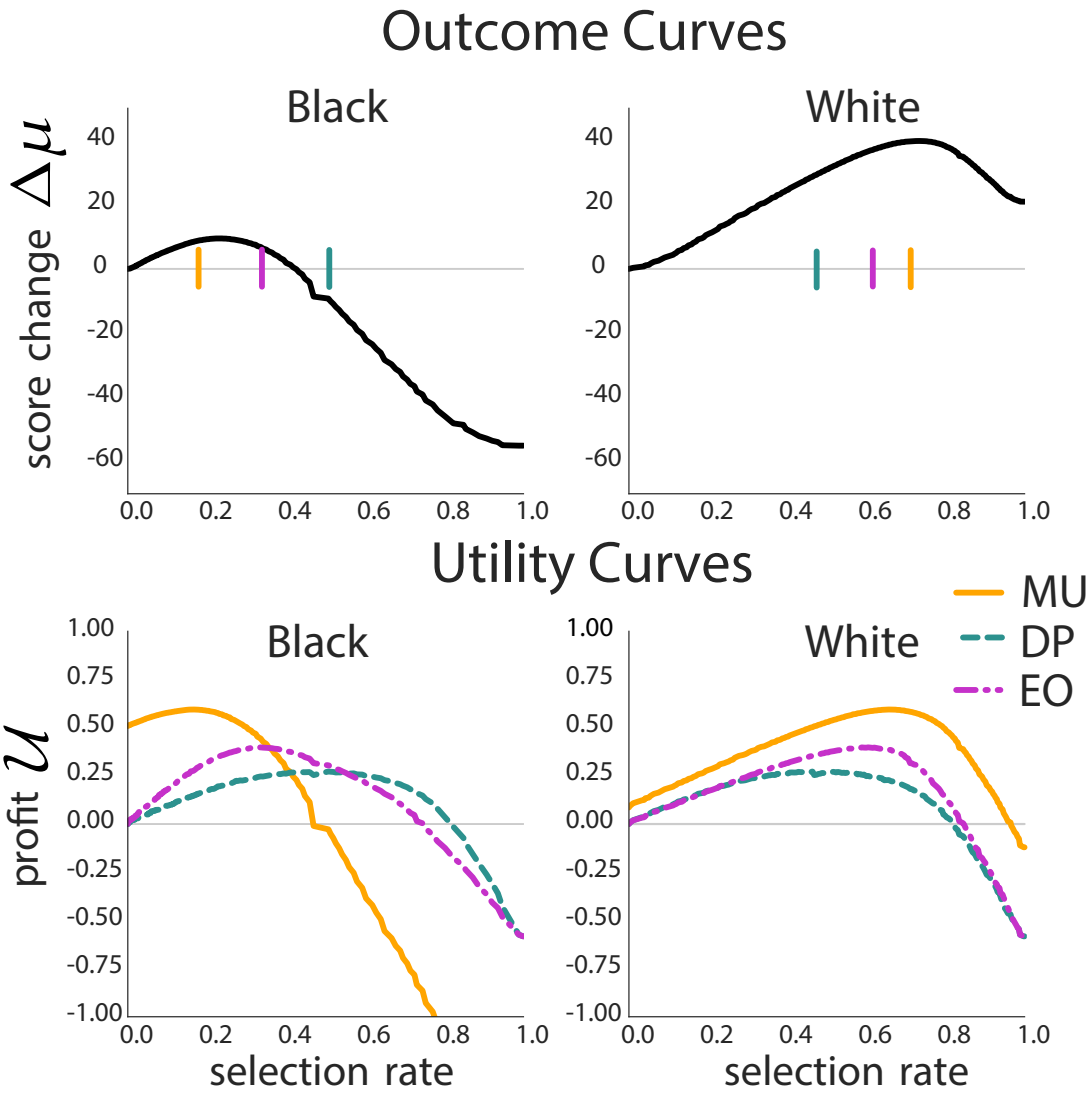


Figure 2.6: The outcome and utility curves are plotted for both groups against the group selection rates. The relative positions of the utility maxima determine the position of the decision rule thresholds. We hold $\frac{u_-}{u_+} = -4$ as fixed.

2.8 Discussion

In this chapter, we presented a model of the welfare impact of a decision rule over one time period. This simple model already suggests the importance of considering longer-term outcomes in the discussion of "fair" machine learning.[3] Without a careful model of delayed outcomes, we cannot foresee the impact a fairness criterion would have if enforced as a constraint on a classification system. However, if such an accurate outcome model is available, we showed that there are more direct ways to optimize for positive outcomes than via existing fairness criteria. We outlined such an outcome-based solution in Section 2.4. Specifically, in the credit setting, the outcome-based solution corresponds to giving out more loans to the protected group in a way that reduces profit for the bank compared to unconstrained profit maximization, but avoids loaning to those who are unlikely to benefit, resulting in a maximally improved group average credit score. The extent to which such a solution could form the basis of successful regulation depends on the accuracy of the available outcome model.

This raises the question if our model of outcomes is rich enough to faithfully capture realistic phenomena. By focusing on the impact that selection has on individuals at a given score, we model the effects for those *not* selected as zero-mean. For example, not getting a loan in our model has no negative effect on the credit score of an individual.[4] This does not mean that wrongful rejection (i.e., a false negative) has no visible manifestation in our model. If a classifier has a higher false negative rate in one group than in another, we expect the classifier to increase the disparity between the two groups (under natural assumptions). In other words, in our outcome-based model, the harm of denied opportunity manifests as growing disparity between the groups. The cost of a false negative could also be incorporated directly into the outcome-based model by a simple modification (see Footnote 1). This may be fitting in some applications where the immediate impact of a false negative to the individual is not zero-mean, but significantly reduces their future success probability.

In essence, the formalism we propose requires us to understand the two-variable causal mechanism that translates decisions to outcomes. This can be seen as relaxing the requirements compared with recent work on avoiding discrimination through causal reasoning that often required stronger assumptions [Kusner et al., 2017, Nabi and Shpitser, 2017, Kilbertus et al., 2017]. In particular, these works required knowledge of how sensitive attributes (such as gender, race, or proxies thereof) causally relate to various other variables in the data. Our model avoids the delicate modeling step involving the sensitive attribute, and instead focuses on an arguably more tangible economic mechanism. Nonetheless, depending on the application, such an understanding might necessitate greater domain knowledge and additional research into the specifics of the application. This is consistent with much scholarship

[3]Follow up work by D'Amour et al. [2020] and Williams and Kolter [2019] have also examined the equilibrium behavior of the presented model, via simulations and theoretical analysis under structural assumptions, respectively.

[4]In reality, a denied credit inquiry may lower one's credit score, but the effect is small compared to a default event.

that points to the context-sensitive nature of fairness in machine learning.

2.9 Omitted proofs

Optimality of Threshold Policies

Proof of Lemma 2.5.1

We begin with the first statement of the lemma. Suppose $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} \boldsymbol{\tau}'_{\mathsf{j}}$. Then there exists a set $\mathcal{S} \subset \mathcal{X}$ such that $\boldsymbol{\pi}_{\mathsf{j}}(x) = 0$ for all $x \in \mathcal{S}$, and for all $x \notin \mathcal{S}$, $\boldsymbol{\tau}_{\mathsf{j}}(x) = \boldsymbol{\tau}'_{\mathsf{j}}(x)$. Thus,

$$\begin{aligned} r_{\boldsymbol{\pi}}(\boldsymbol{\tau}_{\mathsf{j}}) - r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}'_{\mathsf{j}}) &= \sum_{x \in \mathcal{X}} \boldsymbol{\pi}_{\mathsf{j}}(x)(\boldsymbol{\tau}_{\mathsf{j}}(x) - \boldsymbol{\tau}'_{\mathsf{j}}(x)) \\ &= \sum_{x \in \mathcal{S}} \boldsymbol{\pi}_{\mathsf{j}}(x)(\boldsymbol{\tau}_{\mathsf{j}}(x) - \boldsymbol{\tau}'_{\mathsf{j}}(x)) = 0\,. \end{aligned}$$

Conversely, suppose that $r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}) = r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}'_{\mathsf{j}})$. Let $\boldsymbol{\tau}_{\mathsf{j}} = \boldsymbol{\tau}_{c,\gamma}$ and $\boldsymbol{\tau}'_{\mathsf{j}} = \boldsymbol{\tau}_{c',\gamma'}$ as in Definition 2.5.1. We now have the following cases:

1. Case 1: $c = c'$. Then $\boldsymbol{\tau}_{\mathsf{j}}(x) = \boldsymbol{\tau}'_{\mathsf{j}}(x)$ for all $x \in \mathcal{X} - \{c\}$. Hence,

   $$0 = r_{\boldsymbol{\pi}}(\boldsymbol{\tau}_{\mathsf{j}}) - r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}'_{\mathsf{j}}) = \boldsymbol{\pi}(x)(\boldsymbol{\tau}_{\mathsf{j}}(c) - \boldsymbol{\tau}'_{\mathsf{j}}(c))\,.$$

   This implies that either $\boldsymbol{\tau}_{\mathsf{j}}(c) = \boldsymbol{\tau}'_{\mathsf{j}}(c)$, and thus $\boldsymbol{\tau}_{\mathsf{j}}(x) = \boldsymbol{\tau}'_{\mathsf{j}}(x)$ for all $x \in \mathcal{X}$, or otherwise $\boldsymbol{\pi}(c) = 0$, in which case we still have $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} \boldsymbol{\tau}'_{\mathsf{j}}$ (since the two policies agree every outside the set $\{c\}$).

2. Case 2: $c \neq c'$. We assume assume without loss of generality that $c' < c \leq C$. Since the policies $\boldsymbol{\tau}_{c',1}$ and $\boldsymbol{\tau}_{c'+1,0}$ are identity for $c' < C$, we may also assume without loss of generality that $\gamma' \in [0, 1)$. Thus for all $x \in \mathcal{S} := \{c', c'+1, \ldots, C\}$, we have $\boldsymbol{\tau}'_{\mathsf{j}}(x) < \boldsymbol{\tau}_{\mathsf{j}}(x)$. This implies that

   $$\begin{aligned} 0 &= r_{\boldsymbol{\pi}}(\boldsymbol{\tau}_{\mathsf{j}}) - r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}'_{\mathsf{j}}) \\ &= \sum_{x \in \mathcal{S}} \boldsymbol{\pi}_{\mathsf{j}}(x)(\boldsymbol{\tau}_{\mathsf{j}}(x) - \boldsymbol{\tau}'_{\mathsf{j}}(x)) \\ &\geq \min_{x \in \mathcal{S}}(\boldsymbol{\tau}_{\mathsf{j}}(c) - \boldsymbol{\tau}'_{\mathsf{j}}(x)) \cdot \sum_{x \in \mathcal{S}} \boldsymbol{\pi}(x)\,. \end{aligned}$$

   Since $\min_{x \in \mathcal{S}}(\boldsymbol{\tau}_{\mathsf{j}}(c) - \boldsymbol{\tau}'_{\mathsf{j}}(x)) > 0$, it follows that $\sum_{x \in \mathcal{S}} \boldsymbol{\pi}_{\mathsf{j}}(x) = 0$, whence $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} \boldsymbol{\tau}'_{\mathsf{j}}$.

Next, we show that $r_{\boldsymbol{\pi}}$ is a bijection from $\mathcal{T}_{\text{thresh}}(\boldsymbol{\pi}) \to [0, 1]$. That $r_{\boldsymbol{\pi}}$ is injective follows immediately from the fact if $r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}) = r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}'_{\mathsf{j}})$, then $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} \boldsymbol{\tau}'_{\mathsf{j}}$. To show it is surjective, we

exhibit for every $\beta \in [0,1]$ a threshold policy $\boldsymbol{\tau}_{c,\gamma}$ for which $r_{\boldsymbol{\pi}_\mathsf{j}}(\boldsymbol{\tau}_{c,\gamma}) = \beta$. We may assume $\beta < 1$, since the all-ones policy has a selection rate of 1.

Recall the definition of the inverse CDF

$$\mathrm{Q}_\mathsf{j}(\beta) := \mathrm{argmax}\{c : \sum_{x=c}^{C} \boldsymbol{\pi}(x) > \beta\} .$$

Since $\beta < 1$, $\mathrm{Q}_\mathsf{j}(\beta) \leq C$. Let $\beta_+ = \sum_{x=\mathrm{Q}_\mathsf{j}(\beta)}^{C} \boldsymbol{\pi}(x)$, and let $\beta_- = \sum_{x=\mathrm{Q}_\mathsf{j}(\beta)+1}^{C} \boldsymbol{\pi}(x)$. Note that by definition, we have $\beta_- \leq \beta < \beta_+$, and $\beta_+ - \beta_- = \boldsymbol{\pi}(\mathrm{Q}_\mathsf{j}(\beta))$. Hence, if we define $\gamma = \frac{\beta-\beta_-}{\beta_+-\beta_-}$, we have

$$r_{\boldsymbol{\pi}_\mathsf{j}}(\boldsymbol{\tau}_{\mathrm{Q}_\mathsf{j}(\beta),\gamma}) = \boldsymbol{\pi}(\mathrm{Q}_\mathsf{j}(\beta))\gamma + \sum_{x=\mathrm{Q}_\mathsf{j}(\beta)+1}^{C} \boldsymbol{\pi}(x) = \beta_- + (\beta_+ - \beta_-)\gamma = \beta_- + \beta - \beta_- = \beta .$$

Proof of Lemma 2.5.4

Given $\boldsymbol{\tau} \in [0,1]^C$, we define the *normal cone* at $\boldsymbol{\tau}$ as $\mathrm{NC}(\boldsymbol{\tau}) := \mathrm{ConicalHull}\{\boldsymbol{z} : \boldsymbol{\tau} + \boldsymbol{z} \in [0,1]^C\}$. We can describe $\mathrm{NC}(\boldsymbol{\tau})$ explicitly as:

$$\mathrm{NC}(\boldsymbol{\tau}) := \{\boldsymbol{z} \in \mathbb{R}^C : \boldsymbol{z}_i \leq 0 \text{ if } \boldsymbol{\tau}_i = 0, \boldsymbol{z}_i \geq 0 \text{ if } \boldsymbol{\tau}_i = 1\} .$$

Immediately from the above definition, we have the following useful identity, which is that for any vector $\boldsymbol{g} \in \mathbb{R}^C$,

$$\langle \boldsymbol{g}, \boldsymbol{z} \rangle \leq 0 \ \forall \boldsymbol{z} \in \mathrm{NC}(\boldsymbol{\tau}), \quad \text{if and only if} \quad \forall x \in \mathcal{X}, \begin{cases} \boldsymbol{\tau}(x) = 0 & \boldsymbol{g}(x) < 0 \\ \boldsymbol{\tau}(x) = 1 & \boldsymbol{g}(x) > 0 \\ \boldsymbol{\tau}(x) \in [0,1] & \boldsymbol{g}(x) = 0 \end{cases} . \tag{2.21}$$

Now consider the optimization problem (2.11). By the first order KKT conditions, we know that for any optimizer $\boldsymbol{\tau}_*$ of the above objective, there exists some $\widehat{\lambda} \in \mathbb{R}$ such that, for all $\boldsymbol{z} \in \mathrm{NC}(\boldsymbol{\tau}_*)$

$$\langle \boldsymbol{z}, \boldsymbol{v} \circ \boldsymbol{\pi} + \widehat{\lambda} \boldsymbol{\pi} \circ \boldsymbol{w} \rangle \leq 0 .$$

By (2.21), we must have that

$$\boldsymbol{\tau}_*(x) = \begin{cases} 0 & \boldsymbol{\pi}(x)(\boldsymbol{v}(x) + \widehat{\lambda}\boldsymbol{w}(x)) < 0 \\ 1 & \boldsymbol{\pi}(x)(\boldsymbol{v}(x) + \widehat{\lambda}\boldsymbol{w}(x)) > 0 \\ \in [0,1] & \boldsymbol{\pi}(x)(\boldsymbol{v}(x) + \widehat{\lambda}\boldsymbol{w}(x)) = 0 \end{cases} .$$

Now $\boldsymbol{\tau}_*(x)$ is not necessarily a threshold policy. To conclude the theorem, it suffices to exhibit a threshold policy $\widetilde{\boldsymbol{\tau}}_*$ such that $\boldsymbol{\tau}_*(x) \cong_{\boldsymbol{\pi}} \widetilde{\boldsymbol{\tau}}_*$. (Note that $\widetilde{\boldsymbol{\tau}}_*(x)$ will also be feasible for the constraint, and have the same objective value; hence $\widetilde{\boldsymbol{\tau}}_*$ will be optimal as well.)

Given $\boldsymbol{\tau}_*$ and $\widehat{\lambda}$, let $c_* = \min\{c \in \mathcal{X} : \boldsymbol{v}(x) + \widehat{\lambda}\boldsymbol{w}(x) \geq 0\}$. If either (a) $\boldsymbol{w}(x) = 0$ for all $x \in \mathcal{X}$ and $\boldsymbol{v}(x)$ is strictly increasing or (b) $\boldsymbol{v}(x)/\boldsymbol{w}(x)$ is strictly increasing, then the modified policy

$$\widetilde{\boldsymbol{\tau}}_*(x) = \begin{cases} 0 & x < c_* \\ \boldsymbol{\tau}_*(x) & x = c_* \\ 1 & x > c_* \end{cases},$$

is a threshold policy, and $\boldsymbol{\tau}_*(x) \cong_{\boldsymbol{\pi}} \widetilde{\boldsymbol{\tau}}_*$. Moreover, $\langle \boldsymbol{w}, \widetilde{\boldsymbol{\tau}}_* \rangle = \langle \boldsymbol{w}, \widetilde{\boldsymbol{\tau}}_* \rangle$ and $\langle \boldsymbol{\pi}, \widetilde{\boldsymbol{\tau}}_* \rangle = \langle \boldsymbol{\pi}, \widetilde{\boldsymbol{\tau}}_* \rangle$, which implies that $\widetilde{\boldsymbol{\tau}}_*$ is an optimal policy for the objective in Lemma 2.5.4.

Proof of Lemma 2.5.5

We shall prove

$$\partial_+ \left(\boldsymbol{\pi}_{\mathsf{j}} \circ r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(\beta) \right) = \boldsymbol{e}_{\mathrm{Q}_{\mathsf{j}}(\beta)} \,, \tag{2.22}$$

where the derivative is with respect to β. The computation of the left-derivative is analogous. Since we are concerned with right-derivatives, we shall take $\beta \in [0, 1)$. Since $\boldsymbol{\pi}_{\mathsf{j}} \circ r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(\beta)$ does not depend on the choice of representative for $r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}$, we can choose a canonical representation for $r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}$. In Section 2.9, we saw that the threshold policy $\boldsymbol{\tau}_{\mathrm{Q}_{\mathsf{j}}(\beta),\gamma(\beta)}$ had acceptance rate β, where we had defined

$$\beta_+ = \sum_{x=\mathrm{Q}_{\mathsf{j}}(\beta)}^{C} \boldsymbol{\pi}(x) \quad \text{and} \quad \beta_- = \sum_{x=\mathrm{Q}_{\mathsf{j}}(\beta)+1}^{C} \boldsymbol{\pi}(x) \,, \tag{2.23}$$

$$\gamma(\beta) = \frac{\beta - \beta_-}{\beta_+ - \beta_-} \,. \tag{2.24}$$

Note then that for each x, $\boldsymbol{\tau}_{\mathrm{Q}_{\mathsf{j}}(\beta),\gamma(\beta)}(x)$ is piece-wise linear, and thus admits left and right derivatives. We first claim that

$$\forall x \in \mathcal{X} \setminus \{\mathrm{Q}_{\mathsf{j}}(\beta)\}, \quad \partial_+ \boldsymbol{\tau}_{\mathrm{Q}_{\mathsf{j}}(\beta),\gamma(\beta)}(x) = 0 \,. \tag{2.25}$$

To see this, note that $\mathrm{Q}_{\mathsf{j}}(\beta)$ is right continuous, so for all ϵ sufficiently small, $\mathrm{Q}_{\mathsf{j}}(\beta + \epsilon) = \mathrm{Q}_{\mathsf{j}}(\beta)$. Hence, for all ϵ sufficiently small and all $x \neq \mathrm{Q}(\beta)$, we have $\boldsymbol{\tau}_{\mathrm{Q}_{\mathsf{j}}(\beta+\epsilon),\gamma(\beta+\epsilon)}(x) = \boldsymbol{\tau}_{\mathrm{Q}_{\mathsf{j}}(\beta+\epsilon),\gamma(\beta+\epsilon)}(x)$, as needed. Thus, Equation (2.25) implies that $\partial_+ \boldsymbol{\pi}_{\mathsf{j}} \circ r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(\beta)$ is supported on $x = \mathrm{Q}_{\mathsf{j}}(\beta)$, and hence

$$\partial_+ \boldsymbol{\pi}_{\mathsf{j}} \circ r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(\beta) = \partial_+ \boldsymbol{\pi}_{\mathsf{j}}(x) \boldsymbol{\tau}_{\mathrm{Q}_{\mathsf{j}}(\beta),\gamma(\beta)}(x)\big|_{x=\mathrm{Q}_{\mathsf{j}}(\beta)} \cdot \boldsymbol{e}_{\mathrm{Q}_{\mathsf{j}}(\beta)} \,.$$

To conclude, we must show that $\partial_+ \boldsymbol{\pi}_{\mathrm{j}}(x) \boldsymbol{\tau}_{\mathrm{Q_j}(\beta),\gamma(\beta)}(x)\big|_{x=\mathrm{Q_j}(\beta)} = 1$. To show this, we have

$$\begin{aligned} 1 &= \partial_+(\beta) \\ &= \partial_+(r_{\boldsymbol{\pi}_{\mathrm{j}}}(\boldsymbol{\tau}_{\mathrm{Q_j}(\beta),\gamma(\beta)})) \quad \text{since} \qquad r_{\boldsymbol{\pi}_{\mathrm{j}}}(\boldsymbol{\tau}_{\mathrm{Q_j}(\beta),\gamma(\beta)}) = \beta \ \forall \beta \in [0,1) \\ &= \partial_+ \left(\sum_{x \in \mathcal{X}} \boldsymbol{\pi}(x) \cdot \boldsymbol{\tau}_{\mathrm{Q_j}(\beta),\gamma(\beta)}(x) \right) \\ &= \partial_+ \boldsymbol{\pi}(x) \cdot \boldsymbol{\tau}_{\mathrm{Q_j}(\beta),\gamma(\beta)}(x)\big|_{x=\mathrm{Q_j}(\beta)} \ , \text{ as needed.} \end{aligned}$$

Characterization of Fairness Solutions

Derivative Computation for `EqOpt`

In this section, we prove Lemma 2.6, which we recall below. Suppose that $\boldsymbol{w}(x) > 0$ for all x. Then the function

$$T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}(\beta) := \langle r^{-1}_{\boldsymbol{\pi}_{\mathrm{j}}}(\beta), \boldsymbol{\pi}_{\mathrm{j}} \circ \boldsymbol{w}_{\mathrm{j}} \rangle$$

is a bijection from $[0,1]$ to $[0, \langle \boldsymbol{\pi}_{\mathrm{j}}, \boldsymbol{w} \rangle]$. We will prove Lemma 2.6 in tandem with the following derivative computation which we applied in the proof of Theorem 2.6.2.

Lemma 2.9.1. *The function*

$$\mathcal{U}_{\mathrm{j}}(t; \boldsymbol{w}_{\mathrm{j}}) := \mathcal{U}_{\mathrm{j}} \left(r^{-1}_{\boldsymbol{\pi}_{\mathrm{j}}} \left(T^{-1}_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}(t) \right) \right)$$

is concave in t and has left and right derivatives

$$\partial_+ \mathcal{U}_{\mathrm{j}}(t; \boldsymbol{w}_{\mathrm{j}}) = \frac{\boldsymbol{u}(\mathrm{Q_j}(T^{-1}_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}(t)))}{\boldsymbol{w}_{\mathrm{j}}(\mathrm{Q_j}(T^{-1}_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}(t)))} \quad \textit{and} \quad \partial_- \mathcal{U}_{\mathrm{j}}(t; \boldsymbol{w}_{\mathrm{j}}) = \frac{\boldsymbol{u}(\mathrm{Q}^+_{\mathrm{j}}(T^{-1}_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}(t)))}{\boldsymbol{w}_{\mathrm{j}}(\mathrm{Q}^+_{\mathrm{j}}(T^{-1}_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}(t)))} \ .$$

Proof of Lemmas 2.6 and 2.9.1. Consider a $\beta \in [0,1]$. Then, $\boldsymbol{\pi}_{\mathrm{j}} \circ r^{-1}_{\boldsymbol{\pi}_{\mathrm{j}}}(\beta)$ is continuous and left and right differentiable by Lemma 2.5.5, and its left and right derivatives are indicator vectors $\boldsymbol{e}_{\mathrm{Q_j}(\beta)}$ and $\boldsymbol{e}_{\mathrm{Q}^+_{\mathrm{j}}(\beta)}$, respectively. Consequently, $\beta \mapsto \langle \boldsymbol{w}_{\mathrm{j}}, \boldsymbol{\pi}_{\mathrm{j}} \circ r^{-1}_{\boldsymbol{\pi}_{\mathrm{j}}}(\beta) \rangle$ has left and right derivatives $\boldsymbol{w}_{\mathrm{j}}(\mathrm{Q}(\beta))$ and $\boldsymbol{w}_{\mathrm{j}}(\mathrm{Q}^+(\beta))$, respectively; both of which are both strictly positive by the assumption $\boldsymbol{w}(x) > 0$. Hence, $T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}(\beta) = \langle \boldsymbol{w}_{\mathrm{j}}, \boldsymbol{\pi}_{\mathrm{j}} \circ r^{-1}_{\boldsymbol{\pi}_{\mathrm{j}}}(\beta) \rangle$ is strictly increasing in β, and so the map is injective. It is also surjective because $\beta = 0$ induces the policy $\boldsymbol{\tau}_{\mathrm{j}} = \mathbf{0}$ and $\beta = 1$ induces the policy $\boldsymbol{\tau}_{\mathrm{j}} = \mathbf{1}$ (up to $\boldsymbol{\pi}_{\mathrm{j}}$-measure zero). Hence, $T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}(\beta)$ is an order preserving bijection with left- and right-derivatives, and we can compute the left and right derivatives of its inverse as follows:

$$\partial_+ T^{-1}_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}(t) = \frac{1}{\partial_+ T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}(\beta)\big|_{\beta = T^{-1}_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}(t)}} = \frac{1}{\boldsymbol{w}_{\mathrm{j}}(\mathrm{Q_j}(T^{-1}_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}(t)))} \ ,$$

and similarly, $\partial_- T^{-1}_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(t) = \frac{1}{\boldsymbol{w}_\mathsf{j}(\mathrm{Q}^+(T^{-1}_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(t)))}$. Then we can compute that

$$\begin{aligned}\partial_+ \mathcal{U}_\mathsf{j}(r_{\boldsymbol{\pi}_\mathsf{j}}(T^{-1}_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(t))) &= \partial_+ \mathcal{U}(r_{\boldsymbol{\pi}_\mathsf{j}}(\beta))\big|_{\beta = T^{-1}_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(t))} \cdot \partial_+ T_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(\sup(t)) \\ &= \frac{\boldsymbol{u}(\mathrm{Q}_\mathsf{j}(T^{-1}_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(t)))}{\boldsymbol{w}_\mathsf{j}(\mathrm{Q}_\mathsf{j}(T^{-1}_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(t)))}\,.\end{aligned}$$

and similarly $\partial_- \mathcal{U}_\mathsf{j}(r_{\boldsymbol{\pi}_\mathsf{j}}(T_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(t))) = \frac{\mathcal{U}(\mathrm{Q}^+_\mathsf{j}(T^{-1}_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(t)))}{\boldsymbol{w}_\mathsf{j}(\mathrm{Q}^+_\mathsf{j}(T^{-1}_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(t)))}$. One can verify that for all $t_1 < t_2$, one has that $\partial_+ \mathcal{U}_\mathsf{j}(r_{\boldsymbol{\pi}_\mathsf{j}}(T^{-1}_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(t_1))) \geq \partial_- \mathcal{U}_\mathsf{j}(r_{\boldsymbol{\pi}_\mathsf{j}}(T^{-1}_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(t_2)))$, and that for all t, $\partial_+ \mathcal{U}_\mathsf{j}(r_{\boldsymbol{\pi}_\mathsf{j}}(T^{-1}_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(t))) \leq \partial_- \mathcal{U}_\mathsf{j}(r_{\boldsymbol{\pi}_\mathsf{j}}(T^{-1}_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(t)))$. These facts establish that the mapping $t \mapsto \mathcal{U}_\mathsf{j}(r_{\boldsymbol{\pi}_\mathsf{j}}(T^{-1}_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(t)))$ is concave. □

Proofs of Main Results

We remark that the proofs in this section rely crucially on the characterizations of the optimal fairness-constrained policies developed in Section 2.6. We first define the notion of CDF domination, which is referred to in a few of the proofs. Intuitively, it means that for any score, the fraction of group B above this is higher than that for group A. It is realistic to assume this if we keep with our convention that group A is the disadvantaged group relative to group B.

Definition 2.9.1 (CDF domination)**.** $\boldsymbol{\pi}_\mathsf{A}$ *is said to be* dominated by $\boldsymbol{\pi}_\mathsf{B}$ *if* $\forall a \geq 1, \sum_{x>a} \boldsymbol{\pi}_\mathsf{A} < \sum_{x>a} \boldsymbol{\pi}_\mathsf{B}$*. We denote this as* $\boldsymbol{\pi}_\mathsf{A} \prec \boldsymbol{\pi}_\mathsf{B}$*.*

Frequently, we shall use the following lemma:

Lemma 2.9.2. *Suppose that* $\boldsymbol{\pi}_\mathsf{A} \prec \boldsymbol{\pi}_\mathsf{B}$*. Then, for all* $\beta > 0$*, it holds that* $\mathrm{Q}_\mathsf{A}(\beta) \leq \mathrm{Q}_\mathsf{B}(\beta)$ *and* $\boldsymbol{u}(\mathrm{Q}_\mathsf{A}(\beta)) \leq \boldsymbol{u}(\mathrm{Q}_\mathsf{A}(\beta))$

Proof. The fact that $\mathrm{Q}_\mathsf{A}(\beta) \leq \mathrm{Q}_\mathsf{B}(\beta)$ follows directly from the definition of monotonicty of $\boldsymbol{u}$ implies that $\boldsymbol{u}(\mathrm{Q}_\mathsf{A}(\beta)) \leq \boldsymbol{u}(\mathrm{Q}_\mathsf{B}(\beta))$. □

Proof of Proposition 2.3.1

The `MaxUtil` policy for group j solves the optimization

$$\max_{\boldsymbol{\tau}_\mathsf{j} \in [0,1]^C} \mathcal{U}_\mathsf{j}(\boldsymbol{\tau}_\mathsf{j}) = \max_{\beta_\mathsf{j} \in [0,1]} \mathcal{U}_\mathsf{j}(r^{-1}_{\boldsymbol{\pi}_\mathsf{j}}(\beta_\mathsf{j}))\,.$$

Computing left and right derivatives of this objective yields

$$\partial_+ \mathcal{U}_\mathsf{j}(r^{-1}_{\boldsymbol{\pi}_\mathsf{j}}(\beta_\mathsf{j})) = \boldsymbol{u}(\mathrm{Q}_\mathsf{j}(\beta)), \qquad \partial_- \mathcal{U}_\mathsf{j}(r^{-1}_{\boldsymbol{\pi}_\mathsf{j}}(\beta_\mathsf{j})) = \boldsymbol{u}(\mathrm{Q}^+_\mathsf{j}(\beta))\,.$$

By concavity, solutions β^* satisfy

$$\begin{aligned} \beta < \beta^* \quad &\text{if} \quad \boldsymbol{u}(\mathrm{Q}_{\mathsf{j}}(\beta)) > 0\,, \\ \beta > \beta^* \quad &\text{if} \quad \boldsymbol{u}(\mathrm{Q}^+_{\mathsf{j}}(\beta)) < 0\,. \end{aligned} \tag{2.26}$$

Therefore, we conclude that the `MaxUtil` policy loans only to scores x s.t. $\boldsymbol{u}(x) > 0$, which implies $\boldsymbol{\Delta}(x) > 0$ for all scores loaned to. Therefore we must have that $0 \leq \Delta\boldsymbol{\mu}^{\texttt{MaxUtil}}$. By definition $\Delta\boldsymbol{\mu}^{\texttt{MaxUtil}} \leq \Delta\boldsymbol{\mu}^*$.

Proof of Corollary 2.3.2

We begin with proving part (a), which gives conditions under which `DemParity` cases relative improvement. Recall that $\overline{\beta}$ is the largest selection rate for which $\mathcal{U}(\overline{\beta}) = \mathcal{U}(\beta_{\mathsf{A}}^{\texttt{MaxUtil}})$. First, we derive a condition which bounds the selection rate $\beta_{\mathsf{A}}^{\texttt{DemParity}}$ from below. Fix an acceptance rate β such that $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta < \min\{\beta_{\mathsf{B}}^{\texttt{MaxUtil}}, \overline{\beta}\}$. By Theorem 2.6.1, we have that `DemParity` selects to group A with rate higher than β as long as

$$g_{\mathsf{A}} \leq g_1 := \frac{1}{1 - \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta))}{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta))}}\,.$$

By (2.26) and the monotonicity of $\boldsymbol{u}$, $\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta)) < 0$ and $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta)) > 0$, so $0 < g_1 < 1$.

Next, we derive a condition which bounds the selection rate $\beta_{\mathsf{A}}^{\texttt{DemParity}}$ from above. First, consider the case that $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} < \overline{\beta}$, and fix β' such that $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} < \beta' < \overline{\beta}$. Then `DemParity` selects group A at a rate $\beta_{\mathsf{A}} < \beta'$ for any proportion g_{A}. This follows from applying Theorem 2.6.1 since we have that $\boldsymbol{u}(\mathrm{Q}^+_{\mathsf{A}}(\beta')) < 0$ and $\boldsymbol{u}(\mathrm{Q}^+_{\mathsf{B}}(\beta')) < 0$ by (2.26) and the monotonicity of $\boldsymbol{u}$.

Instead, in the case that $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > \overline{\beta}$, fix β' such that $\overline{\beta} < \beta' < \beta_{\mathsf{B}}^{\texttt{MaxUtil}}$. Then `DemParity` selects group A at a rate less than β' as long as

$$g_{\mathsf{A}} \geq g_0 := \frac{1}{1 - \frac{\boldsymbol{u}(\mathrm{Q}^+_{\mathsf{A}}(\beta'))}{\boldsymbol{u}(\mathrm{Q}^+_{\mathsf{B}}(\beta'))}}\,.$$

By (2.26) and the monotonicity of $\boldsymbol{u}$, $0 < g_0 < g_1$. Thus for $g_{\mathsf{A}} \in [g_0, g_1]$, the `DemParity` selection rate for group A is bounded between β and β', and thus `DemParity` results in relative improvement.

Next, we prove part (b), which gives conditions under which `EqOpt` cases relative improvement. First, we derive a condition which bounds the selection rate $\beta_{\mathsf{A}}^{\texttt{EqOpt}}$ from below. Fix an acceptance rate β such that $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta$ and $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > G^{(\mathsf{A}\to\mathsf{B})}(\beta)$. By Theorem 2.6.2, `EqOpt` selects group A at a rate higher than β as long as

$$g_{\mathsf{A}} > g_3 := \frac{1}{1 - \frac{1}{\kappa} \cdot \frac{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))}{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))} \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{A}}(\beta))}}\,.$$

By (2.26) and the monotonicity of $\boldsymbol{u}$, $\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta)) < 0$ and $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta))) > 0$, so $g_3 > 0$.

Next, we derive a condition which bounds the selection rate $\beta_{\mathsf{A}}^{\texttt{EqOpt}}$ from above. First, consider the case that there exists β' such that $\beta' < \overline{\beta}$ and $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} < G^{(\mathsf{A}\to\mathsf{B})}(\beta')$. Then `EqOpt` selects group A at a rate less than this β' for any g_{A}. This follows from Theorem 2.6.2 since we have that $\boldsymbol{u}(\mathrm{Q}^+_{\mathsf{A}}(\beta')) < 0$ and $\boldsymbol{u}(\mathrm{Q}^+_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta'))) < 0$ by (2.26) and the monotonicity of $\boldsymbol{u}$.

In the other case, fix β' such that $\beta < \beta' < \overline{\beta}$ and $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > G^{(\mathsf{A}\to\mathsf{B})}(\beta')$. By Theorem 2.6.2, `EqOpt` selects group A at a rate lower than β' as long as

$$g_{\mathsf{A}} > g_2 := \frac{1}{1 - \frac{1}{\kappa} \cdot \frac{\boldsymbol{\rho}(\mathrm{Q}^+_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta')))}{\boldsymbol{u}(\mathrm{Q}^+_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta')))} \frac{\boldsymbol{u}(\mathrm{Q}^+_{\mathsf{A}}(\beta'))}{\boldsymbol{\rho}(\mathrm{Q}^+_{\mathsf{A}}(\beta'))}} \,.$$

By (2.26) and the monotonicity of $\boldsymbol{u}$, $0 < g_2 < g_3$. Thus for $g_{\mathsf{A}} \in [g_2, g_3]$, the `EqOpt` selection rate for group A is bounded between β and β', and thus `EqOpt` results in relative improvement.

Proof of Corollary 2.3.3

Recall our assumption that $\beta > \beta_{\mathsf{A}}^{\texttt{MaxUtil}}$ and $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > \beta$. As argued in the above proof of Corollary 2.3.2, by (2.26) and the monotonicity of $\boldsymbol{u}$, $\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta)) < 0$ and $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta)) > 0$. Applying Theorem 2.6.1, `DemParity` selects at a higher rate than β for any population proportion $g_{\mathsf{A}} \le g_0$, where $g_0 = 1/(1 - \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta))}{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta))}) \in (0, 1)$. In particular, if $\beta = \beta_0$, which we defined as the harm threshold (i.e. $\Delta\boldsymbol{\mu}_{\mathsf{A}}(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta_0)) = 0$ and $\Delta\boldsymbol{\mu}_{\mathsf{A}}$ is decreasing at β_0), then by the concavity of $\Delta\boldsymbol{\mu}_{\mathsf{A}}$, we have that $\Delta\boldsymbol{\mu}_{\mathsf{A}}(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta_{\mathsf{A}}^{\texttt{DemParity}})) < 0$, that is, `DemParity` causes active harm.

Proof of Corollary 2.3.4

By Theorem 2.6.2, `EqOpt` selects at a higher rate than β for any population proportion $g_{\mathsf{A}} \le g_0$, where $g_0 = 1/(1 - \frac{1}{\kappa} \cdot \frac{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))}{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))} \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{A}}(\beta))})$. Using our assumptions $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > G^{(\mathsf{A}\to\mathsf{B})}(\beta)$ and $\beta > \beta_{\mathsf{A}}^{\texttt{MaxUtil}}$, we have that $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta))) > 0$ and $\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta)) < 0$, by (2.26) and the monotonicity of $\boldsymbol{u}$. This verifies that $g_0 \in (0, 1)$. In particular, if $\beta = \beta_0$, then by the concavity of $\Delta\boldsymbol{\mu}_{\mathsf{A}}$, we have that $\Delta\boldsymbol{\mu}_{\mathsf{A}}(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta_{\mathsf{A}}^{\texttt{EqOpt}})) < 0$, that is, `EqOpt` causes active harm.

Proof of Corollary 2.3.5

Applying Theorem 2.6.1, we have

$$-\frac{1 - g_{\mathsf{A}}}{g_{\mathsf{A}}} \boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta)) < \boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta)) \implies \beta_{\texttt{DemParity}} > \beta \,.$$

Applying Theorem 2.6.2, we have:

$$\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))\cdot\frac{\langle\boldsymbol{\rho},\boldsymbol{\pi}_{\mathsf{B}}\rangle}{\langle\boldsymbol{\rho},\boldsymbol{\pi}_{\mathsf{A}}\rangle}\cdot\frac{\boldsymbol{\rho}(\mathrm{Q}^{+}_{\mathsf{A}}(\beta))}{\boldsymbol{\rho}(\mathrm{Q}^{+}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))}<-\frac{1-g_{\mathsf{A}}}{g_{\mathsf{A}}}\boldsymbol{u}(\mathrm{Q}^{+}_{\mathsf{A}}(\beta))\implies\beta_{\texttt{EqOpt}}<\beta\,.$$

By Corollaries 2.3.3 and 2.3.4, choosing $g_{\mathsf{A}}<g_2:=1/(1-\frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta))}{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta))})$ and $g_{\mathsf{A}}>g_1:=1/(1-\frac{1}{\kappa}\cdot\frac{\boldsymbol{\rho}(\mathrm{Q}^{+}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))}{\boldsymbol{u}(\mathrm{Q}^{+}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))}\frac{\boldsymbol{u}(\mathrm{Q}^{+}_{\mathsf{A}}(\beta))}{\boldsymbol{\rho}(\mathrm{Q}^{+}_{\mathsf{A}}(\beta))})$ satisfies the above.

It remains to check that $g_1<g_2$. Since we assumed $\beta>\sum_{x>\boldsymbol{\mu}_{\mathsf{A}}}\boldsymbol{\pi}_{\mathsf{A}}$, we may apply Lemma 2.9.3 to verify this.

Thus we indeed have sufficient conditions for $\beta_{\texttt{DemParity}}>\beta>\beta_{\texttt{EqOpt}}$. In particular, if $\beta=\beta_0$, then by the concavity of $\Delta\boldsymbol{\mu}_{\mathsf{A}}$, we have that $\Delta\boldsymbol{\mu}_{\mathsf{A}}(r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(\beta^{\texttt{EqOpt}}_{\mathsf{A}}))>0$, that is, `EqOpt` causes improvement, and $\Delta\boldsymbol{\mu}_{\mathsf{A}}(r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(\beta^{\texttt{DemParity}}_{\mathsf{A}}))<0$, that is, `DemParity` causes active harm.

Lastly, because $\beta_{\texttt{DemParity}}>\beta_{\texttt{EqOpt}}$, it is always true that $\Delta\boldsymbol{\mu}_{\mathsf{A}}(r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(\beta^{\texttt{DemParity}}_{\mathsf{A}}))>0\implies\Delta\boldsymbol{\mu}_{\mathsf{A}}(r^{-1}_{\boldsymbol{\pi}_{\mathsf{A}}}(\beta^{\texttt{EqOpt}}_{\mathsf{A}}))>0$, using the concavity of the outcome curve.

Lemma 2.9.3 (Comparison of `DemParity` and `EqOpt` selection rates)**.** *Fix $\beta\in[0,1]$. Suppose $\boldsymbol{\pi}_{\mathsf{A}},\boldsymbol{\pi}_{\mathsf{B}}$ are identical up to a translation with $\boldsymbol{\mu}_{\mathsf{A}}<\boldsymbol{\mu}_{\mathsf{B}}$. Also assume $\boldsymbol{\rho}(x)$ is affine in x. Denote $\kappa=\frac{\langle\boldsymbol{\rho},\boldsymbol{\pi}_{\mathsf{B}}\rangle}{\langle\boldsymbol{\rho},\boldsymbol{\pi}_{\mathsf{A}}\rangle}$. Then,*

$$\beta>\sum_{x>\boldsymbol{\mu}_{\mathsf{A}}}\boldsymbol{\pi}_{\mathsf{A}}$$

implies $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))\cdot\kappa\cdot\frac{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{A}}(\beta))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))}<\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta))$.

Proof. If we have $\beta>\sum_{x>\boldsymbol{\mu}_{\mathsf{A}}}\boldsymbol{\pi}_{\mathsf{A}}$, by lemma 2.9.4, we must also have $\frac{\boldsymbol{\mu}_{\mathsf{B}}}{\boldsymbol{\mu}_{\mathsf{A}}}<\frac{\mathrm{Q}_{\mathsf{B}}(\beta_0)}{\mathrm{Q}_{\mathsf{A}}(\beta_0)}$. This implies $\kappa=\frac{\sum_x\boldsymbol{\pi}_{\mathsf{B}}(x)\boldsymbol{\rho}(x)}{\sum_x\boldsymbol{\pi}_{\mathsf{A}}(x)\boldsymbol{\rho}(x)}<\frac{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}(\beta))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{A}}(\beta_0))}$ by linearity of expectation and linearity of $\boldsymbol{\rho}$. Therefore,

$$\kappa\cdot\frac{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{A}}(\beta))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}(\beta_0))}<1\tag{2.27}$$

Further, using $G^{(\mathsf{A}\to\mathsf{B})}(\beta)>\beta$ from lemma 2.9.4 and the fact that $\frac{\boldsymbol{u}(x)}{\boldsymbol{\rho}(x)}$ is increasing in x, we have $\frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))}<\frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}(\beta))}$. Therefore, $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))\cdot\kappa\cdot\frac{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{A}}(\beta_0))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta_0)))}<\kappa\cdot\frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}(\beta))}\cdot\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{A}}(\beta))<\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta))$ where the last inequality follows from (2.27).

□

We use the following technical lemma in the proof of the above lemma.

Lemma 2.9.4. *If $\boldsymbol{\pi}_{\mathsf{A}},\boldsymbol{\pi}_{\mathsf{B}}$ that are identical up to a translation with $\boldsymbol{\mu}_{\mathsf{A}}<\boldsymbol{\mu}_{\mathsf{B}}$, then*

$$G(\beta)>\beta\ \ \forall\ \beta\,,\tag{2.28}$$

$$\beta>\sum_{x>\boldsymbol{\mu}}\boldsymbol{\pi}_{\mathsf{A}}\implies\frac{\boldsymbol{\mu}_{\mathsf{B}}}{\boldsymbol{\mu}_{\mathsf{A}}}<\frac{\mathrm{Q}_{\mathsf{B}}(\beta)}{\mathrm{Q}_{\mathsf{A}}(\beta)}\,.\tag{2.29}$$

Proof. For (2.28), observe that $\mathrm{TPR}_{\mathsf{A}} = \boldsymbol{\rho}(\boldsymbol{\mu}_{\mathsf{A}}) < \mathrm{TPR}_{\mathsf{B}} = \boldsymbol{\rho}(\boldsymbol{\mu}_{\mathsf{B}})$. For any β, we can write $\mathrm{Q}_{\mathsf{B}}(\beta) = \boldsymbol{\mu}_{\mathsf{B}} + c$ and $\mathrm{Q}_{\mathsf{A}}(\beta) = \boldsymbol{\mu}_{\mathsf{A}} + c$ for some c, since $\boldsymbol{\pi}_{\mathsf{A}}, \boldsymbol{\pi}_{\mathsf{B}}$ that are identical up to translation by $\boldsymbol{\mu}_{\mathsf{A}} - \boldsymbol{\mu}_{\mathsf{B}}$. Thus, by computation, we can see that for $\mathrm{Q}(\beta) < \boldsymbol{\mu}$, $\partial_+ G^{(\mathsf{A}\to\mathsf{B})}(\beta) > 1$ and for $\mathrm{Q}(\beta) < \boldsymbol{\mu}$, $\partial_+ G^{(\mathsf{A}\to\mathsf{B})}(\beta) < 1$. Since $G^{(\mathsf{A}\to\mathsf{B})}$ is monotonically increasing on $[0,1]$, we must have $G^{(\mathsf{A}\to\mathsf{B})}(\beta) > \beta$ for every $\beta \in [0,1]$.

For (2.29), we have $\beta > \sum_{x>\boldsymbol{\mu}} \boldsymbol{\pi}_{\mathsf{A}}$, we can again write $\mathrm{Q}_{\mathsf{B}}(\beta) = \boldsymbol{\mu}_{\mathsf{B}} - c$ and $\mathrm{Q}_{\mathsf{A}}(\beta) = \boldsymbol{\mu}_{\mathsf{A}} - c$, for some $c > 0$. Then it is clear than we have $\frac{\boldsymbol{\mu}_{\mathsf{B}}}{\boldsymbol{\mu}_{\mathsf{A}}} < \frac{\mathrm{Q}_{\mathsf{B}}(\beta)}{\mathrm{Q}_{\mathsf{A}}(\beta)}$. □

Proof of Corollary 2.3.6

Proof. $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta_{\mathsf{B}}^{\texttt{MaxUtil}}$ implies $g_{\mathsf{A}} \cdot \boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta_{\mathsf{A}}^{\texttt{MaxUtil}})) + g_{\mathsf{B}} \cdot \boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta_{\mathsf{A}}^{\texttt{MaxUtil}})) > 0$, which by Theorem 2.6.1, implies $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta_{\mathsf{A}}^{\texttt{DemParity}}$.

$\mathrm{TPR}_{\mathsf{A}}(\boldsymbol{\tau}^{\texttt{MaxUtil}}) > \mathrm{TPR}_{\mathsf{B}}(\boldsymbol{\tau}^{\texttt{MaxUtil}})$ implies $G^{(\mathsf{A}\to\mathsf{B})}(\beta_{\mathsf{A}}^{\texttt{MaxUtil}}) > \beta_{\mathsf{B}}^{\texttt{MaxUtil}}$ and so $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta_{\mathsf{A}}^{\texttt{MaxUtil}}))) < 0$. Therefore by Theorem 2.6.2, we have that $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} > \beta_{\mathsf{A}}^{\texttt{EqOpt}}$. □

We now give a very simple example of $\boldsymbol{\pi}_{\mathsf{A}} \prec \boldsymbol{\pi}_{\mathsf{B}}$ where Theorem 3.5 holds. The construction of the example exemplifies the more general idea of using large in-group inequality in group A to skew the true positive rate at `MaxUtil`, making $\mathrm{TPR}_{\mathsf{A}}(\boldsymbol{\tau}^{\texttt{MaxUtil}}) > \mathrm{TPR}_{\mathsf{B}}(\boldsymbol{\tau}^{\texttt{MaxUtil}})$.

Example 2.9.1 (`EqOpt` causes relative harm)**.** *Let $C = 6$, and let the utility function be such that $\boldsymbol{u}(4) = 0$. Suppose $\boldsymbol{\pi}_{\mathsf{A}}(5) = 1 - 2\epsilon, \boldsymbol{\pi}_{\mathsf{A}}(1) = 2\epsilon$ and $\boldsymbol{\pi}_{\mathsf{B}}(5) = 1 - \epsilon, \boldsymbol{\pi}_{\mathsf{B}}(3) = \epsilon$.*

We can easily check that $\boldsymbol{\pi}_{\mathsf{A}} \prec \boldsymbol{\pi}_{\mathsf{B}}$. However, for any $\epsilon \in (0, 1/4)$, we have that $\mathrm{TPR}_{\mathsf{B}}(\boldsymbol{\tau}^{\texttt{MaxUtil}}) = \frac{5(1-\epsilon)}{5(1-\epsilon)+3\epsilon} < \mathrm{TPR}_{\mathsf{A}}(\boldsymbol{\tau}^{\texttt{MaxUtil}}) = \frac{5(1-2\epsilon)}{5(1-2\epsilon)+2\epsilon}$.

Proof of Proposition 2.4.1

Denote the upper quantile function under $\widehat{\boldsymbol{\pi}}$ as $\widehat{\mathrm{Q}}$. Since $\widehat{\boldsymbol{\pi}} \prec \boldsymbol{\pi}$, we have $\widehat{\mathrm{Q}}(\beta) \le \mathrm{Q}(\beta)$. The conclusion follows for `MaxUtil` and `DemParity` from Theorem 2.6.1 by the monotonicity of $\boldsymbol{u}$.

If we have that $\mathrm{TPR}_{\mathsf{A}}(\boldsymbol{\tau}) > \widehat{\mathrm{TPR}_{\mathsf{A}}}(\boldsymbol{\tau})\ \forall\ \tau$, that is, the true TPR dominates estimated TPR, the conclusion for `EqOpt` follows from Theorem 2.6.2, by the same argument as in the proof of Corollary 2.3.6.

Proof of Proposition 2.4.2

By Proposition 2.5.6, $\beta^* = \mathrm{argmax}_\beta\, \Delta\boldsymbol{\mu}_{\mathsf{A}}(\beta)$ exists and is unique. $\beta_0 = \max\{\beta \in [\beta_{\mathsf{A}}^{\texttt{MaxUtil}}, 1] : \mathcal{U}(\beta_{\mathsf{A}}^{\texttt{MaxUtil}}) - \mathcal{U}_{\mathsf{A}}(\beta) \le \delta\}$ which exists and is unique, by the continuity of $\Delta\boldsymbol{\mu}_{\mathsf{A}}$ and Proposition 2.5.6.

Chapter 3

Disparate Equilibria of Algorithmic Decision Making

3.1 Introduction

In the last chapter, we found that popular definitions of algorithmic fairness do not take into account longer-term impact, even in one time period. In this chapter, we examine how the long-term effectiveness of algorithmic decision making depends on societal level *dynamics*. On one hand, deployed decision making models are updated periodically to assure high performance on the target distribution. On the other hand, deployed models can reshape the underlying populations thus biasing how the model is updated in the future. This complex interplay between algorithmic decisions, individual-level responses, and exogeneous societal forces can lead to pernicious long term effects that reinforce or even exacerbate existing social injustices [Crawford, 2017, Whittaker et al., 2018]. Harmful feedback loops have been observed in automated decision making in several contexts including recommendation systems [Pariser, 2011, Conover et al., 2011, Chaney et al., 2018], predictive policing [Ensign et al., 2018], admission decisions [Lowry and Macpherson, 1988, Barocas and Selbst, 2016], and credit markets [Fuster et al., 2022, Aneja and Avenancio-Leon, 2019]. These examples underscore the need to better understand the dynamics of algorithmic decision making, in order to align decisions made about people with desirable long-term societal outcomes.

Automated decision-making algorithms rely on observable features to predict some variable of interest. In a setting such as hiring, decision making models *assess* features such as scores on standardized tests, resume, and recommendation letters, to identify individuals that are *qualified* for the job. However, equally qualified people from different demographic groups tend to have different features, due to implicit societal biases (e.g., letter writers describe competent men and women differently), gaps in resources (e.g., affluent students can afford different extra-curriculars) and even distinct tendencies in self-description (e.g., gender can be inferred from biographies [De-Arteaga et al., 2019]). Therefore, a model's ability to identify qualified individuals can widely vary across different groups.

The deployed model's ability to identify qualified members of a group affects an individual's incentive to *invest* in their qualification. This is because one's decision to acquire qualification—not observed directly by the algorithm—comes at a cost. Moreover, individuals that are identified by the model as qualified (whether or not they are truly qualified) receive a reward. Consequently, people invest in acquiring qualifications only when their expected reward from the assessment model beats the investment cost.

Rational individuals are aware that upon investing they would develop features that are similar to those of qualified individuals in their group, so they gauge their own expected reward from investing by the observed rewards of their group.[1] If qualified people from one group are not duly identified and rewarded, fewer people from that group are incentivized to invest in qualifications in the future. This reduces the overall fraction of qualified people in that group, or the *qualification rate*. As the assessment model is updated to maximize overall institutional profit on the new population distribution, it may perform even more poorly on qualified individuals from a group with relatively low qualification rate, further reducing the group's incentive to invest.

To understand and mitigate the challenges to long-term welfare and fairness posed by such dynamics, we propose a formal model of sequential learning and decision-making where at each round a new batch of individuals rationally decide whether to invest in acquiring qualification and the institution updates its assessment rule (a classifier) for assessing and thus rewarding individuals. We study the long-term behavior of these dynamics by characterizing their equilibria and comparing these equilibria based on several metrics of social desirability. Our model can be seen as an extension of Coate and Loury [1993]'s widely cited work to explicitly address heterogeneity in observed features across groups. While Coate and Loury [1993]'s model focuses on a single-dimensional feature space, i.e., scores, and assessment rules that act as thresholds on the score, our model considers general, possibly high-dimensional, feature spaces and arbitrary assessment rules, which are typical in high-stakes domains such as hiring and admissions.

We find that two major obstacles to obtaining desirable long-term outcomes are heterogeneity across groups and lack of realizability within a group. *Realizability*—the existence of a (near) perfect way to assess qualifications of individuals from visible features—leads to equilibria that are (near) optimal on several metrics, such as the resulting qualification rates, their uniformity across groups, and the institution's utility. We study (near) realizability and the lack thereof in Sections 3.3 and 3.5 respectively. *Heterogeneity across groups*, i.e., lack of a single assessment rule that perfectly assesses individuals from all groups, necessitates tradeoffs in the quality of equilibria across different groups. We study heterogeneity, as well as interventions for mitigating its negative effects, in Section 3.4. In Section 3.6, we empirically study a more challenging setting where the groups are heterogeneous as well as highly non-realizable, via simulations with a FICO credit score dataset [US Federal Reserve, 2007] that has been widely used for illustration in the algorithmic fairness literature.

Interventions. To mitigate the aforementioned tradeoffs, we consider two common inter-

[1] Strong group identification effects can also be seen in empirical studies [Hoxby and Avery, 2013].

ventions: decoupling the decision policy by group and subsidizing the cost of investment, especially when the cost distribution inherently differs by group. Our model of dynamics sheds a different light on these interventions, complementary to previous work. We show that decoupling [Dwork et al., 2018]—using group-specific assessment rules—achieves optimal outcomes when the problem is realizable within each group, but can significantly hurt certain groups when the problem is non-realizable and there exist multiple equilibria after decoupling. In particular, decoupling can hurt a group with low initial qualification rate if the utility-maximizing assessment rule for a single group is more disincentivizing to individuals than a joint assessment rule, thereby reinforcing the status quo and preventing the group from reaching an equilibrium with higher qualification rate.

We also study subsidizing individuals' investment cost (e.g. subsidizing tuition for a top high school), especially when the cost distribution is varied across different groups. We find that these subsidies increase the qualification rate of the disadvantaged group at equilibrium, regardless of realizability. We note that our subsidies, which affect the qualification of individuals directly, are different than those studied under strategic manipulation [Hu et al., 2019] that involve subsidizing individual's cost to manipulate their features *without changing the underlying qualification* (e.g. subsidizing SAT exam preparation without changing the student's qualification for college) and could have adverse effects on disadvantaged groups. Instead, our theoretical findings resonates with extensive empirical work in economics on the effectiveness of subsidizing opportunities for a disadvantaged group to directly improve their outcomes, such as moving to better neighborhoods to access better educational and environmental resources [Chetty et al., 2016].

Related work. The work presented in this chapter is related to a rich body of work on algorithmic fairness in dynamic settings [Liu et al., 2018, Hu and Chen, 2018a, Hashimoto et al., 2018, Zhang et al., 2019, Mouzannar et al., 2019], strategic classification [Hu et al., 2019, Milli et al., 2019, Kleinberg and Raghavan, 2019], as well as statistical discrimination in economics [Arrow, 1973, Coate and Loury, 1993, Arrow, 1998]. We present a detailed discussion of the similarities and differences in Section 3.7.

3.2 A Dynamic Model of Algorithmic Decision Making

In this section we introduce a model of automated decision making with feedback. We first introduce the notation used throughout the paper and then describe the details of the interactions between individuals and an institution, and the resulting dynamical system.

Notation

We consider an instance space $\mathcal{X}$, where $X \in \mathcal{X}$ denotes the features of an individual that are observable by the institution. We also consider a label space $\mathcal{Y} = \{0, 1\}$ where $Y = 1$ indicates that an individual has the qualifications desired by the institution and $Y = 0$

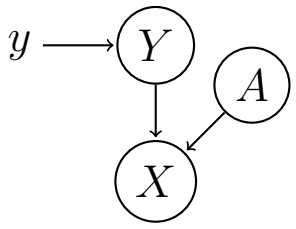


Figure 3.1: Causal graph for the individual investment model. The individual intervenes on the node for qualification, Y—this corresponds to $\text{do}(Y = y)$)—which then affects the distribution of their features X, depending on the group A.

otherwise. We denote the set of all protected/group attributes by $\mathcal{A}$ where $A \in \mathcal{A}$ denotes an individual's protected attribute. We denote the group proportions by $n_a \coloneqq \mathbb{P}(A = a)$ for all $a \in \mathcal{A}$. Furthermore, we denote the qualification rate in group $a \in \mathcal{A}$ by $\pi_a \coloneqq \mathbb{P}(Y = 1 \mid A = a)$. An individual from group $A = a$ who has acquired label $Y = y$ (to become qualified or not)[2] receives features X distributed according to $\mathbb{P}(X = x \mid Y = y, A = a)$. This is illustrated in Figure 3.1.

We also consider a set of parameters Θ that are used for assessing qualifications. We use $\hat{Y}_\theta \in \mathcal{Y}$ parameterized by $\theta \in \Theta$ to denote the *assessed qualification* of an individual. We assume that $\hat{Y}_\theta$ only depends on the features X, which may or may not contain A or its proxies. In later sections, we will also discuss interventions that allow us to use $\hat{Y}_\theta$ that explicitly depends on group membership A. We respectively define the *true positive rate* and *false positive rate* of $\theta \in \Theta$ on group $a \in \mathcal{A}$ by

$$\text{TPR}_a(\theta) = \mathbb{P}(\hat{Y}_\theta = 1 \mid Y = 1, A = a), \text{ and}$$
$$\text{FPR}_a(\theta) = \mathbb{P}(\hat{Y}_\theta = 1 \mid Y = 0, A = a).$$

Model Description

Individual's Rational Response We consider a setting where an individual decides whether to acquire qualifications, that is, to invest in obtaining label $Y = 1$, prior to observing their feature X. The decision to acquire qualification depends on the qualification assessment rule $\theta \in \Theta$ currently implemented by the institution. We will characterize the groups' qualification rates as the *best-response* to θ by function $\pi^{br}(\theta) = (\pi_a^{br}(\theta))_{a \in \mathcal{A}}$.

To get label $Y = 1$ an individual has to pay a cost $C > 0$. In any group, C is distributed randomly according to the cumulative distribution function (CDF), $G(\cdot)$.[3] After deciding

[2]This can be seen as the individual performing a do-intervention on Y [see e.g., Pearl, 2009]. Thus we may write $\text{do}(Y = 1)$ for making the decision to acquire qualifications. Our model (Figure 3.1) assumes that Y is not the child of any node, so we have $\mathbb{P}(\cdot \mid \text{do}(Y = y)) = \mathbb{P}(\cdot \mid Y = y)$. Hence we drop the do-operator whenever we condition on Y.

[3]For the rest of this work, unless otherwise stated, we assume that the distribution of costs, G, is the

whether to acquire qualifications, an individual gets features X and is assessed by θ. An individual (from any group and regardless of actual qualification) receives a payoff of $w > 0$ if they are assessed to be qualified and payoff of 0 otherwise. Therefore, the expected utility an individual from group a receives from acquiring qualification $Y = 1$ is $w\mathbb{P}[\hat{Y}_\theta = 1|Y = 1, A = a] - C = w\mathrm{TPR}_a(\theta) - C$ whereas the expected utility for not acquiring the qualification is $w\mathbb{P}[\hat{Y}_\theta = 1|Y = 0, A = a] = w\mathrm{FPR}_a(\theta)$. Given the qualification assessment parameter $\theta \in \Theta$, an individual from group a acquires qualification if and only if the benefit outweighs the costs, that is

$$w(\mathrm{TPR}_a(\theta) - \mathrm{FPR}_a(\theta)) > C. \tag{3.1}$$

Then each group's qualification rate as a function of a qualification assessment parameter θ is

$$\begin{aligned}\pi_a^{br}(\theta) \coloneqq \mathbb{P}(Y = 1 \mid A = a) &= \mathbb{P}(C < w(\mathrm{TPR}_a(\theta) - \mathrm{FPR}_a(\theta))) \\ &= G(w(\mathrm{TPR}_a(\theta) - \mathrm{FPR}_a(\theta))).\end{aligned}$$

Institution's Rational Response We consider an institution that has to choose a qualification assessment parameter for accepting individuals to maximize its utility. We assume that the institution gains $p_{\mathrm{TP}} > 0$ for accepting a qualified individual and loses $c_{\mathrm{FP}} > 0$ for accepting an unqualified individual. Then the expected utility of the institution for applying parameter θ is

$$\begin{aligned}& p_{\mathrm{TP}}\mathbb{P}(\hat{Y}_\theta = 1, Y = 1) - c_{\mathrm{FP}}\mathbb{P}(\hat{Y}_\theta = 1, Y = 0) \\ = \;& p_{\mathrm{TP}} \sum_{a \in \mathcal{A}} \mathrm{TPR}_a(\theta)\pi_a n_a - \sum_{a \in \mathcal{A}} c_{\mathrm{FP}}\mathrm{FPR}_a(\theta)(1 - \pi_a)n_a.\end{aligned}$$

This illustrates that the utility maximizing parameter is a function of $\pi = (\pi_a)_{a \in \mathcal{A}}$, i.e., the rate of qualification in each group. We denote this function by $\theta^{br}(\pi)$, defined as follows:

$$\theta^{br}(\pi) := \operatorname*{argmax}_{\theta \in \Theta} \; p_{\mathrm{TP}} \sum_{a \in \mathcal{A}} \mathrm{TPR}_a(\theta)\pi_a n_a - \sum_{a \in \mathcal{A}} c_{\mathrm{FP}}\mathrm{FPR}_a(\theta)(1 - \pi_a)n_a.$$

To ensure the above object (and the resulting dynamics) are well-defined, when multiple parameters θ achieve the optimal utility we assume that $\theta^{br}(\pi)$ is uniquely defined using a fixed and well-defined tie-breaking function.

Throughout this paper we assume that the institution has exact knowledge of many quantities such as $\mathrm{TPR}_a(\theta)$, $\mathrm{FPR}_a(\theta)$, and n_a. In a nutshell, we assume that we have infinitely many samples from the underlying distributions. We discuss this further in Section 3.8, and leave the finite sample version of these results to future work.

Although we choose not to focus on game-theoretical aspects in this work, we note that our model can be thought of as a large game [Kalai, 2004] or a game with a continuum of players [Schmeidler, 1973].

same for every group. In Section 3.4 and 3.6, we will consider the implications of having different cost distributions by group.

Dynamical System and Equilibria. We are primarily interested in the evolution of qualification rate, π, over time. Given a current rate of qualification π the assessment parameter used by the institution in the next step is $\theta^{br}(\pi)$, which in turn leads to a qualification rate of $\pi^{br}(\theta^{br}(\pi))$ in the next step. Therefore, we define a dynamical system for a given initial state $\pi(0)$ such that at time t we are in state $\pi(t) = \Phi(\pi(t-1))$, where $\Phi = \pi^{br} \circ \theta^{br}$.

We say that the aforementioned dynamical system is at equilibrium if $\pi = \Phi(\pi)$. Equivalently, we are at an equilibrium if $\pi = \lim_{n\to\infty} \Phi^n(\pi(0))$ is well-defined for some $\pi(0)$, where Φ^n is an n-fold composition of Φ. We call such values of π *equilibria*, or equivalently, *fixed points* of Φ.

In general, Φ may have multiple fixed points that demonstrate different characteristics. We therefore compare the fixed points of Φ on several metrics of societal importance.

1. *Stability:* We say that an equilibrium π^* is stable if there is a non-zero measure set of initial states $\pi(0) \in [0, 1]$ for which $\pi^* = \lim_{n\to\infty} \Phi^n(\pi(0))$. In particular, if there exists a neighborhood around π^* such that all points converge to π^* under the dynamics, we say that π^* is *locally stable*. As such, stable fixed points are robust to small perturbations in the qualification rate, which can occur due to random measurement errors.

2. *Qualification Rate of Group a:* Recall that the qualification rate, π_a, is the fraction of individuals in group a who invested in qualifications. Since it is more desirable to have a high qualification rate in each group, we may compare equilibria based on π_a. We refer to $G(w)$ as the optimal qualification rate in group a, which is the maximum achievable qualification rate corresponding to the perfect assessment rule.[4]

3. *Balance:* We may be interested in equilibria where the qualification rate is similar across groups, that is, to prioritize equilibria with smaller $\max_{a_1,a_2\in\mathcal{A}} |\pi^*_{a_1} - \pi^*_{a_2}|$. When this quantity is 0 we say that π^* is *fully balanced*.

4. *Institutional utility:* We may compare equilibria based on their corresponding institution utility.

Examples From the Real World

Let us instantiate our model in the setting of two important applications from the real world.

College Admissions Consider the college admission setting, where X corresponds to the features that the college can observe, e.g., a candidate's test scores and letters of recommendation. Y indicates whether the candidate meets the qualifications required to succeed in the program. C is the cost of investing in the qualifications, e.g., the money and opportunity cost of studying or taking additional courses to obtain the required qualifications. A candidate

[4] If group a has a group-specific cost distribution, G_a, then we refer to $G_a(w)$ as the optimal qualification rate in group a.

from group a will develop features from distribution $\mathbb{P}[X = x|Y = y, A = a]$, where $y = 1$ indicates a qualified candidate. The differences in the feature distribution between groups can be attributed to several factors such as resources that are available to different groups, e.g., letters of recommendations for qualified female and male candidates often emphasize different traits. θ is the decision parameter used by the college, e.g., $\hat{Y}_\theta = 1$ when the candidate has SAT score of > 1400 and an excellent recommendation letter. The college accepts applicants by trading off between the utility gain, p_{TP}, of admitting qualified candidates and utility cost, c_{FP}, of admitting an unqualified candidates. The candidate is incentivized to acquire the qualifications for the college based on the long term benefit (described in Equation (3.1)) that depends on their expected gain w from completing a college degree and how likely it is to be admitted to college for a qualified or unqualified member of the group the candidate belongs to.

Hiring Consider the hiring setting, where X corresponds to the features that the firm can observe, e.g., a candidate's CV. Y indicates whether the applicant meets the qualifications required by the firm, e.g., having the required knowledge and the ability to work in a team. C is the cost of acquiring the qualifications required by the firm, e.g., the (monetary and opportunity) cost of acquiring a college degree or working on a team project. Parameter θ is the hiring parameter used by the firm, e.g., $\hat{Y}_\theta = 1$ when the applicant has a software engineering degree and two years of experience. The firm accepts candidates according to utility maximization involving p_{TP}, the profit from hiring a qualified candidate, and c_{FP}, the cost of hiring an unqualified candidate e.g., the loss in productivity or the the cost to replace the employee. The candidate is incentivized to acquire the qualifications for the job based on factors including their expected salary w and how likely it is to be hired by the firm given how the firm has hired qualified or unqualified candidates from the group the candidate belongs to (Eq. (3.1)).

We also consider a stylized example of lending in Section 3.6.

3.3 Importance of (Near) Realizability

We start our theoretical investigation of dynamic algorithmic decision making with the classical model of realizability. In the theory of machine learning, a distribution is called realizable if there is a decision rule in the set Θ whose error on the distribution is 0. Analogously, we call a setting *realizable* when there is a decision rule $\theta^{\text{opt}} \in \Theta$ that perfectly classifies every individual from every group, that is $\text{TPR}_a(\theta^{\text{opt}}) = 1$ and $\text{FPR}_a(\theta^{\text{opt}}) = 0$ for all $a \in \mathcal{A}$. Realizability is a widely used assumption and is the basis of seminal works such as Boosting [Freund and Schapire, 1997]. At a high level, realizability corresponds to the assumption that there is an unknown ground truth assessment rule, for example, in a hypothetical setting where x includes all the information that is sufficient for assessing one's qualification, and the chosen set of decision rules is rich enough to contain it.

In static realizable applications of machine learning, the goal is to (approximately) recover θ^{opt} from data. We show that in the our dynamic setting, under realizability, the unique non-zero equilibrium of Φ is where individuals respond to θ^{opt}. Furthermore, each group attains their optimal qualification rate at this equilibrium.

Proposition 3.3.1 (Perfect classfication)**.** *If there exists $\theta \in \Theta$ such that $\text{TPR}_a(\theta) = 1$ and $\text{FPR}_a(\theta) = 0$ for all $a \in \mathcal{A}$, then there is a unique non-zero equilibrium with $\pi_a^* = G(w)$ for all $a \in \mathcal{A}$.*

While realizability is a common assumption in the theory of machine learning, it rarely captures the subtleties that exist in automated decision making in practice. Next, we consider a mild relaxation of realizability and consider a setting where a near-perfect decision rule $\theta \in \Theta$ exists such that $\text{TPR}_a(\theta) \geq 1-\epsilon$ and $\text{FPR}_a(\theta) \leq \epsilon$. As we show (and prove in Section 3.9), when there is a single near-realizable group the main message of Proposition 3.3.1 remains effectively the same. That is, all equilibria that are reachable from initial points that are not too extreme approximately maximize the group's qualification rate.

Theorem 3.3.2 (Equilibria under near-realizability)**.** *Let $|\mathcal{A}| = 1$ and assume that $p_{\text{TP}} = c_{\text{FP}} = 1$. Assume that for fixed $\epsilon \in (0,1)$, $s \in (0, 1/2)$, G is L_G-Lipschitz with property that $1 - s \geq G(w) \geq s + \frac{L_G w \epsilon}{s}$, and there is $\theta \in \Theta$ such that*

$$\text{TPR}(\theta) \geq 1 - \epsilon \text{ and } \text{FPR}(\theta) \leq \epsilon.$$

Then for any initial investment $\pi(0) \in [s, 1-s]$, $\pi^* = \lim_{n\to\infty} \Phi^n(\pi(0))$ is such that

$$\pi^* \geq G(w(1 - \epsilon/s)).$$

A nice aspect of the above results is that the assumption of realizability or near-realizability can be validated from the data. That is, the decision maker can compute whether there is $\theta \in \Theta$ such that $\text{TPR}(\theta) \geq 1 - \epsilon$ and $\text{FPR}(\theta) \leq \epsilon$. If so, then the decision maker can rest assured that the dynamical system is on the path towards achieving near optimal investment by the individuals. Another nice aspect of these results is the characterization of the equilibria in terms of the CDF of the cost distribution. This allows us to use this framework for studying interventions that change the cost function directly. One such intervention is subsidizing the cost for individuals so that the cumulative distribution function of the cost, given by $G(x)$, is increased by a sufficient amount at every cost level x. The following corollary, proved in Section 3.9, shows that under this kind of subsidy, the equilibria reached by the dynamics will have higher qualification rate than any fixed point of the dynamics before subsidy, as long as the initial points are not too extreme. As we are considering different cost distribution functions in the following corollary, we denote the dynamics corresponding to cost distribution function G as Φ_G.

Corollary 3.3.3 (Subsidizing the cost of investment)**.** *Let* $|\mathcal{A}| = 1$ *and assume that* $p_{\text{TP}} = c_{\text{FP}} = 1$*. Assume that for fixed* $\epsilon \in (0,1)$ *,* $s \in (0, 1/2)$*, G is* L_G*-Lipschitz with property that* $1 - s \geq G(w) \geq s + \frac{L_G w \epsilon}{s}$*, and there is* $\theta \in \Theta$ *such that*

$$\text{TPR}(\theta) \geq 1 - \epsilon \text{ and } \text{FPR}(\theta) \leq \epsilon.$$

Let $\pi^* > 0$ *be a fixed point of the dynamics* Φ_G*. Suppose* $\bar{G}$ *is a strictly increasing,* $L_{\bar{G}}$*-Lipschitz CDF such that* $1 - s \geq \bar{G}(x) \geq s + \frac{L_{\bar{G}} w \epsilon}{s}$ *and* $\bar{G}(x(1 - \epsilon/s)) \geq G(x)$ *for all* x *in the domain of* G*. Then for any initial investment* $\pi(0) \in [s, 1 - s]$*, there exists a* $\bar{\pi} \geq \pi^*$*, such that* $\bar{\pi} = \lim_{n \to \infty} \Phi^n_{\bar{G}}(\pi(0))$*.*

3.4 Group Realizability

In this section, we investigate how the nature of equilibria evolves as the assumption of realizability is relaxed to allow for heterogeneity across groups. Specifically, we consider the case where there exists a perfect assessment rule for each group, but not when the groups are combined. We call this "group-realizability". Our results illustrate that without realizability or near-realizability, the utility-maximizing assessment rule can be very sensitive to the relative qualification rates in different groups, resulting in *multiple* equilibria, at which groups may experience disparate outcomes.

In sections 3.4 and 3.4, we study group-realizability under two different and complementary settings. The first setting considers features that are drawn from a multivariate Gaussian distribution and assumes that in each group the qualified individuals are perfectly separated from unqualified ones by a group-specific hyperplane. This is a benign setting where no group is inherently disadvantaged — group features and performance of assessment rules are symmetric up to a reparameterization of the space. The second setting considers features that are uniformly distributed scalar scores and assumes that qualified and unqualified individuals in a group are separated by a group-specific threshold, where one is higher than the other. - This model captures the natural setting where the feature (score) and assessment rules inherently favor one group, e.g., SAT scores are known to be skewed by race [Card and Rothstein, 2007]. We use the aforementioned stylized settings to demonstrate the salient characteristics of equilibria that one might anticipate under group-realizability. We find that stable equilibria tend to favor one group or the other. This is especially surprising in the multivariate Gaussian case where the two groups are identical up to a change in the representation of the space. We also study the existence of balanced equilibria, where both groups acquire qualification at the same rate. We find that when balanced equilibria exist they tend to be unstable, that is, no initial qualification rate (except for the balanced equilibrium itself) will converge to the balanced equilibrium under the dynamics.

We consider two natural interventions in overcoming the challenges of group-realizability as outlined above. As group-realizability poses even greater challenges when the costs of investment are unequally distributed *between* groups, in Section 3.4 we consider the impact of subsidizing the cost of acquiring qualification for one group. In Section 3.4, we consider

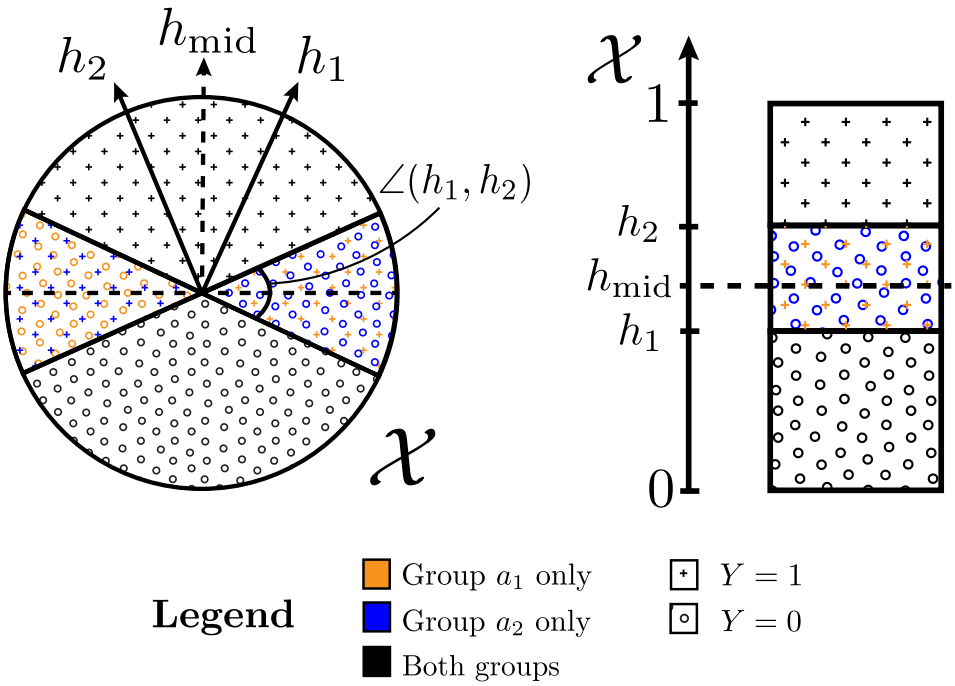


Figure 3.2: Equilibria in the Multivariate Gaussian case (left) and the Uniform case (right)

the impact of decoupling, that is, we allow the institution to use different assessment rules for different groups assuming the group attributes are available. This is in contrast to the typical setting where institutions are constrained to using the same assessment rule across all groups, which may be the case when data on the protected attribute is not available or when the use of protected attributes for assessment is regulated.

Uniformly Distributed Scalar X

We consider $\mathcal{X} = [0, 1]$, the class of assessment paramters $\Theta = [0, 1]$, and assessment decision $\hat{Y}_h = \mathbf{1}\{X > h\}$ for all $h \in \Theta$ that represent all threshold decision policies. Consider two groups a_1, a_2. Let X be a score that is uniformly distributed over $[0, 1]$ where in group a_i those with score more than h_i are qualified and those with score at most h_i are unqualified. This is depicted in Figure 3.2 (right). Formally,

$$\mathbb{P}(X = x \mid Y = y, A = a_i) = \begin{cases} \mathbf{1}\{x > h_i\}/(1 - h_i) & \text{for } y = 1 \text{ and} \\ \mathbf{1}\{x \leq h_i\}/h_i & \text{for } y = 0 \end{cases}.$$

We make the following assumption to simplify notation.

Assumption 2. *We assume $n_{a_1} \cdot p_{\text{TP}} = n_{a_2} \cdot c_{\text{FP}}$. We also assume that the cost for acquiring qualifications is uniformly distributed on $[0, 1]$ (i.e. $G(c) = c$) in both groups.*[5]

[5]Our results also generalize to the setting where the CDF for the cost $G : [0, 1] \to [0, 1]$ is an arbitrary strictly increasing function.

We show that when w is in a certain range, there are two *unbalanced stable* equilibria corresponding to assessment parameters h_1 or h_2, which respectively lead to the optimal qualification rate for groups a_1 or a_2 but low qualification rate for the other group. There is also a more *balanced* but *unstable* equilibrium at some threshold h_{mid} between h_1 and h_2. Outside of this range of w, there is only one equilibrium in which one of the groups achieves its optimal qualification rate. These findings are summarized in the following two propositions.

Proposition 3.4.1. *Define* $g :- \frac{(1-h_1)(-wh_2^2+h_2(1-h_1)-wh_1(1-h_1))}{w((1-h_1)^2-h_2^2))}$. *Note that* $g \in (0, h_2 - h_1)$ *for any* w. *Let* $w \in (w_l, w_u)$ *where*

$$w_l := \frac{(1-h_1)^2}{(1-h_2)h_2 + (1-h_1)^2},\ w_u := \frac{h_2(1-h_1)}{h_2^2 + h_1(1-h_1)}. \tag{3.2}$$

Given Assumption 2, there exists two stable equilibria at

$$h = h_1, \qquad \pi_{a_1} = w, \qquad \pi_{a_2} = w \cdot \frac{h_1}{h_2},\ \textit{and} \tag{3.3}$$

$$h = h_2, \qquad \pi_{a_1} = w \cdot \frac{1-h_2}{1-h_1}, \qquad \pi_{a_2} = w, . \tag{3.4}$$

and a unique non-zero unstable equilibrium at

$$h = h_{\text{mid}} := h_1 + g, \quad \pi_{a_1} = w \cdot \frac{1-h_1-g}{1-h_1}, \quad \pi_{a_2} = w \cdot \frac{h_1+g}{h_2}.$$

When $w = 1 - h_1$, *the unstable equilibrium is fully balanced.*

Proposition 3.4.2. *Given Assumption 2 when* $w < w_l$ *there exists one stable equilibrium defined by Equation 3.4, and when* $w > w_u$ *there exists one stable equilibrium defined by Equation 3.3.*

The details of the proofs are presented in Section 3.9. At a high level, if the wage is not too low or too high, both thresholds h_1 and h_2 correspond to stable equilibria, at which either group a_1 or a_2 is perfectly classified. The equilibrium corresponding to h_{mid}, where the classifier has the same true positive and false positive rates in both groups, is unstable and subsequently harder to achieve.

In Table 3.1, we compare these equilibria in terms of metrics introduced in Section 3.2, under the assumptions of Proposition 3.4.1. We use standard notation $\succ$ and $\sim$ to denote preference and indifference respectively. For example, we find that in terms of balance in qualification rates, the stable equilibrium associated with h_1 is more balanced that the stable equilibrium associated with h_2, but both are always less balanced unstable equilibrium associated with h_{mid}. Details of the computation are deferred to Table 3.3 in Section 3.9.

	Ranking of Equilibria
Stability	h_1, h_2 are stable. h_{mid} is unstable
Qualification rate of group a_1	$h_1 \succ h_{\text{mid}} \succ h_2$
Qualification rate of group a_2	$h_2 \succ h_{\text{mid}} \succ h_1$
Balance of qualification rates	$h_{\text{mid}} \succ h_1 \succ h_2$
Institution's Utility	no ranking

Table 3.1: Comparison of equilibria for uniform features. In this table we refer to each equlibria using the associated threshold decision policy.

Multivariate Gaussian X

We consider $\mathcal{X} = \mathbb{R}^d$ and $\Theta = S_{d-1}$, where S_{d-1} is the set of d-dimensional unit vectors. Let $\hat{Y}_h = \mathbf{1}\{X^\top h \geq 0\}$ for all $h \in \Theta$ denote separating hyperplane policies and $\angle_{h,h'} \coloneqq \frac{1}{\pi}\arccos(\frac{h^\top h'}{\|h\|\|h'\|})$ denote the angle between two vectors, normalized by the constant π. We consider two groups a_1 and a_2 associated respectively with vectors h_1 and h_2, such that $\angle_{h_1,h_2} \neq 0$. We assume the groups have equal size, i.e., $n_{a_1} = n_{a_2}$. For each group, the feature distribution is a d-dimensional spherical Gaussian centered at the origin such that the qualified individuals are in halfspace $\mathbf{1}\{X^\top h_i \geq 0\}$ and the unqualified individuals in halfspace $\mathbf{1}\{X^\top h_i < 0\}$. Formally, for $x \in \mathbb{R}^d$ and $i \in \{1, 2\}$,

$$\mathbb{P}(X = x \mid Y = y, A = a_i) = \begin{cases} 2\phi(x)\mathbf{1}\{x^\top h_i \geq 0\} & \text{for } y = 1 \text{ and} \\ 2\phi(x)\mathbf{1}\{x^\top h_i < 0\} & \text{for } y = 0, \end{cases}$$

where $\phi(x)$ is the density of the spherical d-dimensional Gaussian. This is depicted in Figure 3.2 (left).

Assumption 3. *We assume that the CDF for the cost of acquiring qualifications is a strictly increasing function $G : [0, 1] \to [0, 1]$ and is the same in both groups.*

As we will see, the relative gain (loss) of the institution for accepting a qualified (unqualified) individual, that is $p_{\text{TP}}/c_{\text{FP}}$, plays a role in the nature of the equilibria. The following proposition characterizes the equilibria when this value is strictly positive, that is, when the benefit of true positives outweighs the cost of false positives. Notably, similar to the previous setting of uniform scores, the current setting also has two stable equilibria that each favor one group at the expense of the other, as well as a balanced equilibrium that is unstable.

Proposition 3.4.3. *Given Assumption 3 and $p_{\mathrm{TP}} > c_{\mathrm{FP}}$, there exists two stable equilibria, at*

$$\begin{aligned} &h = h_1, && \pi_{a_1} = G(w) && \pi_{a_2} = G\left(w \cdot (1 - 2\angle_{h_1,h_2})\right), \\ &h = h_2, && \pi_{a_1} = G\left(w(1 - 2\angle_{h_1,h_2})\right) && \pi_{a_2} = G(w). \end{aligned}$$

There is a unique non-zero unstable equilibrium at

$$h = h_{\mathrm{mid}}, \qquad \pi_{a_1} = G\left(w(1 - \angle_{h_1,h_2})\right) \qquad \pi_{a_2} = G\left(w(1 - \angle_{h_1,h_2})\right),$$

where $h_{\mathrm{mid}} := \frac{h_1 + h_2}{\|h_1 + h_2\|}$.

Let us briefly comment on the high level proof idea and defer the full argument to Section 3.9. Since $p_{\mathrm{TP}} > c_{\mathrm{FP}}$, the institution cares more about accepting true positives than avoiding false positives. Therefore, the utility-maximizing h is determined by the group that has a higher qualification rate and thus has a higher fraction of positives — this is h_1 (resp. h_2) whenever $\pi_{a_1} > \pi_{a_2}$ (resp. $\pi_{a_1} < \pi_{a_2}$). When qualification rates are equal between the two groups, the institution maximizes its utility at any h that is a convex combination of h_1 and h_2, but the unique h that would induce equal qualification rate is $h = h_{\mathrm{mid}}$, where the classifier has the same true positive and false positive rates in both groups.

An unfortunate implication of this result is that the dynamics will always converge to an unbalanced qualification rate, except when the initial levels of investment are exactly the same. Even though a fully balanced equilibrium exists, it is unstable and therefore not robust to small perturbations in either the qualification rates or the classifier, which in practice is unavoidable given sampling noise.

In Table 3.2, we compare these equilibria in terms of metrics introduced in Section 3.2. For example, we find that in terms of institutional utility, the stable equilibria associated with h_1 and h_2 are equally preferred, and are both strictly preferred to the unstable equilibrium associated with h_{mid}. This implies that the institution has no incentive at all to keep the dynamics at the unstable equilibrium, even though it induces balanced investment. Exact values are deferred to Table 3.4 in Section 3.9.

	Ranking of Equilibria
Stability	h_1, h_2 are stable. h_{mid} is unstable
Qualification rate of group a_1	$h_1 \succ h_{\text{mid}} \succ h_2$
Qualification rate of group a_2	$h_2 \succ h_{\text{mid}} \succ h_1$
Balance of qualification rate	$h_{\text{mid}} \succ h_1 \sim h_2$
Institution's Utility	$h_1 \sim h_2 \succ h_{\text{mid}}$

Table 3.2: Comparison of equilibria for Multivariate Gaussian features. In this table we refer to each equlibria using the associated hyperplane.

Interestingly, when $p_{\text{TP}} < c_{\text{FP}}$, there are no stable equilibria; instead there exists a stable limit cycle between h_1 and h_2. This is stated informally in the following proposition.

Proposition 3.4.4. *Given Assumption 3 and $p_{\text{TP}} < c_{\text{FP}}$, there exists no stable equilibria. Instead there exists a limit cycle and one non-trivial unstable equilibrium.*

Intuitively, the cycle is caused by misaligned incentives between the institution and the individuals. Since the institution finds false positives more costly than false negatives, it prefers the hyperplane that classifies more false positives correctly. At each time step, it will choose the hyperplane associated with the group that has a lower qualification rate, prompting that group to invest more in the next time step. Strikingly, even a simple group-realizable model involving multivariate Gaussian distributions demonstrates a large range of limiting behaviors. In Section 3.6, we also observe the existence of limit cycles in simulations with real data distributions.

Different Costs of Investment by Group

Thus far we have assumed that all groups have the same distribution of the cost of investment, G. In reality, the cost of investment may be distributed differently in each group; a disadvantaged group might on average experience higher (monetary or opportunity) costs. For example, low income families who may have to take out loans to pay for college tuition incur high interest rates. This is a compelling setting that reflects deep-seated disparities in access to opportunity between demographic groups in the real world; an analogous setting has been considered by works on strategic classification, where the costs for manipulating features is posited to differ across groups [Hu et al., 2019, Milli et al., 2019].

In this section, we consider the ramifications of differences in investment cost across groups, focusing on the setting of Section 3.4. We show that the disadvantage from having higher costs is amplified under group-realizability. Specifically, suppose that group a_1 (resp.

a_2) has costs distributed according to cumulative distribution function G_1 (resp. G_2), and that group a_1 is disadvantaged in terms of costs. The following result observes that if G_1 sufficiently dominates G_2, then there exists no stable equilibrium that encourages optimal investment from group a_1 and no equilibrium that is balanced for both groups, in sharp contrast to the characterization in Proposition 3.4.3. The proof is deferred to Appendix 3.9.

Proposition 3.4.5. *Consider the multi-variate Gaussian setting of Section 3.4. Suppose G_1 and G_2 are such that $G_1(w) < G_2(w(1 - 2\angle_{h_1,h_2}))$, then there exists a single non-trivial equilibrium at h_2, which is also stable. The level of investment by group a_1 (resp. a_2) is $G_1(w(1 - 2\angle_{h_1,h_2})$ (resp. $G_2(w)$).*

Effect of subsidies In this situation, an intervention that would effectively raise the equilibrium level of investment by the disadvantaged group is to subsidize the cost of investment. In particular, as long as we replace G_1 with a stochastically dominated distribution $\bar{G}_1$ such that $\bar{G}_1 > G_2(w(1 - 2\angle_{h_1,h_2}))$, under the new dynamics Φ^{sub}, h_1 will again be a stable equilibrium, and there will also exist a more balanced, unstable equilibrium at $h = \bar{h}_{\text{mid}}$, which is some convex combination of h_1 and h_2. At all equilibria of Φ^{sub}, group a_1 will have higher levels of investment than $G_1(w(1 - 2\angle_{h_1,h_2}))$.

However, this improvement may come at a cost to the advantaged group, since Φ^{sub} has multiple equilibria and some of them have group a_2 investing less than $G_2(w)$. Still one might argue that the equilibria of Φ^{sub} are more equitable, since the dynamics without subsidies always result in optimal investment by group a_2 and low investment by group a_1.

Decoupling the Assessment Rule by Group

The models we studied in Sections 3.4 and 3.4 suggest that applying the same, or "joint", assessment rule to heterogeneous groups results in undesirable trade-offs—between balance, stability, and other metrics—at all equilibria, even though there exists a perfect assessment rule for each group separately.

Decoupling the classifier by group is a natural intervention in this setting. Namely, the institution may choose a group-specific $\theta_a \in \Theta$ to assess individuals from group $a \in \mathcal{A}$, assuming that the group attribute information is available. This corresponds to choosing θ_a that maximizes the utility that the institution derives from each group separately. Thus we now consider the *decoupled dynamics* Φ^{dec} where the institution uses group-specific assessment rules, i.e., for all $a \in \mathcal{A}$

$$\theta_a^{br}(\pi_a) \coloneqq \operatorname*{argmax}_{\theta_a \in \Theta} \; p_{\text{TP}} \text{TPR}_a(\theta_a)\pi_a - c_{\text{FP}} \text{FPR}_a(\theta_a)(1 - \pi_a).^6 \tag{3.5}$$

[6] As when we defined the joint dynamics (Section 3.2), when the argmax is not unique, we assume ties are broken according to a fixed and well-defined order.

As in the standard joint setting individuals still acquire qualification according to their group utility as follows

$$\pi_a^{br}(\theta_a) := G(w(\mathrm{TPR}_a(\theta_a) - \mathrm{FPR}_a(\theta_a))).$$

We denote by $\pi^{\mathrm{dec}} \in [0,1]^{|\mathcal{A}|}$ the equilibria of the decoupled dynamics, $\Phi^{\mathrm{dec}} = \left(\pi_a^{br} \circ \theta_a^{br}\right)_{a\in\mathcal{A}}$. It is not hard to see that decoupling is helpful in a group-realizable setting. That is, the qualification rates of the decoupled equilibrium π^{dec} *Pareto-dominates* the qualification rates of all equilibria π under a joint assessment rule, whenever group-realizability holds.

Proposition 3.4.6 (Decoupling)**.** *Consider a group-realizable setting, that is, for every $a \in \mathcal{A}$, there exists a perfect assessment rule $\theta_a^{\mathrm{opt}} \in \Theta$ such that $\mathrm{TPR}_a(\theta_a^{\mathrm{opt}}) - \mathrm{FPR}_a(\theta_a^{\mathrm{opt}}) = 1$. Then Φ^{dec} has a unique stable equilibrium π^{dec}, where $\pi_a^{\mathrm{dec}} = G(w)$. Moreover, for any equilibrium π of the joint dynamics Φ, $\pi_a^{\mathrm{dec}} \geq \pi_a$ for all $a \in \mathcal{A}$. Furthermore, if there is no perfect assessment rule, i.e.,*

$$\max_{\theta\in\Theta} \sum_{a\in\mathcal{A}} n_a(\mathrm{TPR}_a(\theta) - \mathrm{FPR}_a(\theta)) < 1,$$

then for some $a \in \mathcal{A}$, $\pi_a^{\mathrm{dec}} > \pi_a$.

This proposition directly follows from Proposition 3.3.1.

Indeed, decoupling always helps in the group-realizable setting—not only does it not decrease any group's equilibrium qualification rate, it also increases the equilibrium qualification rate of at least one group when realizability across all groups does not hold. In Sections 3.5 and 3.6 we examine decoupling in the absence of group-realizability and see that those cases are not as clear-cut. When group-realizability does not hold, in some cases decoupling is still helpful while in others it can significantly harm one group.

3.5 Beyond group-realizability: Multiple equilibria within group

We have so far considered settings where the learning problem is realizable (or almost realizable) within each group. This is a common assumption in various prior works, such as Hu et al. [2019]. As we saw in Section 3.4, there may be multiple undesirable equilibria when a joint assessment rule is used in a group-realizable setting, but these undesirable equilibria disappear in the decoupled dynamics.

In many application domains, realizability does not hold even at a group level. That is to say, no assessment rule in Θ can perfectly separate qualified and unqualified individuals even within one group. This may be due to the fact that mapping individuals to the visible feature space $\mathcal{X}$ involves loss of information or there may be other sources of stochasticity in the domain [Corbett-Davies and Goel, 2018], making it impossible to provide a high accuracy

assessment of individuals' qualifications. A key consequence of the lack of realizability is that even for a single group, the optimal classifier now can vary greatly with π_a, the group's qualification rate. As a result, our guarantees about the near-optimality of stable equilibria (Theorem 3.3.2) no longer hold, and there could exist multiple stable equilibria each corresponding to a different qualification rate within a group. In this section, we investigate the existence of bad equilibria for a single group and its implications on decoupling when the learning problem is not group-realizable. For the rest of this section, we consider a single group, i.e., $|\mathcal{A}| = 1$ and suppress a in the notation.

In the following proposition (proved in Section 3.9), we characterize conditions under which multiple equilibria exists for arbitrary feature spaces and assessment rules. This is a generalization of a classical result from Coate and Loury [1993] that considers a one-dimensional feature space; we restate and prove the classical result as a consequence of Proposition 3.5.1 in Section 3.9.

Proposition 3.5.1 (Multiple equilibria in arbitrary feature spaces)**.** *Let Φ be as defined in Section 3.2. For any qualification rate π, let*

$$\beta(\pi) \coloneqq \mathrm{TPR}(\theta^{br}(\pi)) - \mathrm{FPR}(\theta^{br}(\pi)),$$

be the difference between true and false positive rates of the institution's utility maximizing assessment rule with respect to π. Assume $\beta(\pi)$ is continuous, the CDF of the cost G is continuous and that there exists $\theta \in \Theta$ such that $\mathbb{P}(\hat{Y}_\theta = 1) = 0$ and $\theta' \in \Theta$ such that $\mathbb{P}(\hat{Y}_{\theta'} = 1) = 1$, i.e., there is a assessment rule that accepts everyone and an assessment rule that rejects everyone. Also suppose the likelihood ratio $\phi(x) := \frac{\mathbb{P}(X=x|Y=0)}{\mathbb{P}(X=x|Y=1)}$ is strictly positive on $\mathcal{X}$.

If $x < G(w\beta(x))$ for some $x \in (0, 1)$, then there exists at least two distinct non-zero equilibria where $\pi = \Phi(\pi)$. If in addition β is differentiable, an equilibrium at π is locally stable whenever $G'(w\beta(\pi)) < |\beta'(\pi)|$, where G' and β' denote the derivatives of G and β respectively.

Proposition 3.5.1 describes conditions under which there exists more than one equilibrium in the dynamics modeled in Section 2. Given a differentiable $\beta(\pi)$, one can always construct a monotonically increasing G, such that the dynamics Φ has any number of locally stable equilibria. The implication of having multiple equilibria is that the dynamics may converge to different equilibrium qualification rates depending on the initial investment, even for a single group. This makes the setting particularly hard to analyze.

Nevertheless, the following result, proved in Section 3.9, shows that even in the non-realizable setting, subsidizing the cost of investment by changing the distribution G to a stochastically dominant distribution $\bar{G}$ will create a new equilibrium that has a higher qualification rate. In other words, subsidies in the non-realizable setting also improve the quality of equilibria. However, the new equilibrium is not guaranteed to be locally stable. We see some ramifications of this empirically in the next section.

Proposition 3.5.2 (Subsidies without realizability)**.** *Suppose $\pi^* > 0$ is an equilibrium for the dynamics Φ_G, where the cost of investment is distributed according to G on $[0, 1]$. Let $\bar{G}$ be a CDF that is stochastically dominated by G, that is, $\bar{G}(x) > G(x)$ for all $x \in (0, 1)$, and both G and $\bar{G}$ are strictly increasing. Then there exists $\bar{\pi} > \pi^*$ such that $\bar{\pi}$ is an equilibrium for $\Phi_{\bar{G}}$.*

3.6 Simulations with non-realizability

In this section we present results from numerical experiments examining the effects of decoupling and subsidies under our model of dynamics, in the absence of group-realizability. We consider a stylized semi-synthetic experiment, based on a widely used FICO credit score dataset from a 2007 Federal Reserve report [US Federal Reserve, 2007]. Importantly, only aggregate statistics were reported and the data we accessed does not contain sensitive or private information. Our modeling assumptions may not be realistic for this dataset (see Section 3.8) and our simulations should not be interpreted as policy recommendations. Instead, these experiments help us illustrate qualitatively the types of dynamics one may find using real world data.

Stylized Model. We describe how our model can be instantiated to a highly stylized example of credit scoring and lending. Assume a loan applicant either has the means to repay a loan or not. If they have the means to repay, they always repay ($Y = 1$); otherwise they always default ($Y = 0$). In order to have the means to repay, applicants must make an *ex ante* investment at the cost of C, whose distribution is $\mathbb{P}(C < c) = G(c)$. This represents costly actions an individual has to take in order to acquire the financial ability to repay loans, e.g. working at a stable job or taking job preparation classes. Applicants from group a who have the means to repay receive credit scores X drawn from f_1^a and those who don't receive credit scores drawn from f_0^a. The decision of the bank is to approve or reject a loan applicant, given their credit scores.

Dataset. FICO scores are widely used in the United States to predict credit worthiness. The dataset, which contains aggregate statistics, is based on a sample of 301,536 TransUnion TransRisk scores from 2003 [US Federal Reserve, 2007] and has been preprocessed by Hardt et al. [2016b]. These scores, corresponding to X in our model, range from 300 to 850. For simplicity, we rescale the scores so that they are between 0 and 1. Individuals were labeled as defaulted if they failed to pay a debt for at least 90 days on at least one account in the ensuing 18-24 month period. The data is also labeled by race, which is the group attribute A that we use. We compute empirical conditional feature distributions $\mathbb{P}(X = x \mid A = a, Y = y)$ from the available data and fit Beta distributions[7] to these to obtain f_0^a, f_1^a.

We treat these distributions *as if* they came from our model as shown in Figure 3.1, for the sole purpose of illustration. This is not to claim that our modeling assumptions hold on this dataset, as discussed earlier. Given the lending domain is complex, our aim is

[7]We simulate 100,000 samples from the empirical PDF (see Figure 3.3) and fit a Beta distribution by maximum likelihood estimation.

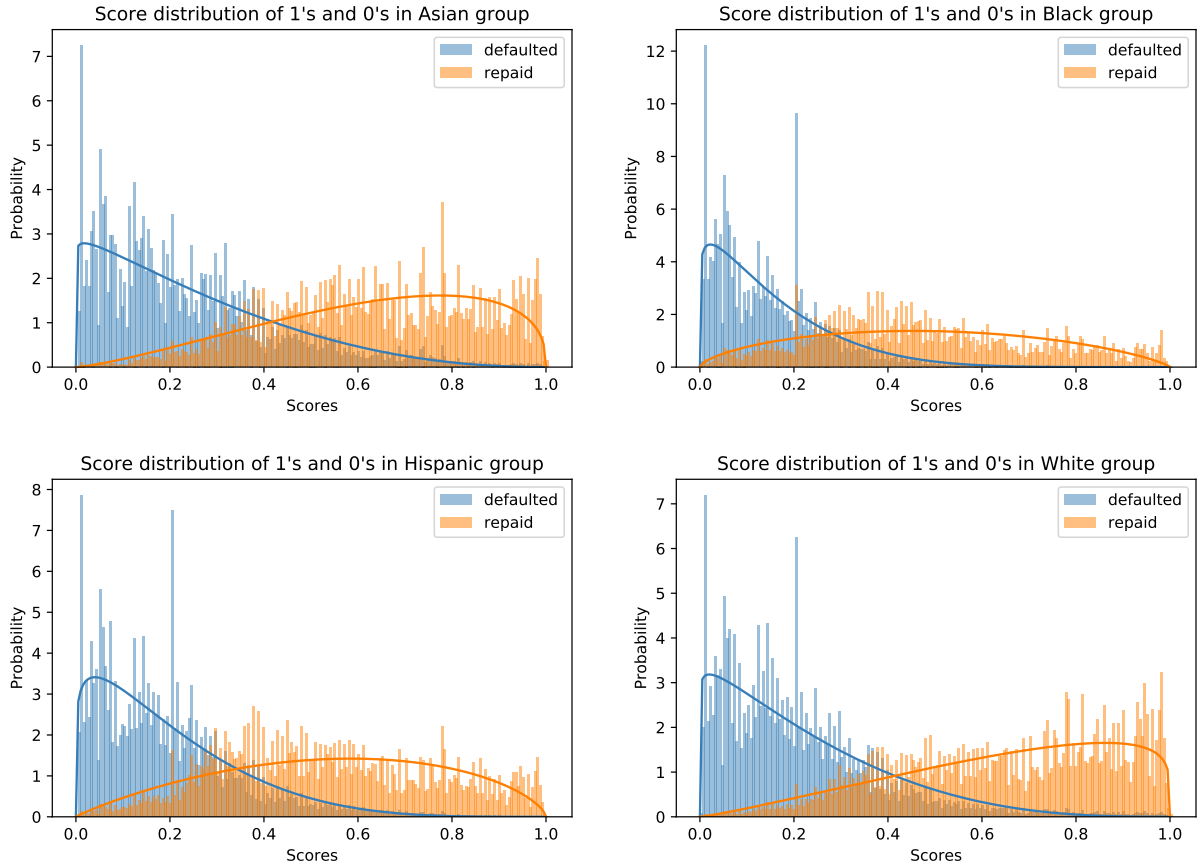


Figure 3.3: Score distributions conditioning on repayment outcome (Y) for different race groups

not to faithfully represent this particular domain with our model, but to simulate feature distributions that exhibit group heterogeneity and non-realizability, hence extending our consideration beyond the idealized settings of Sections 3.3 and 3.4.

Figure 3.3 shows the histograms as well as the fitted Beta distributions for f_0^a, f_1^a, where a is the race attribute. It is clear that group-realizability does not hold even approximately, since there is significant overlap in the distributions of credit scores for people who repaid and for people who did not repay.

Decoupling and Multiple Stable Equilibria

Although decoupling is guaranteed to improve the qualification rate at equilibrium over using a joint decision rule for every group (Sections 3.3 and 3.4), this is not necessarily true in the non-realizable setting. In fact, even when G is the uniform distribution on $[0, 1]$ in all groups (i.e. the cost of investment C is uniformly distributed on $[0, 1]$, as we considered in Section 3.4), decoupling did not benefit all groups. As can be seen from Figure 3.6 in Appendix 3.9, while the White and Asian groups had a higher qualification rate after

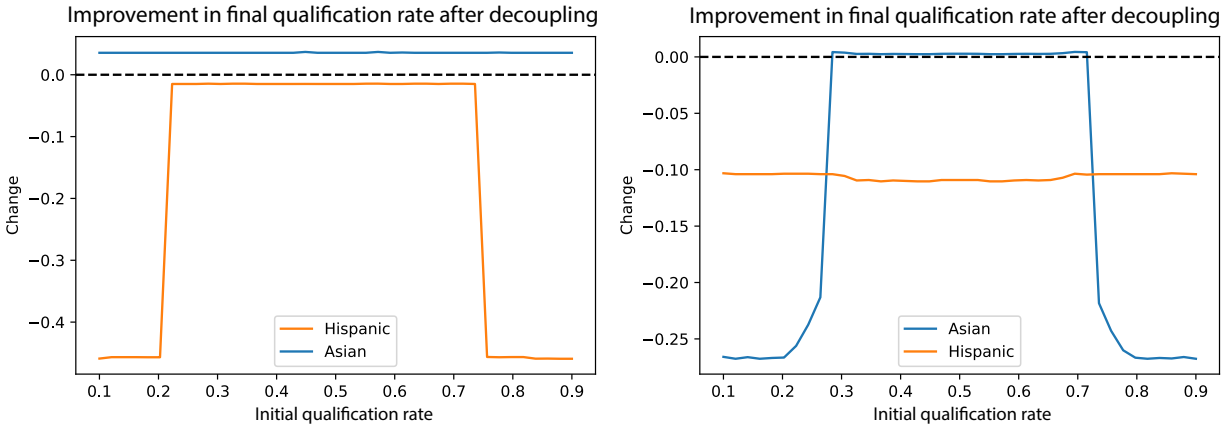


Figure 3.4: Effects of decoupling in presence of multiple equilibria. We vary the initial qualification rate in the x-axis.

decoupling, the Black and Hispanic groups saw their equilibrium qualification rate decrease. On the other hand, the effects of decoupling were small in this case (less than 3 percent points difference in the final qualification rate).

We now show that the effect of decoupling can be drastic depending on G. Recall that in Section 3.5, we showed that multiple equilibria, with possibly vastly different qualification rates, may exist under the non-realizable setting even when there is a only single group. In general the existence of multiple equilibria depends on properties of G, that is, how the cost of investment is distributed in a group. In Figure 3.4, we show the change in equilibrium investment level after decoupling for an experiment with two groups, Asian and Hispanic. The two plots each correspond to a different bimodal Gaussian distribution for G, truncated to $[0, 1]$, that have been chosen such that the decoupled dynamics have multiple stable equilibria for the Hispanic (right) and the Asian (left) respectively[8].

In both plots, we can see that the effects of decoupling depend on the initial qualification rate. If the initial qualification rate was too low, or too high, the decoupled dynamics converge to an equilibrium where one of the groups invest in qualifications at a much lower level than they would under the joint dynamics.[9]

Subsidizing the Cost of Investment

In this experiment, we consider if subsidizing the cost of investment of one group by changing G improves their new equilibrium qualification rate, under both decoupled and joint dynamics. Specifically, we vary the cost of investment in the Black group.

[8]The right (resp. left) plot is generated using a bimodal normal distribution for G with modes at 0.57 and 0.74 (resp. at 0.57 and 0.63).

[9]See Figure 3.7 in Appendix 3.9 for the converged qualification rates of both groups.

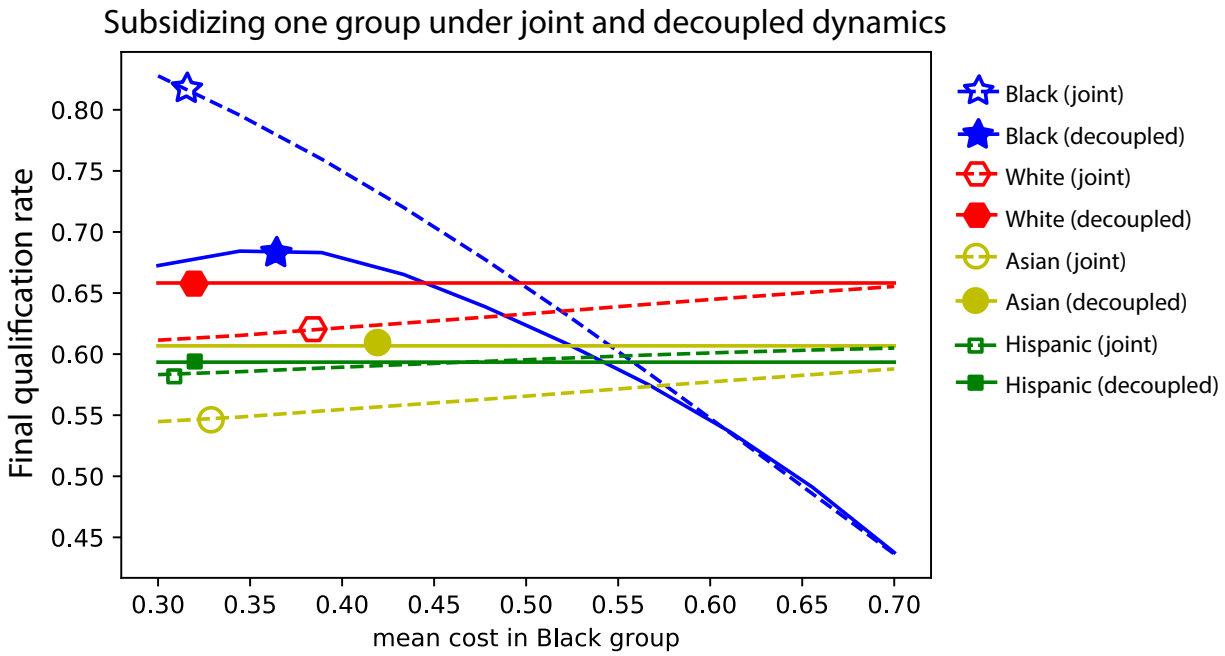


Figure 3.5: Effects of raising the average cost of investment, by varying the mean of G on the x-axis.

We use a truncated normal distribution for G and vary its mean (on the x-axis) for a single group, while keeping the other groups' G unchanged (mean of 0.6).

Figure 3.5 shows that subsidizing the cost of investment is effective in raising the equilibrium investment level of a group, both in the joint learning and decoupled learning case. Interestingly, large amounts of subsidy for a single group reduced the equilibrium investment levels of other groups. As also suggested by theoretical results in section 3.4, subsidizing the qualification rate of one group does sometimes entail a tradeoff in the qualification rates of other possibly more advantaged groups.

Interestingly, lowering the mean cost of investment in the Black group below 0.35 caused the final qualification rate to decrease. This is not a contradiction of Proposition 3.5.2, which argues that equilibria improve under subsidies but does not guarantee that the dynamics will converge to the improved equilibrium. In this case, the decoupled dynamics for the Black group (where the mean cost of investment is 0.30) actually converged to a limit cycle and the final qualification rate in the plot is an average of the points in the limit cycle. Limit cycles are a challenging object to study in dynamical systems and game theory. While we have commented on their existence in a simple model in group-realizable setting of Section 3.4, we leave their implications in the general non-realizable setting to future work.

3.7 Related Work

We follow a growing line of work on how machine learning algorithms interact with human actors in a dynamic setting, with the goal of understanding and mitigating disparate impact.

Recent work examine the long-term impact of group fairness criteria [see e.g., Barocas et al., 2018, Chapter 2] on automated decision making systems: Liu et al. [2018] show that static fairness criteria fail to account for the delayed impact that decisions have on the welfare of disadvantaged groups. In the context of hiring, however, Hu and Chen [2018a] find that applying the demographic parity constraint in a temporary labor market achieves an equitable long-term equilibrium in the permanent labor market by raising worker reputations.

Prior work on the fairness of machine learning has examined tradeoffs between fairness criteria [Kleinberg et al., Chouldechova, 2017], as well as the incompatibility between risk minimization and fairness criteria [Liu et al., 2019], assuming that the qualification rates differ across groups. These results concern the static setting, whereas we highlight the fact that qualification rates tend to change in response to the decision rules.

Another line of work [Hashimoto et al., 2018, Zhang et al., 2019] analyzes a dynamic model where users respond to errors made by an institution by leaving the user base uniformly at random, and demonstrate how the risk-minimizing approach to machine learning can amplify representation disparity over time. This is complementary to our work which models individuals as rational decision makers who may or may not have the incentive to acquire the positive label. In particular, Hashimoto et al. [2018] show that equilibria with equal user representation from all groups can be unstable, and that robust learning can stabilize such equilibria. Unlike in our model, the user representation model does not distinguish between positive and negative labels, and thus do not distinguish between false negative and false positive errors. This is a crucial distinction in high-stakes decision making as different error types present asymmetric incentives for individuals, as explained in Section 3.2; for example, a high false positive rate in hiring would encourage under-qualified job applicants.

Hu et al. [2019] and Milli et al. [2019] study the disparate impact of being robust towards strategic manipulation [see e.g., Hardt et al., 2016a], where individuals respond to machine learning systems by manipulating their features to get a better classification. In contrast to our model (Figure 3.5), their setting models the individual as intervening directly on their features, X, and this is assumed to have no effect on their qualification Y. This assumption applies to features that are easy to 'game' (e.g. scores on standardized tests can be improved by test preparation classes) but is less applicable to features that more directly correspond to *investment* in one's qualifications (e.g. taking AP courses in high school). Hu et al. [2019] also show that subsidizing the costs of the disadvantaged group to strategically manipulate their features can sometimes lead to harmful effects. Kleinberg and Raghavan [2019] and Khajehnejad et al. [2019] study decision policies that incentivize individuals to directly manipulate their features X to optimize particular notions of utility.

Our work is also related to the topic of statistical discrimination in economics [Phelps, 1972, Arrow, 1973, 1998], which studies how disparate market outcomes at equilibrium can arise from imperfect information. This line of work often involves wage discrimination,

whereas we assume the wage is fixed and standard for all groups. Coate and Loury [1993] proposed a model of rational individual investment in the labor market under a fixed wage and showed that affirmative action may lead to an undesirable equilibrium where one group still invests sub-optimally. The model in our work is most closely related to their model, with two key distinctions: 1) We allow features X to be multi-dimensional, whereas Coate and Loury [1993] assumes that X is a one-dimensional 'noisy signal'. 2) We consider the case where the conditional feature distributions, $\mathbb{P}(X = x \mid Y = y, A = a)$, differ by groups whereas Coate and Loury [1993] assumes that the groups are identically distributed. Under our models, if the conditional feature distributions were shared across groups, then *any* hiring policy will result in fully balanced equilibria where all groups have the same qualification rate and are hired at the same rate. This does not corroborate with reality, where conditional feature distributions do in fact differ across groups and we routinely observe institutions applying the same model to all individuals only to see obviously discriminatory outcomes [Dastin, 2019]. By modeling feature heterogeneity across groups, we find it necessarily leads to disparate equilibria.

Recently, Mouzannar et al. [2019] studied the equilibria of qualification rates under a generic class of dynamics, focusing on contractive maps and the effects of affirmative action. In contrast, our work motivates a model of dynamics based on rational investment, and this typically leads to non-contractive dynamics. We are both interested in balanced equilibria, which they termed 'social equality'.

Finally, our work studies two interventions for finding more desirable equilibria: decoupling the classifier and subsidizing the cost of investment. Several works, including Dwork et al. [2018] and Ustun et al. [2019], have studied decoupled classifiers in the static classification setting. Our work sheds light on when such interventions are useful in the dynamic decision making setting.

3.8 Discussion and Future Work

In this chapter, we have made the following contributions:

1. We proposed a dynamic model of decision making where individuals invest rationally based on the current assessment rule. We studied the properties of equilibria under these dynamics.

2. We showed that common properties of real data, namely heterogeneity across groups and the lack of realizability, lead to undesirable tradeoffs at equilibria, resulting in long term outcomes that disadvantage one or more groups.

3. We considered two interventions—decoupling and subsidizing the cost of investment—and showed that they have a significant impact on the nature of equilibria both in theory and in numerical experiments.

We now discuss the limitations of the current work and avenues for future research. Questions related to sampling and its ramifications for the nature of equilibria are challenging

and warrant further study. This work assumed that the institution can estimate the true and false positive rates over the entire population, even though it really can only observe the qualification of candidates *after* hiring them. This is known as the selective labeling problem, which could introduce bias. In theory, unbiased estimates can be achieved by a small degree of random sampling and appropriate reweighting [see e.g., Kilbertus et al., 2019], but this is still a large problem in practice that requires domain-specific knowledge and solutions [De-Arteaga et al., 2018, Kallus and Zhou, 2018].

Our model assumed that individuals make a rational decision to invest and can affect their qualification Y directly. This assumption could be reasonable in settings like hiring, for example, where investing to acquire skills usually leads to increased competence. In some settings, however, individuals may be unable to effectively intervene on Y. For example, a business loan applicant who is a good business operator could still default on their loan due to external economic shocks or other forms of disadvantage that have not been taken into account. In this case, the current model does not fully capture the complex societal processes that lead to a positive outcome. Our work nonetheless shows that even in an idealized setting where individuals can effectively and rationally intervene on their outcome labels Y, underlying factors such as heterogeneity across groups and non-realizability already lead to undesirable tradeoffs at equilibrium. We leave the extensions of the current model beyond rational individual investment to future work.

3.9 Omitted proofs and supplementary material

Proof of Theorem 3.3.2

For any $\pi \in [s, 1-s]$, consider the profit-maximizing classifier,

$$\theta^{br}(\pi) = \operatorname*{argmax}_{\theta \in \Theta} \ \mathrm{TPR}(\theta) \cdot \pi - \mathrm{FPR}(\theta) \cdot (1-\pi).$$

Let $\pi(0)$ be the initial qualification rate. For ease of notation, denote $\theta^* = \theta^{br}(\pi(0))$. We examine the new qualification rate π_1 under the best response model θ^*. Since there exists a θ such that $\mathrm{TPR}(\theta) \geq 1-\epsilon$ and $\mathrm{FPR}(\theta) \leq \epsilon$, we have $\mathrm{TPR}(\theta) \cdot \pi - \mathrm{FPR}(\theta) \cdot (1-\pi) \geq \pi - \epsilon$. It follows that

$$\mathrm{TPR}(\theta^*) \cdot \pi - \mathrm{FPR}(\theta^*) \cdot (1-\pi) = \pi(\mathrm{TPR}(\theta^*) - \mathrm{FPR}(\theta^*)) + (2\pi - 1)\mathrm{FPR}(\theta^*) \geq \pi - \epsilon$$

Rearranging this inequality gives the following lower bound on $\mathrm{TPR}(\theta^*) - \mathrm{FPR}(\theta^*)$:

$$\mathrm{TPR}(\theta^*) - \mathrm{FPR}(\theta^*) \geq \frac{\pi - \epsilon - (2\pi - 1)\mathrm{FPR}(\theta^*)}{\pi} \tag{3.6}$$

For $\pi < 1/2$, it follows from (3.6) that

$$\begin{aligned} \mathrm{TPR}(\theta^*) - \mathrm{FPR}(\theta^*) &\geq \frac{\pi - \epsilon}{\pi} && \text{(using } \mathrm{FPR}(\theta^*) \geq 0\text{)} \\ &\geq 1 - \frac{\epsilon}{s} && (3.7) \end{aligned}$$

For $\pi > 1/2$, we have

$$\mathrm{FPR}(\theta^*) \leq \frac{\pi \mathrm{TPR}(\theta^*) - \pi + \epsilon}{1 - \pi} \leq \frac{\epsilon}{1 - \pi}.$$

Substituting this into (3.6) gives:

$$\mathrm{TPR}(\theta^*) - \mathrm{FPR}(\theta^*) \geq \frac{\pi - \epsilon - (2\pi - 1)\frac{\epsilon}{1-\pi}}{\pi} = 1 - \frac{\epsilon}{1 - \pi} \geq 1 - \frac{\epsilon}{s} \tag{3.8}$$

Therefore the new qualification rate π_1 satisfies $\pi_1 > G(w(1 - \epsilon/s))$.

Notice that $\pi_1 \leq G(w) \leq 1 - s$ and $\pi_1 > G(w(1 - \epsilon/s)) \geq G(w) - L_G w\epsilon/s \geq s$, so we may repeat the argument to conclude that the qualification rate in the limit must be greater than $G(w(1 - \epsilon/s))$.

Proof of Corollary 3.3.3

From Theorem 3.3.2 applied to investment level $\bar{G}$, we can conclude $\bar{\pi} \geq \bar{G}(w(1 - \epsilon/s))$. By assumption $\bar{G}(w(1-\epsilon/s)) \geq G(w)$, thus it remains to show $G(w) \geq \pi$. However, this follows immediately from the fact that $1 \geq \mathrm{TPR}(\theta) - \mathrm{FPR}(\theta)$, $\forall \theta \in \Theta$.

Supplementary material and proofs for Section 3.4

Proof of Proposition 3.4.1 . First consider the policy $\hat{Y}_1 = \mathbf{1}\{X > h_1\}$. Given this policy, we have $\pi_{a_1} = G(w)$ and $\pi_{a_2} = G(w(1 - \frac{h_2 - h_1}{h_2})) = G\left(w \cdot \frac{h_1}{h_2}\right)$. $\hat{Y}_1$ is optimal for this π if the gain from true positives in group a offsets the loss from false positives in group a_2 for $X \in [h_1, h_2]$, i.e. we need

$$\frac{G(w) \cdot n_{a_1} \cdot p_{\mathrm{TP}}}{1 - h_1} > \frac{\left(1 - G\left(w \cdot \frac{h_1}{h_2}\right)\right) \cdot n_{a_2} \cdot c_{\mathrm{FP}}}{h_2}. \tag{3.9}$$

Now consider the policy $\hat{Y}_2 = \mathbf{1}\{X > h_2\}$. Given this policy, we have $\pi_{a_1} = G\left(w \cdot \frac{1-h_2}{1-h_1}\right)$ and $\pi_{a_2} = G(w)$. $\hat{Y}_2$ is optimal for this π if the gain from true positives in group a fails to offset the loss from false positives in group a_2, for $X \in [h_1, h_2]$, i.e. we need

$$\frac{G\left(w \cdot \frac{1-h_2}{1-h_1}\right) \cdot n_a \cdot p_{\mathrm{TP}}}{1 - h_1} < \frac{(1 - G(w)) \cdot n_{a_2} \cdot c_{\mathrm{FP}}}{h_2}. \tag{3.10}$$

Direct computation shows that (3.9) and (3.10) are satisfied as long as w lies in the interval

$$\left(\frac{h_2(1-h_1)}{h_2^2+h_1(1-h_1)}, \frac{(1-h_1)^2}{(1-h_2)h_2+(1-h_1)^2}\right).$$

Both equilibria above are stable since (3.9) and (3.10) hold with strict inequality. For all small enough perturbations to (π_{a_1}, π_{a_2}), h_1 (or h_2) will still remain as the profit maximizing threshold.

There exists an equilibrium at $h = h_1 + g$ if we have

$$\frac{G(w \cdot \frac{1-h_1-g}{1-h_1})}{1-h_1} = \frac{1-G(w \cdot \frac{h_1+g}{h_2})}{h_2} \tag{3.11}$$

Direct computation shows that the above equation is satisfied by a unique value of

$$g = \frac{(1-h_1)(-wh_2^2+h_2(1-h_1)-wh_1(1-h_1))}{w((1-h_1)^2-h_2^2))},$$

and that $\pi_{a_1} = \frac{w(1-h_1-g)}{1-h_1} = \pi_{a_2} = \frac{w(h_1+g)}{h_2}$ if $w = 1-h_1$.

□

For an illustration of Proposition 3.4.1, consider an example where $n_{a_1} \cdot p_{\text{TP}} = n_{a_2} \cdot c_{\text{FP}}$, $G(c) = c$ for $c < 1$ (uniformly distributed cost of investment), and $h_1 = 0.4$, $h_2 = 0.8$, which gives $h_1/h_2 = 0.5$. Let $w = 0.6$. We compute:

$$\frac{G(w)}{1-h_1} = 1, \frac{1-G\left(w \cdot \frac{h_1}{h_2}\right)}{h_2} = 0.875, \frac{G\left(w \cdot \frac{1-h_2}{1-h_1}\right)}{1-h_1} = 1/3, \frac{1-G(w)}{h_2} = 1/2,$$

and check that these satisfy the assumptions.

In this example, note that the first equilibrium has qualification rate 0.6 and 0.3 for groups a and a_2 respectively, while the second equilibrium has qualification rate 0.2 and 0.6 respectively. The first equilibrium might be more desirable since there is higher qualification rate overall, though neither equilibrium has equal qualification rates across the two groups.

Comparison of equilibria We can compare the equilibria described in Proposition 3.4.1 in terms of metrics shown in Table 3.3. There is no fixed ranking for the Institution's utility; instead the ranking varies depending on the values of h_1, h_2, and w.

Lemma 3.9.1 (Skill acquisition in group a_1)**.** *Let $w \in (w_l, w_u)$, as defined in* (3.2)*. Then for h_1, h_2, as defined in Proposition 3.4.1, we have*

$$w \cdot \frac{1-h_2}{1-h_1} < \frac{(w-h_2)(1-h_1)}{(1-h_1)^2-h_2^2} < w \tag{3.12}$$

Equilibrium h	h_1	h_2	$h_m := h_1 + g$	**R**anking
Stability	Stable	Stable	Unstable	-
Qualification rate in group a_1, π_{a_1}	w	$w \cdot \frac{1-h_2}{1-h_1}$	$\frac{(w-h_2)(1-h_1)}{(1-h_1)^2-h_2^2}$	$h_1 \succ h_m \succ h_2$ (Lem. 3.9.1)
Qualification rate in group a_2, π_{a_2}	$w \cdot \frac{h_1}{h_2}$	w	$\frac{(1-h_1)^2-wh_2}{(1-h_1)^2-h_2^2}$	$h_2 \succ h_m \succ h_1$ (Lem. 3.9.2)
Balance in qualification rate $\lvert\pi_{a_1} - \pi_{a_2}\rvert$	$w \cdot \frac{h_2-h_1}{h_2}$	$w \cdot \frac{h_2-h_1}{1-h_1}$	$\frac{\lvert 1-h_1-w\rvert}{h_2-(1-h_1)}$	$h_m \succ h_1 \succ h_2$ (Lem. 3.9.3)

Table 3.3: Comparison of equilibria for uniform scores. In this table we refer to each equlibria using the associated threshold decision policy.

Proof. First we show that $\frac{(w-h_2)(1-h_1)}{(1-h_1)^2-h_2^2} < w$ for all $w \in (w_l, w_u)$. It suffices to show that

$$\left(1 - \frac{h_2}{w}\right)(1-h_1) \leq (1-h_1)^2 - h_2^2$$

holds for $w = w_u$, since the LHS is strictly increasing in w. We may check by computation that the above in fact holds with equality.

Next we show that $\frac{(w-h_2)(1-h_1)}{(1-h_1)^2-h_2^2} > w \cdot \frac{1-h_2}{1-h_1}$ for all $w \in (w_l, w_u)$. This amounts to showing that

$$\left(1 - \frac{h_2}{w}\right)(1-h_1)^2 \geq (1-h_2)((1-h_1)^2 - h_2^2)),$$

for $w = w_l$, since the LHS is strictly increasing in w. We may check by computation that the above in fact holds with equality. □

Lemma 3.9.2 (Qualification rate in group a_2)**.** *Let $w \in (w_l, w_u)$, as defined in* (3.2)*. Then for h_1, h_2, as defined in Proposition 3.4.1, we have*

$$w \cdot \frac{h_1}{h_2} < \frac{(1-h_1)^2 - wh_2}{(1-h_1)^2 - h_2^2} < w \tag{3.13}$$

Proof. First we show that $\frac{(1-h_1)^2-wh_2}{(1-h_1)^2-h_2^2} < w$ for all $w \in (w_l, w_u)$. It suffices to show that

$$\frac{(1-h_1)^2}{w} - h_2 \leq (1-h_1)^2 - h_2^2)$$

holds for $w = w_l$, since the LHS is strictly decreasing in w. We may check by computation that the above in fact holds with equality.

Next we show that $w \cdot \frac{h_1}{h_2} < \frac{(1-h_1)^2 - wh_2}{(1-h_1)^2 - h_2^2}$ for all $w \in (w_l, w_u)$. This amounts to showing that

$$\frac{h_2(1-h_1)^2}{w} - h_2^2 \geq h_1((1-h_1)^2 - h_2^2)$$

for $w = w_u$, since the LHS is strictly decreasing in w. We may check by computation that the above in fact holds with equality. □

Lemma 3.9.3 (Unstable Equilibrium is the most balanced)**.** *Let $w \in (w_l, w_u)$, as defined in* (3.2)*. Then for h_1, h_2, as defined in Proposition 3.4.1, we have*

$$\frac{|1-h_1-w|}{h_2-(1-h_1)} < w \cdot \frac{h_2-h_1}{h_2} < w \cdot \frac{h_2-h_1}{1-h_1} \tag{3.14}$$

Proof. First note that since $h_2 > 1-h_1$ by assumption, we have $w \cdot \frac{h_2-h_1}{h_2} < w \cdot \frac{h_2-h_1}{1-h_1}$, so it suffices to show that

$$\frac{|\frac{1-h_1}{w} - 1|}{h_2-(1-h_1)} < \frac{h_2-h_1}{h_2}$$

for all $w \in (w_l, w_u)$. Consider the first case: $w \in (w_l, 1-h_1]$. We want to show

$$\frac{\frac{1-h_1}{w} - 1}{h_2-(1-h_1)} < \frac{h_2-h_1}{h_2}. \tag{3.15}$$

Since the LHS is decreasing in w, it suffices to show that (3.15) holds for $w = w_l$. By computation, we have following:

$$\frac{\frac{1-h_1}{w_l} - 1}{h_2-(1-h_1)} < \frac{h_2-h_1}{h_2} \iff (h_2+1)(h_2-h_1)(h_1+h_2-1) > 0, \tag{3.16}$$

which is indeed satisfied since we have $h_2 > h_1$ and $h_2 > 1-h_1$ by assumption.

Now consider the second case: $w \in (w_l, 1-h_1)$. We want to show

$$\frac{1 - \frac{1-h_1}{w}}{h_2-(1-h_1)} < \frac{h_2-h_1}{h_2}. \tag{3.17}$$

Since the LHS is increasing in w, it suffices to show that (3.17) holds for $w = w_u$. By computation, we have following:

$$\frac{1 - \frac{1-h_1}{w_u}}{h_2-(1-h_1)} < \frac{h_2-h_1}{h_2} < \frac{h_2-h_1}{h_2} \iff (h_2-h_1)(h_1+h_2-1) > 0, \tag{3.18}$$

which is indeed satisfied since we have $h_2 > h_1$ and $h_2 > 1-h_1$ by assumption.

□

Supplementary material and proofs for Section 3.4

Proof of Proposition 3.4.3 . Denote $r := \frac{p_{\text{TP}}}{c_{\text{FP}}}$. For any hyperplane h, we may compute the true positive rate and false positive rate for a group with hyperplane h_i as follows:

$$\text{TPR}_{a_i} = 1 - \angle_{h,h_i}, \quad \text{FPR}_{a_i} = \angle_{h,h_i}. \tag{3.19}$$

Therefore, for any investment levels (π_{a_1}, π_{a_2}), the firm solves the following profit maximization problem:

$$\begin{aligned} h^* &= \operatorname*{argmax}_{h \in S_{d-1}} \; r\pi_{a_1}(1 - \angle_{h,h_1}) - (1 - \pi_{a_1})\angle_{h,h_1} + r\pi_{a_2}(1 - \angle_{h,h_2}) - (1 - \pi_{a_2})\angle_{h,h_2} \\ &= \operatorname*{argmax}_{h \in S_{d-1}} \; (1 - r)\pi_{a_1}\angle_{h,h_1} + (1 - r)\pi_{a_2}\angle_{h,h_2} - (\angle_{h,h_1} + \angle_{h,h_2}). \end{aligned}$$

The last term $\angle_{h,h_1} + \angle_{h,h_2}$ is minimized whenever h is in the convex hull of h_1 and h_2. Then it is clear that for $r > 1$, the profit is maximized at $h = h_1$ whenever $\pi_{a_1} > \pi_{a_2}$, at $h = h_2$ whenever $\pi_{a_1} < \pi_{a_2}$, and at any h in the convex hull of h_1 and h_2 whenever $\pi_{a_1} = \pi_{a_2}$.

To conclude that $h = h_1$ and $h = h_2$ are indeed stable equilibria, we check the best response qualification rates by both groups at $h = h_1$ and $h = h_2$ satisfies the optimality conditions. For $h = h_1$, we have $\pi^{br}_{a_1}(h_1) = G(w)$ and $\pi^{br}_{a_2}(h_1) = G\left(w \cdot (1 - 2\angle_{h_1,h_2})\right)$, and indeed $\pi^{br}_{a_1} > \pi^{br}_{a_2}$, by the monotonicity of G. For $h = h_2$, we have $\pi^{br}_{a_1}(h_2) = G\left(w \cdot (1 - 2\angle_{h_1,h_2})\right)$ and $\pi^{br}_{a_2}(h_2) = G(w)$, and indeed $\pi^{br}_{a_1} < \pi^{br}_{a_2}$.

Now we identify the unstable equilibrium, which is $h = h_{\text{mid}}$, because the qualification rate is indeed equal for both groups under this policy, i.e., we have

$$\pi^{br}_{a_1} = G(w(1 - 2\angle_{h_1,h_{\text{mid}}})) = G(w(1 - \angle_{h_1,h_2})) = G(w(1 - 2\angle_{h_2,h_{\text{mid}}})) = \pi^{br}_{a_2}.$$

When both groups are investing at this rate, we may assume that the institution's best response involves breaking ties among all utility-maximizing hyperplanes to choose h_{mid}. This ensures that the dynamics are well-defined. Notice that this is an unstable equilibrium, since this is the unique value of h such that $\pi^{br}_{a_2}(h) = \pi^{br}_{a_2}(h)$, and any deviation from equal qualification rates will change the profit-maximizing hyperplane to h_1 or h_2. □

Proof of Proposition 3.4.4 . Following the proof of Proposition 3.4.3, we find that for $p_{\text{TP}} < c_{\text{FP}}$, the profit is maximized at $h = h_1$ whenever $\pi_{a_1} < \pi_{a_2}$, at $h = h_2$ whenever $\pi_{a_1} > \pi_{a_2}$, and at any h in the convex hull of h_1 and h_2 whenever $\pi_{a_1} = \pi_{a_2}$.

Checking the qualification rate at $h = h_1$ (resp. $h = h_2$), we find that $\pi^{br}_{a_1}(h_1) = G(w)$ and $\pi^{br}_{a_2}(h_1) = G\left(w \cdot (1 - \angle_{h_1,h_2})\right)$, so $\pi^{br}_{a_1}(h_1) > \pi^{br}_{a_2}(h_1)$. Similarly, we have $\pi^{br}_{a_1}(h_2) < \pi^{br}_{a_2}(h_2)$. This implies there is a 2-point limit cycle at $h = h_1$ and $h = h_2$.

As before, the unstable equilibrium is at $h = h_{\text{mid}}$, because the qualification rate is indeed equal for both groups under this policy. Notice that this is an unstable equilibrium, since this is the unique value of h such that $\pi^{br}_{a_2}(h) = \pi^{br}_{a_2}(h)$. □

Equilibrium h	h_1	h_2	h_{mid}	Ranking by metric
Stability	Stable	Stable	Unstable	-
Qualification rate in group a_1, π_{a_1}	w	$w(1-2\angle)$	$w(1-\angle)$	$h_1 \succ h_{\text{mid}} \succ h_2$
Qualification rate in group a_2, π_{a_2}	$w(1-2\angle)$	w	$w(1-\angle)$	$h_2 \succ h_{\text{mid}} \succ h_1$
Balance in qualification rate $\lvert\pi_{a_1} - \pi_{a_2}\rvert$	$2w\angle$	$2w\angle$	$w\angle$	$h_{\text{mid}} \succ h_1 \sim h_2$
Institution's utility	$p_{\text{TP}}w(2-3\angle) + 2(p_{\text{TP}} - c_{\text{FP}})w\angle^2$	$p_{\text{TP}}w(2-3\angle) + 2(p_{\text{TP}} - c_{\text{FP}})w\angle^2$	$p_{\text{TP}}w(2-3\angle) + (p_{\text{TP}} - c_{\text{FP}})w\angle^2$	$h_1 \sim h_2 \succ h_{\text{mid}}$

Table 3.4: Comparison of equilibria for Multivariate Gaussian features

Comparison of equilibria In Table 3.4, we compare the equilibria described in Proposition 3.4.3 on several metrics. We use $\angle$ to denote $\angle_{h_1,h_2}$.

Supplementary material for Section 3.4

Proof of Proposition 3.4.5 . First notice that by assumption, $h = h_1$ cannot be at equilibrium. It is easy to check that $G_1(w(1-2\angle_{h_1,h_2})) \leq G_1(w) < G_2(w(1-2\angle_{h_1,h_2})) \leq G_2(w)$, so $h = h_2$ is still at a stable equilibrium. Now, for any h that is a convex combination of h_1 and h_2, we have $G_1(w(1-2\angle_{h,h_1})) \leq G_1(w) < G_2(w(1-2\angle_{h_1,h_2})) \leq G_2(w(1-2\angle_{h,h_2}))$, implying that $G_1(w(1-2\angle_{h,h_1})) \neq G_2(w(1-2\angle_{h,h_2}))$ for all h that maximize institutional utility, so no other fixed points exist. □

Supplementary material and proofs for Section 3.5

We first prove Proposition 3.5.1.

Proof of Proposition 3.5.1. When $\pi = 1$, the institution's best response is to accept everyone regardless of their features, so $\beta(1) = \text{TPR}(\theta') - \text{FPR}(\theta') = 1 - 1 = 0$.[10]

[10]Recall that $\text{FPR}(\theta) := \mathbb{P}(\hat{Y}_\theta = 1 \mid do(Y = 0))$, which is well-defined even when $\mathbb{P}(Y = 0) = 0$.

Note that $\frac{1-\pi}{\pi} \to \infty$ as $\pi \to 0$. This, together with the fact that $\phi(x)$ is strictly positive means that there must exist $\bar{\pi} > 0$ such that $\frac{p_{\text{TP}}}{c_{\text{FP}}} < \frac{1-\bar{\pi}}{\bar{\pi}}\phi(x)$ for all $x \in \mathcal{X}$. Therefore, for all $\pi \leq \bar{\pi}$, the institution's best response is to accept no one regardless of their features, so we have that $\beta(\pi) = \text{TPR}(\theta) - \text{FPR}(\theta) = 0 - 0 = 0$ for $\pi \leq \bar{\pi}$.

Since $G(0) = 0$, we have that $\bar{\pi} > G(w\beta(\bar{\pi})) = 0$ and $1 > G(w\beta(1)) = 0$. By assumption there exists $x < G(w\beta(x))$ for some $x \in (0, 1)$, and by the above discussion, we must have $x \in (\bar{\pi}, 1)$. Hence there must be at least 2 solutions to $\pi = \Phi(\pi)$ in $(\bar{\pi}, 1)$ and in particular they are non-zero. The condition for local stability follows directly from chain rule. □

The following result from Coate and Loury [1993] establishes conditions under which multiple equilibria exists for a single group when the features $\mathcal{X} = [0, 1]$ represent a score and the assessment rule is a threshold function. For completeness, we show that it can be derived as a consequence of Proposition 3.5.1.

Proposition 3.9.4 (Proposition 1 of Coate and Loury [1993])**.** *Consider the case where $\mathcal{X} = [0, 1]$ is a space of one-dimensional scores, $\Theta = [0, 1]$, and $\hat{Y}_\theta = \mathbf{1}\{X > \theta\}$ for all $\theta \in \Theta$. Denote the conditional score CDFs as*

$$F_1(x) \coloneqq \mathbb{P}(X < x \mid Y = 1),\ F_0(x) \coloneqq \mathbb{P}(X < x \mid Y = 0).$$

Let $f_1(x), f_0(x)$ be the point densities of F_1 and F_0, respectively. Let $\phi(x) \coloneqq \frac{f_0(x)}{f_1(x)}$ be the likelihood ratio at x. Let $r \coloneqq \frac{p_{\text{TP}}}{c_{\text{FP}}}$ be the ratio of net gain to loss for the firm. Assume $\phi(x)$ is strictly decreasing — i.e. as score increases, candidate is more likely to be skilled — continuous and strictly positive on $[0, 1]$. Further assume that $G(c)$ is continuous and $G(w(F_0(\theta) - F_1(\theta))) > \frac{\phi(\theta)}{r+\phi(\theta)}$ for some $\theta \in (0, 1)$. Then there exists at least two distinct non-zero equilibria.

Proof of Proposition 3.9.4. Note that for any θ, $\text{TPR}(\theta) - \text{FPR}(\theta) = F_0(\theta) - F_1(\theta)$. Therefore, the group's qualification rate in response to assessment parameter θ is

$$\pi^{br}(\theta) = G(w(F_0(\theta) - F_1(\theta))). \tag{3.20}$$

Since $\phi(x)$ is strictly decreasing, the utility maximizing assessment rule θ in response to the qualification rate π is

$$\theta^{br}(\pi) = \inf\left\{x \in [0, 1] : r \geq \frac{1-\pi}{\pi} \cdot \phi(x)\right\}. \tag{3.21}$$

Since $\phi(x)$ is continuous and strictly positive, we must also have that F_0, F_1 are continuous, and so in particular $\beta(\pi) = F_0(\theta^{br}(\pi)) - F_1(\theta^{br}(\pi))$ is continuous. By assumption, G is continuous and there exists $x \in (0, 1)$ such that $x < G(w\beta(x))$ since $\theta^{br}(\pi)$ is surjective. Therefore, the claim follows from Proposition 3.5.1. □

Proposition 3.5.2 follows from the monotonicity of the CDF, G.

Proof of Proposition 3.5.2. By assumption, we have that $\bar{G}^{-1}(\pi^*) < G^{-1}(\pi^*)$, so $\bar{G}^{-1}(\pi^*) < w\beta(\pi^*)$. Since $\bar{G}^{-1}(1) > \beta(1)$, we must have $\bar{G}^{-1}(\bar{\pi}) = w\beta(\pi)$ for some $\bar{\pi} \in (\pi^*, 1)$. □

Supplementary material for Section 3.6

We collect here additional figures for Section 3.6.

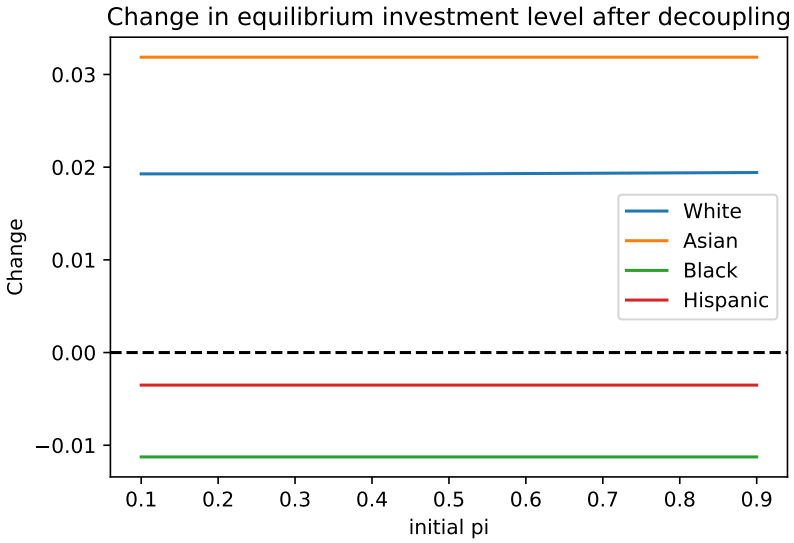


Figure 3.6: Effects of decoupling without multiple equilibria. G is the uniform distribution on $[0, 1]$ for all groups and the reward is $w = 1$. The decoupled equilibria are unique for this choice of G.

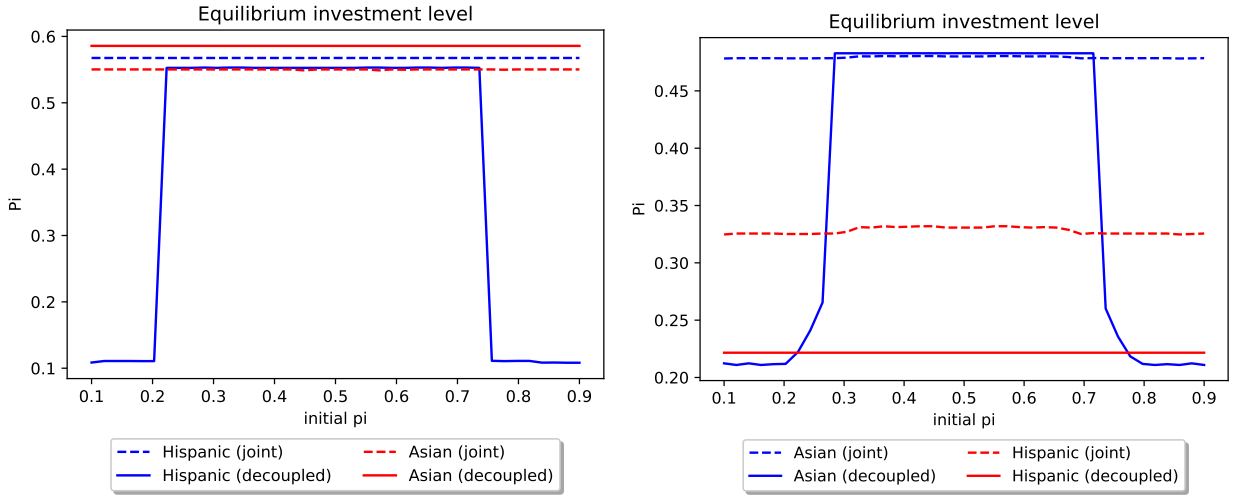


Figure 3.7: Effects of decoupling in presence of multiple equilibria. We vary the initial level of investment in the x-axis. A different bimodal Gaussian distribution G was used to generated each plot.

Chapter 4

Competing Bandits in a Centralized Matching Market

4.1 Introduction

In previous chapters, we examined algorithmic decision making and its potentially disparate impact on populations over time. In our stylized models, institutional decision makers employ decision rules that minimize current population risk, without explicitly accounting for future outcomes and data collection. In this chapter and next, we turn to the different setting of *online learning*, where decision makers are repeatedly confronted with the same decisions under uncertainty, and must simultaneously optimize for current outcomes—"exploitation"—and future outcomes through data collection—"exploration".

We focus on the multi-arm bandit (MAB) problem, a core machine-learning problem in which there are K actions giving stochastic rewards, and the learner must discover which action gives maximal expected reward [Bubeck and Cesa-Bianchi, 2012, Lai and Robbins, 1985a, Lattimore and Szepesvari, 2019, Thompson, 1933]. The bandit problem highlights the fundamental tradeoff between exploration and exploitation. Regret bounds quantify this tradeoff. We study an *economic* version of the problem in which there are *multiple* players solving a bandit problem, and there is *competition*—if two or more players pick the same arm, only one of the players is given a reward.[1] We assume that the arms have a *preference* ordering over the players—a key point of departure from the line of work on multi-player bandits with collisions [Bubeck et al., 2020b, Cesa-Bianchi et al., 2016, Liu and Zhao, 2010, Shahrampour et al., 2017]—and this ordering is unknown a priori to the players.

We are motivated by problems involving two-sided markets that link producers and consumers or workers and employers, where each side sees the other side via a recommendation system, and where there is scarcity on the supply side (for example, a restaurant has a lim-

[1]Note that Mansour et al. [2018] and Aridor et al. [2019] have used the term "competing bandits" for a different problem formulation where a user can choose between two different bandit algorithms; this differs from our setting where multiple learners compete over scarce resources.

ited number of seats, a street has a limited capacity, or a worker can attend to one task at a time). The overall goal is an economic one—we wish to find a stable matching between producers and consumers. (In the end of this section, we illustrate an application in online labor markets.) In the context of two-sided markets the arms' preferences can be explict, e.g. when the arms represent entities in the market with their own utilities for the other side of the market, or implicit, e.g. when the arms represent resources their "preferences" encode the skill levels of the players in securing those resources.

To determine the appropriate notions of equilibria in our multi-player MAB model, we turn to the literature on stable matching in two-sided markets Gale and Shapley [1962], Gusfield and Irving [1989], Roth and Sotomayor [1990], Knuth [1997], Roth [2008]. Since its introduction by Gale and Shapley [1962], the stable matching problem has had high practical impact, leading to improved matching systems for high-school admissions and labor markets Roth [1984], house allocations with existing tenants Abdulkadiroglu and Sonmez [1999], content delivery networks Maggs and Sitaraman [2015], and kidney exchanges Roth et al. [2005].

In spite of these advances, standard matching models tend to assume that entities in the market know their preferences over the other side of the market. Models that allow unknown preferences usually assume that preferences can be discovered through one or few interactions Ashlagi et al. [2017a], e.g., one interview per candidate in the case of medical residents market Roth [1984], Roth and Sotomayor [1990]. These assumptions do not capture the statistical uncertainty inherent in problems where data informs preferences. We defer an in-depth discussion of related work to Chapter 5, Section 5.8.

In contrast, our work is motivated by modern matching markets which operate at scale and require repeated interactions between the two sides of the market, leading to exploration-exploitation tradeoffs. We consider two-sided markets in which entities on one side of the market do not know their preferences over the other side, and develop matching and learning algorithms that can provably attain a stable market outcome in this setting. Our contributions are as follows:

- We introduce a new model for understanding two-sided markets in which one side of the market does not know its preferences over the other side, but is allowed multiple rounds of interaction. Our model combines work on multi-armed bandits with work on stable matchings. In particular, we define two natural notions of regret, based on stable matchings of the market, which quantify the exploration-exploitation trade-off for each individual player.

- We extend the Explore-then-Commit (ETC) algorithm for single player MAB to our multi-player setting. We prove $\mathcal{O}(\log(n))$ problem-dependent upper bounds on the regret of each player.

- In addition to the known limitations of ETC for single player MAB, in Section 4.3 we discuss other issues with ETC in the multi-player setting. To address these issues we introduce a centralized version of the well-known upper confidence bound (UCB) algorithm.

We prove that centralized UCB achieves $\mathcal{O}(\log(n))$ problem-dependent upper bounds on the regret of each player. Moreover, we show that centralized UCB is approximately incentive compatible in a long-term sense.

Most of the above results can be extended to the case where arms also have uncertain preferences over players in a straightforward manner. For the sake of simplicity, we focus on the setting where one side of market initiates the exploration and leave extensions of our results to future work.

Online labor markets Our model is applicable to matching problems that arise in online labor markets (e.g., Upwork and Taskrabbit for freelancing, Handy for housecleaning) and online crowdsourcing platforms (e.g., Amazon Mechanical Turks). In this case, the employers, each with a stream of similar tasks to be delegated, can be modeled as the players, and the workers can be modeled as the arms. For an employer, the mean reward received from each worker when a task is completed corresponds to how well the task was completed (e.g., did the Turker label the picture correctly?). This differs for each worker due to differing skill levels, which the employer does not know a priori and must learn by exploring different workers. A worker has preferences over different types of tasks (e.g., based on payment or prior familiarity the task) and can only work on one task at a time; hence they will pick their most preferred task to complete out of all the tasks that are offered to them.

4.2 Problem setting

We denote the set of N players by $\mathcal{N} = \{p_1, p_2, \ldots, p_N\}$ and the set of K arms by $\mathcal{K} = \{a_1, a_2, \ldots, a_K\}$. We assume $N \leq K$. At time step t, each player p_i selects an arm $m_t(i)$, where $m_t \in \mathcal{K}^N$ is the vector of all players' selections.

When multiple players select the same arm only one player is allowed to pull the arm, according to the arm's preferences via a mechanism we detail shortly. Then, if player p_i successfully pulls arm $m_t(i)$ at time t, they are said to be *matched* to $m_t(i)$ at time t and they receive a stochastic reward $X_{i,m_t}(t)$ sampled from a 1-sub-Gaussian distribution with mean $\mu_i(m_t(i))$.

Each arm a_j has a fixed known ranking π_j of the players, where $\pi_j(i)$ is the rank of player p_i. In other words, π_j is a permutation of $[N]$ and $\pi_j(i) < \pi_j(i')$ implies that arm a_j prefers player p_i to player $p_{i'}$. If two or more players attempt to pull the same arm a_j, there is a *conflict* and only the top-ranked player successfully pulls the arm to receive a reward; the other player(s) $p_{i'}$ is said to be *unmatched* and does not receive any reward, that is, $X_{i',m_t}(t) = 0$. As a shorthand, the notation $p_i \succ_j p_{i'}$ means that arm a_j prefers player p_i over $p_{i'}$. When arm a_j is clear from context, we simply write $p_i \succ p_{i'}$. Similarly, the notation $a_j \succ_i a_{j'}$ means that p_i prefers arm a_j over $a_{j'}$, i.e. $\mu_i(j) > \mu_i(j')$.

Given the full preference rankings of the arms and players, arm a_j is called a *valid match* of player p_i if there exists a stable matching according to those rankings such that a_j and

p_i are matched. We say a_j is the *optimal match* of player p_i if it is the most preferred valid match. Similarly, we say a_j is the *pessimal match* of player p_i if it is the least preferred valid match. Given complete preferences, the Gale-Shapley (GS) algorithm [Gale and Shapley, 1962] finds a stable matching after repeated proposals from one side of the market to the other. The matching returned by the GS algorithm is always optimal for each member of the proposing side and pessimal for each member of the non-proposing side [Knuth, 1997].

We denote by $\overline{m}$ and $\underline{m}$ the functions from $\mathcal{N}$ to $\mathcal{K}$ that define the optimal and pessimal matchings of the players according to the true preferences of the players and arms. Then, it is natural to define the *player-optimal stable regret* of player p_i as

$$\overline{R}_i(n) := n\mu_i(\overline{m}(i)) - \sum_{t=1}^{n} \mathbb{E}X_{i,m_t}(t), \tag{4.1}$$

because when the arms' mean rewards are known the GS algorithm outputs the optimal matching $\overline{m}$, and in online learning, regret is generally defined so that the reward of the player is as good as the reward of playing the best action in hindsight at every time step. However, as we show in the sequel, there is a desirable class of centralized algorithms which cannot achieve sublinear player-optimal stable regret. Therefore, we also consider the *player-pessimal stable regret* defined by

$$\underline{R}_i(n) := n\mu_i(\underline{m}(i)) - \sum_{t=1}^{n} \mathbb{E}X_{i,m_t}(t). \tag{4.2}$$

Throughout we assume that the players cannot observe each other's rewards or confidence intervals for the arms' mean rewards. Now, we described a so-called "centralized" setting that determines how players may interact with arms—via a coordinating platform. In Chapter 5, we discuss a "decentralized" setting, where such a platform is absent.

Centralized: At each time step the players are required to send a ranking of the arms to a matching platform. Then, the platform decides the action vector m_t. In this work we consider two platforms. The first platform (shown in on the left of Table 4.1) outputs a random assignment for a number of time steps and then computes the player-optimal stable matching according to the players' preferences. The second platform (shown on the right of Table 4.1) takes in the player's preferences at each time step and outputs a stable matching between the players and arms. Both platforms ensure that there will be no conflicts between the players. The first platform corresponds to an explore-then-commit strategy. When the second platform is used the players must rank arms in a way which enables exploration and exploitation. We show that ranking according to upper confidence bounds yields $\mathcal{O}(\log(n))$ player-pessimal stable regret.

input: h, and the preference ranking π_j of all arms $a_j \in \mathcal{K}$, the horizon length n

for $t = 1, \dots, T$ **do**
  if $t \leq hK$ **then**
    $m_t(i) \leftarrow a_{t+i-1 \pmod K+1}$, $\forall i$.
  else if $t = hK + 1$ **then**
    Receive rankings $\widehat{r}_{i,t}$ from all p_i.
    Compute player-optimal stable matching $m_t(i)$ according to $\widehat{r}_{i,t}$ and π_j.
  else
    $m_t(i) \leftarrow m_{hK+1}(i)$, $\forall i$.

input: the preference ranking π_j of all arms $a_j \in \mathcal{K}$

for $t = 1, \dots, T$ **do**
  Receive rankings $\widehat{r}_{i,t}$ from all p_i.
  Compute player-optimal stable matching m_t according to all $\widehat{r}_{i,t}$ and π_j.

Table 4.1: (*left*) Explore-then-Commit Platform. (*right)* Gale-Shapley Platform.

4.3 Multi-player bandits with a platform

Centralized Explore-then-Commit

In this section we give a guarantee for the explore-then-commit planner defined in Algorithm 4.1(left). At each iteration, each player p_i updates their mean reward for arm j to be

$$\widehat{\mu}_{i,j}(t) = \frac{1}{T_{i,j}(t)} \sum_{s=1}^{t} \mathbf{1}\{m_s(i) = j\} X_{i,m_s}(s), \tag{4.3}$$

where $T_{i,j}(t) = \sum_{s=1}^{t} \mathbf{1}\{m_t(i) = j\}$ is the number of times player p_i successfully pulled arm a_j. At each time step, player p_i ranks the arms in decreasing order according to $\widehat{\mu}_{i,j}(t)$ and sends the resulting ranking $\widehat{r}_{i,t}$ to the platform. As seen in Table 4.1, for the first hK time steps, the platform assigns players to arms cyclically, ensuring that each player samples every arm h times. We now provide a regret analysis of centralized ETC. The proof is deferred to Section 4.3.

Theorem 4.3.1. *Suppose all players rank arms according to the empirical mean rewards* (4.3) *and submit their rankings to the explore-then-commit platform. Let* $\overline{\Delta}_{i,j} = \mu_i(\overline{m}(i)) - \mu_i(j)$, $\overline{\Delta}_{i,\max} = \max_j \overline{\Delta}_{i,j}$, *and* $\Delta = \min_{i \in [N]} \min_{j:\, \overline{\Delta}_{i,j} > 0} \overline{\Delta}_{i,j} > 0$. *Then, the expected player-optimal regret of player* p_i *is upper bounded by*

$$\overline{R}_i(n) \leq h \sum_{j=1}^{K} \overline{\Delta}_{i,j} + (n - hK) \overline{\Delta}_{i,\max} NK \exp\left(-\frac{h\Delta^2}{4}\right). \tag{4.4}$$

In particular, if $h = \max\left\{1, \frac{4}{\Delta^2}\log\left(1+\frac{n\Delta^2 N}{4}\right)\right\}$, *we have*

$$\overline{R}_i(n) \leq \max\left\{1, \frac{4}{\Delta^2}\log\left(1+\frac{n\Delta^2 N}{4}\right)\right\}\sum_{j=1}^{K}\overline{\Delta}_{i,j} + \frac{4K\overline{\Delta}_{i,\max}}{\Delta^2}\log\left(1+\frac{n\Delta^2 N}{4}\right). \quad (4.5)$$

This result shows that centralized ETC achieves $\mathcal{O}(\log(n))$ player-optimal stable regret when the number of exploration rounds is chosen apriopriately. As is the case for single player ETC, centralized ETC requires knowledge of both the horizon n and the minimum gap Δ [see, e.g., Lattimore and Szepesvari, 2019, Chapter 6]. However, a glaring difference between the the settings is that in the latter the regret of each player scales with $1/\Delta^2$, where Δ is the minimum reward gap between the optimal match and a suboptimal arm across all players. In other words, the regret of an player might depend on the suboptimality gap of other players. Example 4.3.1 shows that this dependence is real in general and not an artifact of our analysis. Moreover, while single player ETC achieves $\mathcal{O}(\sqrt{n})$ problem-independent regret, Example 4.3.1 shows that centralized ETC does not have this desirable property. Finally, $\sum_{j=1}^{K}\overline{\Delta}_{i,j}$ could be negative for some players. Therefore, some players can have negative player-optimal regret, an effect that never occurs in the single player MAB problem.

Example 4.3.1 (The dependence on $1/\Delta^2$ cannot be improved in general)**.** *Let* $\mathcal{N} = \{p_1, p_2\}$ *and* $\mathcal{K} = \{a_1, a_2\}$ *with true preferences:*

$$p_1 \colon a_1 \succ a_2 \qquad a_1 \colon p_1 \succ p_2$$
$$p_2 \colon a_2 \succ a_1 \qquad a_2 \colon p_1 \succ p_2.$$

The player-optimal stable matching is given by $\overline{m}(1) = 1$ *and* $\overline{m}(2) = 2$. *Both* a_1 *and* a_2 *prefer* p_1 *over* p_2. *Therefore, at the end of the exploration stage* p_1 *is matched to their top choice arm while* p_2 *is matched to the remaining arm. In order for* p_2 *to be matched to their optimal arm,* p_1 *must correctly determine that they prefer* a_2 *over* a_1. *The number of exploration rounds would then have to be* $\Omega(1/\overline{\Delta}_{1,2}^2)$ *where* $\overline{\Delta}_{1,2} = \mu_1(2) - \mu_1(2)$. *Hence, when* $\overline{\Delta}_{1,2} \leq 1/\sqrt{n}$, *the regret of* p_2 *is* $\Omega(n\overline{\Delta}_{2,1})$. *Figure 4.1a depicts this effect empirically; we observe that a smaller gap* $\overline{\Delta}_{1,2}$ *causes* p_1 *to have larger regret.*

Proof of Theorem 4.3.1

First we present two instructive lemmas that are used in the proof of Theorem 4.3.1, Throughout the remainder of this section, we say the ranking $\widehat{r}_{i,t}$ submitted by p_i at time t is *valid* if whenever an arm a_j is ranked higher than $\overline{m}(i)$, i.e. $\widehat{r}_{i,j}(t) < \widehat{r}_{i,\overline{m}(i)}(t)$, it follows that $\mu_i(j) > \mu_i(\overline{m}(i))$.

Lemma 4.3.2. *If all the players submit valid rankings to the planner, then the GS-algorithm finds a match* m *such that* $\mu_i(m(i)) \geq \mu_i(\overline{m}(i))$ *for all players* p_i.

Proof. First we show that true player optimal matching $\overline{m}$ is stable according to the rankings submitted by the players when all those rankings are valid. Let a_j be an arm such that $\widehat{r}_{i,j}(t) < \widehat{r}_{i,\overline{m}(i)}(t)$ for an player p_i. Since $\widehat{r}_{i,t}$ is valid, it means p_i prefers a_j over $\overline{m}(i)$ according to the true preferences also. However, since $\overline{m}$ is stable according to the true preferences, arm a_j must prefer player $\overline{m}^{-1}(j)$ over p_i, where $\overline{m}^{-1}(j)$ is a_j's match according to $\overline{m}$ or the emptyset if a_j does not have a match. Therefore, according to the ranking $\widehat{r}_{i,t}$, p_i has no incentive to deviate to arm a_j because that arm would reject her. Now, since $\overline{m}$ is stable according to the rankings $\widehat{r}_{i,t}$, we know that the GS-algorithm will output a matching which is at least as good as $\overline{m}$ for all players according to the rankings $\widehat{r}_{i,t}$. Since all the rankings are valid, it follows that the GS-algorithm will output a matching m which is as least as good as $\overline{m}$ according to the true preferences also, i.e., $\mu_i(m(i)) > \mu_i(\overline{m}(i))$. □

Lemma 4.3.3. *Consider the player p_i and let $\overline{\Delta}_{i,j} = \mu_i(\overline{m}(i)) - \mu_i(j)$ and $\overline{\Delta}_{i,\min} = \min_{j:\, \overline{\Delta}_{i,j}>0} \overline{\Delta}_{i,j}$. Then, if p_i follows the Explore-then-Commit platform (see Table 4.1(a)), we have*

$$\mathbb{P}(\widehat{r}_{i,hK} \text{ is invalid }) \leq K e^{-\frac{h\overline{\Delta}_{i,\min}^2}{2}}.$$

Proof. Throughout this proof we denote $t = hK$ as a shorthand. In order for the ranking $\widehat{r}_{i,t}$ to not be valid there must exist an arm a_j such that $\mu_i(\overline{m}(i)) > \mu_i(j)$, but $\widehat{r}_{i,j}(t) < \widehat{r}_{i,\overline{m}(i)}(t)$. This can happen only when $\widehat{\mu}_{i,j}(t) \geq \widehat{\mu}_{i,\overline{m}(i)}(t)$. The probability of this event is equal to

$$\begin{aligned}\mathbb{P}\left(\widehat{\mu}_{i,j}(t) \geq \widehat{\mu}_{i,\overline{m}(i)}(t)\right) &= \mathbb{P}\left(\widehat{\mu}_{i,\overline{m}(i)}(t) - \mu_i(\overline{m}(i)) - \widehat{\mu}_{i,j}(t) + \mu_i(j) \leq \mu_i(j) - \mu_i(\overline{m}(i))\right) \\ &\leq \mathbb{P}\left(\widehat{\mu}_{i,\overline{m}(i)}(t) - \mu_i(\overline{m}(i)) - \widehat{\mu}_{i,j}(t) + \mu_i(j) \leq \overline{\Delta}_{i,\min}\right).\end{aligned}$$

Since each player pulls each arm exactly h times during the exploration stage and since the rewards from each arm are 1-sub-Gassian, we know that $\widehat{\mu}_{i,j'}(t) - \mu_i(j') - \widehat{\mu}_{i,j}(t) + \mu_i(j)$ is $\sqrt{2/h}$-sub-Gaussian. Therefore,

$$\mathbb{P}\left(\widehat{\mu}_{i,j}(t) \geq \widehat{\mu}_{i,\overline{m}(i)}(t)\right) \leq e^{-\frac{h\Delta_i^2}{4}}.$$

The conclusion follows by a union bound over all possible arms a_j. □

Proof of Theorem 4.3.1. During the exploration stage each player p_i pulls each arm a_j exactly h times. Therefore, the expected player-optimal stable regret of player p_i after the first hK time steps is exactly equal to $h\sum_{j=1}^{K} \overline{\Delta}_{i,j}$ (note that $\overline{\Delta}_{i,j}$ might be negative for some values of j). The player-optimal stable regret p_i from time $hK+1$ to time n is at most $(n-hK)\overline{\Delta}_{i,\max}$. However, from Lemma 4.3.2 we know that p_i can incurr positive regret only if there exists a player who submits an invalid ranking at time $hK+1$. Lemma 4.3.3, together with a union bound over all players, ensures that the probability there exists a player who submits an invalid ranking is at most $N \exp\left(-\frac{h\Delta^2}{4}\right)$. This completes the proof. □

Centralized UCB

In the previous section we saw that centralized ETC achieves $\mathcal{O}(\log(n))$ player-optimal regret for all players. However, centralized ETC must know the horizon n and the minimum gap Δ between an optimal arm and a suboptimal arm. While knowing the horizon n is feasible in certain scenarios, knowing Δ is not plausible. It is known that single player ETC achieves $\mathcal{O}(n^{2/3})$ when the number of exploration rounds is chosen deterministically without knowing Δ, and there are also known methods for adaptively choosing the number of exploration rounds so that single player ETC achieves $\mathcal{O}(\log(n))$ Lattimore and Szepesvari [2019]. However, in our setting, the $\mathcal{O}(n^{2/3})$ guarantee does not hold because the suboptimality gaps of one player affect the regret of other players, and the known adaptive stopping times cannot be implemented because the platform does not observe the players' rewards. Therefore, it is necessary to find methods which do not need to know Δ.

Another drawback of centralized ETC is that it requires players to learn concurrently. It thus does not take prior knowledge of preferences into account and forces that player to explore arms which might be suboptimal for them. The Gale-Shapley Platform shown in Table 4.1(right) resolves this problem, always outputting the player-optimal matching given the rankings received from the players. We derive an upper bound on the regret in this setting when all players use upper confidence bounds to rank arms. In Section 4.3 we show this method is incentive compatible.

Before proceeding with the analysis we define more precisely the UCB method employed by each player and also introduce several technical concepts. At each time step the platform matches player p_i with arm $m_t(i)$. Each player p_i successfully pulls arm $m_t(i)$, receives reward $X_{i,m_t}(t)$, and updates their empircal mean for $m_t(i)$ as in (4.3). They then compute the upper confidence bound

$$u_{i,j}(t) = \begin{cases} \infty & \text{if } T_{i,j}(t) = 0, \\ \widehat{\mu}_{i,j}(t) + \sqrt{\frac{3\log t}{2T_i(t-1)}} & \text{otherwise.} \end{cases} \tag{4.6}$$

Finally, each player p_i orders the arms according to $u_{i,j}(t)$ and computes the ranking $\widehat{r}_{i,t+1}$ so that a higher upper confidence bound means a better rank, e.g. $\arg\max_j u_{i,j}(t)$ is ranked first in $\widehat{r}_{i,t+1}$.

Let m be an injective function from the set of players $\mathcal{N}$ to the set of arms $\mathcal{K}$; hence m is the matching where $m(i)$ is the match of player i. Then, let $T_m(t)$ be the number of times matching m is played by time t. For a matching m to be played at time t it must be stable according to the current preference rankings of the players and the fixed rankings of the arms, i.e. according to $\widehat{r}_{i,t}$ for all $p_i \in \mathcal{N}$ and π_j for all $a_j \in \mathcal{K}$. We call such matchings *achievable*. We say a matching is *truly stable* if it is stable according to the true preferences induced by the mean rewards of the arms. For player p_i and arm p_ℓ we consider the set $M_{i,\ell}$ of non-truly stable, achievable matchings m such that $m(i) = \ell$. Let $\underline{\Delta}_{i,\ell} = \mu_i(\underline{m}(i)) - \mu_i(\ell)$.

Then, since any truly-stable matching yields regret smaller or equal than zero for all

players, we can upper bound the regret of player i as follows:

$$\underline{R}_i(n) \leq \sum_{\ell:\, \underline{\Delta}_{i,\ell}>0} \underline{\Delta}_{i,\ell} \left(\sum_{m \in M_{i,\ell}} \mathbb{E}T_m(n) \right). \tag{4.7}$$

For any matching m that is non-truly stable there must exist an player p_j and an arm a_k, different from arm $m(j)$, such that the pair (p_j, a_k) is a *blocking pair* according to the true preferences μ, i.e. $\mu_j(k) > \mu_j(m(j))$ and arm a_k is either unmatched or $\pi_k(j) < \pi_k(m^{-1}(k))$. We say the triplet $(p_j, a_k, a_{k'})$ is blocking when p_j is matched with $a_{k'}$ and the pair (p_j, a_k) is blocking. Let $B_{j,k,k'}$ be the set of all matches blocked by the triplet $(p_j, a_k, a_{k'})$. Given a set S of matchings, we say a set Q of triplets $(p_j, a_k, a_{k'})$ is a *cover* of S if

$$\bigcup_{(p_j,a_k,a_{k'})\in Q} B_{j,k,k'} \supseteq S.$$

Let $\mathcal{C}(S)$ denote the set of covers of S. Also, let $\Delta_{j,k,k'} = \mu_j(k) - \mu_j(k')$. Now we state our result.

Theorem 4.3.4. *When all players rank arms according to the upper confidence bounds* (5.12) *and submit their preferences to the Gale-Shapley Platform, the regret of player p_i up to time n satisfies*

$$\underline{R}_i(n) \leq \sum_{\ell:\, \underline{\Delta}_{i,\ell}>0} \underline{\Delta}_{i,\ell} \left[\min_{Q \in \mathcal{C}(M_{i,\ell})} \sum_{(p_j,a_k,a_{k'})\in Q} \left(5 + \frac{6\log(n)}{\Delta^2_{j,k,k'}} \right) \right].$$

Theorem 4.3.4 offers a problem-dependent $\mathcal{O}(\log(n))$ upper bound guarantee on the player-pessimal stable regret of each player p_i. Similarly to the case of centralized ETC, the regret of one player depends on the suboptimality gaps of other players. However, we saw in Section 4.3 that centralized ETC achieves $\mathcal{O}(\log(n))$ player-optimal stable regret, a stronger notion of regret. Example 5.11.1 shows that centralized UCB cannot yield sublinear player-optimal stable regret in general.

Example 4.3.2 (Centralized UCB does not achieve sublinear player-optimal stable regret)**.** *Let $\mathcal{N} = \{p_1, p_2, p_3\}$ and $\mathcal{K} = \{a_1, a_2, a_3\}$, with true preferences given by:*

$$\begin{aligned} p_1 &: a_1 \succ a_2 \succ a_3 & \qquad a_1 &: p_2 \succ p_3 \succ p_1 \\ p_2 &: a_2 \succ a_1 \succ a_3 & \qquad a_2 &: p_1 \succ p_2 \succ p_3 \\ p_3 &: a_3 \succ a_1 \succ a_2 & \qquad a_3 &: p_3 \succ p_1 \succ p_2. \end{aligned}$$

The player-optimal stable matching is (p_1, a_1), (p_2, a_2), (p_3, a_3). When p_3 incorrectly ranks $a_1 \succ a_3$ and the other two players submit their correct rankings, the Gale-Shapley Platform outputs the matching (p_1, a_2), (p_2, a_1), (p_3, a_3). In this case p_3 will never correct their mistake because they never get matched with a_1 again, and hence their upper confidence bound for a_1 will never shrink. Figure 4.1b illustrates this example; the optimal regret for p_1 and p_2 is seen to be linear in n.

Proof of Theorem 4.3.4. Let $L_{j,k,k'}(n)$ be the number of times player p_j pulls arm $a_{k'}$ when the triplet $(p_j, a_k, a_{k'})$ is blocking the matching selected by the platform. Then, by definition

$$\sum_{m \in B_{j,k,k'}} T_m(n) = L_{j,k,k'}(n). \tag{4.8}$$

By the definition of a blocking triplet we know that if p_j pulls $a_{k'}$ when $(p_j, a_k, a_{k'})$ is blocking, they must have a higher upper confidence bound for $a_{k'}$ than for a_k. In other words, we are trying to upper bound the expected number of times the upper confidence bound on $a_{k'}$ is higher than that of the better arm a_k when we have the guarantee that each time this event occurs $a_{k'}$ is successfully pulled. Therefore, standard analysis for the single player UCB [e.g., Bubeck and Cesa-Bianchi, 2012, Chap. 2] shows that

$$\mathbb{E}L_{j,k,k'}(n) \leq 5 + \frac{6\log(n)}{\Delta^2_{j,k,k'}}. \tag{4.9}$$

The conclusion follows from equations (4.7) and (4.8). □

To better understand the guarantee of Theorem 4.3.4 we consider two examples in which the markets have a special structure which enables us to simplify the upper bound on the regret. Moreover, in Corollary 4.3.5 we consider the a worst case upper bound over possible coverings of matchings.

Example 4.3.3 (Global preferences)**.** *Let $\mathcal{N} = \{p_1, \cdots, p_N\}$ and $\mathcal{K} = \{a_1, \cdots, a_K\}$. We assume the following preferences: $p_i : a_1 \succ \cdots \succ a_K$ and $a_j : p_1 \succ \cdots \succ p_N$. In other words all players have the same ranking over arms, and all arms have the same ranking over players. Hence, the unique stable matching is (p_1, a_1), (p_2, a_2), ..., (p_N, a_N). Moreover, for any p_i and a_ℓ we can cover the set of matchings $M_{i,\ell}$ with the triplets (p_i, a_k, a_ℓ) for all k with $1 \leq k \leq i$. Then, Theorem 4.3.4 implies* (4.10) *once we observe that $\Delta_{i,k,\ell} \geq \underline{\Delta}_{i,\ell}$ for all $k \leq i$.*

$$\underline{R}_i(n) \leq 5i \sum_{\ell=i+1}^{K} \underline{\Delta}_{i,\ell} + \sum_{\ell=i+1}^{K} \frac{6i\log(n)}{\underline{\Delta}_{i,\ell}}. \tag{4.10}$$

Figure 4.1c illustrates this example empirically, displaying the optimal (also pessimal) regret of 5 out of 20 players. The 1st-ranked player has sublinear regret, consistent with (4.10)*, while the 20th-ranked player has negative regret and our upper bound is indeed 0.*

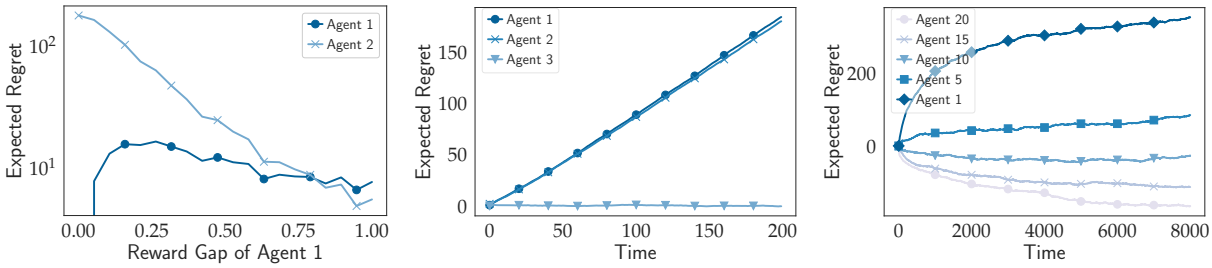
Example 4.3.4 (Unique pairs)**.** *Let $\mathcal{N} = \{p_1, \cdots, p_N\}$ and $\mathcal{K} = \{a_1, \cdots, a_N\}$ and assume that player p_i prefers arm a_i the most and that arm a_i prefers player p_i the most. Therefore, the unique stable matching is (p_1, a_1), (p_2, a_2), ..., (p_N, a_N). Then, we can cover each set $M_{i,\ell}$ with the triplet (p_i, a_i, a_ℓ). Therefore, Theorem 4.3.4 implies* (4.11)*; note that the right-hand side is identical to the guarantee for single player UCB:*

$$\underline{R}_i(n) \leq 5 \sum_{\ell \neq i}^{K} \underline{\Delta}_{i,\ell} + \sum_{\ell \neq i}^{N} \frac{6\log(n)}{\underline{\Delta}_{i,\ell}}. \tag{4.11}$$

Corollary 4.3.5. *Let* $\Delta = \min_i \min_{j,j'} |\mu_i(j) - \mu_i(j')|$. *When all players follow the centralized UCB method, the regret of* p_i *can be upper bounded as follows*

$$\underline{R}_i(n) \leq \max_{\ell} \Delta_{i,\ell} \left(6NK^2 + 12 \frac{NK \log(n)}{\Delta^2} \right).$$

Proof We consider the covering (j, k, k') composed of all possible triples with $\mu_j(k) > \mu_j(k')$. Then, Theorem 4.3.4 implies the result because $\sum_{k' \colon \mu_j(k') < \mu_j(k)} \frac{1}{\Delta^2_{j,k,k'}} \leq \sum_{\ell=1}^{K} \frac{1}{\ell^2 \Delta^2} \leq \frac{2}{\Delta^2}$.



(a) Example 4.3.1 (b) Example 5.11.1 (c) Example 4.3.3

Figure 4.1: The empirical performance of centralized UCB in the settings described in Examples 4.3.1, 5.11.1, and 4.3.3. The experimental details for each figure is given below.

Figure 4.1a. This figure represents an empirical evaluation of Example 4.3.1. In this setting, there are two agents and two arms. Player p_2 receives Gaussian rewards from the arms a_1, a_2 with means 0 and 1 respectively and variance 1. Player p_1 receives Gaussians rewards Δ and 0 from the arms a_1 and a_2. Both arms prefer p_1 over p_2. Figure 4.1a shows the regret of each agent as a function of Δ when we run centralized UCB with horizon 400 and average over 100 trials.

Figure 4.1b. This figure represents an empirical evaluation of Example 4.3.1. The rewards of the arms for each agent are Gaussian with variance 1. The have mean rewards of the arms are set so that the preference structure shown in Example 4.3.1 is satisfied. For agents p_1 and p_2, the gap in mean rewards between consecutive arms is 1. For agent p_3 the gap in mean reward between arms a_1 and a_3 is 0.05. Figure 4.1b shows the performance of centralized UCB, averaged over 100 trials, as a function of the horizon.

Figure 4.3.3. This figure represents an empirical evaluation of Example 4.3.3 when there are 20 agents and 20 arms. The rewards of the arms are Guassian with variance 1. The mean reward gap between consecutive arms is 0.1. Figure 4.1b shows the performance of centralized UCB, averaged over 50 trials, as a function of the horizon.

Honesty and Strategic Behavior

Classical results show that in the player-proposing GS algorithm, no single player can improve their match by misrepresenting their preferences, assuming that the other players and arms submit their true preferences [Roth, 1982, Dubins and Freedman, 1981]. The result generalizes to coalition of players. Moreover, when there is a unique stable matching, the Dubins-Freedman Theorem says that no arms or players can benefit from misrepresenting their preferences [Dubins and Freedman, 1981].

The ETC Platform does not allows players to choose which arms to explore. In this case, the classical results on honesty in player-proposing GS apply; the players are incentivized to submit the rankings according to their current mean estimates. When players have some degree of freedom to explore over multiple rounds, it is no longer clear if any players, or arms, can benefit from misrepresenting their preferences in some of the rounds. In general, one player's preferences can influence not only the matches of other players, but also their reward estimates. One might be able to improve their regret by capitalizing on the ranking mistakes of other players. The possibilities for long-term strategic behavior are more diverse than in the single-round setting.

We now show that when all players except one submit their UCB-based preferences to the GS Platform, the remaining player has an incentive to also submit preferences based on their UCBs, so long as they do not have multiple stable arms.

First, we establish the following lemma, which is an upper bound on the expected number of times the remaining player can pull an arm that is better than their optimal match, regardless of what preferences they might have submitted to the platform.

Lemma 4.3.6. *Let $T_l^i(n)$ be the number of times an player i pulls an arm l such that the mean reward of l for i is greater than i's optimal match. Then*

$$\mathcal{E}[T_l^i(n)] \leq \min_{Q \in \mathcal{C}(M_{i,\ell})} \sum_{(j,k,k') \in Q} \left(5 + \frac{6\log(n)}{\Delta^2_{j,k,k'}}\right) \tag{4.12}$$

Proof. If player i is matched with arm l in any round, the matching m must be unstable according true preferences. We claim that there must exist a blocking triplet (j, k, k') where $j \neq i$.

Arguing by contradiction, we suppose otherwise, that all blocking triplets in m only involve player i. By Theorem 4.2 in Abeledo and Rothblum [1995], we can go from the matching m to a μ-stable matching, by iteratively *satisfying* block pairs in a 'gender consistent' order O. To satisfy a blocking pair (k, j), we break their current matches, if any, and match (k, j) to get a new matching. Doing so, player i can never get a worse match than l or become unmatched as the algorithm proceeds, so the matching remains unstable—a contradiction. Hence there must exist a $j \neq i$ such that j is part of a blocking triplet in m. In particular, player j must be submitting its UCB preferences.

The result then follows from the identity

$$\mathcal{E}[T_l^i(n)] = \sum_{m \in M_{i,\ell}} \mathbb{E}T_m(n),$$

and Equation 4.9 □

Lemma 4.3.6 directly implies the following lower bound on the remaining player's optimal regret.

Proposition 4.3.7. *Suppose all players other than p_i submit preferences according to the UCBs* (5.12) *to the GS Platform. Then the following upper bound on player i's optimal regret holds:*

$$\overline{R}_i(n) \geq \sum_{\ell:\, \overline{\Delta}_{i,l}<0} \overline{\Delta}_{i,l} \left[\min_{Q \in \mathcal{C}(M_{i,\ell})} \sum_{(j,k,k') \in Q} \left(5 + \frac{6\log(n)}{\Delta^2_{j,k,k'}} \right) \right]. \tag{4.13}$$

Therefore, there is no sequence of preferences that an player can submit to the GS Platform that would give them negative optimal regret greater than $\mathcal{O}(\log n)$ in magnitude. When there is a unique stable matching, Proposition 4.3.7 shows that no player can gain significantly in terms of stable regret by submitting preferences other than their UCB rankings.

When there exist multiple stable matchings, however, Proposition 4.3.7 leaves open the question of whether any player can submit a sequence of preferences that achieves super-logarithmic negative *pessimal* regret for themselves, when all other players are playing their UCB preferences. In other words, can an player do significantly better than its pessimal stable arm, by possibly deviating from their UCB rankings? This is an interesting direction for future work.

Chapter 5

Competing Bandits in a Decentralized Matching Market

5.1 Introduction

In the last chapter, we focused on a *centralized* setting in which the players are able to communicate with a central platform that computes matchings for the entire market. We defined a notion of regret called *stable regret*, which is the average reward a player obtains less the rewards achieved under a stable matching with respect to the true preferences of the market. We showed that an algorithm that combines the upper confidence bound principle from the bandit literature [Lai and Robbins, 1985b] with the Gale-Shapley algorithm from the matching market literature [Gale and Shapley, 1962] can achieve low stable regret.

In this chapter, we discuss a decentralized version of the problem, where the actions of the players cannot be coordinated by a central platform, and our goal is to find a viable algorithm for the decentralized case. The decentralized setting is arguably a more useful formulation in practice. Indeed, most online marketplaces are decentralized, that is, there is no central clearinghouse and players are unable to coordinate their actions with each other directly. However, players may observe limited information about past matchings, such as their own conflicts.

New theoretical challenges arise in the decentralized setting, in both the design and the analysis of algorithms. Given that players may use past matchings to inform their current play (e.g., to avoid conflicts), a player who has statistical uncertainty about their preferences over arms may impose externalities on other players not only at the current time step but also into the future. In essence, the decentralized formulation more fully exposes the challenges of the economic and learning aspects of the problem.

We propose a solution for the decentralized version of the two-sided matching bandit problem. Although a version of multiplayer Explore-Then-Commit can be extended to the decentralized setting, the stable regret attained is suboptimal (Section 5.10). Our primary contribution is a new multiplayer bandit algorithm, *Decentralized Conflict-Avoiding Upper*

Confidence Bound (CA-UCB), that is guaranteed to yield for all players a stable regret that grows polylogarithmically with the number of rounds of interaction between players and arms, also known as the time horizon, n. In particular, to prove this regret guarantee we roughly showed that the market converges to a stable matching at a polylogarithmic rate. When the arms have the same preferences over players we offer a better guarantee. In this case we prove that the stable regret grows at most logarithmically with the time horizon. Informally, we can state our results as follows.

Theorem 5.1.1 (Informal main results)**.** *Suppose we have a market with N players and K arms, with arbitrary preferences, and let Δ be the minimum absolute gap between the mean rewards of different arms. Then, if all players run the CA-UCB algorithm for T steps, the probability that the market is unstable at time T is $\mathcal{O}(\log(T)^2/T)$ (see Theorem 5.5.1). Moreover, the players' stable regret satisfies*

$$R(n) = \mathcal{O}\left(\rho^{N^4}\frac{\log(n)^2}{\Delta^2}\right), \text{ for some } \rho > 1. \qquad \text{(Corollary 5.5.2)}$$

When the arms have the same preferences over players, the players' stable regret satisfies

$$R(n) = \mathcal{O}\left(N^2 K\frac{\log(n)}{\Delta^2}\right). \qquad \text{(Theorem 5.4.1)}$$

Moreover, if $N-1$ players implement the CA-UCB algorithm, the remaining player cannot significantly improve their regret by running a different algorithm (Proposition 5.6.1).

The CA-UCB algorithm is simple and does not require communication between players. There are two features of this algorithm that enable players to avoid conflicts. Firstly, when implementing this algorithm a player observes the actions of other players in the previous round and avoids attempting an arm if that arm was previously pulled by a better player for it. Secondly, players randomly decide whether to choose the same arm as at the previous time step or to make a new decision. When players implement our method conflicts can still occur, but our analysis shows that the expected number of conflicts would be small.

The rest of the chapter is organized as follows: In Section 5.2, we review the matching bandits problem, following the presentation in the previous chapter, and fully specify the decentralized setting that is our focus. In Section 5.3, we motivate and introduce the algorithm that is the subject of our regret analyses in Sections 5.4 and 5.5. In Section 5.6, we discuss the incentive compatibility of this algorithm, showing one positive and one negative result. Our theoretical guarantee on the performance of CA-UCB exhibits an exponential dependence on the size of the market. In Section 5.7 we show empirically that this dependence is an artifact of our analysis; CA-UCB performs much better in practice than these results suggest. In Section 5.8, we survey the related literatures, and in Section 5.9, we present a thorough discussion of our results, as well as avenues for future work. In Section 5.10, we analyze a suboptimal algorithm based on explore-then-commit for the decentralized setting, for comparison; in Section 5.11, we include omitted examples and proofs.

5.2 Problem Setting

We consider a multiplayer multi-armed bandit problem with N players and K stochastic arms, with $N \leq K$. We denote the set of players by $\mathcal{N} = \{p_1, p_2, \ldots, p_N\}$ and the set of arms by $\mathcal{K} = \{a_1, a_2, \ldots, a_K\}$. At time step t, each player p_i attempts to pull an arm $m_t(i) \in \mathcal{K}$.

When multiple players attempt to pull the same arm, only one player will successfully pull the arm, according to the arm's preferences via a mechanism we detail shortly. Then, if player p_i successfully pulls arm $m_t(i)$ at time t, they are said to be *matched* to $m_t(i)$ at time t and they receive a stochastic reward, $X_{i,m_t}(t)$, sampled from a 1-sub-Gaussian distribution with mean $\mu_i(m_t(i)) > 0$.

For each player p_i we assume $\mu_i(j) \neq \mu_i(j')$ for all distinct arms, a_j and $a_{j'}$. If $\mu_i(j) > \mu_i(j')$, we say that player p_i *truly prefers* a_j to $a_{j'}$, and denote this as $a_j \succ_{p_i} a_{j'}$.

Each arm a_j has a fixed, known, and strict preference ordering over all the players, $\succ_{a_j}$. In other words, $p_i \succ_{a_j} p_{i'}$ indicates that arm a_j prefers player p_i to player $p_{i'}$. If two or more players attempt to pull the same arm a_j, there is a *conflict* and only the most preferred player successfully pulls the arm to receive a reward; the other player(s) $p_{i'}$ is said to be *unmatched* and does not receive any reward, that is, $X_{i',m_t}(t) = 0$.

A *stable matching* [Gale and Shapley, 1962] of players and arms is one where no pair of player and arm would prefer to be matched with each other over their respective matches. Given the full preferences of the arms and players, arm a_j is called a *achievable match* of player p_i if there exists a stable matching according to those preferences such that a_j and p_i are matched. We say a_j is the *optimal match* of player p_i if it is the most preferred achievable match. Similarly, we say a_j is the *pessimal match* of player p_i if it is the least preferred achievable match. We denote by $\overline{m}$ and $\underline{m}$ the functions from $\mathcal{N}$ to $\mathcal{K}$ that define the optimal and pessimal matches of a player according to the true preferences of the players and arms.

In the decentralized matching setting, a notion of *stable regret*, as introduced in Chapter 4, is useful for analyzing the performance of learning algorithms. We consider a player's *player-pessimal stable regret*, where the baseline for comparison is the mean reward of the arm that is the player's pessimal match.[1] It is defined as follows for player p_i:

$$\underline{R}_i(n) := n\mu_i(\underline{m}(i)) - \sum_{t=1}^{n} \mathbb{E}X_{i,m_t}(t). \tag{5.1}$$

The above notion of stable regret considers regret from the perspective of the players only, that is, we are primarily interested in how the players perform with respect to their

[1]We can define analogously the *player-optimal stable regret* corresponding to the player's optimal match, denoted $\overline{R}_i(n)$. The player-pessimal stable regret and player-optimal stable regret tend to coincide in many real-world markets, such as in unbalanced random matching markets [Ashlagi et al., 2017b] where the stable matching is essential unique. This is as well the case when players are globally ranked. In this work, we focus on the player-pessimal stable regret.

stable arms over time. Focusing on the welfare of one side of the market is consistent with the stable matching literature, in particular that on school choice, where one side of the market (the schools) are said to have "priorities", rather than "preferences", for the other side of the market (the students), and it is the students' welfare that is of primary interest [Abdulkadiroglu and Snmez, 2003, Abdulkadiroglu et al., 2006].[2] Recently, [Cen and Shah, 2021] studied fairness and social welfare in the context of matching markets.

In order to fully specify the problem we need to clarify what information the players have access to. We consider the following decentralized setting:

Decentralized with Conflict Information At each round, each player attempts to pull an arm, with the choice of arm based on only their rewards and observations from previous rounds. At the end of the round, all players can observe the winning player for each arm. They can see their own rewards only if they successfully pull an arm. They cannot see the rewards of other players. We also assume that all players know, for each arm, which players are ranked higher than themselves.[3]

5.3 Algorithm: Decentralized Conflict-Avoiding UCB

In the single-player multi-armed bandit (MAB) the player must explore different arms in order to identify the arms with the highest mean payoff. At the same time, the player must keep selecting arms that seem to give high payoff in order to accumulate a large reward over time. The upper confidence bounds (UCB) algorithm offers an elegant solution to this exploration-exploitation dilemma. As the name suggests, UCB maintains upper confidence bounds on the arms' mean payoffs and selects the arm with the largest upper confidence bound. Then, the UCB algorithm updates the upper confidence bound corresponding to the selected arm according to the reward observed.

In the aforementioned decentralized model, however, a player cannot implement UCB obliviously of other players' actions given the possibility of conflicts. Let us discuss this issue from the perspective of player p_1. Suppose p_1 chooses arm a_1, and suppose player p_2 chooses a_1 at the same time. Then, if a_1 prefers p_2 over p_1, a conflict arises and player p_1 receives no reward. In addition to not receiving a reward, in this case, player p_1 does not learn anything new about the distribution of rewards offered by arm a_1. Therefore, in the decentralized case players must balance exploration and exploitation while avoiding conflicts that they would lose.

[2]We thank a reviewer for pointing out this connection to the economics literature.

[3]This assumption allows for a cleaner analysis of our algorithm. Our results can be generalized to the setting where players do not know this information initially because the arms know their own preferences and the conflicts between players are resolved deterministically. It is sufficient for each player to assume in the beginning that they are the most preferred player by every arm. Then, each lost conflict reveals which players are more preferred by which arms. This procedure would introduce at most KN^2 conflicts.

Algorithm 1 CA-UCB with random delays

Input: $\lambda \in [0, 1)$

for $t = 1, \ldots, T$ **do**
  for $i = 1, \ldots, N$ **do**
    if $t = 1$ **then**
      Set upper confidence bound to ∞ for all arms.
      Sample an index $j \sim 1, \ldots, K$ uniformly at random. Sets $A_t^{(i)} \leftarrow a_j$.
    else
      Draw $D^{(i)}(t) \sim Ber(\lambda)$ independently.
      if $D^{(i)}(t) = 0$ **then**
        Update plausible set $S^{(i)}(t)$ for player p_i:

$$S^{(i)}(t) := \{a_j : p_i \succ_{a_j} p_k \text{ or } p_i = p_k, \text{ where } \bar{A}^{(k)}(t-1) = a_j\}.$$

        Pulls $a \in S^{(i)}(t)$ with maximum upper confidence bound. Sets $A_t^{(i)} \leftarrow a$.
      else
        Pulls $A_{t-1}^{(i)}$. Sets $A_t^{(i)} \leftarrow A_{t-1}^{(i)}$.
    if p_i wins conflict **then**
      Update upper confidence bound for arm $A_t^{(i)}$.

To see intuitively how p_1 can achieve such conflict avoidance let us assume that there are only two players and that all arms prefer p_2. Then, from the perspective of p_2, the problem is identical with the single-player MAB problem and therefore p_2 can achieve small regret by using the standard UCB method. Since p_2 aims to minimize their own regret, p_2 will sample the arm that gives them the highest mean payoff most of the time. More precisely, there can be at most $\mathcal{O}(\log(T))$ time steps when p_2 does not sample the best arm for themself.

On the other hand, p_1 must minimize the number of times they select the same arm as p_2 because they would lose the conflicts with p_2. Because most of the time player p_2 chooses the best arm for themselves, the following simple heuristic allows player p_1 to avoid choosing the same arm as p_2 most of the time: player p_1 should not select the arm p_2 chose at the previous time step.

It turns out that this conflict-avoidance heuristic, combined with the UCB method, gives rise to an algorithm that provably achieves small regret for all players. We call this method *Decentralized Conflict-Avoiding Upper Confidence Bound*, or *CA-UCB* for short, and detail it in Algorithm 1. Before introducing our algorithm, let us first introduce some notation for the players' actions. We use $A^{(i)}(t)$ to denote the player p_i's attempted arm at time t, and $\bar{A}^{(i)}(t)$ to denote the player i's successfully pulled arm at time t. When the player fails to pull an arm successfully because of a lost conflict, we have $\bar{A}^{(i)}(t) = \emptyset$.

According to Algorithm 1, at each time step t each player p_i independently samples a biased Bernoulli random variable $D^{(i)}(t)$ with mean $\lambda \in [0, 1)$. When $D^{(i)}(t)$ comes up 1, the

player chooses the same arm as they did at the previous time step. We will soon return to explain the rationale behind staying on the same arm as the previous time step with some probability. For now, let us focus on the case where $D^{(i)}(t)$ comes up 0.

When the Bernoulli random variable $D^{(i)}(t)$ comes up 0, the player constructs a *plausible set* of arms that includes all arms except those that the player would not have been able to pull successfully at the previous time step. In other words, the player p_i will consider an arm plausible, only if in the previous time step $t-1$, the arm was not pulled by a player that the arm strictly prefers to p_i. Then, the player chooses the arm in the plausible set with the highest upper confidence bound, which is updated as in the single-player UCB method. We formally define the upper confidence bound in Equation 5.12 of Section 5.4.

We refer to the parameter λ as the *delay probability*. When $\lambda = 0$ the actions of the players that implement CA-UCB are deterministic functions of the history up to that point. This property has no impact on the algorithm's convergence when the players are globally ranked (i.e., all arms have the same preferences), as shown in Section 5.4. However, for more general preference structures, if all players implement CA-UCB with delay probability zero, they can enter into infinite loops. The following simple example showcases this failure mode.

Example 5.3.1 (2-player globally ranked arms)**.** *Consider the following setting with two players and two arms:*

$$\begin{aligned} p_1 : a_1 \succ a_2 \qquad & a_1 : p_1 \succ p_2 \\ p_2 : a_1 \succ a_2 \qquad & a_2 : p_2 \succ p_1. \end{aligned}$$

In this case the unique stable matching is $(p_1, a_1), (p_2, a_2)$.

Suppose both players in Example 5.3.1 implement CA-UCB with zero probability of delay. Through a random initialization of CA-UCB it is possible that both players select arm a_1 at the first time step. Then, p_2 loses the conflict and at the next step will choose a_2, which is the only arm in their plausible set. On the other hand, the UCB of player p_1 for arm a_2 is positive infinity at this point because they have not pulled it yet. Hence, p_1 attempts to pull a_2 at the second time step. Since a_2 prefers p_2, p_1 loses the conflict and their UCB for arm a_2 remains infinite. The same argument shows that both players will keep choosing the same arm, alternating between a_1 and a_2. As long as they stay in this cycle, both players experience a constant stable regret. We showcase another example of when deterministic conflict-avoiding might fail in Appendix 5.11.

To break such cycles CA-UCB incorporates randomness via the delay probability. As we will see, for arbitrary preferences and delay probability $\lambda \in (0, 1)$, the CA-UCB algorithm achieves $\mathcal{O}(\log(T)^2)$ regret, with the hidden constant depending on λ, the gap between mean rewards, and the number of players and arms. On the other hand, the size of the regret that we obtain depends exponentially on the number of players, regardless of the choice of λ. We can obtain stronger results by making additional assumptions on the structure of preferences. In particular, if the players are globally ranked, then we obtain a polynomial dependence on the number of players; moreover, we obtain $\mathcal{O}(\log(T))$ regret. We begin with this specialized setting in Section 5.4 and turn to the general case in Section 5.5.

5.4 Globally Ranked Players

In this section, we prove regret bounds for the CA-UCB algorithm, Algorithm 1, without random delays (i.e., with $\lambda = 0$). We assume all arms have the same preferences over players, whereas each player may have arbitrary preferences over arms. This preference structure is made precise in the following assumption.

Assumption 4 (Globally ranked players)**.** *We assume the players are globally ranked: for any p_i, $p_{i'}$ where $i < i'$, and any arm a_j, we have $p_i \succ_{a_j} p_{i'}$.*

In other words, more preferred players have lower indices. Under this assumption, there is a unique stable matching in the market. By re-indexing the arms we can assume without loss of generality that the stable player-arm pairs are $\{(p_i, a_i)\}_{i=1}^N$. Under such an indexing, the following critical property holds: for any player p_i and any arm a_j with $j > i$, p_i must prefer a_i over a_j; that is, $a_i \succ_{p_i} a_j$. Also, since the stable matching is unique, there is a single notion of stable regret, that is, for any player p_k, we have $R_k(n) := \underline{R}_k(n) = \overline{R}_k(n)$.

Our goal in this section is to prove an upper bound on the stable regret of a player, taking into account their ranking in the market. We use the following notation to denote the gaps in mean rewards of arms for players p_i, p_j:

$$\Delta_j^{(i)} := \mu_i(i) - \mu_i(j) \quad \text{and} \quad \Delta_\emptyset^{(i)} := \mu_i(i). \tag{5.2}$$

We use $\Delta^2 := \min_{i<j} |\Delta_j^{(i)}|^2$ to denote the minimum squared gap.

Theorem 5.4.1 (Stable regret under globally ranked players)**.** *Suppose each player runs Algorithm 1 with $\lambda = 0$. The following regret bound holds for any player p_k and any horizon $T \geq 2$:*

$$R_k(n) \leq 6k^2 \left(\frac{\log n}{\Delta^2} + 1 \right) \cdot \left((K-k)\Delta_\emptyset^{(k)} + k \sum_{i : a_k \succ_{p_k} a_i} \Delta_i^{(k)} \right). \tag{5.3}$$

This result shows that the stable regret of any player in the market is logarithmic in the horizon n, matching the known lower bound for single-player stochastic bandits [Lai and Robbins, 1985b]. Moreover, the regret scales cubically with the rank of the player and linearly with the number of arms. It is useful to compare this result to the corresponding stable regret in the centralized setting (a direct corollary of Theorem 4.3.4), also under Assumption 4:

$$R_k(n) \leq 6k \sum_{l=k+1}^{K} \left(\Delta_l^{(k)} + \frac{\log n}{\Delta_l^{(k)}} \right). \tag{5.4}$$

We see that in the centralized setting, the dependence on the rank k is linear instead of cubic. Moreover, the dependence on the reward gap is reduced to $\sum_{i>k} 1/\Delta_i^{(k)}$, which matches the optimal dependence on the reward gaps in the classical single-player bandit problem

[Lai and Robbins, 1985b]. In the decentralized setting where players are globally ranked, Sankararaman et al. [2020] showed a instance dependent lower bound suggesting that the dependence on $1/\Delta^2$ cannot be improved upon in general. We further discuss lower bounds in Section 5.9.

Before we proceed to the proof of Theorem 5.4.1, we introduce the following notation, and establish two technical lemmas.

- $A^{(k)}(t) \in [K]$ is the arm attempted by p_k at time t;
- $\bar{A}^{(k)}(t) \in [K] \cup \{\emptyset\}$ is outcome of p_k's attempt at time t;
- $T_i^{(k)}(t)$ is the total number of attempts by p_k of a_i up to time t;
- $T_{k,i}(t)$ is the total number of successful attempts by p_k of a_i up to time t.

The following events are central to our analysis:

$$\Lambda_l^{(j)}[t] = \left\{\bar{A}^{(j)}(t) = a_l, a_j \in S^{(j)}(t)\right\}. \tag{5.5}$$

In plain language, $\Lambda_l^{(j)}[t]$ denotes the event in which a player p_j chooses to pull an arm a_l over a stable matching arm a_j that belongs to the plausible set at time t.

The next lemma shows that if a player p_k pulls a suboptimal arm a_i (with $i > k$) at time t, then there must be some same or better-ranked player p_j (with $j \leq k$), who, though having its matching arm a_j in their plausible set, chose to pull a suboptimal arm a_l (with $l > j$) at some time t' between times $t - k$ and t.

Lemma 5.4.2 (Suboptimal pulls)**.** *For any player p_k and arm a_i such that $a_k \succ_{p_k} a_i$,*

$$\left\{\bar{A}^{(k)}(t) = a_i\right\} \subseteq \Lambda_i^{(k)}[t] \bigcup \left(\bigcup_{1 \leq j < l \leq k} \bigcup_{t-k \leq t' < t} \Lambda_l^{(j)}[t'] \right). \tag{5.6}$$

Proof. The key to the proof is the following observation. Suppose the event $\left\{\bar{A}^{(k)}(t) = a_i\right\}$ takes place. Then, one of the two things must happen:

- $a_k \in S^{(k)}(t)$, in which case the event $\Lambda_i^{(k)}[t]$ occurs by definition.
- $a_k \notin S^{(k)}(t)$, in which case some better-ranked player, say p_u with $u < k$, must have pulled the arm a_k at time $t - 1$ according to the definition of Algorithm 1.

This observation translates to the following assertion: for any player p_k and arm a_i where $i \neq k$, we have

$$\left\{\bar{A}^{(k)}(t) = a_i\right\} \subseteq \Lambda_i^{(k)}[t] \bigcup \left(\bigcup_{u < k} \left\{\bar{A}^{(u)}(t-1) = a_k\right\} \right). \tag{5.7}$$

We can now prove the lemma by induction on k.

Base case $k = 1$: This is trivially true, due to the fact that the top-ranked player p_1 has all the arms in their plausible set at all times t, and thus, for any arm a_i,

$$\left\{\bar{A}^{(1)}(t) = a_i\right\} = \left\{\bar{A}^{(1)}(t) = a_i, a_i \in S^{(1)}(t)\right\} = \Lambda_i^{(1)}[t].$$

Induction step: We assume (5.6) for all $k < m$ and prove it also holds for $k = m$. Let arm a_i be such that $a_m \succ_{p_m} a_i$. By equation (5.7), we have

$$\left\{\bar{A}^{(m)}(t) = a_i\right\} \subseteq \Lambda_i^{(m)}[t] \bigcup \left(\bigcup_{u<m} \left\{\bar{A}^{(u)}(t-1) = a_m\right\} \right). \tag{5.8}$$

By our assumptions we know that $a_u \succ_{p_u} a_m$ when $u < m$. Consequently, we can apply the induction hypothesis for player p_u, with $u < m$, and arm a_m and time $t-1$, to obtain that

$$\left\{\bar{A}^{(u)}(t-1) = a_m\right\} \subseteq \Lambda_m^{(u)}[t-1] \bigcup \left(\bigcup_{1\le j<l\le u} \bigcup_{t-u\le t'<t-1} \Lambda_l^{(j)}[t'] \right).$$

Taking the union over $u < m$ on both sides yields the inclusion

$$\bigcup_{u<m} \left\{\bar{A}^{(u)}(t-1) = a_m\right\} \subseteq \bigcup_{1\le j<l\le m} \bigcup_{t-m\le t'<t} \Lambda_l^{(j)}[t']. \tag{5.9}$$

By substituting equation (5.9) into equation (5.8), we obtain the conclusion. □

The next lemma tells a similar story as Lemma 5.4.2; it shows that when p_k has a conflict, there must be some better player p_j, with $j < k$, who chooses to pull a suboptimal arm a_l at some time t' between times $t-k$ and t although they have the matching arm a_j in their plausible set.

Lemma 5.4.3 (Conflicts)**.** *For any player p_k, we have the inclusion*

$$\left\{\bar{A}^{(k)}(t) = \emptyset\right\} \subseteq \bigcup_{\substack{1\le j<k \\ j<l\le K}} \bigcup_{t-k\le t'\le t} \Lambda_l^{(j)}[t']. \tag{5.10}$$

Proof. Player p_k can have a conflict on any of the arms a_1, a_2, …, a_K. We have

$$\left\{\bar{A}^{(k)}(t) = \emptyset\right\} = \bigcup_{l=1}^{K} \left\{\bar{A}^{(k)}(t) = \emptyset, A^{(k)}(t) = a_l\right\}.$$

For all $m \ge k$ we observe that p_k can have a conflict on a_l only if there is a player p_j with $j < k$ who successfully pulls arm a_m at time t. In this case we have

$$\left\{\bar{A}^{(k)}(t) = \emptyset, A^{(k)}(t) = a_m\right\} \subseteq \bigcup_{j<k} \left\{\bar{A}^{(j)}(t) = a_m\right\}.$$

We can then apply Lemma 5.4.2 to each event $\left\{\bar{A}^{(j)}(t) = a_m\right\}$.

We now have to analyze the events $\left\{\bar{A}^{(k)}(t) = \emptyset, A^{(k)}(t) = a_m\right\}$ with $m < k$. Since

$$\left\{\bar{A}^{(k)}(t) = \emptyset, A^{(k)}(t) = a_m\right\} \subseteq \left\{A^{(k)}(t) = a_m\right\},$$

it suffices to prove by induction that

$$\bigcup_{m=1}^{k-1} \left\{A^{(k)}(t) = a_m\right\} \subseteq \bigcup_{\substack{1\le j<k \\ j<l\le K}} \bigcup_{t-k\le t'\le t} \Lambda_l^{(j)}[t']. \tag{5.11}$$

The base case $k = 1$ is obvious since the left-hand side is the empty set. Now, we assume the induction hypothesis holds for all $k < k'$ and we prove it for $k = k'$. If $\left\{A^{(k')}(t) = a_m\right\}$ holds, we know that p_m at time $t-1$ did not attempt to pull a_m. They either attempted to pull an arm $a_{m'}$ with $m' > m$ or with $m' < m$. In the former case, the induction step follows from Lemma 5.4.2. In the latter case, we can apply our induction hypothesis. The result follows. □

The final ingredient we need to prove Theorem 5.4.1 is the UCB argument for a single player. This is given in the following display. For completeness, we provide an elementary proof in Section 5.11.

Lemma 5.4.4 (UCB bound)**.** *Suppose we use the following upper confidence bounds in Algorithm 1:*

$$u_{i,j}(t) = \begin{cases} \infty & \text{if } T_{i,j}(t) = 0, \\ \widehat{\mu}_{i,j}(t) + \sqrt{\frac{3\log t}{2T_{i,j}(t-1)}} & \text{otherwise.} \end{cases} \tag{5.12}$$

Then, for any player p_i, arms a_j, a_k, such that $a_j \prec_i a_k$, we have, for $n > 0$:

$$\sum_{t=1}^{n} \mathbb{P}\left(\{u_{i,j}(t) > u_{i,k}(t)\} \cap \left\{\bar{A}^{(i)}(t) = j\right\}\right) \le \frac{6}{\Delta^2}\log(T) + 6.$$

Proof of Theorem 5.4.1. We bound the regret of player p_k. By definition, their regret is

$$R_k(n) \le \Delta_{\emptyset}^{(k)} \cdot \mathcal{E}\left[T_{k,\emptyset}(n)\right] + \sum_{i:a_k \succ_{p_k} a_i} \Delta_i^{(k)} \cdot \mathcal{E}\left[T_{k,i}(n)\right], \tag{5.13}$$

where, because of our assumption on the indexing of arms, the last summation can also be written simply as a sum over all $i \in \{k+1, \ldots, K\}$.

Upper bounding $\mathcal{E}[T_{k,i}(n)]$. By definition,

$$\mathcal{E}\left[T_{k,i}(n)\right] = \sum_{t=1}^{n} \mathbb{P}\left(\bar{A}^{(k)}(t) = a_i\right). \tag{5.14}$$

We now bound the probability $\mathbb{P}\left(\bar{A}^{(k)}(t) = a_i\right)$ for each t. Lemma 5.4.2 yields

$$\mathbb{P}(\bar{A}^{(k)}(t) = a_i) \leq \mathbb{P}(\Lambda_i^{(k)}[t]) + \sum_{1\leq j<l\leq k} \sum_{t-k\leq t'<t} \mathbb{P}(\Lambda_l^{(j)}[t']). \tag{5.15}$$

Summing from $t = 1$ to n, and using equation (5.14), we obtain the bound

$$\begin{aligned} \mathcal{E}\left[T_{k,i}(n)\right] &\leq \sum_{1\leq t\leq T} \mathbb{P}(\Lambda_i^{(k)}[t]) + \sum_{1\leq t\leq n} \sum_{1\leq j<l\leq k} \sum_{t-k\leq t'\leq t} \mathbb{P}(\Lambda_l^{(j)}[t']) \\ &\leq \sum_{1\leq t\leq T} \mathbb{P}(\Lambda_i^{(k)}[t]) + (k+1) \sum_{1\leq j<l\leq k} \sum_{1\leq t\leq n} \mathbb{P}(\Lambda_l^{(j)}(t)). \end{aligned} \tag{5.16}$$

Recall that for all players p_j and arms a_l with $l > j$, and time $t > 0$,

$$\Lambda_l^{(j)}(t) \subseteq \{u_{j,l}(t) > u_{j,i}(t)\} \cap \left\{\bar{A}^{(j)}(t) = l\right\}.$$

Therefore, using Lemma 5.4.4, we can show that the following upper bound holds:

$$\sum_{1\leq t'\leq n} \mathbb{P}(\Lambda_l^{(j)}(t')) \leq 6\left(\frac{\log n}{|\Delta_l^{(j)}|^2} + 1\right). \tag{5.17}$$

Substituting equation (5.17) into equation (5.16) yields the bound

$$\mathcal{E}\left[T_{k,i}(n)\right] \leq 6\left(\frac{\log n}{|\Delta_i^{(k)}|^2} + 1\right) + 6(k+1) \sum_{1\leq j<l\leq k} \left(\frac{\log n}{|\Delta_l^{(j)}|^2} + 1\right) \leq 6k^3 \left(\frac{\log n}{\Delta^2} + 1\right). \tag{5.18}$$

Recall that $\Delta^2 = \min_{i\neq j} |\Delta_j^{(i)}|^2$.

Upper bounding $\mathcal{E}\left[T_{k,\emptyset}(n)\right]$. By definition,

$$\mathcal{E}\left[T_{k,\emptyset}(n)\right] = \sum_{t=1}^{n} \mathbb{P}\left(\bar{A}^{(k)}(t) = \emptyset\right). \tag{5.19}$$

Lemma 5.4.3 and a derivation mutatis mutandis to the argument from equation (5.15) to (5.18) yields

$$\mathcal{E}\left[T_{\emptyset,i}(n)\right] \leq 6(k+1) \sum_{\substack{1\leq j<k \\ j<l\leq K}} \left(\frac{\log n}{|\Delta_l^{(j)}|^2} + 1\right) \leq 6k^2(K-k)\left(\frac{\log n}{\Delta^2} + 1\right). \tag{5.20}$$

Substitute (5.18) and (5.20) into (5.13) to complete the proof of the theorem. □

5.5 Arbitrary Preferences on Both Sides of the Market

In this section, we analyze the convergence of Algorithm 1 under arbitrary preference lists for both sides of the market. Note that in this setting, the stable matching may not be unique. We consider throughout the randomized version of Algorithm 1, with delay probability $\lambda > 0$.

Without the assumption of shared preferences among the arms, the analysis of the convergence of Algorithm 1 becomes more challenging. In fact, it is not obvious that Algorithm 1, or any other algorithm, can achieve sublinear player regret against the pessimal stable matching for any set of preferences. As seen in Example 5.11.1 in Appendix 5.11, decentralized coordination among players can be difficult even in small markets with only three players. In order to prove the regret bound in Section 5.4, we relied heavily on the structure conferred by the global ranking of players. Without this particular structure, we have to appeal to more general results about stable matching. This generality also comes at a cost: the regret bound we prove in this section is polylogarithmic in the horizon and has an exponential dependence on the number of players.

Before introducing the main result, we first present some essential notation. Recall that $\mathcal{N} = \{p_i\}_{i=1}^N$ denotes the set of players, and $\mathcal{K} = \{a_i\}_{i=1}^K$ denotes the set of arms. We denote the attempted actions (i.e., arms) at time t as

$$m_t : \mathcal{N} \mapsto \mathcal{K}, \text{ where } m_t(p_i) := A^{(i)}(t).$$

We note that m_t in general does not have to be a matching between players and arms, because two or more players may attempt to pull the same arm. However, whenever there are no conflicts, m_t is indeed a matching (an injective map) between players and arms, so we can distinguish the set of attempted actions that coincide with a stable matching. We thus refer to $m : \mathcal{N} \mapsto \mathcal{K}$ as *stable* if m indeed coincides with a stable matching between players and arms.

We denote the set of stable attempted actions as

$$M^* := \{M \mid M : \mathcal{N} \mapsto \mathcal{K}, M \text{ is stable}\}.$$

Let $\Delta = \min_{i,j,k} |\mu_i(j) - \mu_i(k)|$ denote the minimum reward gap between any two arms for any player. We also define the constant $\varepsilon := (1-\lambda)\lambda^{N-1}$, which depends on the delay probability λ.

Our goal in this section is to prove the following upper bound on the probability that the market is in an unstable configuration when running the algorithm. More formally, we bound the sum, over t, of probabilities that the attempted actions at time t yield an unstable matching. Understanding how this quantity depends on the horizon and various problem parameters enables us to provide a general regret bound for Algorithm 1.

Theorem 5.5.1 (Convergence to stability of Algorithm 1 for arbitrary preferences)**.** *Let $N, K \geq 2$, $T \geq 2$, and suppose we run Algorithm 1 with delay probability $\lambda \in (0,1)$. Then,*

$$\sum_{t=1}^{T} \mathbb{P}(m_t \neq M^*) \leq 24 \cdot \frac{N^5 K^2}{\varepsilon^{N^4+1}} \log(T) \left(\frac{1}{\Delta^2} \log(T) + 3 \right). \tag{5.21}$$

As a corollary of Theorem 5.5.1, we have the following upper bound on the pessimal stable regret of any player.

Corollary 5.5.2 (Pessimal stable regret of Algorithm 1 for arbitrary preferences)**.** *The following inequality holds for the agent-pessimal regret of player p_k up to time n:*

$$\underline{R}_k(n) \leq 24 \cdot \max_{a_\ell \in \mathcal{K}} \underline{\Delta}_{k,\ell} \left(\frac{N^5 K^2}{\varepsilon^{N^4+1}} \log(T) \left(\frac{1}{\Delta^2} \log(T) + 3 \right) \right),$$

where $\underline{\Delta}_{k,\ell} = \max\{\mu_k(\underline{m}(k)) - \mu_k(\ell), \mu_k(\underline{m}(k))\}$.

In short, we find that the stable regret of Algorithm 1 is $\mathcal{O}((\log n)^2)$. Unlike in previous sections where we derived player-specific stable regret bounds that depended on the ranking of the player, or the ranking of their stable arm, in the current setting the players have no particular ranking. Corollary 5.5.2 is derived from a general bound on the probabilities that the matching of the entire market is unstable.

Proof sketch We begin by sketching the main ideas in the proof of Theorem 5.5.1. There are two main technical ingredients that are new to the current section: the first is the observation that in the event that each player's UCB rankings of the arms in their plausible set are correct (colloquially we refer to this event as "no statistical mistakes"), and the previous matching was stable, then running one step of Algorithm 1 will preserve the stability of the matching with probability one. This is established in Lemma 5.5.3. Therefore, if the matching at time t is unstable, it must be either be that some player had incorrect UCB rankings, or there were no statistical ranking mistakes but the matching at time $t-1$ was unstable.

In Lemma 5.5.4, we generalize this statement to consider histories of arbitrary length. That is, if a matching at time t is unstable, it must either be that some player had incorrect UCB rankings over the last h time steps, or there were no ranking mistakes in all the last h time steps but the matchings reached were unstable.

As in Section 5, we know how to upper bound the probability that a player had incorrect UCB rankings when running Algorithm 1 with $\lambda > 1$. Recall that this entailed a simple adaptation of the single-player UCB argument (Lemma 5.4.4). The new problem we face is that of controlling the probability that there were no ranking mistakes but the matchings in all the last h time steps were unstable. It turns out that a classical result from the stable matching literature [Abeledo and Rothblum, 1995] gives us a way to argue that this probability is exponentially small in the length of the history considered (Lemma 5.5.7). Intuitively,

we are using the fact that Algorithm 1, when there are no ranking mistakes, is essentially resolving *blocking pairs*—pairs of players and arms that would prefer to be matched with each other over their current matches—in a randomized fashion, but following an order that is consistent with player preferences (Lemma 5.5.6). This is crucial for establishing that Algorithm 1 will always reach a stable matching with enough steps, as long as there are no ranking mistakes.

Finally, our analysis needs to balance the tradeoff inherent in the choice of the length of history considered, h. If we consider a longer history length, there can be many ranking mistakes made in this window, hence contributing to a higher probability of an unstable matching. On the other hand, a longer history length with no ranking mistakes means that there is a higher probability that a stable matching can be reached. By choosing h to depend on the time step t, we are able to achieve a $\log(n)^2$ dependence on the horizon n in the final bound (5.21).

Before presenting the technical lemmas, we first rigorously define the events of interest that were alluded to in the proof sketch.

1. Let E_t denote the event that, for every player, the arm that has the highest mean reward in their plausible set coincides with the arm with the highest UCB in their plausible set at time t:

$$E_t := \bigcap_{p_i \in \mathcal{N}} \left\{ \operatorname*{argmax}_{a_j \in S^{(i)}(t)} \mu_i(j) = \operatorname*{argmax}_{a_j \in S^{(i)}(t)} u_{i,j}(t) \right\}. \tag{5.22}$$

   Let E_t^c denote the complement of this event.

2. Let $F_{j,k}^{(i)}(t)$ denote the event that player p_i's UCB for arm a_j is greater than their UCB for arm a_k at time t:

$$F_{j,k}^{(i)}(t) = \{u_{i,j}(t) > u_{i,k}(t)\}.$$

The following lemma shows that if the current matching is stable, then one step of Algorithm 1 under the event defined in (5.22) preserves the stability of the current matching.

Lemma 5.5.3 (Preservation of Stability)**.** *Assume $m_t \in M^*$. Then $m_{t+1} \in M^*$ on the event E_{t+1}.*

Proof. We show $m_{t+1} = m_t$ on event E_{t+1}. Let $m = m_t$. Assume E_{t+1} happens. Suppose, for a contradiction, that some player p attempts an arm $a \neq m(p)$ at time $t+1$. Let p' be the player that a is matched to at time t, that is, $M(a) = p'$, if a is matched at time t, and let $p' = \emptyset$, otherwise. Note that since p is matched with $m(p)$ at time t, $m(p)$ must belong to the plausible set of p at time $t+1$ by definition of the algorithm. Since p attempts $a \neq m(p)$ at $t+1$, this implies that (i) p truly prefers a over $m(p)$ by definition of E_{t+1} and (ii) a truly prefers p over p', since a must be in the plausible set of p at time $t+1$. Thus (p, a) are a blocking pair for the matching m, contradicting the assumption that $m \in M^*$. Thus we have shown $m_{t+1} = m_t = m \in M^*$. □

In the next lemma, we apply Lemma 5.5.3 repeatedly to show that the event that the current matching is unstable can be decomposed into prior events that occurred up to K steps in the past. Specifically, if the current matching is unstable, then either the UCB ranking of arms were wrong at some point in the history of length K (that is, (5.22) was false), or the matching was unstable for K consecutive steps even though (5.22) was true in all K steps.

Lemma 5.5.4 (Inclusion for unstable matching event)**.** *We have the following inclusion that holds for any* $0 \leq K < t-1$*:*

$$\{m_t \notin M^*\} \subseteq \left(\bigcup_{s=0}^{K} E_{t-s}^c\right) \bigcup \left(\bigcap_{s=0}^{K} (E_{t-s} \cap \{m_{t-s-1} \notin M^*\})\right).$$

Proof. This is an immediate consequence of Lemma 5.5.3. In fact, Lemma 5.5.3 shows

$$\{m_t \notin M^*\} \subseteq E_t^c \cup (E_t \cap \{m_{t-1} \notin M^*\}). \tag{5.23}$$

This shows Lemma 5.5.4 holds for $K = 0$. A simple induction argument shows that Lemma 5.5.4 holds for general $K > 0$, $K < t-1$. □

Lemma 5.5.4 suggests that in order to derive an upper bound on the probability that m_t is unstable, we can separately bound the probabilities of the event that the UCB ranking of arms has an error, and the event that the matching was unstable for K consecutive steps even though UCB rankings were correct in all K steps. The following lemma addresses the former.

Lemma 5.5.5 (Probability of ranking error event)**.** *The following inequality holds for any* $t > 0$*:*

$$\mathbb{P}(E_t^c) \leq \varepsilon^{-1} \cdot \sum_{(i,j,k),:a_j \prec_i a_k} \mathbb{P}(F_{j,k}^{(i)}(t) \cap \bar{A}^{(i)}(t) = j).$$

Proof. The key is the following observation. That E_t^c happens implies the existence of some player p_i and arms a_j, a_k in their plausible set at time t, such that while the arm a_j achieves the highest UCB with respect to player p_i, the player truly prefers arm a_k over a_j, Hence, this implies

$$\mathbb{P}(E_t^c) \leq \sum_{(i,j,k),:a_j \prec_i a_k} \mathbb{P}(u_{i,j}(t) > u_{i,k}(t) \cap \{j = \operatorname*{argmax}_{j'} u_{i,j'}(t)\}).$$

Recall $F_{j,k}^{(i)}(t) = \{u_{i,j}(t) > u_{i,k}(t)\}$. Lemma 5.5.5 now follows if we can show

$$\mathbb{P}(F_{j,k}^{(i)}(t) \cap \{j = \operatorname*{argmax}_{j'} u_{i,j'}(t)\}) \leq \varepsilon^{-1} \cdot \mathbb{P}(F_{j,k}^{(i)}(t) \cap \bar{A}^{(i)}(t) = j).$$

To see this, note that the player p_i will successfully pull a_j if player p_i doesn't draw a random delay and all the rest of the players draw the random delay (meaning they all attempt the same arm as they attempted in the last round). By independence of the random draws, this event happens with probability at least $\varepsilon = (1-\lambda)\lambda^{N-1}$. □

Having established this lemma, we can now easily apply the UCB argument as given in Lemma 5.4.4 to bound the relevant quantity, $\sum_{t=1}^{T} \mathbb{P}(E_t^c)$.

We proceed to analyze the probability of the event that the matching was unstable for K consecutive steps even though UCB rankings were correct in all K steps. Essentially, this requires us to establish how quickly the decentralized conflict-avoiding procedure converges to a stable matching when there are no statistical errors in the rankings of arms. To do so, we invoke a result from the stable matching literature [Abeledo and Rothblum, 1995]. First, we introduce the notion of a blocking pair that is *player-consistent*.

Definition 1 (Player-consistent blocking pair)**.** *A blocking pair (p_i, a_j) in a matching μ is* player-consistent *if*

$$a_j \succ_{p_i} a_k \quad \text{for any } k \text{ such that } (p_i, a_k) \text{ is a blocking pair in } \mu. \tag{5.24}$$

In other words, if player p_i most prefers the a_j out of all the arms that prefer p_i over the player that they are matched to in μ, then the blocking pair (p_i, a_j) is player-consistent. Notice that in Algorithm 1, at time t, if the UCB rankings are accurate, then each player p_i (who did not draw a random delay) will attempt precisely the arm a_j where (p_i, a_j) is a player-consistent blocking pair in the matching μ induced by the previous attempted actions m_{t-1}, by the definition of the plausible set.

We also require the following definition of *resolving* a blocking pair, in the context of running one step of Algorithm 1.

Definition 2 (Resolution of blocking pair)**.** *Given attempted actions $m_t \notin M^*$ and a blocking pair (p_i, a_j) in the matching induced by m_t, we say that m_{t+1} is obtained by resolving (p_i, a_j), if $m_{t+1}(p_i) = a_j$ and $m_{t+1}(p) = m_t(p)$ for all $p \in \mathcal{N}, p \neq p_i$.*

We are ready to establish a key result—that there is a strictly positive probability that a single player-consistent blocking pair is resolved in one step of Algorithm 1.

Lemma 5.5.6 (Positive probability of resolving a single blocking pair)**.** *Assume m_{t-1} is unstable. Let (p_i, a_j) be a blocking pair in m_{t-1} that is player-consistent. Condition on the event E_t. Then, with probability at least $\varepsilon = (1-\lambda)\lambda^{N-1}$, (p_i, a_j) is the only blocking pair to be resolved at time t, i.e.,*

$$\mathbb{P}\left(m_t(p_i) = a_j, m_t(p) = m_{t-1}(p)\ \forall p \in \mathcal{N}, p \neq p_i \mid E_t\right) \geq \varepsilon. \tag{5.25}$$

Proof. Assume E_t holds. Let (p_i, a_j) be any blocking pair that is player-consistent. First, we show p_i has probability at least λ of pulling the arm a_j conditioned on all of the other players attempting the same arm as they pulled at time t. Indeed, since (p_i, a_j) is a blocking pair of m_{t-1}, it means that a_j is in the plausible set of p_i at time t. As E_t occurs, and a_j is the top choice among all the arms in p_i's plausible set, the player p_i has probability at least λ of attempting a_j, and will be successful if all other players stay on the same arm as they pulled at time t. Second, independently, each of the rest of the $N-1$ players have probability at least $(1-\lambda)$ of attempting the same arm that they attempted at time $t-1$. Together, this proves equation (5.25). □

Now we can finally show that the event that the matching was unstable for K consecutive steps even though UCB rankings were correct in all K steps happens with a probability that is exponentially small in K, as stated formally in the lemma below.

Lemma 5.5.7 (Probability of not reaching a stable matching)**.** *For any $0 \leq K < t-1$, the following inequality holds:*

$$\mathbb{P}\left(\bigcap_{s=0}^{K}(\{m_{t-s-1} \notin M^*\} \cap E_{t-s})\right) \leq (1-\varepsilon^{N^4})^{\lfloor K/N^4 \rfloor}. \tag{5.26}$$

Proof. The result is a direct consequence of Lemma 5.5.6 and the theorem below.

Theorem 5.5.8 (Theorem 4.2 in Abeledo and Rothblum [1995])**.** *Given any unstable matching μ_0, there exists a sequence of blocking pairs of length at most N^4 such that resolving the sequence of blocking pairs reaches a stable matching. Moreover, this sequence of blocking pairs results from resolving blocking pairs in a* player-consistent order*, that is, any blocking pair (p_i, a_j) resolved in the current matching μ is player-consistent with respect to the matching μ.*

We now prove Lemma 5.5.7 using Lemma 5.5.6 and Theorem 5.5.8.

1. We first show Lemma 5.5.7 holds when $K = N^4$. Let $E = \cap_{s=0}^{K} E_{t-s}$. Condition on the event that E happens. Condition on the matching $\mu = m_{t-K-1}$. By Theorem 5.5.8 and Lemma 5.5.6, we know that with probability at least ε^{N^4}, a stable matching will be reached within N^4 steps of the algorithm. Since this holds for arbitrary $\mu = m_{t-K-1}$, we obtain
$$\mathbb{P}\left(\bigcap_{s=0}^{K}\{m_{t-s-1} \notin M^*\} \mid E\right) \leq 1-\varepsilon^{N^4}.$$
Thus, we have
$$\mathbb{P}\left(\bigcap_{s=0}^{K}(\{m_{t-s-1} \notin M^*\} \cap E_{t-s})\right) \leq \mathbb{P}\left(\bigcap_{s=0}^{K}\{m_{t-s-1} \notin M^*\} \mid E\right) \leq 1-\varepsilon^{N^4}.$$

2. We next generalize the result to $K > N^4$. This is straightforward, as the random seeds x in Algorithm 1 are mutually independent for any non-overlapping blocks of N^4 steps.

□

Note that in order for this bound to be meaningful, we require $K \gg \varepsilon^{-N^4} N^4$.
Finally, we are now fully equipped to prove the main result of this section.

Proof. **of Theorem 5.5.1** Let $0 \leq h_t < t$ be a time window that we are free to choose in a way that depends on the time t. By Lemma 5.5.4 and the union bound, we have

$$\mathbb{P}\left(m_t \notin M^*\right) \leq \mathbb{P}\left(\bigcap_{s=0}^{h_t}(E_{t-s} \cap \{m_{t-s-1} \notin M^*\})\right) + \sum_{s=0}^{h_t}\mathbb{P}(E_{t-s}^c).$$

Let $g_t = \lfloor h_t/N^4 \rfloor$. Lemmas 5.5.5 and 5.5.7 immediately yield the following:

$$\mathbb{P}\left(m_t \notin M^*\right) \leq (1-\varepsilon^{N^4})^{g_t} + \varepsilon^{-1}\sum_{s=0}^{h_t}\sum_{(i,j,k),:a_j \prec_i a_k}\mathbb{P}(F_{j,k}^{(i)}(t-s) \cap \bar{A}^{(i)}(t-s) = j).$$

Summing these inequalities over t up to T, we obtain

$$\begin{aligned}\sum_{t=1}^{T}\mathbb{P}\left(m_t \notin M^*\right) &\leq \sum_{t=1}^{T}(1-\varepsilon^{N^4})^{g_t} + \varepsilon^{-1}\sum_{t=1}^{T}\sum_{s=0}^{h_t}\sum_{(i,j,k),:a_j \prec_i a_k}\mathbb{P}(F_{j,k}^{(i)}(t-s) \cap \bar{A}^{(i)}(t-s) = j) \\ &= \sum_{t=1}^{T}(1-\varepsilon^{N^4})^{g_t} + \varepsilon^{-1}\sum_{(i,j,k),:a_j \prec_i a_k}\sum_{s=0}^{h_T}\sum_{\substack{t:s\leq h_t \\ 1\leq t\leq T}}\mathbb{P}(F_{j,k}^{(i)}(t-s) \cap \bar{A}^{(i)}(t-s) = j)\end{aligned} \tag{5.27}$$

We seek upper bounds for the terms on the right-hand side. Focus on the second term in equation (5.27). Recall the standard UCB Lemma (e.g., Lemma 5.4.4):

$$\sum_{\substack{t:s\leq h_t \\ 1\leq t\leq T}}\mathbb{P}(F_{j,k}^{(i)}(t-s) \cap \bar{A}^{(i)}(t-s) = j) \leq 6 \cdot \left(\frac{1}{\Delta^2}\log(T) + 1\right), \text{ for each } s.$$

Substituting this bound into equation (5.27) yields

$$\sum_{t=1}^{T}\mathbb{P}\left(m_t \notin M^*\right) \leq \sum_{t=1}^{T}(1-\varepsilon^{N^4})^{g_t} + 6\varepsilon^{-1}NK^2\left(h_T+1\right)\left(\frac{1}{\Delta^2}\log(T)+1\right), \tag{5.28}$$

where we have used the fact that there are at most NK^2 triplets (i,j,k) such that $a_j \prec_i a_k$. We now choose a specific sequence (h_t) to optimize the upper bound. Let $B \geq 1$ be determined later. Set $h_t = \min\{t, B\} - 1$. With this choice of h_t, and after some elementary computations, we can bound the first term in equation (5.28) by

$$\sum_{t=1}^{T}(1-\varepsilon^{N^4})^{g_t} \leq 3 \cdot \sum_{t=1}^{T}\exp(-h_t\varepsilon^{N^4}/N^4) \leq 6 \cdot \left(T\exp\left(-\frac{B\varepsilon^{N^4}}{2N^4}\right) + \frac{N^4}{\varepsilon^{N^4}}\right).$$

The second term in equation (5.28) is bounded by $6\varepsilon^{-1}BNK^2\left(\frac{1}{\Delta^2}\log(T)+1\right)$, since $h_T < B$ by definition. Consequently, these two bounds lead to the following (that holds for all B)

$$\sum_{t=1}^{T}\mathbb{P}\left(m_t \notin M^*\right) \leq 6\cdot\left(T\exp\left(-\frac{B\varepsilon^{N^4}}{2N^4}\right)+\frac{N^4}{\varepsilon^{N^4}}+\frac{1}{\varepsilon}NK^2B\left(\frac{1}{\Delta^2}\log(T)+1\right)\right).$$

By carefully setting $B = 2\left\lceil\frac{N^4}{\varepsilon^{N^4}}\log\left(T\right)\right\rceil$, we obtain the final bound as desired

$$\sum_{t=1}^{T}\mathbb{P}\left(m_t \notin M^*\right) \leq 24\cdot\frac{N^5K^2}{\varepsilon^{N^4+1}}\log(T)\cdot\left(\frac{1}{\Delta^2}\log(T)+3\right).$$

□

5.6 Strategy and Incentive Compatibility

In this section, we examine the CA-UCB algorithm from the perspective of incentive compatibility.

Thus far we have given stable regret guarantees for each player, when all players follow the same algorithm, whether assuming a global ranking of players (Theorem 5.4.1), or without making assumptions on the market's preferences (Theorem 5.5.1). Given these results, a natural question to consider, in the decentralized setting, is whether the players are indeed incentivized to run the same algorithm as everyone else. In other words, could any single player benefit from running a different algorithm, when all other players are running Algorithm 1?

A positive result for globally ranked players

In the setting of Section 5.4, when players are globally ranked, we can show that the gains from deviating are limited. The following proposition gives an lower bound on the stable regret of the deviating player that scales logarithmically in the horizon n, for any algorithm that they run. This implies that the time-averaged gains from deviating must vanish quickly as learning progresses.

Proposition 5.6.1 (Incentive compatibility under globally ranked players)**.** *Under Assumption 4, suppose that all players other than player p_k run Algorithm 1 with $\lambda = 0$, and p_k can run any algorithm. The following lower bound on player p_k's stable regret holds:*

$$R_k(n) \geq 6k^2(K-k)\left(\frac{\log n}{\Delta^2}+1\right)\left(\min_{j:\Delta_j^{(k)}<0}\Delta_j^{(k)}\right). \tag{5.29}$$

This result follows from a simple application of the same arguments that we developed to prove Theorem 5.4.1. The key idea is as follows. A deviating player that is rank k in the market can successfully pull an arm a_i that they prefer to their stable arm, only if the better-ranked player p_i is not pulling their stable arm a_i in the same round. This can only happen if p_i or a better-ranked player had a mistake in their UCB rankings and pulled a suboptimal arm within the last k rounds, since all players other than p_k are indeed following the CA-UCB algorithm. The gains to deviating are limited for player p_k when all the arms have the same preferences, precisely because p_k cannot affect the actions of better ranked players. A complete proof can be found in Appendix 5.11.

A negative result

Given that we have a general stable regret guarantee for arbitrary preferences, established in Section 5.5, one might ask if there also exists a general incentive compatibility result for Algorithm 1. Unfortunately, the answer is a negative one. The following proposition shows, by way of counterexample, that there can be no blanket incentive compatibility guarantee for Algorithm 1 without making additional assumptions, such as on the preference structure.

Proposition 5.6.2. *Consider the market of three players and three arms with preferences as given in Example 5.11.2. When two players p_1 and p_2 run Algorithm 1 with any $\lambda \in (0, 1/4)$, there exists a sequence of actions $\{A^{(3)}(t)\}_{t=1\ldots n}$ for player p_3 such that p_3's stable regret can be upper bounded as:*

$$R_3(n) \leq -C_1 \cdot n + C_2 \left(\frac{1}{\Delta^2} \log(T) + 1 \right), \tag{5.30}$$

where C_1 and C_2 are constants that depend only on $\lambda, \Delta_1^{(3)}, \Delta_\emptyset^{(3)}$. Moreover, there exists $\Delta_1^{(3)}, \Delta_\emptyset^{(3)}$ such that C_1 is strictly positive.

The above upper bound on the deviating player p_3's stable regret shows that there exists a set of preferences and arm reward gaps such that a player could make significant gains over their stable arm by not running Algorithm 1. We defer the full description of Example 5.11.2 and the proof of Proposition 5.6.2 to Section 5.11. In this example, p_3 has stable arm a_3 but prefers a_1. Because the arms have idiosyncratic preferences (as opposed to shared preferences), p_3 could pull a suboptimal arm in order to 'trick' p_1 into not attempting a_1 two rounds later, by exploiting the conflict avoidance mechanism; p_3 can then successfully pull a_1 for one round, with some probability. As long as the reward for p_3 from a_1 is large enough, p_3 is guaranteed a strictly negative stable regret that is linear in the horizon n.

We have shown that Algorithm 1 is not incentive compatible in the fully general setting. It therefore remains an open question whether there exists an algorithm with low stable regret, under arbitrary preferences, that also has an incentive compatibility guarantee under the same.

5.7 Simulation experiments for random preferences

In our theoretical analysis we considered two cases: markets in which the players are globally ranked (i.e. all arms have the same preferences over players) and markets with arbitrary preferences. For the first case Theorem 5.4.1 we were able to prove a regret upper bound that resembles the guarantee derived in the centralized case in Chapter 4. However, in the case of general markets our guarantee (Theorem 5.5.1) has an exponential dependence on the size of the market.

In this section, through empirical evaluations we show that the true performance of our proposed method is likely better than our guarantee suggests for markets with randomly drawn preferences. More precisely, we perform two sets of simulations. In the first set, we investigate how the average regret and market stability depend on the size of the market in balanced markets—markets with an equal number of players and arms—with preferences drawn from a distribution that will be specified later. We find that empirically the algorithm converges more slowly for larger number of players as expected, though the dependence on the number of players, N, appears to be significantly better than the exponential dependence appearing in Theorem 5.5.1.

In the second set of experiments, we vary the heterogeneity of the players' preferences. We perform this experiment because one might expect that in markets in which different players have the same preferences there would be more conflicts (since different players have an incentive to attempt the same arms). Despite this intuition, our simulations show that CA-UCB performs equally well in markets with different level of heterogeneity. To sum up, our simulations show that not only is Theorem 5.5.1 overly pessimistic, but that CA-UCB avoids conflicts equally well in different markets.

For all experiments we use Algorithm 1 with delay probability $\lambda = 0.1$. We now present the details of our simulations.

Varying the size of the market. We examine balanced markets of size $N \in \{5, 10, 15, 20\}$, and sample each player's and arm's ordinal preferences uniformly at random. For all players the reward gaps between consecutively ranked arms are chosen to be equal to $\Delta = 1$, regardless of the market size. The rewards are normally distributed with unit variance. We sampled ten markets as such, and run Algorithm 1 once on each market.

For each market size N, we plot the mean, over ten markets, of the following two quantities: (i) the maximum average regret among players, $\max_{k \in \mathcal{N}} R_k(n)$, and (ii) the averaged market stability $\sum_{t=1}^{T} \mathbb{P}(m_t \notin M^*)$ for horizon n up to 5000. As can be seen in Figure 5.1, both the average regret and the market stability converge more slowly for larger markets. However, the dependence on N appears to be much better than exponential.

Varying the heterogeneity of the players' preferences. We examine balanced markets of size 10, and sample each arm's ordinal preferences uniformly at random. To sample the mean rewards $\mu_k(i)$ of arm a_i for player p_k we rely on random utility model used by

Ashlagi et al. [2017b], with a slight modification:

$$\begin{aligned} x_i &\overset{i.i.d.}{\sim} \text{Uniform}([0,1]) \\ \varepsilon_{i,k} &\overset{i.i.d.}{\sim} \text{Logistic}(0,1) \\ \overline{\mu}_i^{(k)} &= \beta x_i + \varepsilon_{i,k} \\ \mu_k(i) &= \#\{j : \overline{\mu}_j^{(k)} \leq \overline{\mu}_i^{(k)}\} \end{aligned}$$

The intermediate utilities $\overline{\mu}_i^{(k)}$ are sampled according to random utility model used by Ashlagi et al. [2017b]. We map these random utilities to $\mu_k(i)$ so that the reward gaps between consecutively ranked arms are kept constant at $\Delta = 1$. The parameter $\beta > 0$ determines the degree of correlation between the players' preferences. As β increases the correlation between the players' preferences also increases. In fact, in the limit as $\beta \to \infty$, all the players share the same preferences with probability 1.

As before, the rewards are normally distributed with unit variance. We sample ten markets for each β value, and plot the maximum average regret among players as well as the averaged market stability for horizon n up to 5000 in Figure 5.2. As can be seen, there is no discernible difference in the convergence of Algorithm 1 in terms of regret or market stability, for markets with different levels of preference heterogeneity.

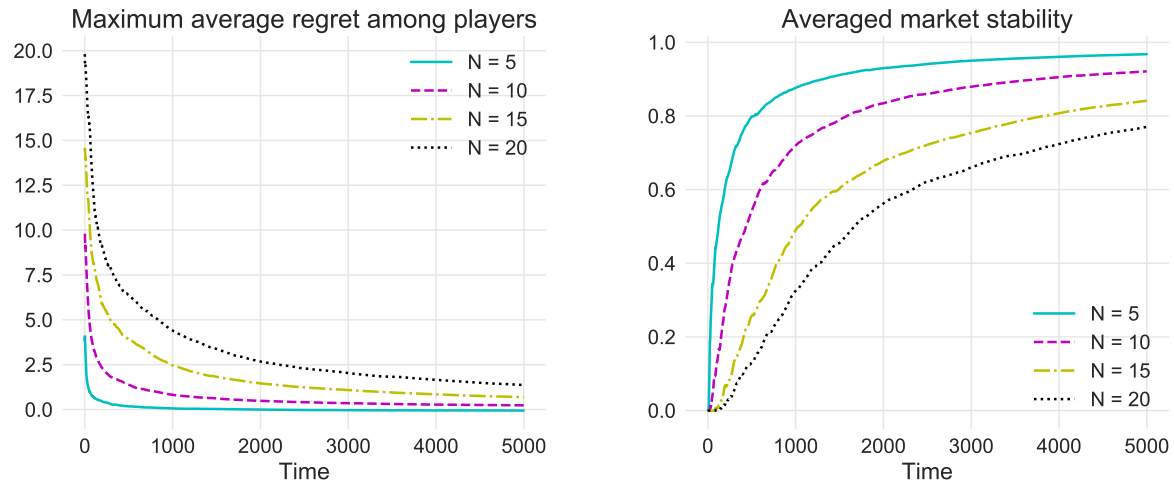


Figure 5.1: Varying the number of players. The plot on the left shows the maximum average regret among players and the plot on the right shows the averaged market stability.

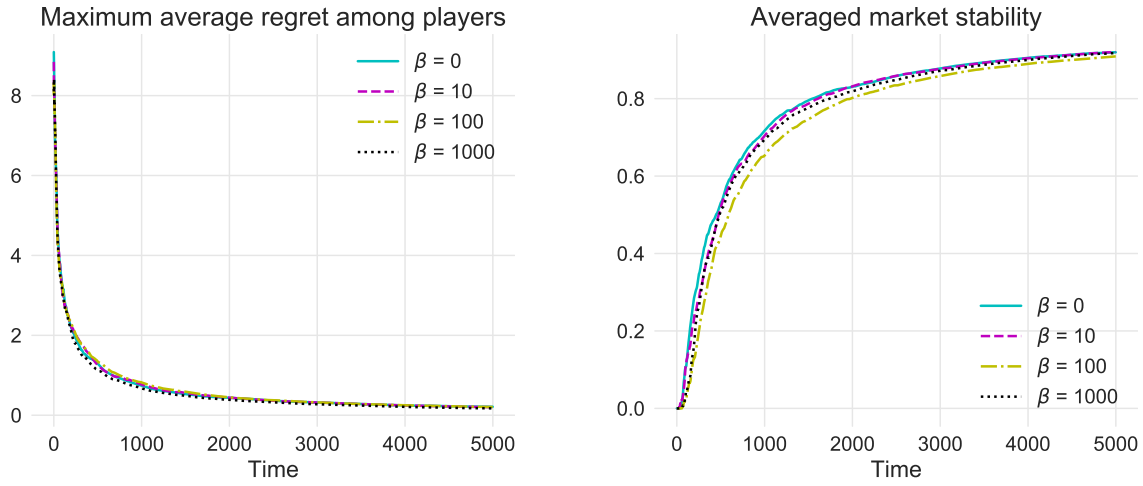


Figure 5.2: Varying the heterogeneity of the players' preferences. The plot on the left shows the maximum average regret among players and the plot on the right shows the averaged market stability. The larger the β parameter, the more correlated the players' preferences are on average.

5.8 Related Work

There has been significant recent interest in stochastic multi-armed bandits problems with multiple, interacting players [Cesa-Bianchi et al., 2016, Shahrampour et al., 2017]. In one formulation, known as *bandits with collision*, multiple players choose from the same set of arms, and if two or more players choose the same arm, no reward is received by any player [Liu and Zhao, 2010, Anandkumar et al., 2011, Avner and Mannor, 2014, Bistritz et al., 2020, Bubeck et al., 2020a,b, Kalathil et al., 2014, Rosenski et al., 2016, Lugosi and Mehrabian, 2018]. In this setting, players are typically assumed to be cooperative, that is, their goal is to maximize the collective reward. Bistritz and Leshem [2018] and Boursier and Perchet [2020] consider the setting where agents have heterogeneous preferences over arms, and the latter work also analyzes the effect of selfish players whose goal is to maximize individual rewards. Avner and Mannor [2016] and Darak and Hanawal [2019] considered a "stable configuration" as a solution concept in the heterogeneous player preference setting; however, because the arms do not have preferences in their setting, their notion of "stability" is distinct from that of two-sided stable matching. Bubeck et al. [2020b] also delineated the optimal rates for the non-stochastic version of the cooperative problem.

In Chapter 4, we introduced a multi-player stochastic multi-armed bandits problem motivated by two-sided matching markets, where arms also have preferences, and in case of collision only the most preferred player receives a reward. Unlike in the aforementioned line of work, where the natural goal is to find a maximum matching between players and

arms, a more appropriate goal here is to find a stable matching. In the centralized setting, where a platform can coordinate the actions of players at each round, our algorithm combining the upper confidence bound method and the deferred acceptance algorithm attains $\mathcal{O}(\log(T)/\Delta^2)$ stable regret, which is order-optimal. In Section 5.10, we analyze a suboptimal algorithm based on explore-then-commit for the decentralized setting. Follow-up work by Sankararaman et al. [2020] on the decentralized setting analyzed an order-optimal algorithm for globally ranked players. A more detailed discussion of this work is in Section 5.9.

The two-sided stable matching problem with preference learning has been studied in other dynamic settings under different assumptions. Given the large space of modeling choices, there has been a flowering of research on two-sided matching models that highlight different challenges introduced by uncertainty and decentralization. One modeling choice is to define arrival and departure processes for market participants, as opposed to analyzing a fixed set of players and arms. Johari et al. [2017] studied a sequential matching problem in which the market participants satisfy certain arrival processes, and the participants on the demand side of the market have a 'type' that is learned through bandit feedback.

Another choice is how one formulates the cost of preference learning. Ashlagi et al. [2017a], which studies the costs of communication and learning for stable matching, formulates preference learning as querying a costly but noiseless choice function. Different players can query their choice functions independently; thus there is no congestion in the preference learning process. Many models studied in the literature on information acquisition in two sided matching [see Lee and Schwarz, 2009, Immorlica et al., 2020, and references therein] also do not capture congestion in the information acquisition stage. In some markets, however, obtaining information about the other side of the market itself could lead to congestion and thus the need for strategic decisions. For example, Roth and Sotomayor [1990, chap. 10] note that graduating medical students go to interviews to ascertain their own preferences for hospitals, but the collection of interviews that a student can schedule is limited. In the model that is studied in the current work, congestion in preference learning is captured by conflicts when two or more players attempt to pull the same arm.

Other models of uncertainty in two-sided matching that do not explicitly consider preference learning have also been studied. In this setting, there has been much interest in decentralized models. For example, Niederle and Yariv [2009] studied a decentralized market game in which firms make directed offers to workers, agents have aligned preferences, and equilibrium outcomes under preference uncertainty are analyzed. Arnosti et al. [2014] employed mean field modeling to analyze the welfare costs of not knowing the availability of agents, as opposed to preferences. Ashlagi et al. [2019] considered providing match recommendations to participants in markets for which both sides of the market propose with some probability, and a successful match occurs only in the case of a mutual proposal. Dai and Jordan [2020] study a single-stage matching problem with uncertain preferences where players learn from historical data and act in a decentralized manner.

Lastly, the empirical aspects of stable matching in decentralized settings have also garnered significant research interest [Das and Kamenica, 2005, Echenique and Yariv, 2012, Pais et al., 2012].

5.9 Discussion

In this section, we discuss the strengths and limitations of Algorithm 1, in the context of broader themes in decentralized matching and multiplayer bandit learning. We also suggest future research directions motivated by our current findings.

Single-phase algorithm One advantage of Algorithm 1 is its simplicity, specifically the fact that it does not involve separate phases or subroutines. Recent work by Sankararaman et al. [2020] studied an algorithm ('UCB-D3') for decentralized matching bandits, assuming *globally ranked players*, that proceeds in phases of exponentially increasing length; each phase comprises of a learning stage, where players choose arms according to their own UCBs, followed by a communication subroutine, where players broadcast their preferred arms to other players. In contrast, our algorithm does not require players to keep track of which phase they are in, or when to begin a subroutine. Not having separate algorithmic phases is desirable because multiple phases requires players to synchronize their transition from one phase to the next. In 'more decentralized' situations this may not be possible. For example, players may enter the market at different times, or leave the market for a number of rounds only to return later [see e.g., Akbarpour et al., 2020]. The CA-UCB algorithm can be run in such cases without modification and is still guaranteed to have small regret.

Dependence of stable regret on market size While both UCB-D3 and our method are guaranteed to achieve $\mathcal{O}(\log(T)/\Delta^2)$ stable regret for globally ranked players, the regret guarantee for UCB-D3 has a better dependence on the number of arms (which upper bounds the number of players). In the worst case, the guarantee on the regret of UCB-D3 depends on the square of the number of arms while the guarantee on the regret of our method depends on the cube of the number of arms. The optimal order-dependence on the rank k and the number of arms K is still an open question, since the lower bound [e.g., Corollary 6 in Sankararaman et al., 2020] and upper bounds currently do not match. Another interesting question is whether UCB-D3's better regret guarantee under these assumptions translates to better performance in practice; an in-depth empirical comparison of UCB-D3 and CA-UCB will be needed and is beyond the scope of the current work.

Random delays Another important feature of Algorithm 1 is the injection of additional randomness through each player's independently drawn random delays. Randomization is key for this algorithm to achieve a $\mathcal{O}(\log(T)^2)$ regret guarantee in the case of *arbitrary* two-sided preferences. Intuitively, the added randomness allows players to escape conflict cycles, as illustrated in Examples 5.3.1 and 5.11.1. Technically, it allows us to leverage a result from Abeledo and Rothblum [1995]) to show that the players must converge to a stable matching (which may not be unique), in a low-regret sense. Nevertheless, repeating one's previous action with a constant probability at every step could be considered wasteful. Are there

other, more efficient ways of utilizing randomness as an implicit coordination mechanism than random delays?

Improving the stable regret under arbitrary preferences While Algorithm 1 is the first method to provably achieve polylogarithmic regret in markets with arbitrary preferences, we believe there is a significant room for the development of better algorithms. In particular, for markets with arbitrary preferences, the regret guarantee for our method depends exponentially on the number of players. This dependence arises because our regret analysis hinges on a reduction to the convergence rate of the corresponding randomized decentralized matching dynamics under *known* preferences. As shown in Ackermann et al. [2008] and Hoffman et al. [2013], existing randomized dynamics for decentralized matching under known preferences have worst-case convergence time that is exponential in the number of market participants. While this may suggest that there is indeed a real computational barrier in the arbitrary preferences setting, it might be possible to improve upon the exponential dependence by considering sub-classes of two-sided preferences or randomly drawn preferences. For example, Algorithm 1 has improved rates if we assume that the players are globally ranked.

It is also not clear that the $\mathcal{O}((\log n)^2)$ dependence on the horizon is optimal in this setting, even though it is unavoidable given our analysis strategy and our algorithm. Obtaining a regret bound that depends polynomially on the number of players and arms *and* has an optimal order dependence on the horizon may require a new algorithm.

Information available to players A player that implements CA-UCB must observe the successful arm pulls of all other players. On one hand, by leveraging this information our algorithm ensures that players avoid conflicts most of the time. On the other hand, it is not clear that such information is absolutely necessary for achieving sublinear regret in general markets. For example, UCB-D3 [Sankararaman et al., 2020], which achieves sublinear regret in the setting of globally ranked players, does not require players to see the actions of other players. However, players must participate in a rank estimation routine, which relies on the assumption that the players are ranked globally.

Conclusion and open questions In this work we have made progress on the problem of stochastic bandits in decentralized matching markets. Still, many open questions remain. We conclude by highlighting the most intriguing directions for future inquiry:

1. *Better algorithms and matching lower bounds.* Even though algorithms such as UCB-D3 [Sankararaman et al., 2020] and CA-UCB have stable regret that is almost order-optimal in the setting of globally ranked players, there is still a lot of room for improvement in the setting of arbitrary preferences. Is there a large class of preferences for which one can show matching upper and lower regret bounds, in terms of the dependence on the horizon, the reward gap, and size of the market?

2. *Incentive compatibility in the decentralized setting.* Unlike in the centralized setting (Chapter 4), where a single algorithm was shown to be incentive compatible given any set of preferences, decentralization appears to pose more challenges for incentive compatibility. As seen in Section 5.6, the randomized conflict avoidance mechanism of Algorithm 1 can be strategically exploited by a deviating player when arm preferences are uncorrelated. How fundamental is this difficulty to the decentralized setting, and can it be overcome by a better algorithm?

5.10 Decentralized Explore-Then-Commit

In the decentralized setting of the matching bandits problem, we propose a simple algorithm based on Explore-Then-Commit (ETC) that can achieve low player-*optimal* regret, albeit at a suboptimal rate. A key observation behind this algorithm is that the Gale-Shapley algorithm can be implemented with simultaneous proposing [see e.g. Roth, 2007, Theorem 1], hence a central platform is not necessary for the players to reach a stable matching. We analyze this simple algorithm in order to motivate the search for more efficient algorithms in the decentralized setting.

Description of the algorithm There are three stages. Stage 1 has HK rounds. In stage 1 ("Exploration"), every K rounds, each player independently samples a random permutation of arms and attempts arms in that order. Agents update the respective sample means of the arms only if the pull was successful. In stage 2 ("Simultaneous Proposing GS"), each round each player attempts the arm with the highest sample mean that they haven't had a conflict on in Stage 2. Stage 2 continues for N rounds. In stage 3 ("Exploitation"), every player keeps pulls the last arm they pulled successfully in stage 2.

In the following result, we analyze the regret of the decentralized ETC.

Proposition 5.10.1 (Regret bound for decentralized ETC)**.** *Consider Decentralized ETC with stage 1 lasting HK rounds. Let $\overline{\Delta}_{i,j}$, $\overline{\Delta}_{i,\max}$ and Δ be defined as before in Theorem 4.3.1. Let $\rho_{N,K} := \left(1 - \frac{1}{K}\right)^{N-1}$. The expected player-optimal regret of player p_i is upper bounded by*

$$\overline{R}_i(n) \leq HK\mu_i(\overline{m}(i)) + (n - HK)\overline{\Delta}_{i,\max} NK \left(2\exp\left(-\frac{H\rho_{N,K}^2}{2}\right) + \exp\left(-\frac{H\rho_{N,K}\Delta^2}{8}\right)\right). \tag{5.31}$$

Proof. Suppose stage 1 lasts HK rounds. Fix the agent. For any particular attempt on arm i, let Y_i denote the event of a successful pull. The probability of a successful pull is bounded from below by the probability of a successful pull for an agent that is the least preferred by the arm attempted.

$$p_i := \mathbb{P}\{Y_i = 1\} \geq \left(1 - \frac{1}{K}\right)^{N-1} =: \rho_{N,K}$$

Let T_i be the number of times arm i was pulled in stage 1. By the independence of the random permutations sampled, we have that $T_i \sim \text{Binomial}(H, p_i)$. We can use a standard tail bound:

$$\mathbb{P}\{T_i \leq t\} \leq \exp\left(-2\frac{(Hp_i - t)^2}{H}\right)$$

Stage 2 gives rise to a matching that is stable according to the order of the average rewards ($\hat{R}_j^i(T)$), after N rounds. This is essentially the Gale-Shapley algorithm but with simultaneous proposals.

Now we bound the probability that this matching is agent optimal according to the true preferences. If any agent j ranks arm k and arm k' wrongly, we must have $\hat{R}_j^k(H) > \hat{R}_j^{k'}(H)$ but $\mu_j(k') > \mu_j(k)$. Therefore, we may bound the probability of a blocking pair using the sub-Gaussianity of $\hat{R}_j^k(H) - \hat{R}_j^{k'}(H)$. Let $A_{k,k'}$ denote the event $\{(\hat{R}_j^{k'} - \hat{R}_j^k) - (\mu_j(k') - \mu_j(k)) \leq -(\mu_j(k') - \mu_j(k))\}$.

$$\begin{aligned}
\mathbb{P}(A_{k,k'}) &= \sum_{j,j'<H} \mathbb{P}(A_{k,k'} \cap T_k = j \cap T_{k'} = j') \\
&= \sum_{j \wedge j' < h} \mathbb{P}(A_{k,k'} \cap T_k = j \cap T_{k'} = j') + \sum_{j \wedge j' \geq h} \mathbb{P}(A_{k,k'} \cap T_k = j \cap T_{k'} = j') \\
&\leq \mathbb{P}(T_k < h) + \mathbb{P}(T_{k'} < h) + \sum_{j \wedge j' \geq h} \mathbb{P}(A_{k,k'} \mid T_k = j, T_{k'} = j') \cdot \mathbb{P}(T_k = j, T_{k'} = j') \\
&\leq \mathbb{P}(T_k < h) + \mathbb{P}(T_{k'} < h) + \sum_{j \wedge j' \geq h} \exp\left(-\frac{(j \wedge j')(\Delta_{k,k'})^2}{4}\right) \mathbb{P}(T_k = j, T_{k'} = j') \\
&\leq \mathbb{P}(T_k < h) + \mathbb{P}(T_{k'} < h) + \exp\left(-\frac{h(\Delta_{k,k'})^2}{4}\right).
\end{aligned}$$

Choosing $h = \frac{1}{2}Hp_i$ gives

$$\mathbb{P}(A_{k,k'}) \leq 2\exp\left(-\frac{Hp_i^2}{2}\right) + \exp\left(-\frac{Hp_i(\Delta_{k,k'})^2}{8}\right) \leq 2\exp\left(-\frac{H\rho_{N,K}^2}{2}\right) + \exp\left(-\frac{H\rho_{N,K}\Delta^2}{8}\right)$$

By Lemma 4.3.2, we only have to consider $k' = \overline{m}(i)$ and k such that $\overline{\Delta}_{i,k} > 0$, so there are at most K such pairs, for each agent.

□

5.11 Omitted examples and proofs

Example 5.11.1

In this section, we present a second counterexample in which CA-UCB without random delays (i.e., $\lambda = 0$) would fail to converge to a stable matching and the players can enter

into a conflict cycle. In this example, neither the arms nor the players are globally ranked. In contrast to Example 5.3.1, the type of coordination failure seen in Example 5.11.1 is unrelated to the failure of the players to learn their rewards. In fact, they can enter into such a cycle even after they have acquired perfect information on all the arms.

Example 5.11.1 (3-player market with non-unique stable matching)**.** *Let the set of players be* $\mathcal{N} = \{p_1, p_2, p_3\}$ *and the set of arms be* $\mathcal{K} = \{a_1, a_2, a_3\}$*, with true preferences given by:*

$$\begin{aligned} p_1 &: a_3 \succ a_2 \succ a_1 & \qquad a_1 &: p_3 \succ p_2 \succ p_1 \\ p_2 &: a_1 \succ a_3 \succ a_2 & \qquad a_2 &: p_1 \succ p_3 \succ p_2 \\ p_3 &: a_2 \succ a_1 \succ a_3 & \qquad a_3 &: p_2 \succ p_1 \succ p_3. \end{aligned}$$

Then the conflict-avoiding algorithm cycles even when the preferences of the players are known. Suppose the players are following Algorithm 1, and their UCB rankings for the arms always coincide with their true preferences. The cycle it enters is as follows:

- *Time* t*:* p_1 *and* p_3 *conflict on* a_2*,* p_1 *wins.*

  p_2 *pulls* a_1*.*

- *Time* $t+1$*:* p_3 *attempts* a_1 *because* a_2 *is not in its plausible set.* p_2 *and* p_3 *conflict on* a_1*,* p_3 *wins.*

  p_1 *pulls* a_3 *because* a_3 *was not pulled by any player at time* t*.*

- *Time* $t+2$*:* p_2 *attempts* a_3 *because* a_1 *is not in its plausible set.* p_1 *and* p_2 *conflict on* a_3*,* p_2 *wins.*

  p_3 *pulls* a_2 *because* a_2 *was not pulled by any player at time* $t+1$

At time $t+3$*, the players attempt the same actions as they did at time* t*, entering into a cycle where there is a conflict at every round henceforth.*

Previous work has found other examples where sequentially resolving blocking pairs in an unstable matching leads to cycling [Knuth, 1997, Roth and Vande Vate, 1990, Abeledo and Rothblum, 1995]. Example 5.11.1 shows that players following the decentralized conflict-avoiding protocol (where more than one blocking pair may be resolved at every time step) can also enter into cycles.

These examples highlight the failure modes of decentralized conflict-avoiding algorithms. One way to escape these failure modes is by introducing randomness, such that the probability of coordination failures becomes exponentially small. This is the motivation for incorporating random delays into Algorithm 1.

Proof of Lemma 5.4.4

Proof. Our proof is essentially identical to the single-agent UCB analysis in Section 2.2 of Bubeck and Cesa-Bianchi [2012]. Assuming that the event

$$F_{j,k}^{(i)}(t) = \{u_{i,j}(t) > u_{i,k}(t)\}$$

is true, then at least one of the three following events must occur:

$$\mathcal{E}_1(t) = \left\{ \widehat{\mu}_{i,j}(t) > \mu_j^{(i)} + \sqrt{\frac{3\log t}{2\bar{T}_j^{(i)}(t)}} \right\},$$
$$\mathcal{E}_2(t) = \left\{ \widehat{\mu}_{i,k}(t) + \sqrt{\frac{3\log t}{2\bar{T}_k^{(i)}(t)}} < \mu_k^{(i)} \right\},$$
$$\mathcal{E}_3(t) = \left\{ \bar{T}_j^{(i)}(t) \leq \frac{6}{\Delta^2}\log(t) \right\}.$$

To see this, suppose that none of three events $\mathcal{E}_1(t)$, $\mathcal{E}_2(t)$ and $\mathcal{E}_3(t)$ occur. Then,

$$\begin{aligned}\widehat{\mu}_{i,k}(t) + \sqrt{\frac{3\log(t)}{2T_{i,k}(t)}} &\geq \mu_i(k) \geq \mu_i(j) + \Delta \geq \mu_i(j) + \sqrt{\frac{6\log(t)}{T_{i,j}(t)}} \\ &\geq \widehat{\mu}_{i,j}(t) + \sqrt{\frac{3\log(t)}{2T_{i,j}(t)}}\end{aligned}$$

which is a contradiction because the left-hand side equals $u_{i,k}(t)$ and the right-hand side equals $u_{i,j}(t)$.

Let $u > 0$ be some value to be chosen later. Then, we have

$$\begin{aligned}\sum_{t=1}^{T} \mathbf{1}\{F_{j,k}^{(i)}(t) \cap \bar{A}^{(i)}(t) = j\} &= \sum_{t=1}^{T} \mathbf{1}\{F_{j,k}^{(i)}(t) \cap \bar{A}^{(i)}(t) = j \cap T_{i,j}(t) \leq u\} \\ &\quad + \sum_{t=1}^{T} \mathbf{1}\{F_{j,k}^{(i)}(t) \cap \bar{A}^{(i)}(t) = j \cap T_{i,j}(t) > u\}.\end{aligned}$$

Therefore, if we choose $u = \frac{6}{\Delta^2}\log(t)$, we obtain

$$\begin{aligned}\sum_{t=1}^{T} \mathbf{1}\{F_{j,k}^{(i)}(t) \cap \bar{A}^{(i)}(t) = j\} &= u + \sum_{t=u+1}^{T} \mathbf{1}\{F_{j,k}^{(i)}(t) \cap \bar{A}^{(i)}(t) = j \cap T_{i,j}(t) > u\} \\ &\leq u + \sum_{t=\lfloor u \rfloor + 1}^{T} \mathbf{1}\{\mathcal{E}_1(t)\} + \sum_{t=\lfloor u \rfloor + 1}^{T} \mathbf{1}\{\mathcal{E}_2(t)\}.\end{aligned}$$

We are left to establish an upper bound on $\mathbb{P}(\mathcal{E}_1(t))$ and $\mathbb{P}(\mathcal{E}_2(t))$. We can do this by a simple application of a union bound and concentration:

$$\begin{aligned}\mathbb{P}(\mathcal{E}_1(t)) &\leq \mathbb{P}\left(\exists s \in \{1,2,\dots t\} \colon \widehat{\mu}_{i,j}(s) + \sqrt{\frac{3\log(t)}{2s}} \leq \mu_j^{(i)}\right) \\ &\leq \sum_{s=1}^{t} \mathbb{P}\left(\widehat{\mu}_{i,j}(s) + \sqrt{\frac{3\log(t)}{2s}} \leq \mu_j^{(i)}\right) \\ &\leq \sum_{s=1}^{t} \frac{1}{t^3} = \frac{1}{t^2},\end{aligned}$$

where the last inequality follows by a standard concentration argument for independent sub-Gaussian random variables. The probability of $\mathcal{E}_2(t)$ occurring can be upper bounded similarly. Then, using $\sum_{t=1}^{\infty} t^2 = \frac{\pi^2}{6}$ yields the conclusion. □

Proof of Proposition 5.6.1

Proof. By definition, player p_k's regret can be lower-bounded as follows:

$$R_k(n) \geq \sum_{i:a_i \succ_{p_k} a_k} \Delta_i^{(k)} \cdot \mathcal{E}\left[T_{k,i}(n)\right] \geq \left(\min_{i:\Delta_i^{(k)}<0} \Delta_i^{(k)}\right) \cdot \sum_{i:a_i \succ_{p_k} a_k} \mathcal{E}\left[T_{k,i}(n)\right]. \tag{5.32}$$

Since $a_i \succ_{p_k} a_k$ implies that $i < k$, we may proceed to upper bound $\sum_{i:i<k} \mathcal{E}\left[T_{k,i}(n)\right]$. We claim that the following inclusion is true:

$$\bigcup_{i=1}^{k-1} \left\{\bar{A}^{(k)}(t) = a_i\right\} \subseteq \bigcup_{\substack{1\leq j<k \\ j<l\leq K}} \bigcup_{t-k\leq t'\leq t} \Lambda_l^{(j)}[t']. \tag{5.33}$$

The argument is as follows. If $\left\{\bar{A}^{(k)}(t) = a_i\right\}$ holds for some i, we know that p_i at time t did not attempt to pull a_i. They either attempted to pull an arm $a_{i'}$ with $i' > i$ or with $i' < i$. Since we know that p_i is running Algorithm 1, in the former case, we can apply Lemma 5.4.2 to player p_i; in the latter case, we can apply equation (5.11), also to player p_i. This establishes equation (5.33).

Since all players p_j with $j < k$ are running Algorithm 1, we may apply Lemma 5.4.4 to yield

$$\sum_{i:i<k} \mathcal{E}\left[T_{k,i}(n)\right] \leq 6(k+1) \sum_{\substack{1\leq j<k \\ j<l\leq K}} \left(\frac{\log n}{|\Delta_l^{(j)}|^2} + 1\right) \leq 6k^2(K-k)\left(\frac{\log n}{\Delta^2} + 1\right). \tag{5.34}$$

Substituting the above into (5.32) yields the desired lower bound. □

Proof of Proposition 5.6.2

Example 5.11.2. *Let the set of players be $\mathcal{N} = \{p_1, p_2, p_3\}$ and the set of arms be $\mathcal{K} = \{a_1, a_2, a_3\}$, with true preferences given by:*

$$\begin{aligned} p_1 &: a_1 \succ a_3 \succ a_2 & a_1 &: p_2 \succ p_1 \succ p_3 \\ p_2 &: a_2 \succ a_1 \succ a_3 & a_2 &: p_3 \succ p_2 \succ p_1 \\ p_3 &: a_1 \succ a_3 \succ a_2 & a_3 &: p_3 \succ p_1 \succ p_2. \end{aligned}$$

The unique stable matching in this case is $(p_1, a_1), (p_2, a_2), (p_3, a_3)$.

Proof. Let $D^{(3)}(t) \overset{i.i.d.}{\sim} Ber(\lambda)$ for any t. The set of actions that player p_3 can play, for $t = 1, \cdots, n$, to get negative stable regret is as follows:

$$A^{(3)}(t) = \begin{cases} a_2 & \text{if } t = 3m-2 \\ a_3 & \text{if } t = 3m-1 \text{ and } D^{(3)}(t) = 0 \\ a_2 & \text{if } t = 3m-1 \text{ and } D^{(3)}(t) = 1 \\ a_1 & \text{if } t = 3m \end{cases}, \text{ for } m \in \mathbb{N}. \tag{5.35}$$

By the definition of p_3's regret, and using the fact that $\Delta_{\emptyset}^{(k)} > \max\{\Delta_1^{(3)}, \Delta_2^{(3)}\} > 0$, we have:

$$R_3(n) \leq \Delta_1^{(3)} \cdot \mathcal{E}[T_{3,1}(n)] + \Delta_{\emptyset}^{(3)} \cdot (T - \mathcal{E}[T_{3,1}(n)]). \tag{5.36}$$

Thus it suffices to lower bound the expected number of times that p_3 successfully attempts a_1.

Define the following events:

$$\begin{aligned} \Omega_t^1 &:= \{F_{3,1}^{(2)}(t) \cap A^{(2)}(t) = a_3\}^c, \\ \Omega_t^2 &:= \{F_{1,2}^{(2)}(t) \cap A^{(2)}(t) = a_1\}^c. \end{aligned}$$

We first show the following inclusion, for any $m \in \mathbb{N}$:

$$\Omega_{3m-1}^1 \cap \Omega_{3m}^2 \cap \{D^{(2)}(3m-1) = D^{(3)}(3m-1) = D^{(2)}(3m) = D^{(1)}(3m) = 0\} \subseteq \{\bar{A}^{(3)}(3m) = a_1\}. \tag{5.37}$$

We can simply check that this holds:

- At time $3m-2$, p_3 attempts and successfully pulls a_2.
- At time $3m-1$, p_2 pulls a_1, since a_2 is not in its plausible set, $D^{(2)}(3m-1) = 0$ and the event Ω_{3m-1}^1 holds. p_3 pulls a_3, since $D^{(3)}(3m-1) = 0$.
- At time $3m$, p_1 does not pull a_1, since a_1 is not in its plausible set and $D^{(1)}(3m) = 0$. p_2 does not pull a_1, since a_2 is in its plausible set, $D^{(2)}(3m) = 0$ and the event Ω_{3m}^2 holds. Thus p_3 successfully pulls a_1.

Taking expectation of (5.37) and rearranging gives

$$\begin{aligned}
&\mathbb{P}\left(\bar{A}^{(3)}(3m) = a_1\right) \\
&\geq 1 - \mathbb{P}\left(\left(\Omega^1_{3m-1} \cap \Omega^2_{3m} \cap \{D^{(2)}(3m-1) = D^{(3)}(3m-1) = D^{(2)}(3m) = D^{(1)}(3m) = 0\}\right)^c\right) \\
&\geq 1 - \left(\mathbb{P}((\Omega^1_{3m-1})^c) + \mathbb{P}((\Omega^2_{3m})^c) + 4\lambda\right), \qquad (5.38)
\end{aligned}$$

where the last inequality follows from a union bound.

It is useful to upper bound the following:

$$\begin{aligned}
&\sum_{m=1}^{\lfloor n/3 \rfloor} \mathbb{P}((\Omega^1_{3m-1})^c) + \mathbb{P}((\Omega^2_{3m})^c) \\
&= \sum_{m=1}^{\lfloor n/3 \rfloor} \mathbb{P}\left(F^{(2)}_{3,1}(3m-1) \cap A^{(2)}(3m-1) = a_3\right) + \mathbb{P}\left(F^{(2)}_{1,2}(3m) \cap A^{(2)}(3m) = a_1\right) \\
&\leq \frac{1}{\lambda(1-\lambda)} \sum_{m=1}^{\lfloor n/3 \rfloor} \mathbb{P}\left(F^{(2)}_{3,1}(3m-1) \cap \bar{A}^{(2)}(3m-1) = a_3\right) + \mathbb{P}\left(F^{(2)}_{1,2}(3m) \cap \bar{A}^{(2)}(3m) = a_1\right) \\
&\leq \frac{1}{\lambda(1-\lambda)} \sum_{t=1}^{n} \mathbb{P}\left(F^{(2)}_{3,1}(t) \cap \bar{A}^{(2)}(t) = a_3\right) + \mathbb{P}\left(F^{(2)}_{1,2}(t) \cap \bar{A}^{(2)}(t) = a_1\right) \\
&\leq \frac{1}{\lambda(1-\lambda)} \cdot 12 \cdot \left(\frac{1}{\Delta^2}\log(T) + 1\right), \qquad (5.39)
\end{aligned}$$

where the last inequality follows from Lemma 5.4.4.

Now, we sum (5.38) over $m = 1, \cdots, \lfloor n/3 \rfloor$ to get:

$$\begin{aligned}
\mathcal{E}[T_{3,1}(n)] &\geq \sum_{m=1}^{\lfloor n/3 \rfloor} \mathbb{P}\left(\bar{A}^{(3)}(3m) = a_1\right) \\
&\geq (1-4\lambda) \cdot \lfloor n/3 \rfloor - \sum_{m=1}^{\lfloor n/3 \rfloor} \mathbb{P}((\Omega^1_{3m-1})^c) + \mathbb{P}((\Omega^2_{3m})^c) \\
&\geq (1-4\lambda) \cdot \lfloor n/3 \rfloor - \frac{1}{\lambda(1-\lambda)} \cdot 12 \cdot \left(\frac{1}{\Delta^2}\log(T) + 1\right),
\end{aligned}$$

where the last two inequalities follow from Equation (5.38) and Equation (5.39).

Thus we have

$$R_3(n) \leq \Delta^{(3)}_1 \cdot \left((1-4\lambda) \cdot \lfloor n/3 \rfloor - \frac{1}{\lambda(1-\lambda)} \cdot 12 \cdot \left(\frac{1}{\Delta^2}\log(n) + 1\right)\right) + \Delta^{(3)}_\emptyset \cdot \left(\frac{2}{3}n\right). \quad (5.40)$$

Note that $\Delta^{(3)}_1 < 0$, and $1 - 4\lambda > 0$ by assumption. Upon rearranging terms, we get the desired result.

□

Bibliography

A. Abdulkadiroglu and T. Snmez. School choice: A mechanism design approach. *American economic review*, 93(3):729–747, 2003.

A. Abdulkadiroglu and T. Sonmez. House allocation with existing tenants. *Journal of Economic Theory*, 88(2):233–260, 1999.

A. Abdulkadiroglu, P. Pathak, A. E. Roth, and T. Sonmez. Changing the boston school choice mechanism. Technical report, National Bureau of Economic Research, 2006.

R. Abebe, S. Barocas, J. Kleinberg, K. Levy, M. Raghavan, and D. G. Robinson. Roles for computing in social change. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, page 252260, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450369367.

H. Abeledo and U. G. Rothblum. Paths to marriage stability. *Discrete Applied Mathematics*, 63:1–12, 10 1995.

H. Ackermann, P. W. Goldberg, V. S. Mirrokni, H. Röglin, and B. Vöcking. Uncoordinated two-sided matching markets. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, pages 256–263, 2008.

M. Akbarpour, S. Li, and S. O. Gharan. Thickness and information in dynamic matching markets. *Journal of Political Economy*, 128(3):783–815, 2020.

A. Anandkumar, N. Michael, A. K. Tang, and A. Swami. Distributed algorithms for learning and cognitive medium access with logarithmic regret. *IEEE Journal on Selected Areas in Communications*, 29(4):731–745, 2011.

A. P. Aneja and C. F. Avenancio-Leon. No Credit For Time Served? Incarceration and Credit-Driven Crime Cycles, 2019.

G. Aridor, K. Liu, A. Slivkins, and Z. S. Wu. Competing bandits: The perils of exploration under competition. *The 20th ACM Conference on Economics and Computation*, 2019.

N. Arnosti, R. Johari, and Y. Kanoria. Managing congestion in decentralized matching markets. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pages 451–451, 2014.

K. J. Arrow. The Theory of Discrimination. In *Discrimination in Labor Markets*, pages 3–33. Princeton University Press, 1973.

K. J. Arrow. What Has Economics to Say about Racial Discrimination? *Journal of Economic Perspectives*, 12(2):91–100, Spring 1998.

I. Ashlagi, M. Braverman, Y. Kanoria, and P. Shi. Communication requirements and informative signaling in matching markets. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, EC '17, pages 263–263, 2017a.

I. Ashlagi, Y. Kanoria, and J. D. Leshno. Unbalanced random matching markets: The stark effect of competition. *Journal of Political Economy*, 125(1):69–98, 2017b.

I. Ashlagi, A. K. Krishnaswamy, R. M. Makhijani, D. Sabán, and K. Shiragur. Assortment planning for two-sided sequential matching markets. *CoRR*, abs/1907.04485, 2019.

O. Avner and S. Mannor. Concurrent bandits and cognitive radio networks. In T. Calders, F. Esposito, E. Hüllermeier, and R. Meo, editors, *Machine Learning and Knowledge Discovery in Databases*, pages 66–81, 2014.

O. Avner and S. Mannor. Multi-user lax communications: A multi-armed bandit approach. In *The 35th Annual IEEE International Conference on Computer Communications*, pages 1–9, 2016.

S. Barocas and A. D. Selbst. Big data's disparate impact. *California Law Review*, 104(671), 2016.

S. Barocas, M. Hardt, and A. Narayanan. *Fairness and Machine Learning*. fairmlbook.org, 2018. `http://www.fairmlbook.org`.

R. Benjamin. Race after technology: Abolitionist tools for the new jim code. *Social forces*, 2019.

I. Bistritz and A. Leshem. Distributed multi-player bandits—A game of thrones approach. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 7222–7232, 2018.

I. Bistritz, T. Z. Baharav, A. Leshem, and N. Bambos. My fair bandit: Distributed learning of max-min fairness with multi-player bandits. In *Proceedings of The 37th International Conference on Machine Learning*, 2020.

E. Boursier and V. Perchet. Selfish robustness and equilibria in multi-player bandits. In J. Abernethy and S. Agarwal, editors, *Proceedings of the 33rd Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 530–581, 2020.

S. Bubeck and N. Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.

S. Bubeck, T. Budzinski, and M. Sellke. Cooperative and stochastic multi-player multi-armed bandit: Optimal regret with neither communication nor collisions. *arXiv preprint arXiv:2011.03896*, 2020a.

S. Bubeck, Y. Li, Y. Peres, and M. Sellke. Non-stochastic multi-player multi-armed bandits: Optimal rate with collision information, sublinear without. In *Proceedings of the 33rd Conference on Learning Theory*, pages 961–987, 2020b.

T. Calders, F. Kamiran, and M. Pechenizkiy. Building classifiers with independency constraints. In *Proc. IEEE ICDMW*, ICDMW '09, pages 13–18, 2009.

D. Card and J. Rothstein. Racial segregation and the black-white test score gap. *Journal of Public Economics*, 91(11):2158 – 2184, 2007. ISSN 0047-2727.

S. H. Cen and D. Shah. Regret, stability, and fairness in matching markets with bandit learners. *arXiv preprint arXiv:2102.06246*, 2021.

N. Cesa-Bianchi, C. Gentile, Y. Mansour, and A. Minora. Delay and cooperation in nonstochastic bandits. In V. Feldman, A. Rakhlin, and O. Shamir, editors, *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 605–622, 23–26 Jun 2016.

A. J. B. Chaney, B. M. Stewart, and B. E. Engelhardt. How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. In *Proceedings of the 12th ACM Conference on Recommender Systems*, RecSys '18, pages 224–232, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-5901-6. doi: 10.1145/3240323.3240370.

R. Chetty, N. Hendren, and L. F. Katz. The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, 106(4):855–902, 2016.

A. Chouldechova. Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. *Big Data*, 5:153–163, 2017.

S. Coate and G. C. Loury. Will affirmative-action policies eliminate negative stereotypes? *The American Economic Review*, 83(5):1220–1240, 1993. ISSN 00028282.

M. Conover, J. Ratkiewicz, M. R. Francisco, B. Goncalves, F. Menczer, and A. Flammini. Political polarization on twitter. *ICWSM*, 133:8996, 2011.

S. Corbett-Davies and S. Goel. The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning. *CoRR*, abs/1808.00023, 2018.

K. Crawford. The trouble with bias. NeurIPS Keynote, 2017.

X. Dai and M. I. Jordan. Learning strategies in decentralized matching markets under uncertain preferences. *arXiv preprint arXiv:2011.00159*, 2020.

A. D'Amour, H. Srinivasan, J. Atwood, P. Baljekar, D. Sculley, and Y. Halpern. Fairness is not static: Deeper understanding of long term fairness via simulation studies. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, page 525534, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450369367.

S. J. Darak and M. K. Hanawal. Multi-player multi-armed bandits for stable allocation in heterogeneous ad-hoc networks. *IEEE Journal on Selected Areas in Communications*, 37 (10):2350–2363, 2019.

S. Das and E. Kamenica. Two-sided bandits and the dating market. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, pages 947–952, 2005.

J. Dastin. Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*, 10 2019.

M. De-Arteaga, A. Dubrawski, and A. Chouldechova. Learning under selective labels in the presence of expert consistency. *Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT/ML)*, 2018.

M. De-Arteaga, A. Romanov, H. Wallach, J. Chayes, C. Borgs, A. Chouldechova, S. Geyik, K. Kenthapadi, and A. T. Kalai. Bias in bios: A case study of semantic representation bias in a high-stakes setting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, pages 120–128, New York, NY, USA, 2019. ACM. ISBN 978-1-4503-6125-5. doi: 10.1145/3287560.3287572.

L. E. Dubins and D. A. Freedman. Machiavelli and the Gale-Shapley Algorithm. *The American Mathematical Monthly*, 88(7):485–494, 1981.

C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*, pages 214–226, 2012.

C. Dwork, N. Immorlica, A. T. Kalai, and M. Leiserson. Decoupled classifiers for group-fair and efficient machine learning. In S. A. Friedler and C. Wilson, editors, *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, volume 81 of *Proceedings of Machine Learning Research*, pages 119–133, New York, NY, USA, 23–24 Feb 2018. PMLR.

F. Echenique and L. Yariv. An experimental study of decentralized matching. 2012.

D. Ensign, S. A. Friedler, S. Neville, C. Scheidegger, and S. Venkatasubramanian. Runaway feedback loops in predictive policing. In *Conference on Fairness, Accountability and Transparency, FAT 2018, 23-24 February 2018, New York, NY, USA*, pages 160–171, 2018.

Executive Office of the President. Big data: A report on algorithmic systems, opportunity, and civil rights. Technical report, White House, May 2016.

D. P. Foster and R. V. Vohra. An economic argument for affirmative action. *Rationality and Society*, 4(2):176–188, 1992.

Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

D. Fudenberg and D. K. Levine. *The theory of learning in games*, volume 2. MIT press, 1998.

A. Fuster, P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther. Predictably unequal? the effects of machine learning on credit markets. *The Journal of Finance*, 77(1):5–47, 2022.

D. Gale and L. S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

D. Gusfield and R. W. Irving. *The Stable Marriage Problem: Structure and Algorithms*. MIT Press, Cambridge, MA, USA, 1989.

M. Hardt, N. Megiddo, C. Papadimitriou, and M. Wootters. Strategic classification. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, ITCS '16, pages 111–122, New York, NY, USA, 2016a. ACM. ISBN 978-1-4503-4057-1.

M. Hardt, E. Price, and N. Srebo. Equality of opportunity in supervised learning. In *Advances in Neural Information Processing Systems*, pages 3315–3323, 2016b.

M. Hardt, E. Price, and N. Srebro. Equality of opportunity in supervised learning. In *Proc. 30th NIPS*, 2016c.

T. Hashimoto, M. Srivastava, H. Namkoong, and P. Liang. Fairness without demographics in repeated loss minimization. In J. Dy and A. Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1929–1938, Stockholm, Sweden, 10–15 Jul 2018. PMLR.

W. D. Heaven. Bias isn't the only problem with credit scoresand no, ai can't help. In *Ethics of Data and Analytics*, pages 300–302. Auerbach Publications, 2022.

M. Hoffman, D. Moeller, and R. Paturi. Jealousy graphs: Structure and complexity of decentralized stable matching. In *Web and Internet Economics*, pages 263–276, 2013.

C. Hoxby and C. Avery. The missing "one-off": The hidden supply of high-achieving, low-income students. *Brookings Papers on Economic Activity*, 1:1–65, 2013.

J. Hu, M. P. Wellman, et al. Multiagent reinforcement learning: theoretical framework and an algorithm. In *ICML*, volume 98, pages 242–250. Citeseer, 1998.

L. Hu and Y. Chen. A short-term intervention for long-term fairness in the labor market. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, pages 1389–1398, Republic and Canton of Geneva, Switzerland, 2018a. International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-5639-8.

L. Hu and Y. Chen. A short-term intervention for long-term fairness in the labor market. In *Proc.* 27*th WWW*, 2018b.

L. Hu, N. Immorlica, and J. W. Vaughan. The disparate effects of strategic manipulation. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, pages 259–268, New York, NY, USA, 2019. ACM. ISBN 978-1-4503-6125-5. doi: 10.1145/3287560.3287597.

N. Immorlica, J. Leshno, I. Lo, and B. Lucier. Information acquisition in matching markets: The role of price discovery. *Available at SSRN*, 2020.

R. Johari, V. Kamble, and Y. Kanoria. Matching while learning. In *ACM Conference on Economics and Computation*, pages 119–119, 2017.

M. Joseph, M. Kearns, J. H. Morgenstern, and A. Roth. Fairness in learning: Classic and contextual bandits. In *Proc.* 30*th NIPS*, pages 325–333, 2016.

E. Kalai. Large robust games. *Econometrica*, 72(6):1631–1665, 2004.

D. Kalathil, N. Nayyar, and R. Jain. Decentralized learning for multiplayer multiarmed bandits. *IEEE Transactions on Information Theory*, 60(4):2331–2345, 2014.

A. Kalev, F. Dobbin, and E. Kelly. Best Practices or Best Guesses? Assessing the Efficacy of Corporate Affirmative Action and Diversity Policies. *American Sociological Review*, 71 (4):589–617, 2006.

N. Kallus and A. Zhou. Residual unfairness in fair machine learning from prejudiced data. In J. Dy and A. Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 2439–2448, Stockholmsmssan, Stockholm Sweden, 10–15 Jul 2018. PMLR.

S. N. Keith, R. M. Bell, A. G. Swanson, and A. P. Williams. Effects of affirmative action in medical schools. *New England Journal of Medicine*, 313(24):1519–1525, 1985.

M. Khajehnejad, B. Tabibian, B. Schölkopf, A. Singla, and M. Gomez-Rodriguez. Optimal decision making under strategic behavior. *CoRR*, abs/1905.09239, 2019.

N. Kilbertus, M. Rojas-Carulla, G. Parascandolo, M. Hardt, D. Janzing, and B. Schölkopf. Avoiding discrimination through causal reasoning. In *In Proc.* 30*th NIPS*, pages 656–666, 2017.

N. Kilbertus, M. Gomez-Rodriguez, B. Schölkopf, K. Muandet, and I. Valera. Improving consequential decision making under imperfect predictions. *CoRR*, abs/1902.02979, 2019.

J. Kleinberg and M. Raghavan. How Do Classifiers Induce Agents to Invest Effort Strategically? In *Proceedings of the 2019 ACM Conference on Economics and Computation*, EC '19, pages 825–844, New York, NY, USA, 2019. ACM. ISBN 978-1-4503-6792-9.

J. M. Kleinberg, S. Mullainathan, and M. Raghavan. Inherent trade-offs in the fair determination of risk scores. *Proceedings of the 8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*.

D. E. Knuth. *Stable Marriage and its Relation to Other Combinatorial Problems*. American Mathematical Society, 1997.

M. J. Kusner, J. R. Loftus, C. Russell, and R. Silva. Counterfactual fairness. In *In Proc.* 30*th NIPS*, pages 4069–4079, 2017.

T. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Adv. Appl. Math.*, 6(1):4–22, Mar. 1985a.

T. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4 – 22, 1985b.

T. Lattimore and C. Szepesvari. *Bandit Algorithms*. Cambridge University Press (To Appear), 2019.

R. S. Lee and M. Schwarz. Interviewing in two-sided matching markets. Technical report, National Bureau of Economic Research, 2009.

M. L. Littman. Markov games as a framework for multi-agent reinforcement learning. In *Machine learning proceedings 1994*, pages 157–163. Elsevier, 1994.

K. Liu and Q. Zhao. Distributed learning in multi-armed bandit with multiple players. *IEEE Transactions on Signal Processing*, 58(11):5667–5681, 2010.

L. T. Liu, S. Dean, E. Rolf, M. Simchowitz, and M. Hardt. Delayed impact of fair machine learning. In J. Dy and A. Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 3150–3158, Stockholm, Sweden, 10–15 Jul 2018. PMLR.

L. T. Liu, M. Simchowitz, and M. Hardt. The implicit fairness criterion of unconstrained learning. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 4051–4060, Long Beach, California, USA, 2019. PMLR.

L. T. Liu, H. Mania, and M. Jordan. Competing bandits in matching markets. In *International Conference on Artificial Intelligence and Statistics*, volume 108, pages 1618–1628, 26–28 Aug 2020a.

L. T. Liu, A. Wilson, N. Haghtalab, A. T. Kalai, C. Borgs, and J. Chayes. The disparate equilibria of algorithmic decision making when individuals invest rationally. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 381–391, 2020b.

L. T. Liu, F. Ruan, H. Mania, and M. I. Jordan. Bandit learning in decentralized matching markets. *Journal of Machine Learning Research*, 22(211):1–34, 2021a. URL `http://jmlr.org/papers/v22/20-1429.html`.

L. T. Liu, S. Wang, R. Abebe, and T. Britton. Lost in translation: Reimagining the machine learning life cycle in education. *Manuscript in preparation. A preliminary version appeared in The 49th Annual Research Conference on Communications, Information, and Internet Policy (TPRC)*, 2021b.

S. Lowry and G. Macpherson. A blot on the profession. *British Medical Journal*, 296(6623): 657–658, 1988.

G. Lugosi and A. Mehrabian. Multiplayer bandits without observing collision information. *arXiv preprint arXiv:1808.08416*, 2018.

M. Madaio, S. L. Blodgett, E. Mayfield, and E. Dixon-Román. Confronting structural inequities in AI for education. *CoRR*, abs/2105.08847, 2021. URL `https://arxiv.org/abs/2105.08847`.

B. M. Maggs and R. K. Sitaraman. Algorithmic nuggets in content delivery. *ACM SIGCOMM Computer Communication Review*, 45(3):52–66, 2015.

Y. Mansour, A. Slivkins, and Z. S. Wu. Competing bandits: Learning under competition. In *9th Innovations in Theoretical Computer Science Conference, ITCS 2018, January 11-14, 2018, Cambridge, MA, USA*, pages 48:1–48:27, 2018.

S. Milli, J. Miller, A. D. Dragan, and M. Hardt. The social cost of strategic classification. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, pages 230–239, New York, NY, USA, 2019. ACM.

H. Mouzannar, M. I. Ohannessian, and N. Srebro. From fair decision making to social equality. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, pages 359–368, New York, NY, USA, 2019. ACM. ISBN 978-1-4503-6125-5.

R. Nabi and I. Shpitser. Fair inference on outcomes. *arXiv:1705.10378v1*, 2017.

M. Niederle and L. Yariv. Decentralized matching with aligned preferences. Technical report, National Bureau of Economic Research, 2009.

J. Pais, A. Pintér, and R. F. Veszteg. Decentralized matching markets: a laboratory experiment. 2012.

E. Pariser. *The Filter bubble: What the Internet is hiding from you.* Penguin, UK, 2011.

J. Pearl. *Causality: Models, Reasoning and Inference.* Cambridge University Press, New York, NY, USA, 2nd edition, 2009. ISBN 052189560X, 9780521895606.

E. Phelps. The statistical theory of racism and sexism. *American Economic Review*, 62: 659–61, 02 1972.

G. Pleiss, M. Raghavan, F. Wu, J. Kleinberg, and K. Q. Weinberger. On fairness and calibration. In *Advances in Neural Information Processing Systems 30*, pages 5684–5693, 2017.

J. Rosenski, O. Shamir, and L. Szlak. Multi-player bandits—A musical chairs approach. In M. F. Balcan and K. Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 155–163, New York, New York, USA, 20–22 Jun 2016. PMLR.

S. Ross and J. Yinger. *The Color of Credit: Mortgage Discrimination, Research Methodology, and Fair-Lending Enforcement.* MIT Press, Cambridge, 2006.

A. E. Roth. The economics of matching: Stability and incentives. *Mathematics of Operations Research*, 7(4):617–628, 1982.

A. E. Roth. The evolution of the labor market for medical interns and residents: A case study in game theory. *Journal of political Economy*, 92(6):991–1016, 1984.

A. E. Roth. Deferred acceptance algorithms: History, theory, practice, and open questions. Working Paper 13225, National Bureau of Economic Research, July 2007.

A. E. Roth. Deferred acceptance algorithms: History, theory, practice, and open questions. *International Journal of Game Theory*, 36(3):537–569, Mar 2008.

A. E. Roth and M. A. O. Sotomayor. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis.* Econometric Society Monographs. Cambridge University Press, 1990.

A. E. Roth and J. H. Vande Vate. Random paths to stability in two-sided matching. *Econometrica*, 58(6):1475–1480, 1990.

A. E. Roth, T. Sönmez, and M. U. Ünver. Pairwise kidney exchange. *Journal of Economic theory*, 125(2):151–188, 2005.

A. Sankararaman, S. Basu, and K. Abinav Sankararaman. Dominate or delete: Decentralized competing bandits with uniform valuation. *arXiv preprint arXiv:2006.15166*, 2020.

D. Schmeidler. Equilibrium points of nonatomic games. *Journal of Statistical Physics*, 7(4): 295–300, Apr 1973.

S. Shahrampour, A. Rakhlin, and A. Jadbabaie. Multi-armed bandits in multi-agent networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2786–2790, 2017.

W. R. Thompson. On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples. *Biometrika*, 25(3-4):285–294, 12 1933.

US Federal Reserve. Report to the congress on credit scoring and its effects on the availability and affordability of credit, 2007.

B. Ustun, Y. Liu, and D. Parkes. Fairness without harm: Decoupled classifiers with preference guarantees. In K. Chaudhuri and R. Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6373–6382, Long Beach, California, USA, 09–15 Jun 2019. PMLR.

M. Whittaker, K. Crawford, G. F. Roel Dobbe, E. Kaziunas, V. Mathur, S. M. West, R. Richardson, J. Schultz, and O. Schwartz. AI Now Report 2018, 2018.

J. Williams and J. Z. Kolter. Dynamic modeling and equilibria in fair decision making. *CoRR*, abs/1911.06837, 2019. URL `http://arxiv.org/abs/1911.06837`.

M. B. Zafar, I. Valera, M. G. Rogriguez, and K. P. Gummadi. Fairness Constraints: Mechanisms for Fair Classification. In *Proc. 20th AISTATS*, pages 962–970. PMLR, 2017.

X. Zhang, M. M. Khalili, C. Tekin, and M. Liu. Long term impact of fair machine learning in sequential decision making: representation disparity and group retention. *CoRR*, abs/1905.00569, 2019.