Tradeoffs between Information, Tractability, and Fairness in Large Matching Markets

Aapeli Vuorinen

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2025

Abstract

Tradeoffs between Information, Tractability, and Fairness in Large Matching Markets

Aapeli Vuorinen

Matching theory is one of the cornerstones of modern algorithmic market design, and is therefore heavily studied by communities in operations research, economics, and theoretical computer science. Such models arise whenever a central planner is tasked with pairing together agents of two distinct types but cannot use money to clear the market; and where each agent has idiosyncratic preferences over the agents of the other type. Natural applications of matching markets are found in the allocation of indivisible goods such as school assignment, medical residency matching, and the allocation of government subsidized housing.

Matching theory and in particular stable matching has seen significant research effort since seminal work by Gale and Shapley algorithmically established the existence of a condition called *stability*, which guarantees that a matching can be found with the property that no participant is incentivized to deviate from it. Many centralized mechanisms for two-sided matching markets have since been shown to enjoy strong theoretical properties, which *ipso facto*, justifies their use in the real world.

However, due to practical constraints—commonly due to the size and complexity of the market at hand—real world applications often resort to simplified models that do not faithfully capture reality. These simplifications lead to the central planner operating without perfect information about the participants or their preferences.

In this thesis, we present three interconnected branches of research exploring tradeoffs between information, tractability, and fairness in large matching markets. We investigate how various limitations on information acquisition and exchange affect the mechanisms, outcomes, and fairness of matching markets.

We are motivated by the matching mechanism that assigns students to public schools in New York City. The unified school district—which encompasses all five boroughs—has since 2003 employed the theory of stable matching to perform this assignment, with students as one side of the market and schools as the other. Consisting of over 1.1 million students, the size and diversity of the market presents a massively interesting real-world object of study.

In the first chapter, we investigate a model where students are restricted in the length of preference lists that they may submit to the market operator. In particular, we study such random instances of the Serial Dictatorship mechanism where students choose d schools uniformly at random from n schools as their preference list, and each school has exactly one seat. Our main result is that if the students primarily care about being matched to any school of their list (as opposed to ending up unmatched), then all students in position $i \leq n$ will prefer markets with longer lists when n is large enough, whereas students after some cutoff $c > n$ (that quickly approaches n as the list length grows) prefer markets with shorter lists. This suggests that markets that are well-approximated by our hypothesis and where the demand of schools does not exceed supply, should be designed with preference lists as long as reasonable.

In the second chapter, we study the impact of systemic bias in school matching by investigating the admissions process to the eight elite public schools (called the Specialized High Schools) in New York City. These schools admit students solely based on their score on a standardized test, but we observe a clear distributional shift in the test scores of disadvantaged students (as defined by the city). To study this shift, we present a stylized model where all students have a *true potential* (representing their innate ability)

sampled independently from the same distribution. While non-disadvantaged students always perform at their true potential, disadvantaged students appear at a *perceived potential* strictly below their innate ability due to some systemic bias. We investigate both theoretically and empirically the impact of such bias on the admissions process, then turn to studying interventions to counter it. These interventions are in the form of *vouchers* targeted at certain disadvantaged students, which we assume give that student the resources they need to perform at their true potential. We measure aggregate mistreatment under various metrics, first investigating optimal deterministic voucher distribution, and then turning to randomized voucher distribution. We additionally present extensive numerical experiments both on a real dataset from New York, as well as on simulated data. We then confirm that our results hold under various relaxations to our stylized model, including moderate levels of model misspecification. Our key takeaway is that resources should be targeted at slightly above average performers instead of the absolute top performers.

In the third chapter, we take the schools' side. Currently the city allows schools to only specify their preferences using a strict preference list over students. While this leads to a simple algorithm, stable matching models may be extended to allow schools to communicate much more rich preference via *choice functions*. We discuss the impact of using general choice functions in the *offline model* of stable matching, establishing that general path-independent and quota-filling choice functions are too large a class to be used in this setting. We propose the class of *Kuhn choice functions*, that arise as maximum-weight matchings in an auxiliary bipartite graph, as a tractable yet rich subclass. We show that such choice functions are amenable to use in the offline model and possess many desirable properties. We further discuss the hierarchy of choice and approximability of various classes of choice functions. Theoretical proofs are complemented by computational results and a discussion on various practical aspects of using choice functions in stable matching.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I want to thank Yuri for being a phenomenal advisor and mentor. There is universal agreement within the department that those that get to work with Yuri are very lucky for it. I don't take this for granted. When we started working together, I was on the brink of dropping out of the PhD. It didn't take long for Yuri's positive attitude and encouragement to bring me robustly back into the non-quitting territory. I have been very fortunate to learn an incredible amount from Yuri and I have had so much fun working together, both through our academic endeavours as well as through our numerous conversations about a wide gamut of different topics.

The IEOR Department at Columbia has been a great place to grow intellectually over the last five years. The department embodies a rare property for academia that I call *stability*: there appears to be no pair of individuals that dislike each other and would not like to see each other succeed. In particular I'd like to thank my fellow PhD students, and the rest of the faculty & staff that make it one of the best academic environments I have had the pleasure of experiencing. Thanks to all the masters students for indirectly funding a good chunk of my PhD.

I want to thank Yoni for an unwavering friendship through the many ups and downs of academic life, for always making time for a call across timezones, and for teaching me so much about empathy and how to always see the best in others and in myself.

I want to thank my family for being there (in the form of a literal 24/7 oncall rotation on our group chat), even when I'm abysmal at staying in touch. I try to make up for it by coming to visit for long stretches at a time, so thanks for bearing with that too. Special shout out to those family members hard at work making more family. I didn't know how much fun the new family would be. Keep it up and keep making more of 'em.

Thanks to my many friends outside Columbia, for keeping me relatively sane.

Finally, thanks to the many awesome folks at Couchers.org who kept me very busy when I allegedly didn't have PhD stuff to do? It's been so much fun.

Dedication

To mum and dad: am I finally a grown-up now?

Introduction

Markets arise whenever supply is aggregated together with demand. They come in many flavors: on couch surfing platforms hosts put up postings expressing interest in welcoming surfers from around the world, in school choice mechanisms cities make seats in public schools available to students (and their parents) to apply to through unified application processes, and in financial markets participants form exchanges to trade standardized securities among themselves.

This thesis studies *two-sided matching markets*. A market is *two-sided* when supply and demand are of two differentiated types: in the couch surfing example the two sides are hosts (who have time to host and a room or couch to spare) and surfers (who are visiting a city and looking to meet and stay with a local). Similarly in the case of school choice markets, the students are fundamentally different from the schools that they are matched to. This rules out markets such as the market of roommates where there is only one type of participant simultaneously representing both supply and demand. *Matching markets* are those markets where we look for a matching—or a pairing—between agents of the two sides, and the outcome is decided by the heterogeneous preferences of agents rather than by prices; such as pairing up hosts with surfers (where compatibility is key), or assigning students to schools (where students have preferences and schools are selective). This rules out markets such as those found in the financial world where buyers and sellers have no preference over whom they wish to transact with.

Agents from both sides participate in such markets in exchange for a guarantee that

the resultant matching will respect their preferences by not matching them to an undesirable partner. This gives rise to the concept of *stability*. In the case where each agent can be matched to at most one other partner, stability states that there be no pair of agents of opposing types such that they both prefer each other to whomever they were matched with. Such a pair, if it exists, is called a *blocking pair*. Agents forming a blocking pair are incentivized to deviate from the proposed solution by forming their own exogenous pairing together, thereby discarding their assigned partner. A market that contains blocking pairs would therefore be inherently unstable as agents could not trust that their assigned partner would honor the matching, which could ultimately lead to participants abandoning the market entirely. Stability—the absence of such blocking pairs—is therefore the fundamental equilibrium concept used within the field.

Matching theory and in particular stable matching has seen significant research effort since the seminal work by Gale and Shapley [1] which introduced this concept of stability in the *marriage model* and algorithmically established the existence of stable solutions via their Deferred Acceptance procedure. Many centralized mechanisms for two-sided matching markets have since been shown to enjoy strong theoretical properties, which *ipso facto* justifies their use in the real world.

However, due to practical constraints—commonly due to the size and complexity of the market at hand—real world applications often resort to simplified models. These simplifications lead to the central planner operating without perfect information about the participants or their preferences.

In this thesis, we present three interconnected branches of research exploring tradeoffs between information, tractability, and fairness within large matching markets. We investigate how various limitations on information acquisition and exchange affect the mechanisms, outcomes, and fairness of matching markets.

Our models are by and large motivated by the New York City public school choice program. The unified school district, which encompasses all five boroughs, has since

2003 employed the theory of stable matching to perform this assignment [2, 3]. Consisting of over 1.1 million students, the size and diversity of the market presents a massively interesting real-world object of study. Students form one side of the market, each compiling and submitting a ranked list of schools they wish to attend. Schools, each with some number of seats, form the other side of the market, ranking the students in the order of whom they wish to admit.

In Chapter 1, we investigate the very common constraint where the length of preference lists is limited. Such constraints arise in many large real-world markets, where agents cannot be expected to produce a ranking of all options on the other side. In particular, we study such random instances of the Serial Dictatorship mechanism where students choose d schools uniformly at random from n schools as their preference list, and each school has exactly one seat.

Our main results shows that if the students primarily care about being matched to any school of their list (as opposed to ending up unmatched), then all students in position $i \leq n$ will prefer markets with longer lists when n is large enough, whereas students after some cutoff $c > n$ (that quickly approaches n as the list length grows) prefer markets with shorter lists. Schools on the other hand will always prefer longer lists in our model. We further investigate the impact of d on the rank of the school that a student gets matched to. We conjecture that our result holds when each school has $q > 1$ seats, and perform numerical experiments to test this conjecture. We also show empirically that our results continue to hold in many cases where students sample schools independently at random but not necessarily uniformly.

Our work suggests that markets that are well-approximated by our hypotheses and where the demand of schools does not exceed supply should be designed with preference lists as long as reasonable, since longer lists would favor all agents.

In Chapter 2, we study the impact of systemic bias in school matching. Our model is motivated by the admissions process to the eight elite public schools in New York City

called the Specialized High Schools, which admit students solely based on their score on the Specialized High School Admissions Test (SHSAT). The Department of Education (DOE) deems some students *disadvantaged* based largely on socio-economic factors. We observe a significant discrepancy between these two groups, in the form of a distributional shift in student test scores. In particular, we find that the distribution of performance between non-disadvantaged and disadvantaged students match closely under a constant multiplicative or additive shift in scores.

This motivates our stylized model of bias, where we posit that disadvantaged students and non-disadvantaged students alike have some *true potential* (representing their innate ability) sampled independently from the same distribution. However, due to headwinds in their educational experience, those students that are disadvantaged perform at a *perceived potential* that is lower than their true potential; whereas the non-disadvantaged students are always perceived at their true potential. We assume that students have a shared preference over the top schools, and based on the DOE data, we study in detail the case where potentials follow the Pareto law.

We investigate both theoretically and empirically the impact of such differences in perceived potential, then turn to interventions to counter it. These interventions are in the form of *vouchers* targeted at certain disadvantaged students. We assume that such vouchers (in the form of supplemental instruction, additional test prep, or even scholarships) give a disadvantaged student the support and resources they need to perform at their true potential. We define the mistreatment of a student as the displacement they experience due to the presence of bias, then define two opposing metrics of aggregate welfare—the maximum mistreatment metric and the positive area under the mistreatment curve (PAUC)—corresponding to the L^∞ and L^1 norms of mistreatment, respectively.

We analytically identify optimal deterministic debiasing sets under both metrics, but observe that these fail to be incentive compatible and individually fair. Motivated by this, we study randomized policies for voucher distribution, and produce such a policy

that is individually fair, incentive compatible and—by targeting average disadvantaged students—can improve aggregate welfare more than any deterministic policy.

We additionally present extensive numerical experiments both on a real dataset from New York for the 2016–17 academic year, as well as on simulated data, which confirm that our results hold even when some of our stylized assumptions do not hold, and under various levels of model misspecification.

The key takeaway from our work is that interventions should be targeted at slightly above average performers instead of the absolute top performers. This is in contrast to common scholarship distribution schemes that tend to award extra resources to only the very best students.

In Chapter 3, we take the schools' side, investigating the impact on matching when schools are not able to communicate their complete preferences to the central planner. Many school districts, such as the one in New York, only allow schools to express their preference over students via a ranked list. This is in contrast to the increasing desire of schools to assemble classes of balanced students from diverse backgrounds to facilitate learning; and a body of theoretical work that has moved to studying *choice functions*, objects that allow schools much greater ability in expressing their preferences.

We show that letting schools express their preference via arbitrary choice functions is practically untenable in the *offline model* adopted in most markets. This motivates our proposal for an elegant subclass called *Kuhn choice functions* that arise as maximum-weight matchings in an auxiliary bipartite graph. We show that such choice functions have many desirable properties, such as being representable in polynomial size, which makes them an appropriate choice for the offline model. We then turn to questions of modeling and recognition, discussing the gaps in what preference can be represented with various classes of choice functions, and discussing the problem of recognizing which class a particular choice function belongs in. We present results on approximation of simpler choice functions with more expressive ones, establishing a hierarchy of choice functions. Our

theoretical results are complemented by numerical computations counting both Kuhn and non-Kuhn choice functions over small cohorts.

We close the chapter by arguing that markets such as the New York City public school market should adopt the more expressive class of preference systems captured by Kuhn choice functions, as it can be easily adopted within the current matching mechanism while yielding significant gains in ability of schools to express rich preference.

Chapter 1: Longer Lists Yield Better Matchings

Joint work with Yuri Faenza.

1.1 Introduction

One of the central objectives of theoretical research in matching markets is to design mechanisms such that the mechanisms and their outcomes provably satisfy desirable properties. Depending on the specific market, the focus may be on concepts such as strategy-proofness, one- or two-sided optimality, or equilibria. These and other properties are often viewed as justifications for applying such mechanisms in practice, under the expectation that the valuable theoretical features will persist in the real world. In many applications, however, the assumptions that are necessary for such properties to hold may not be satisfied. For example, one can compute a stable assignment efficiently in a static two-sided market, but in the very common real-world application arising in public school allocation, a substantial number of agents often enter or leave the market after the first assignment has been decided (see for example [4]). The goal of the central planner therefore becomes less well-defined and, depending on the mathematical formalization, may lead to computationally hard problems (see for example [5, 6, 7, 8]).

In this chapter, we focus on one of the main restrictions encountered in real-world markets: imposing a limit on the maximum length of preference lists of agents. Such limits are common in school markets: while Gale and Shapley's Deferred Acceptance algorithm [1] assumes that students are allowed to list all schools they deem acceptable in their preference list, education departments traditionally impose a strict maximum limit on the number of schools an applicant may list[1]. This restriction is motivated by

[1]These include programs in Spain and Hungary [9], Australia [10], as well as some cities in the United

practical concerns such as the additional burden that acquiring more information poses to schools and students alike. However, the effect of restricting the length of preference lists is significant on the properties of the mechanism and of its outcome. For instance, while the original Deferred Acceptance mechanism is strategy-proof for the proposing side, its implementation with bounded-size lists leaves room for strategic behavior of students. This is not merely a theoretical concern, but has practical implications [9] which are well known to central planners. For instance, until the school year 2024/2025, the Department of Education of New York City limited[2] the number of schools in student preference lists to 12 and suggested that the applicants be strategic by reserving some of the slots for "safe schools" (schools that are less desirable or where the student has high priority, and as a result, a high probability of being accepted) [11].

When deciding the length of preference lists in a matching market, it is therefore essential to carefully balance practical concerns with the potential efficiency losses that can result from limiting information exchange. A correct estimation of the efficiency loss induced by short(er) lists can therefore guide a market designer in striking the right trade-off.

1.1.1 Our Contributions

In this chapter, we focus on the impact that the length of preference lists has on the quality of the output matching in the Serial Dictatorship mechanism for two-sided matching markets. In our theoretical model, we assume that the proposing side (students) have preference lists drawn uniformly at random and that the disposing side (schools) have only one seat (we later discuss how these assumptions can be relaxed in our computational experiments).

The concept of "quality of a matching" can be defined in multiple ways; we mostly

States.

[2]This rule was changed for the school year 2025/2026, and applicants can now list as many schools as they want.

focus on the probability that a student will be matched to any school of their preference list as opposed to remaining unmatched. This is justified by the fact that if a student cannot be matched to any school of their preference list, they completely lose control of the school they are matched to (in many markets, such students are assigned a remaining seat arbitrarily at the whim of the central planner). This seems by far the worse outcome for a student. We moreover study the probability that a student is matched to their top k choices, a popular measure of quality of a matching[3].

1. Comparative analysis of balanced markets. In our model, the probability p_i that a student i is matched to any school in their preference list is determined by two opposite effects. On one hand, longer preference lists mean that the student at hand has a larger list of acceptable schools, causing an higher p_i. On the other hand, the longer the preference lists, the more schools will have been matched to students with priority higher than i, decreasing p_i. Our main result is that under the Serial Dictatorship mechanism, when the number of schools n is large enough, the former effect dominates for all students in positions $i \leq n$: that is, p_i increases with the length of lists. This is our Theorem 1. In particular, if the demand of schools does not exceed supply, all students will prefer longer lists. It is not hard to see that in such a market, all schools are matched with higher probability when preference lists are longer (see Lemma 2). Therefore our results suggest that markets that are well-approximated by these hypotheses should be designed with preference lists as long as reasonable, since longer lists favor all participating agents.

2. Comparative analysis of general markets. Interestingly, it is not true in general that longer lists will result in a higher p_i for every student. Indeed, for n large enough, all students in position $i \geq \lceil 1.22n \rceil$ will be matched with a higher probability when the length of preference lists is 1 (students randomly choose one school to apply to), as opposed to when the length of preference lists is 2. This case is discussed in Section 1.2.3 and

[3]For instance, the National Resident Matching Program, while using the Deferred Acceptance algorithm, explicitly reports the number of applicants matched to their first choice [12], while the Boston Mechanism directly aims at maximizing the number of students matched to their first choice.

Figure 1.3. More generally, for all $d < \ell$ and n large enough, there exists a cutoff $c(d) > 1$, such that every student in position $i > n \cdot c(d)$ is matched with higher probability when preference lists are of length d, as opposed to when preference lists are of length ℓ; furthermore, as $d \to \infty$, $c(d) \to 1$. This is our Theorem 3.

3. Absolute bounds on the probability of being matched. While the previous results compare the probabilities of a given student getting matched to any school in their list between markets with different list length, they do not give us any information on the absolute probability. We show in Theorem 4 that as the length of preference lists increases, the probability that the student in position $i = n$ is matched (recall that n is the number of schools) quickly approaches $1/2$. This is therefore also a lower bound (resp., an upper bound) to the probability of a student $i \leq n$ (resp., $i \geq n$) being matched. In Lemma 5, we further prove bounds on the probability that a student in position $i \leq n$ gets matched to one of their top-k choices as a function of the length of the market.

4. Numerical results. In Section 1.4, we numerically study two extensions of the model. We first focus on the case when the preference lists are not sampled uniformly at random, but rather, schools are sampled i.i.d. from one of 5 distributions that place different weights on different schools. With the exception of a "degenerate" distribution, numerical experiments confirm our main result even when the uniformity assumption is relaxed: that every student in position $i \leq n$ will be matched with higher probability when lists are longer. We then focus on the case when schools have $q > 1$ available seats each, and confirm via simulations, that in this case too, students in balanced markets appear to always prefer longer lists.

1.1.2 Organization of the Chapter

We conclude this introductory section with further pointers to the literature. Section 1.2 is devoted to introducing the main models and ideas, and formally stating many results (including those discussed above) without proofs. In particular, in Section 1.2.1,

we introduce the main (discrete) model that we investigate in this chapter. To analyze its properties for n large enough, it will be useful to consider a continuous model that can be interpreted as a limit of the discrete model. This continuous model is introduced in Section 1.2.2. In Section 1.2.3 we state the connections between the two models, as well as some relevant properties of the continuous model. We moreover discuss some implications of our results for Random Serial Dictatorship and extend our models to the case of schools having multiple seats in Section 1.2.4. Last, we discuss the relevance of our model and our hypothesis in Section 1.2.5.

Proofs of results stated in Section 1.2 appear in Section 1.3: we start with a limit theorem rigorously establishing the connection between the continuous and the discrete models (Section 1.3.1), followed by proofs of properties of the continuous market (Section 1.3.2). Results from Section 1.3.1 and Section 1.3.2 will allow to deduce properties of the discrete market (Section 1.3.3 and Section 1.3.4).

Numerical experiments testing the validity of our theoretical results when some assumptions are relaxed appear in Section 1.4. We conclude in Section 1.5.

1.1.3 Related Literature

Random models for matching markets have a long and rich history. The model that is closest to ours is the one by [13], where the authors consider random preference lists and assume that one of the two sides has preference lists of bounded length, providing an investigation of the model for large markets. In particular, [13] show that with high probability and market size large enough, the core (i.e., the set of stable matchings) has small size. [14] study a similar model, investigating the change in the quality of the matchings in the core as a function of the preference list length. Similar questions have also been answered in the case when preference lists are complete and markets are balanced [15, 16] or unbalanced [17, 18].

The effect of short lists on the behavior of agents and on the quality of the output

matching has been the subject of experimental studies [10, 19, 9]. While specifics differ, they all share the common message that a loss of efficiency is experienced when agents are asked to report preference lists of bounded length, as opposed to preference lists of arbitrary length.

Many theoretical studies have been devoted to Serial Dictatorship and Random Serial Dictatorship, with goals different from ours—see, for instance, [20, 21, 22, 5, 23] and the references within the survey article [24].

To investigate our discrete model, we introduce a continuous model that can be thought of as a limit of the former. Continuous or semi-continuous models for two-sided markets have recently received quite some attention, since they often allow for tighter analysis, see, e.g., [25, 26, 27]. In particular, the work in [26] which builds on earlier work in [28] is closely aligned in its methods with our work. They similarly introduce a description of the limit of a discrete market via a solution to an initial value problem; then utilize a technical argument similar[4] to our Theorem 6 to show convergence of certain quantities relating to that market. Their differential equation is very general and allows for arbitrary distributions and varying lengths of preferences across students. In contrast, our main technical contribution is to carefully construct an initial value problem for our specific case that produces a simple interpretation connected to the preference of students and can be readily analyzed to yield insights for the original discrete model. We see our comprehensive technical analysis of the resulting differential equations that yields an answer to an interesting and practical question as one of our key contributions.

The outcome of the Serial Dictatorship mechanism in our model can also be understood as the application of a randomized version of the greedy algorithm on an online bipartite matching problem where, as usual, nodes from one side of the graph are given, while the others arrive one at the time, have degree exactly d, and are matched to one

[4]Indeed, the pointwise convergence in probability of our limit is a corollary of their work. We provide a simpler proof for a slightly stronger result (uniformly and pathwise) via an application of a functional law of large numbers.

of their currently unmatched neighbors uniformly at random (or discarded if no such neighbor exists). Starting from the seminal work by [29], many versions of online matching problems in bipartite graphs have been studied (see [30, 31] for recent surveys). To the best of our knowledge, such models mostly focus on the competitive ratio of global objective functions (such as the number of matches or some profit or cost function associated to the matching), while our node-by-node analysis of the probability of being matched appears to be new.

1.1.4 Notation

We write $\mathbb{N} = \{1, 2, 3, \dots\}$ and, for $k \in \mathbb{N}$, $[k] = \{1, 2, \dots, k\}$. We use $n \in \mathbb{N}$ for the number of schools, $d \in \mathbb{N}$ and $\ell \in \mathbb{N}$ for lengths of preference lists, and $m \in \mathbb{N}$ as the number of students (in cases where we do not assume an infinite list of students). Superscripts denote properties that are fixed for the market (except in the continuous realm where we place the d of $x_d(t)$ in the subscript for convenience), and subscripts are used for running indices. We use $i \in \mathbb{N}$ for students, $j \in [n]$ for schools. Random variables are denoted by uppercase letters.

1.2 Models and Results

1.2.1 A Discrete Random Market Model

Model description. Consider a random market consisting of $n \in \mathbb{N}$ schools each with exactly one seat, and infinitely many students indexed by $i = 1, 2, 3, \dots$. Each student is endowed with a strict *preference list*, consisting of $d \leq n$ schools, chosen independently and uniformly at random from the set of n schools. We call d the *preference list length*, and it is a central quantity of interest in this chapter. If a occurs before b in this preference list, the student strictly prefers being matched to a rather than b. If a school c does not appear in a student's preference list, the student finds this school unacceptable and prefers remaining unmatched to being matched to that school. A school that appears in a student's

preference list is *acceptable* to that student.

Students are matched to schools via a *Serial Dictatorship* mechanism as follows. Students are listed in the order they are indexed; and each student in their turn picks their most preferred school that has an available seat. If a student finds none of the remaining schools acceptable, then that student goes unmatched. For a student $i \in [m]$, we denote by $K_i^{n,d}$ the random variable denoting the position of the school they get matched to within their preference list, setting $K_i^{n,d} = \infty$ if the student is unmatched (so $K_i^{n,d} \in \{\infty, 1, \dots, d\}$). We define the random vector $\mathcal{K} = \{K_1^{n,d}, K_2^{n,d}, K_3^{n,d}, \dots\}$ as the realization of one run of the mechanism.

A student i wishes to minimize $K_i^{n,d}$. That is, they want to be ranked to the most preferred school in their list. On the other hand, not getting matched at all ($K_i^{n,d} = \infty$) is by far worse than getting matched to a school that the student finds acceptable but ranks lower. Define $M_i^{n,d} = \mathbb{1}_{\{K_i^{n,d} < \infty\}}$ to be the indicator random variable for the event that student i gets matched at all. The probability that they get matched to any school in their preference list is then $\mathbb{P}(M_i^{n,d} = 1)$.

For a given market with a fixed number of schools, only d is chosen by the central planner. It is therefore natural to ask what impact varying d has on the distributional properties of $\mathcal{K}$, and on the distribution of individual students' ranks (i.e. $K_i^{n,d}$), as well as their probabilities of getting matched to any school, $\mathbb{P}(M_i^{n,d} = 1)$.

Main results. As our main result[5], we show that for a number n of schools large enough, every student in position $i \leq n$ will be matched to a school with higher probability in markets with longer lists. The formal statement is as follows.

Theorem 1. *Let $d, \ell \in \mathbb{N}$ with $\ell > d$. For every n large enough and $i \leq n$, we have*

$$\mathbb{P}\left(M_i^{n,\ell} = 1\right) > \mathbb{P}\left(M_i^{n,d} = 1\right). \tag{1.1}$$

[5]All proofs of the results from this subsection appear in Section 1.3.3 and Section 1.3.4.

As we show in the following lemma, the probability of being matched increases with the length of preference lists for each school. For a school $j \in [n]$, define $H_j^{n,d}(i)$ to be the indicator random variable for the event that school j gets matched to some student just before student i has had their turn. Then the following holds.

Lemma 2. *Let $d, \ell, n \in \mathbb{N}$ with $d \leq \ell \leq n$. For every $j \in [n]$ and every $i = 1, 2, \ldots,$ we have:*

$$\mathbb{P}\left(H_j^{n,\ell}(i) = 1\right) \geq \mathbb{P}\left(H_j^{n,d}(i) = 1\right).$$

Theorem 1 and Lemma 2 imply that if the number of students does not exceed the number of schools, then for n large enough, all agents will be matched with higher probability in the market where preference lists are longer. Interestingly, Theorem 1 does not necessarily hold for i larger than n: as we discuss in Section 1.2.3, for any n large enough, students in position $i = \lceil 1.22n \rceil$ and beyond have a higher probability of being matched when preference lists are of length 1 as opposed to length 2. A similar phenomenon happens also for longer preference lists, as formalized by the next result.

Theorem 3. *Let $d, \ell \in \mathbb{N}$ with $\ell \geq d$. There exists a cutoff $c(d) > 1$ such that for every n large enough, and for all $i > n \cdot c(d)$,*

$$\mathbb{P}\left(M_i^{n,\ell} = 1\right) < \mathbb{P}\left(M_i^{n,d} = 1\right).$$

Furthermore, $c(d) \to 1$ as $d \to \infty$.

While Theorem 1 allows us to compare the relative value of $\mathbb{P}(M_i^{n,d} = 1)$ for different list lengths d, it does not give us any information on the absolute probability itself. This probability decreases with i and eventually converges to 0 as i becomes large enough (since all schools will have been taken), so a natural question is to ask how this probability behaves for moderate values of i. It turns out that as d increases, $\mathbb{P}(M_n^{n,d} = 1)$ quickly approaches $1/2$. This is significant for all students $i \leq n$ (since $\mathbb{P}(M_i^{n,d} = 1)$ decreases in i)

and for all students when the number of students is comparable to the number of schools. Recall that $K_i^{n,d}$ is the rank of the school that student i gets matched to in their preference list. We formalize this in the next result.

Theorem 4. *Let $d \in \mathbb{N}$. For every n large enough, we have*

$$\frac{d}{2d+1} \leq \mathbb{P}\left(M_n^{n,d} = 1\right) \leq \frac{2d}{4d+1}. \tag{1.2}$$

In particular,

$$\lim_{d\to\infty} \lim_{n\to\infty} \mathbb{P}(M_n^{n,d} = 1) = \frac{1}{2}.$$

We additionally prove the following lemma that bounds the change in probability of a given student getting matched to one of their top-k schools for $k \leq d$.

Lemma 5. *Let $d \in \mathbb{N}$. For every n large enough, $k \leq d$ and $i \leq n$, we have*

$$\mathbb{P}\left(K_i^{n,d} \leq k\right) - \mathbb{P}\left(K_i^{n,d+1} \leq k\right) \leq \left(\frac{d+2}{2d+3}\right)^{k/(d+1)} - \left(\frac{2d+1}{4d+1}\right)^{k/d}.$$

The dynamics of the discrete model. To analyze the discrete model, it is useful to study the number of students matched to any school just before it is student i's turn, given by

$$T_i^{n,d} = \sum_{j=1}^{i-1} \mathbb{1}_{\{K_j^{n,d} < \infty\}}.$$

Since each school has exactly one seat, $T_i^{n,d}$ coincides with the number of schools matched to the students $\{1, \ldots, i-1\}$. The number of schools that remain unmatched before student i's turn is therefore $n - T_i^{n,d}$.

We remark on one property of the random market, called the *principle of deferred decisions*. This has proved useful in analyzing other markets, see for example [13]. Observe that the preference list of student i does not play a role in the output until exactly the

i-th round of Serial Dictatorship, when it is the turn of student i to pick their favorite remaining school. We therefore do not need to specify i's preference list until this moment, allowing us to defer this decision until the list is needed. In particular, note that the distribution of $K_i^{n,d}$ depends only on $T_i^{n,d}$: we only need to know the number of schools that remain unmatched at the start of the turn to know the distribution of the position that the student gets matched to. This is further illustrated below.

Suppose it is the i-th student's turn, and the students prior to i have been matched to k schools, so $T_i^{n,d} = k$. It is then straightforward to compute the probability that a list of d schools chosen uniformly at random from the set of n schools will overlap with any of the unmatched $n - k$ schools, which is given by

$$\mathbb{P}\left(M_i^{n,d} = 1 \mid T_i^{n,d} = k\right) = \begin{cases} 1 - \frac{\binom{k}{d}}{\binom{n}{d}}, & k \geq d, \\ 1, & \text{otherwise.} \end{cases} \tag{1.3}$$

It is immediately clear from this formula that the probability of getting matched depends only on n, d, and $T_i^{n,d}$. Furthermore, all else held constant, $\mathbb{P}(M_i^{n,d} = 1 \mid T_i^{n,d} = k)$ decreases as $T_i^{n,d}$ increases. The principle of deferred decisions allows us to connect our discrete model to a continuous one, which we discuss next.

1.2.2 A Continuous Market Model

We now introduce a continuous market model, which we will show in the next section to be equivalent to the limit (uniformly in probability) of the previously introduced discrete market model.

Description. Consider a (deterministic) continuous matching model consisting of a unit mass of schools and a continuum of students represented by the interval $[0, \infty)$ and indexed by $t \geq 0$. Let $d \geq 1$ be a parameter of the market, and define $x_d(t)$ to be the proportion of schools matched to students in $[0, t)$. We now define the market by the

initial value problem

$$x_d(0) = 0, \quad x_d'(t) = 1 - x_d(t)^d. \tag{1.4}$$

The interpretation of the initial value problem (1.4) is as follows: start with $x_d(0) = 0$ and traverse students in order from $t = 0$. At the time of student t, they get matched to schools at rate $1 - x_d(t)^d$, which is also then the rate by which $x_d(t)$ increases at t. We note $x_d'(t)$ depends only on the proportion of schools matched until that point, and that $x_d(t)$ is smooth in its argument and takes values in $[0, 1]$.

Dynamics. The dynamics of the continuous model are conceptually simple. Note that x_d is everywhere strictly increasing, and $x_d(t) \to 1$ as $t \to \infty$, but $x_d'(t) \to 0$ as $t \to \infty$. Further, one immediately observes that for $t \approx 0$, $x_d'(t) \approx 1$: at the start, students are getting matched at the highest possible rate. The parameter d is a measure of intensity of matching: if the proportion of schools matched were fixed, smaller d would yield a slower rate of matching.

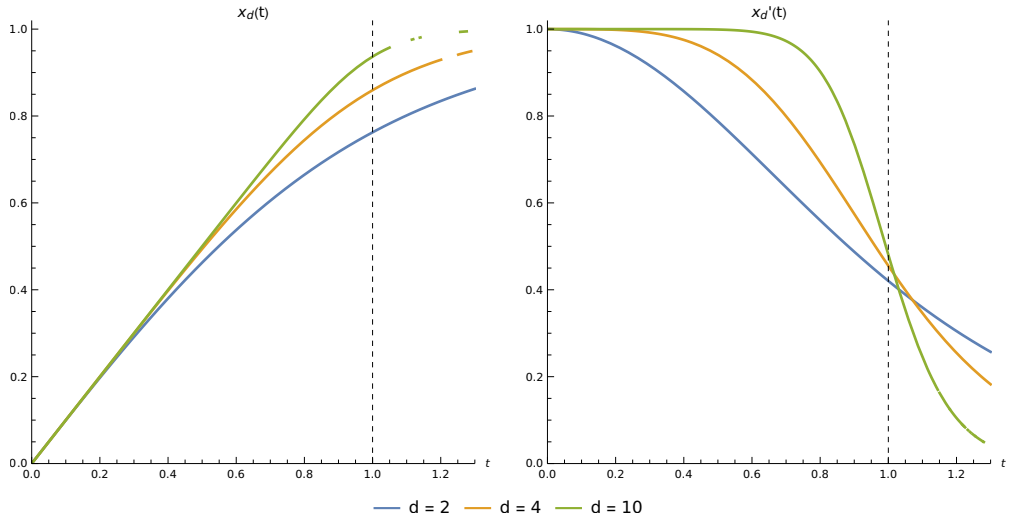


Figure 1.1: Plots of x_d and x_d' for various values of d: note how the values for x_d' cross just after $t = 1$.

Similarly to the discrete model (see the discussion in Section 1.1.1), we see that there are again two forces at play: for small d, the rate of getting matched is smaller, but simultaneously the proportion of schools taken up by earlier students is smaller, so later

students may prefer small d as it leaves them some schools to possibly get matched to. Indeed, for t exceeding 1, the probability of getting matched quickly vanishes as d increases. On the other hand, a large d increases the rate of getting matched, but also reduces the proportion of schools remaining by the time of student t's turn.

We remark that although the description of the initial value problem in (1.4) might look deceptively straightforward, actually analyzing it proves to be challenging. The differential equation is highly non-linear, a special case of the *Chini type* differential equation studied at least since 1924 [32], to which there is no known analytical solution for general d [33]. To circumvent this, we must find implicit ways of proving properties of the differential equation that do not require finding an analytical solution explicitly.

1.2.3 From Continuous to Discrete

In this section we discuss the main theorem that connects the continuous and discrete markets, then state our main results within the continuous realm, and finally describe how they carry over to the discrete case.

Connection to the discrete model. We now give an intuition on how the discrete market with fixed list length d approaches the continuous market when the number of schools $n \to \infty$. We make this limit rigorous via a functional law of large numbers in Theorem 6.

To see the connection, consider the discrete market described in Section 1.2.1 for a fixed list length d, and suppose we are at the turn of student $t = i/n$ for some fixed n. Recall that $T_i^{n,d}$ is the number of schools taken by students $\{1, \dots, i-1\}$. In Lemma 12, we show that

$$\mathbb{P}\left(M_i^{n,d} = 1\right) = 1 - (T_i^{n,d}/n)^d + O(d^2/n). \tag{1.5}$$

Now $T_i^{n,d}/n$ is the proportion of schools taken by students $\{1, \dots, i-1\}$ and so is an analogue of $x_d(t)$ (with $t = i/n$) in the continuous model, which denotes the proportion of

schools taken by students in $[0, t)$. In fact, Theorem 6 proves exactly that $T_i^{n,d}/n \to x_d(i/n)$ as $n \to \infty$. In the limit as $n \to \infty$, the last term vanishes so when student $t = i/n$ has their turn, they get matched to a school with probability $x_d'(t) = 1 - x_d(t)^d$. Indeed, we show in Lemma 7 that $\mathbb{P}(M_i^{n,d} = 1) \to x_d'(i/n)$ in probability as $n \to \infty$. In the continuous model d is therefore the analogue of the list length (but is now relaxed to be any real number in $[1, \infty)$).

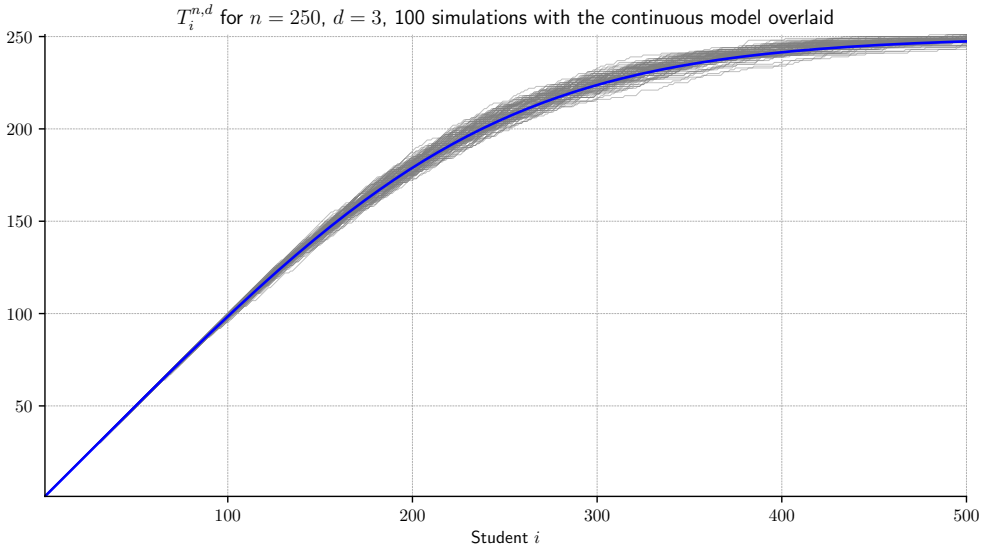


Figure 1.2: 100 simulations of the discrete model with the continuous model overlaid.

The following theorem, proved in Section 1.3.1, describes rigorously how the continuous market is a limit of the discrete market, showing that $T_i^{n,d}/n \to x_d(i/n)$ as $n \to \infty$ in a very strong sense.

Theorem 6. *Fix $d \in \mathbb{N}$ and define $p(x) = 1 - x^d$ for $x \in [0, 1]$. For $t \geq 0$, let $T_{\lfloor tn \rfloor}^{n,d}$ be the random variable denoting the number of schools matched by the first $\lfloor tn \rfloor$ students when there are n schools in the discrete random market with list length d. Then $n^{-1}T_{\lfloor tn \rfloor}^{n,d} \to x_d(t)$ uniformly in probability as $n \to \infty$, where $x_d(t)$ is the unique solution satisfying $x_d(0) = 0$ and $x_d'(t) = p(x_d(t))$ for $t \geq 0$. That is, for all $s \geq 0$ and $\varepsilon > 0$, as $n \to \infty$,*

$$\mathbb{P}\left(\sup_{t \in [0,s]} \left| n^{-1}T_{\lfloor tn \rfloor}^{n,d} - x_d(t) \right| \geq \varepsilon \right) \to 0. \tag{1.6}$$

Moreover, as $n \to \infty$ for all $r \geq 1$, we have

$$\mathbb{E}\left(\left|n^{-1}T^{n,d}_{\lfloor tn \rfloor} - x_d(t)\right|^r\right) \to 0. \tag{1.7}$$

In particular, (1.6) implies that as $n \to \infty$, the number of schools taken prior to student $t = i/n$ converges in probability to the continuous market defined by the initial value problem (1.4). Not only does the proportion of schools taken converge in probability for every point $t \geq 0$ and its mean converge to the continuous solution (by taking $r = 1$ in (1.7)), but additionally on a sample path level, the maximum deviations of the random market around the continuous solution vanish in probability. This is a very strong form of pathwise convergence. The theorem additionally posits that the initial value problem (1.4) defining the continuous market has a unique solution satisfying the boundary data and that this solution extends to all time $t \geq 0$.

The next lemma rigorously connects the match rates of the discrete and continuous models in this limit. Recall that $M^{n,d}_i$ is the indicator random variable for whether student i gets matched to any school in the discrete random market with n schools and preference lists of length d.

Lemma 7. *For all $d \in \mathbb{N}$, $t \geq 0$, we have $\mathbb{P}(M^{n,d}_{\lfloor tn \rfloor} = 1) \to x'_d(t)$ as $n \to \infty$.*

To understand how the list length d affects $\mathbb{P}(M^{n,d}_i = 1)$, we therefore need to understand how the d parameter affects $x'_d(t)$. That is, if for all $\ell > d$, $x'_\ell(t) \geq x'_d(t)$ for some $t \geq 0$, then student $t = i/n$ prefers lists of length ℓ to those of length d for large enough n. Student t then always prefers longer lists to shorter ones. Similarly if for some $t \geq 0$, we have $x'_\ell(t) < x'_d(t)$, then student t prefers shorter lists. Our task of understanding match probability therefore becomes one of understanding $x'_d(t)$ in the continuous market.

The case of $d = 1$ and $d = 2$. For $d = 1$ and $d = 2$, we can in fact compute analytic solutions to the initial value problem (1.4) defining the continuous market. For the case

of $d = 1$, we unsurprisingly get

$$x_1(t) = 1 - e^{-t}, \qquad x_1'(t) = e^{-t},$$

and for $d = 2$ we get

$$x_2(t) = \frac{e^t - e^{-t}}{e^t + e^{-t}}; \qquad x_2'(t) = \frac{4e^{2t}}{(e^{2t}+1)^2}.$$

The latter may also be written $x_2(t) = \tanh(t)$ and $x_2'(t) = \operatorname{sech}(t)^2$ in terms of the hyperbolic functions. With these analytic expressions, one can solve to find $x_2'(t) \geq x_1'(t)$ if and only if $t \leq 1.219$. This means that for students before this cutoff, they prefer lists of length 2 to lists of length 1, and vice versa for larger t (see Figure 1.3). We wish to highlight two key phenomena from this example. Firstly, all students in $t \in [0, 1]$ have a higher probability of getting matched to any school with longer lists compared to shorter lists, and secondly, there is a cutoff (after $t = 1$) where this behavior reverses and students prefer shorter lists.

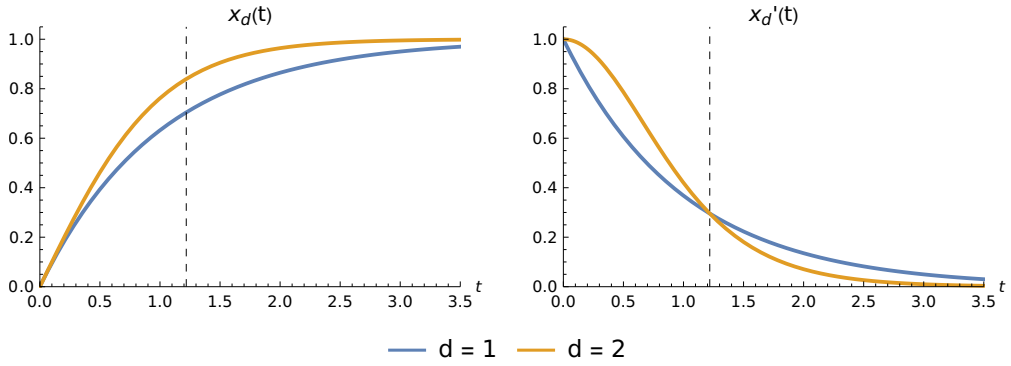


Figure 1.3: Plots of x_d and x_d' for $d = 1$ and $d = 2$. Note that students before the highlighted cutoff at $t = 1.219$ have higher probability of being assigned to any school with longer lists, and vice versa for students after the cutoff.

The case of general d. In the case of general d, the initial value problem (1.4) cannot be solved analytically, so we cannot hope to analyze the market dynamics directly in

the same way as for $d = 1$ and $d = 2$. Nonetheless, we are able to prove these results implicitly, and our main results show that for general d these two phenomena still hold: all students in $t \in [0, 1]$ have a higher probability of getting matched to any school with larger d, and there is some cutoff $c(d) > 1$ so that for $t > c(d)$, this preference reverses and those students prefer shorter lists. We furthermore show that this cutoff approaches 1 as $d \to \infty$, so for large d, all students prefer longer lists if and only if the market is balanced. These two results are formalized for the continuous realm in the form of the following two theorems.

Theorem 8. *For all $\ell > d \geq 1$ and for all $t \in (0, 1]$, $x'_\ell(t) > x'_d(t)$.*

Theorem 9. *For all $d \geq 1$, there exists $c(d) > 1$ such that for all $\ell > d$ and $t \geq c(d)$, $x'_\ell(t) < x'_d(t)$. Furthermore, $c(d) \to 1$ as $d \to \infty$.*

We later combine these two results with the limit results in Theorem 6 and Lemma 7 to bring them back to the discrete market as Theorem 1 and Theorem 3, respectively.

1.2.4 Implications and Extensions

Random Serial Dictatorship. In the Random Serial Dictatorship mechanism, the order in which students propose is chosen randomly from all possible permutations of students before the mechanism starts. When applied to our setting, it introduces an additional layer of randomness, on top of the random preferences of students.

Formally, consider a modification of our model, where there are n schools each with one seat, and $m \in \mathbb{N}$ students ordered randomly. Each student has a strict preference list of d schools chosen uniformly at random from the set of n schools. For $d, n \in \mathbb{N}$, we denote by $R^{n,d}$ the indicator random variable for the event that a student is matched to some school under the Random Serial Dictatorship algorithm (note that this probability does not depend on the specific student, while it depends on m, but we are omitting this dependency for the sake of brevity). The following is an easy corollary of Lemma 2.

Corollary 10. *Let $d, \ell, n \in \mathbb{N}$ with $n \geq \ell \geq d$. We have*

$$\mathbb{P}\left(R^{n,\ell} = 1\right) \geq \mathbb{P}\left(R^{n,d} = 1\right).$$

Proof. By symmetry, we have $\mathbb{P}(R^{n,d} = 1) = \frac{n}{m}\mathbb{P}(H_1^{n,d}(m) = 1)$. By Lemma 2, $\mathbb{P}(H_1^{n,d}(m) = 1)$ increases with d so $\mathbb{P}(R^{n,d} = 1)$ also increases with d. □

Schools with multiple seats. In this section we discuss an extension of our model to the case where schools each have exactly $q = 1, 2, 3, \ldots$, thereby relaxing the assumption of a single seat per shcool. Like in the simpler case, we are interested in the probability that students $i = 1, 2, 3, \ldots$ get matched to any school as they each have their turn and are (possibly) matched to one of their remaining acceptable schools. We briefly describe this extension here and leave a detailed discussion with proofs to Appendix A.2.

Formally fix $q \in \mathbb{N}$ and let n be the number of schools, this time each with q seats. As student $i = 1, 2, \ldots$ arrives, they randomly sample d schools as their preference list. We denote by $M_i^{n,d,q}$ the indicator random variable of whether this student gets matched to any school in this list. In addition to keeping track of the number of schools with no seats remaining, $T_i^{n,d,q}$, we now also define the vector $S_i = (S_i^0, S_i^1, S_i^2, \ldots, S_i^{q-1})$ counting the number of schools with $k = 0, 1, \ldots, q-1$ seats taken by the start of the i-th turn (we do not include the market parameters in the notation of S_i for brevity). We discuss the specifics of the market dynamics in Appendix A.2, where we also prove Lemma 99 which is an analogue of Theorem 6 in the case of multiple seats, rigorously connecting this discrete market to a continuous one. It shows that $n^{-1}T_{\lfloor tn \rfloor}^{n,d,q} \to x_d(t)$ and $n^{-1}S_{\lfloor tn \rfloor} \to y_d(t)$ uniformly in probability as $n \to \infty$ where $x_d(t)$ and $y_d(t) = (y_d^0(t), \ldots, y_d^{q-1}(t))$ are given

by the initial value problem (we denote derivative with respect to t by dot for clarity)

$$\begin{aligned} y_d^0(0) &= 1, \qquad \dot{y}_d^0(t) = -\frac{1-x_d(t)^d}{1-x_d(t)} y_d^0(t), \\ y_d^k(0) &= 0, \qquad \dot{y}_d^k(t) = \frac{1-x_d(t)^d}{1-x_d(t)} (y_d^{k-1}(t) - y_d^k(t)), \qquad k = 1, 2, \ldots, q-1, \\ x_d(0) &= 0, \qquad \dot{x}_d(t) = \frac{1-x_d(t)^d}{1-x_d(t)} y_d^{q-1}(t). \end{aligned} \tag{1.8}$$

Again, this differential equation is analytically intractable for general values of d and q, but we provide an explicit solution for $d = 1$ and arbitrary q in Appendix A.2, then provide a connection between that solution and the solution for arbitrary parameters. We also have Lemma 11, an analogue of Lemma 7 connecting the continuous solution to the match probability.

Lemma 11. *For all $d \in \mathbb{N}$, $t \geq 0$, we have $\mathbb{P}(M_{\lfloor tn \rfloor}^{n,d,q} = 1) \to 1 - x_d(t)^d$ as $n \to \infty$.*

In the case of $q = 1$ we observe interesting behavior near the point where the market is balanced, at $i \approx n$. Now with $q > 1$ we would look for this at $i \approx nq$ due to the higher number of seats. Indeed, based on numerical experiments (some of which are reported in Section 1.5), we conjecture that the following extension of Theorem 1 holds. Let $d, \ell \in \mathbb{N}$ with $\ell \geq d$. For every n large enough and $i \leq nq$, we have

$$\mathbb{P}\left(M_i^{n,\ell,q} = 1\right) \geq \mathbb{P}\left(M_i^{n,d,q} = 1\right).$$

The missing piece to prove Conjecture 1.2.4 is a result like Theorem 8 in the case of $q \geq 1$. We defer the rest of the discussion of multiple seats to Appendix A.2.

1.2.5 Discussion of the Discrete Model

Models where agents have preference lists that are sampled uniformly at random from the list of all permutations (possibly of bounded size, as in ours) form a very common theoretical assumption for the study of two-sided matching markets, see for example [17,

13, 15, 16, 18]. One of the reasons for the popularity of these models is that they often lead to tractable expressions. Results can then be verified to hold numerically for models with correlated preferences, or this can motivate further analytical study of such more cumbersome markets. Our approach follows this general line of inquiry, with our main results (Theorem 1 and Theorem 3) proved for the case where lists are sampled uniformly at random, and verified numerically in Section 1.4 for preference lists sampled from more complex distributions.

The choice of Serial Dictatorship for the mechanism investigated in this chapteris motivated by multiple facts. First, its relevance: it is broadly studied in theory (including its randomized version, see for example [5, 23]) and applied in practice to many settings, most famously house allocation problems [34]. A Serial Dictatorship mechanism also naturally arises in the case of two-sided stable matching problems where schools have a shared (homogeneous) and strict preference list over the students, for instance when school preference is dictated by test scores, or by a lottery system where the lottery number determines priority [26]. In such cases, the Deferred Acceptance mechanism reduces to Serial Dictatorship, with students arriving according to the shared order and picking their favorite remaining school. Both of these scenarios are common in school districts. For instance, in the New York primary school matching, the Department of Education has very limited information on students and so assigns priorities largely at random, and even completely at random in cases such as city-wide gifted and talented schools (where the admissible students are restricted to those who are considered gifted or talented). On the other hand, the NYC specialized high schools are by law required to only rank students uniquely by their score on the standardized Specialized High School Admissions Test (SHSAT) [35].

Another reason to study the Serial Dictatorship mechanism is that it endows students with an obvious order, allowing granular statements to be made about each student uniquely (such as in our case discussing outcomes based on order). In many other

mechanisms where students do not have a natural order, students become indistinguishable, and one can hope for at most aggregate level statements and results. For example in school matching, one may wish to make statements about "better students", but if schools do not rank all students in the same order, it becomes difficult to distinguish good students from bad students in the way we are able to do so.

1.3 Proofs of Main Results

1.3.1 Connections Between the Discrete and Continuous Markets

In this section we prove the following theorem, which rigorously establishes the discrete market converging to the continuous market as $n \to \infty$.

Theorem 6. *Fix $d \in \mathbb{N}$ and define $p(x) = 1 - x^d$ for $x \in [0,1]$. For $t \geq 0$, let $T^{n,d}_{\lfloor tn \rfloor}$ be the random variable denoting the number of schools matched by the first $\lfloor tn \rfloor$ students when there are n schools in the discrete random market with list length d. Then $n^{-1}T^{n,d}_{\lfloor tn \rfloor} \to x_d(t)$ uniformly in probability as $n \to \infty$, where $x_d(t)$ is the unique solution satisfying $x_d(0) = 0$ and $x_d'(t) = p(x_d(t))$ for $t \geq 0$. That is, for all $s \geq 0$ and $\varepsilon > 0$, as $n \to \infty$,*

$$\mathbb{P}\left(\sup_{t \in [0,s]} \left| n^{-1}T^{n,d}_{\lfloor tn \rfloor} - x_d(t) \right| \geq \varepsilon \right) \to 0. \tag{1.6}$$

Moreover, as $n \to \infty$ for all $r \geq 1$, we have

$$\mathbb{E}\left(\left| n^{-1}T^{n,d}_{\lfloor tn \rfloor} - x_d(t) \right|^r \right) \to 0. \tag{1.7}$$

We begin by computing an approximation to the probability of a student getting matched in the discrete case.

Lemma 12 (Probability of getting matched)**.** *In the discrete market with n schools, we have*

$$\left| \mathbb{P}\left(M_i^{n,d} = 1 \mid T_i^{n,d} = k \right) - \left(1 - (k/n)^d \right) \right| \leq d^2/n.$$

That is, when k schools have been matched prior to student i, then the probability that student i gets matched to any school is approximated by $1 - (k/n)^d$.

Proof. Recall that when it is student i's turn, they get matched to any school with probability given in (1.3), which for $k \geq d$ states $\mathbb{P}\left(M_i^{n,d} = 1 \mid T_i^{n,d} = k\right) = 1 - \binom{k}{d}/\binom{n}{d}$. This is the complement of the probability that a random list of d schools sampled from n without replacement will overlap entirely with the list of k taken schools. Consider now an alternative procedure where student i samples a list of d schools *with replacement*, where each school has a probability of k/n of being already taken by previous students. The probability of sampling a list with at least one school not taken by previous students conditioned on exactly k schools having been matched to students $1, \ldots, i-1$ would be $1 - (k/n)^d$.

It remains to bound the difference between sampling the preference lists with replacement and sampling them without replacement. Observe that the outcomes differ exactly in the case that we choose at least one pair of schools that are the same while sampling with replacement. Otherwise the outcomes coincide. Formally, consider the probability space generated by the student sampling with replacement, assuming k schools are matched to students $1, \ldots, i-1$. Let E be the event that student i gets matched to any school and let X be the event that the sample contains a repeated school. We have $\mathbb{P}\left(M_i^{n,d} = 1 \mid T_i^{n,d} = k\right) = \mathbb{P}(E \mid X^c)$, and $\mathbb{P}(E) = 1 - (k/n)^d$. Now with the law of total probability, write

$$\begin{aligned}
|\mathbb{P}(E) - \mathbb{P}(E \mid X^c)| &= |\mathbb{P}(E \mid X)\,\mathbb{P}(X) + \mathbb{P}(E \mid X^c)\,(1 - \mathbb{P}(X)) - \mathbb{P}(E \mid X^c)| \\
&= \mathbb{P}(X)\,|\mathbb{P}(E \mid X) - \mathbb{P}(E \mid X^c)| \\
&\leq \mathbb{P}(X)\,.
\end{aligned}$$

Finally we bound $\mathbb{P}(X)$, the probability that a sample drawn with replacement contains duplicate items. Observe that for each of the $d(d-1)/2$ pairs of positions in the sample there is exactly a $1/n$ probability that that pair is the same. Using the union bound, this

yields $\mathbb{P}(X) \leq \frac{d(d-1)}{2n} \leq d^2/n$. □

Our aim is to show that the discrete market converges to the continuous one as n grows large. We assume here that reader is familiar with the theory of ordinary differential equations and Markov processes. A gentle introduction to the relevant background can be found in Appendix A.1.

To connect the behavior of the discrete market with that of the continuous market, we apply the following theorem. See [36, Theorem 17.3.1] for this version and [37] for a proof.

Theorem 13 (A Functional Law of Large Numbers)**.** *Suppose $\{J_t^n\}_{n=1,2,\ldots}$ is a family of continuous time Markov chains on finite state spaces where for $t \geq 0$, J_t^n takes values in $\mathcal{S}^n \subseteq \mathbb{Z}$ and has transition rate matrix $Q^n = (q^n(i,j); i,j \in \mathcal{S}^n)$. Suppose that there is a subset $\mathcal{D} \subseteq \mathbb{R}$ and a family $\{f^n\}_{n=1,2,\ldots}$ of functions with $f^n : \mathcal{D} \times \mathbb{Z} \to \mathbb{R}$ which are bounded and continuous in the first argument such that, for each $i \in \mathcal{S}^n$ and $k \in \mathbb{Z} \setminus \{0\}$ such that $i + k \in \mathcal{S}^n$, we have*

$$q^n(i, i+k) = n f^n(i/n, k). \tag{1.9}$$

Define $g^n(x) = \sum_{k\in\mathbb{Z}} k f^n(x,k)$ for $x \in \mathcal{D}$, and suppose there exists a Lipschitz continuous function $g : \mathcal{D} \to \mathbb{R}$ such that the g^n converge uniformly to g on $\mathcal{D}$. Suppose $\lim_{n\to\infty} n^{-1} J_0^n = x_0$ for some x_0. Then there exists a unique deterministic trajectory $x(t)$ satisfying $x(0) = x_0$ and $x'(t) = g(x(t))$, $x(t) \in \mathcal{D}$, $t \in [0,T]$, and $\{n^{-1}J_t^n\}_{t\geq 0}$ converges uniformly in probability on $[0,T]$ to $x(t)$.

We are now ready to prove Theorem 6.

Proof of Theorem 6. As in the statement of the theorem, fix $d \geq 1$ and define $p(x) = 1 - x^d$ for $x \in [0,1]$. For $t \geq 0$, define $X_t^n = T_{\lfloor nt \rfloor}^{n,d}$. Letting $\Delta = n^{-1}$ for convenience, note that $\{X_t^n\}_{t=0,\Delta,2\Delta,\ldots}$ is a discrete time Markov chain on the state space $\mathcal{S}^n = \{0, 1, \ldots, n\}$ with

transition probabilities

$$\mathbb{P}\left(X^n_{t+\Delta} = j \mid X^n_t = i\right) = \begin{cases} p^n_i, & j = i+1, \\ 1 - p^n_i, & j = i, \\ 0, & \text{otherwise.} \end{cases}$$

where $p^n_i = 1 - (i/n)^d \pm O(d^2/n)$ by Lemma 12.

Starting from $\{X^n_t\}_{t=0,\Delta,2\Delta}$, we will now construct a sequence of continuous time Markov chains that satisfy the conditions of Theorem 13, then bring back the result to the discrete time chain.

Associate to each discrete time chain $\{X^n_t\}_{t=0,\Delta,2\Delta,\ldots}$, a coupled continuous time Markov chain $\{J^n_t\}_{t\geq 0}$ as follows: let H^n_t be a homogeneous Poisson process with rate $\lambda = n$, and let $\{J^n_t\}_{t\geq 0}$ be defined by $J^n_t = X^n_{\Delta H^n_t}$. This is a well-known embedding technique (see Theorem 96 in Appendix A.1 for a statement and discussion) and yields the transition rate matrix

$$q^n(i,j) = \begin{cases} np^n_i, & j = i+1, \\ -np^n_i, & j = i, \\ 0, & \text{otherwise.} \end{cases}$$

Set $f^n(x,1) = p^n_{xn}$ and $f^n(x,\cdot) = 0$ otherwise. We now verify that we can apply Theorem 13. First, let $\mathcal{S}^n = \{0,1,2,\ldots,n\}$ and $\mathcal{D} = [0,1]$. Note q^n, f^n satisfy (1.9): we only need to check $k = 1$ in (1.9), which we verify as

$$q^n(i,i+1) = np^n_i = nf^n(i/n,1).$$

For each n, f^n is clearly bounded and continuous in the first argument. We moreover

have for all $x \in [0,1]$

$$\begin{aligned} g^n(x) &= \sum_{k\in\mathbb{Z}} k\, f^n(x,k) \\ &= p^n_{xn} \\ &= p(x) \pm O(d^2/n), \qquad \text{(by Lemma 12)} \end{aligned}$$

hence $g^n \to p$ uniformly, since $|g^n(x) - p(x)| = O(d^2/n)$ is independent of x. Observe that p is Lipschitz-continuous with Lipschitz-constant d. Since $J^n_0 = 0$, we have $n^{-1}J^n_0 = 0$.

We can therefore apply Theorem 13 and deduce that $\{n^{-1}J^n_t\}_{t\geq 0}$ converges uniformly in probability to the unique solution $x_d(t)$ of the initial value problem (1.4).

To carry over convergence uniformly in probability to the discrete Markov chain $\{X^n_t\}_{t=0,\Delta,2\Delta,\ldots}$, fix some $s \geq 0$. Recall that, for $\tau \geq 0$, H^n_τ is the count of events in the underlying Poisson process up to time τ. Since every event of H^n_t corresponds to to either a unit jump in $\{X^n_t\}_{t=0,\Delta,2\Delta,\ldots}$ (that is, $X^n_{t+\Delta} = X^n_t + 1$) or no transition (that is, $X^n_{t+\Delta} = X^n_t$), we have for each $t \geq 0$.

$$\begin{aligned} \left|X^n_t - J^n_t\right| &= \left|X^n_t - X^n_{\Delta H^n_t}\right| \\ &\leq \left|\lfloor nt \rfloor - H^n_t\right| \\ &\leq 1 + \left|nt - H^n_t\right|. \end{aligned}$$

This yields

$$\begin{aligned} \sup_{t\in[0,s]} \left|n^{-1}X^n_t - x_d(t)\right| &\leq \sup_{t\in[0,s]} \left|n^{-1}X^n_t - n^{-1}J^n_t\right| + \sup_{t\in[0,s]} \left|n^{-1}J^n_t - x_d(t)\right| \\ &\leq \frac{1}{n} + \sup_{t\in[0,s]} \left|t - n^{-1}H^n_t\right| + \sup_{t\in[0,s]} \left|n^{-1}J^n_t - x_d(t)\right|. \end{aligned} \tag{1.10}$$

Note that $n^{-1}H_t^n - t$ is a martingale[6], and applying Doob's martingale inequality (see Theorem 98 of Appendix A.1) we have

$$\mathbb{P}\left(\sup_{t\in[0,s]}\left|n^{-1}H_t^n - t\right| \geq \varepsilon\right) \leq \varepsilon^{-2}\mathbb{E}\left((n^{-1}H_s^n - s)^2\right)$$
$$= \frac{s}{\varepsilon^2 n}. \tag{1.11}$$

To see that the latter inequality holds, observe $\mathbb{E}\left(n^{-1}H_s^n\right) = s$, so $\mathbb{E}\left((n^{-1}H_s^n - s)^2\right)$ equals the variance of $n^{-1}H_s^n$. Since H_s^n has a Poisson distribution with rate ns, (1.11) follows.

Note that the second (by (1.11)) and third (by Theorem 13) terms of the right-hand side of (1.10) vanish in probability as $n \to \infty$. So we must have that $n^{-1}T_{\lfloor tn\rfloor}^{n,d} \to x_d(t)$ uniformly in probability as $n \to \infty$. The convergence in r-mean follows since $n^{-1}T_{\lfloor tn\rfloor}^{n,d}$ is bounded, see Lemma 90 in Appendix A.1. □

We connect the solution to preferences of students via the following lemma.

Lemma 14. *For all $d \in \mathbb{N}$, $t \geq 0$, we have $\mathbb{P}(M_{\lfloor tn\rfloor}^{n,d} = 1) \to x_d'(t)$ as $n \to \infty$.*

Proof. Fix $d \in \mathbb{N}$, and $t, \varepsilon > 0$. We will show that there exists $N \in \mathbb{N}$ such that for $n > N$,

$$\left|\mathbb{P}\left(M_{\lfloor tn\rfloor}^{n,d} = 1\right) - x_d'(t)\right| < \varepsilon.$$

To proceed, observe that Lemma 12 implies that there exists N_1 such that for $n > N_1$ and all $x \in [0,1]$,

$$\left|\mathbb{P}\left(M_{\lfloor tn\rfloor}^{n,d} = 1 \mid n^{-1}T_{\lfloor tn\rfloor}^{n,d} = x\right) - (1 - x^d)\right| \leq \frac{\varepsilon}{4}.$$

Next, note that $1 - x^d$ is continuous in x, so there exists some $\delta > 0$ such that for all y with

[6]For $s \geq t$, using the independence of increments property and the fact that $\mathbb{E}\left(H_{s-t}^n\right) = n(s-t)$, we have

$$\mathbb{E}\left(n^{-1}H_s^n - s \mid H_t^n\right) - (n^{-1}H_t^n - t) = \mathbb{E}\left(n^{-1}(H_s^n - H_t^n) \mid H_t^n\right) - (s-t) = \mathbb{E}\left(n^{-1}(H_{s-t}^n)\right) - (s-t) = 0.$$

$|y - x| \leq \delta$, $\left|(1 - x^d) - (1 - y^d)\right| < \varepsilon/4$. Combining these two facts, we have that

$$\left|\mathbb{P}\left(M^{n,d}_{\lfloor tn \rfloor} = 1 \mid n^{-1}T^{n,d}_{\lfloor tn \rfloor} = y, |y - x| \leq \delta\right) - (1 - x^d)\right| \leq \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \frac{\varepsilon}{2}.$$

By Theorem 6, there exists now $N_2 \in \mathbb{N}$ such that for $n > N_2$,

$$\mathbb{P}\left(\left|n^{-1}T^{n,d}_{\lfloor tn \rfloor} - x_d(t)\right| > \delta\right) < \frac{\varepsilon}{2}.$$

Let $N = \max\{N_1, N_2\}$, then putting all these together with the law of total probability, we have for $n > N$

$$\begin{aligned}
\left|\mathbb{P}\left(M^{n,d}_{\lfloor tn \rfloor} = 1\right) - x'_d(t)\right| &= \Bigg|\mathbb{P}\left(M^{n,d}_{\lfloor tn \rfloor} = 1 \mid \left|n^{-1}T^{n,d}_{\lfloor tn \rfloor} - x_d(t)\right| \leq \delta\right)\mathbb{P}\left(\left|n^{-1}T^{n,d}_{\lfloor tn \rfloor} - x_d(t)\right| \leq \delta\right) \\
&\qquad + \mathbb{P}\left(M^{n,d}_{\lfloor tn \rfloor} = 1 \mid \left|n^{-1}T^{n,d}_{\lfloor tn \rfloor} - x_d(t)\right| > \delta\right)\mathbb{P}\left(\left|n^{-1}T^{n,d}_{\lfloor tn \rfloor} - x_d(t)\right| > \delta\right) - x'_d(t)\Bigg| \\
&\leq \left|\mathbb{P}\left(M^{n,d}_{\lfloor tn \rfloor} = 1 \mid n^{-1}T^{n,d}_{\lfloor tn \rfloor} = y, |y - x_d(t)| \leq \delta\right) + \mathbb{P}\left(\left|n^{-1}T^{n,d}_{\lfloor tn \rfloor} - x_d(t)\right| > \delta\right) - x'_d(t)\right| \\
&< \left|1 - x_d(t)^d + \frac{\varepsilon}{2} + \frac{\varepsilon}{2} - x'_d(t)\right| \\
&= \varepsilon,
\end{aligned}$$

as required, since $\varepsilon > 0$ was arbitrary. □

1.3.2 Continuous Market

In this section we prove that the match rate in the continuous market increases with d for all students $t \in [0, 1]$, and that there exists a cutoff $c(d) > 1$ approaching 1 as $d \to \infty$ such that the match rate decreases in d for students $t \geq c(d)$. Significant technical portions of the proofs are deferred to Appendix A.3. We begin by introducing an integral equation that will aid our analysis.

Lemma 15. *A function $x_d(t)$ is a solution to the initial value problem* (1.4) *if and only if it solves*

the integral equation

$$t = \int_0^{x_d(t)} \frac{1}{1-u^d}\, du. \tag{1.12}$$

Proof. Note that $x_d(0) = 0$, and by implicitly differentiating (1.12) with respect to t, we recover the required condition on the derivative of $x_d(t)$ with respect to t, that is, $x_d'(t) = 1 - x_d(t)^d$. □

We next prove a technical lemma that relates the sign of $\frac{\partial}{\partial d} x_d'(t)$ to the sign of a certain integral, then defer the rest of the technicalities to Appendix A.3 where we bound the actual integrals.

Lemma 16. *Let $x_d(t)$ be defined for all $d \geq 1$ and $t > 0$ as in the initial value problem* (1.4)*. Then*

$$\frac{\partial}{\partial d} x_d'(t) \cdot \int_0^{x_d(t)} \frac{1+\log u}{1-u^d}\, du < 0. \tag{1.13}$$

That is, $\frac{\partial}{\partial d} x_d'(t)$ has opposite sign to the integral.

Proof. We use the notation $x(d,t)$ and simplify to x when it is clear from the context. With this notation, the initial value problem (1.4) becomes $\frac{\partial}{\partial t} x(d,t) = 1 - x^d$ with $x(d,0) = 0$. Note that x is twice continuously differentiable on its domain so its partial derivatives commute.

Using now the Leibniz integral rule to implicitly differentiate the integral representation in (1.12) with respect to d, we obtain

$$\begin{aligned} 0 &= \frac{\partial}{\partial d}\left(\int_0^{x(d,t)} \frac{1}{1-u^d}\, du\right) \\ &= \int_0^{x(d,t)} \frac{\partial}{\partial d}\left(\frac{1}{1-u^d}\right) du + \frac{1}{1-x(d,t)^d}\frac{\partial x}{\partial d}. \end{aligned}$$

Rearranging for $\frac{\partial x}{\partial d}$ and computing the derivative in the integrand yields

$$\begin{aligned}\frac{\partial x}{\partial d} &= -(1-x(d,t)^d)\int_0^{x(d,t)} \frac{u^d \log u}{(1-u^d)^2}\,du. \\ &= -\frac{\partial x}{\partial t}\int_0^{x(d,t)} \frac{u^d \log u}{(1-u^d)^2}\,du.\end{aligned} \tag{1.14}$$

Note that $\frac{\partial x}{\partial t}$ is strictly positive, and the integrand is negative (since $0 < u < x \leq 1$, so $\log u \leq 0$). This means $\frac{\partial x}{\partial d} > 0$ (intuitively, fixing t and increasing d will increase the number of schools taken). Next compute

$$\begin{aligned}\frac{\partial^2 x}{\partial d \partial t} &= \frac{\partial}{\partial d}(1-x(d,t)^d) \\ &= -x^d \log x - d x^{d-1}\frac{\partial x}{\partial d} \\ &= -x^{d-1}(1-x^d)\left(\frac{x\log x + d\frac{\partial x}{\partial d}}{1-x^d}\right).\end{aligned}$$

Expressing the last multiplier as two integrals, we get

$$\begin{aligned}\frac{x\log x + d\frac{\partial x}{\partial d}}{1-x^d} &= \frac{x\log x}{1-x^d} - \int_0^x \frac{du^d \log u}{(1-u^d)^2}\,du \\ &= \int_0^x \left(\frac{1}{1-u^d} + \frac{du^d \log u}{(1-u^d)^2} + \frac{\log u}{1-u^d}\right)du - \int_0^x \frac{du^d \log u}{(1-u^d)^2}\,du \\ &= \int_0^x \frac{1+\log u}{1-u^d}\,du.\end{aligned}$$

We therefore have

$$\frac{\partial^2 x}{\partial d \partial t} = -x^{d-1}(1-x^d)\left(\int_0^x \frac{1+\log u}{1-u^d}\,du\right).$$

For $t > 0$, we have $-x^{d-1}(1-x^d) < 0$, which completes the proof. □

With this lemma, we can then prove Theorem 8 and Theorem 9 by bounding the appropriate integral, whose technicalities we defer to Appendix A.3. We quote these theo-

rems here and give a brief note on how we prove them.

Theorem 8. *For all $\ell > d \geq 1$ and for all $t \in (0,1]$, $x'_\ell(t) > x'_d(t)$.*

We prove Theorem 8 by showing that for $d \geq 1$ and $t \in [0,1]$,

$$\int_0^{x(d,t)} \frac{1+\log u}{1-u^d}\, du < 0.$$

This then, with Lemma 16 implies $\frac{\partial}{\partial d} x'_d(t) > 0$ so $x'_d(t)$ is strictly increasing in d and for any $\ell > d$, we have $x'_\ell(t) > x'_d(t)$ with $t \in (0,1]$ which completes the proof.

Theorem 9. *For all $d \geq 1$, there exists $c(d) > 1$ such that for all $\ell > d$ and $t \geq c(d)$, $x'_\ell(t) < x'_d(t)$. Furthermore, $c(d) \to 1$ as $d \to \infty$.*

Similarly, to prove Theorem 9, we show that for any $d \geq 1$ one can appropriately construct a $c(d) > 1$ such that $c(d) \to 1$ as $d \to \infty$ and for all $t \geq c(d)$,

$$\int_0^{x(d,t)} \frac{1+\log u}{1-u^d}\, du > 0.$$

Again, this shows that $x'_d(t)$ is strictly decreasing in d, so that for any $\ell > d$, we must have $x'_d(t) > x'_\ell(t)$ if $t \geq c(d)$ (since $c(d)$ is decreasing so also $t \geq c(\ell)$).

1.3.3 Discrete Market: Probability of Being Matched

In this section we prove the main results in the discrete market, often bringing back results from the continuous market via an application of Lemma 7.

The following theorem succinctly states that all students in a balanced market (where $i \leq n$) prefer longer lists for n large.

Theorem 1. *Let $d, \ell \in \mathbb{N}$ with $\ell > d$. For every n large enough and $i \leq n$, we have*

$$\mathbb{P}\left(M_i^{n,\ell} = 1\right) > \mathbb{P}\left(M_i^{n,d} = 1\right). \tag{1.1}$$

Proof. This follows directly from Theorem 8 and Lemma 7. □

Lemma 17. *Let $d, \ell, n \in \mathbb{N}$ with $d \leq \ell \leq n$. For every $j \in [n]$ and every $i = 1, 2, \ldots,$ we have:*

$$\mathbb{P}\left(H_j^{n,\ell}(i) = 1\right) \geq \mathbb{P}\left(H_j^{n,d}(i) = 1\right).$$

Proof. We have

$$\begin{aligned}
\mathbb{P}\left(H_j^{n,d}(i) = 1\right) &= \frac{1}{n}\sum_{r=1}^{n} \mathbb{P}\left(H_r^{n,d}(i) = 1\right) && \text{(by symmetry)} \\
&= \frac{1}{n}\sum_{r=1}^{n} \mathbb{E}(H_r^{n,d}(i)) \\
&= \frac{1}{n}\mathbb{E}\left(\sum_{r=1}^{n} H_r^{n,d}(i)\right) \\
&= \frac{1}{n}\mathbb{E}\left(T_i^{n,d}\right).
\end{aligned}$$

Indeed, the random variable $\sum_{r=1}^{n} H_r^{n,d}(i)$ counts the number of schools matched to a student in positions $\{1, 2, \ldots, i-1\}$ in the the discrete market with n schools, where every student has a preference list of length d. This equals the number of students from $\{1, 2, \ldots, i-1\}$ that are matched in that market, which is precisely $T_i^{n,d}$. Recall from (1.3) that

$$\mathbb{P}\left(M_i^{n,d} = 1 \mid T_i^{n,d} = k\right) = \begin{cases} 1 - \frac{\binom{k}{d}}{\binom{n}{d}}, & k \geq d, \\ 1, & \text{otherwise.} \end{cases}$$

It is clear from this expression that the left hand side monotonically increases in d: for all $i = 1, 2, 3, \ldots,$ a longer list length increases the probability that student i is matched to any school. Since $T_i^{n,d} = \sum_{\tilde{\imath}=1}^{i-1} M_{\tilde{\imath}}^{n,d}$ we deduce that $T_i^{n,\ell} \geq T_i^{n,d}$ in the sense of stochastic dominance, and therefore in expectation. □

We next state a lemma that provides explicit bounds for the rate at which student $t = 1$

gets matched to a school, which we prove in Appendix A.3.

Lemma 18. *For $d \geq 1$, we have $\left(\frac{2d+1}{4d+1}\right)^{1/d} \leq x(d,1) \leq \left(\frac{d+1}{2d+1}\right)^{1/d}$.*

Theorem 3. *Let $d, \ell \in \mathbb{N}$ with $\ell \geq d$. There exists a cutoff $c(d) > 1$ such that for every n large enough, and for all $i > n \cdot c(d)$,*

$$\mathbb{P}\left(M_i^{n,\ell} = 1\right) < \mathbb{P}\left(M_i^{n,d} = 1\right).$$

Furthermore, $c(d) \to 1$ as $d \to \infty$.

Proof. This follows directly from Theorem 9 and Lemma 7. □

We finally prove a corollary of Lemma 18 that is an intriguing and surprising result.

Theorem 4. *Let $d \in \mathbb{N}$. For every n large enough, we have*

$$\frac{d}{2d+1} \leq \mathbb{P}\left(M_n^{n,d} = 1\right) \leq \frac{2d}{4d+1}. \tag{1.2}$$

In particular,

$$\lim_{d\to\infty} \lim_{n\to\infty} \mathbb{P}(M_n^{n,d} = 1) = \frac{1}{2}.$$

Proof. From Lemma 7 and (1.4), we have $\mathbb{P}\left(M_n^{n,d} = 1\right) \to x_d'(1) = 1 - x_d(1)^d$. The claim then follows from the bounds in Lemma 18. □

1.3.4 Discrete Market: Impact on Rank

In this section, we prove the following lemma.

Lemma 19. *Let $d \in \mathbb{N}$. For every n large enough, $k \leq d$ and $i \leq n$, we have*

$$\mathbb{P}\left(K_i^{n,d} \leq k\right) - \mathbb{P}\left(K_i^{n,d+1} \leq k\right) \leq \left(\frac{d+2}{2d+3}\right)^{k/(d+1)} - \left(\frac{2d+1}{4d+1}\right)^{k/d}.$$

Proof. Fix $d \in \mathbb{N}$ and $k \leq d$. Similar to Lemma 12, it is easy to show that probability of getting matched to the first k schools is given by

$$\mathbb{P}\left(K_i^{n,d} \leq k \mid T_i^{n,d} = k\right) = 1 - (k/n)^k \pm O(d^2/n).$$

By Theorem 6, we have that $n^{-1}T_i^{n,d} \xrightarrow{\mathcal{P}} x_d(i/n)$, so

$$\mathbb{P}\left(K_i^{n,d} \leq k\right) - \mathbb{P}\left(K_i^{n,d+1} \leq k\right) \to x_{d+1}(i/n)^k - x_d(i/n)^k.$$

Consider the function $x_{d+1}(t)^k - x_d(t)^k$ for $t \in [0,1]$, and observe that

$$\begin{aligned}
\frac{\partial}{\partial t}\left(x_{d+1}(t)^k - x_d(t)^k\right) &= k(x_{d+1}(t)^{k-1}x'_{d+1}(t) - x_d(t)^{k-1}x'_d(t)) \\
&\geq k(x_d(t)^{k-1}x'_d(t) - x_d(t)^{k-1}x'_d(t)) \\
&= 0.
\end{aligned}$$

The second last line follows from Theorem 8 since $x'_{d+1}(t) \geq x'_d(t)$ and $x_{d+1}(t) \geq x_d(t)$ (from the former since $x_{d+1}(0) - x_d(0) = 0$).

We therefore have that $x_{d+1}(i/n)^k - x_d(i/n)^k$ is increasing in i, so applying Lemma 18, we have

$$\begin{aligned}
\max_{i \leq n}\left(x_{d+1}(i/n)^k - x_d(i/n)^k\right) &= x_{d+1}(1)^k - x_d(1)^k \\
&\leq \left(\frac{d+2}{2d+3}\right)^{k/(d+1)} - \left(\frac{2d+1}{4d+1}\right)^{k/d},
\end{aligned}$$

which completes the proof. □

1.4 Numerical Experiments

We perform numerical experiments to evaluate how well our results generalize beyond the one-to-one Serial Dictatorship mechanism with uniformly random preferences.

Schools not being sampled uniformly at random. In our first set of numerical experiments we investigate the impact of students not picking their preference list uniformly at random from all schools. In particular, we now assume there is some probability distribution $p : \{1, \dots, n\} \to [0, 1]$ with $\sum_{j=1}^{n} p(j) = 1$, where $p(j)$ dictates the probability that a student samples school j at their turn. The case of uniform sampling occurs when $p \equiv 1/n$.

We experiment with five different distributions for p: a uniform distribution (p_1), two types of Pareto-like distributions where students' preference of schools is concentrated at certain schools (p_2 and p_3 with low, and high concentration respectively), a setting of two classes (high demand and low demand) of schools (p_4), and finally a "degenerate" distribution where students sample from the first half of schools almost exclusively (p_5). Note that we would expect the last distribution to behave as if we had half as many schools, so that students with $i \leq n/2$ would prefer longer lists and those arriving sometime after this cutoff preferring shorter lists. The distributions are as follows:

$$p_1(j) = \frac{1}{n}, \qquad p_2(j) \propto \frac{1}{2 + j/n}, \qquad p_3(j) \propto \frac{1}{(1 + j/n)^{10}},$$

$$p_4(j) \propto 1 + 3\mathbb{1}_{\{j \leq 200\}}, \qquad p_5(j) = \begin{cases} 2/n - 1/100, & j \leq 500, \\ 1/100, & j > 500. \end{cases}$$

We perform all numerics with $n = 1000$ schools, and compare the cases of d being $1, 2, 4, 10$, and 20. All experiments are done with 100 000 repetitions.

Figure A.2 in Appendix A.4 shows the outcome of these experiments. We graph first the distribution of $p(\cdot)$ for all schools on the left, then the proportion of schools taken by students up to student i in the middle, and the probability that a given student i is matched to a school on the right.

Our main thesis—that all students in balanced markets prefer long lists—holds under

all distributions other than the degenerate one (which is almost equivalent to having half as many schools). In particular, $\mathbb{P}(M_i^{n,d} = 1)$ increases in d under all distributions other than the degenerate, and in that latter case, students soon prefer shorter lists as suggested by our Theorem 3. We conclude that our main results, Theorem 1 and Theorem 3, seem to be robust to i.i.d. sampling via reasonable non-uniform distributions.

Schools with multiple seats. In Section 1.2.4 and Appendix A.2 we discuss the extension of the discrete and continuous models to the case when schools have $q \in \mathbb{N}$ seats with $q > 1$. We performed numerical experiments to verify Conjecture 1.2.4. To do so, we computed numerical solutions to the differential equations describing the continuous model using the Euler method with step size 5×10^{-5}, and compared the probability of getting matched for successive values of $d \in \mathbb{N}$. We verified that the conjecture holds for all pairs of $q = 1, 2, \ldots, 20$ and $d = 1, 2, \ldots, 15$. This leads us to believe that the conjecture holds for all $q \in \mathbb{N}$ and $d \geq 1$. In Appendix A.4 we show some plots of the resultant output.

1.5 Conclusions

In this chapter, we investigated the impact of truncating preference lists in a two-sided matching market where students choose schools following a Serial Dictatorship order. Our main result is that if the market is balanced, all agents increase their probability of being matched when lists are longer, and if the market is not balanced students after the balance point quickly prefer shorter lists. These results are shown for preference lists that are sampled uniformly at random and school quotas equal to 1 and then shown to hold numerically for more complex preferences and larger quotas. We believe these valuable insights can be used to support the expansion of the length of preference lists, which are often set to small values. Investigating similar questions for other mechanisms is a relevant open question.

Chapter 2: Mitigating the Impact of Systemic Bias in School Choice

Joint work with Yuri Faenza, Swati Gupta, and Xuan Zhang.

2.1 Introduction

Disparity in opportunities plays a major role in access to education at different levels of the educational pipeline [38]. It is known that outcomes of middle school admissions dictate high school admissions, which in turn impact pathways to higher education [39]. Selection starts much earlier however, with gifted and talented programs screening students as young as 4 years old, but with these tests often seeing few students from ethnic minorities succeeding [40]. Our work is motivated by high school admissions in large public school districts such as New York City (NYC). NYC has an extensive public school system with current enrollment of over one million students, where every year roughly 80,000 students wish to join one of the 700 high school programs. By far, the most sought after public schools are the so-called Specialized High Schools (SHSs) which by law select candidates solely based on their score on the Specialized High School Admissions Test (SHSAT) [41]. Such scores are known to be impacted by socioeconomic status of students [42] and test preparation received in middle schools [39, 43]. Since ethnic minorities tend to cluster in middle schools of lower quality [44], they are already at a disadvantage in high school admissions, which is then reflected as under-representation in higher education programs [45]. The results is a massive filtering effect in high school admissions: 50% (resp. 80%) of students admitted to the SHSs come from only the top 5% (resp. 15%) of middle schools [39].

The goal of this work is to investigate data-driven interventions at the middle school

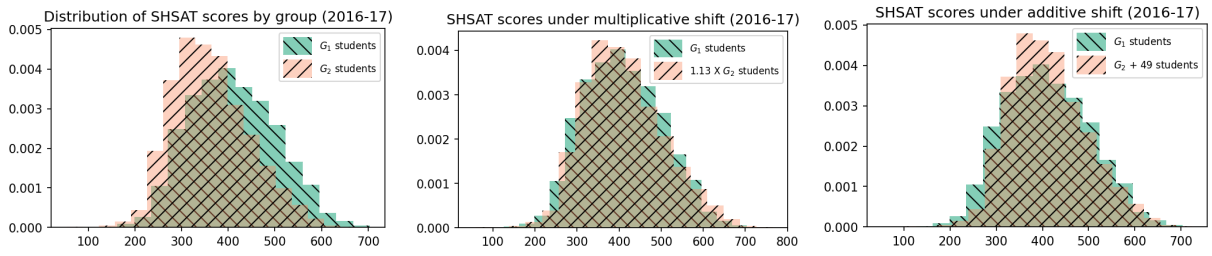
level to reduce this filtering effect. An extensive literature has focused on doing so by proposing changes to admissions policies themselves (see Section 2.1.3). That approach however, has multiple downsides. For one, simply "fixing" the admissions process to boost under-represented students does not fundamentally prepare those students to perform well once admitted. Another downside of such admissions policies is that they are seen as unfair by many, and there are significant political and legislative hurdles to implementing criteria that take the disparate backgrounds of students into account during the admissions process. For example, in 2003, an attempt by the University of Michigan to add 12 points for "diversity" on a 150 point scale in an effort to promote admissions of underrepresented ethnic minorities was met with a lawsuit, which was ultimately decided *not* in favor of the university [46]. Similarly, a 2019 plan supported by the then mayor of New York City to eliminate the SHSAT—criticized by some as inequitable due to unequal access to test preparation—failed to gain enough support, and was not approved by the New York State Senate [47]. A possibly more appropriate mechanism could therefore be for the city to afford free test preparation assistance to some under-represented students to help them compete more favorably with those students with ample resources.

In this work, we take a completely different operations perspective. We focus on centralized pre-admission interventions, a fundamentally meritocratic approach that does not involve an unfair or legally dubious change in the admissions criteria. We introduce a matching model of schools and students where some students (that we call *disadvantaged*) are not evaluated at their true potential, but at a strictly lower level. We then investigate both theoretically and empirically the impact of such differences in treatment, and investigate interventions to counter it. These interventions are in the form of *vouchers* targeted at certain disadvantaged students, affording them access to supplemental instruction: thereby providing them real support in order to perform closer to their innate ability. Our main contribution is a randomized policy for voucher allocation that is individually fair, incentive compatible and (by targeting average disadvantaged students)

can substantially reduce the mistreatment they experience, as measured by various metrics. We next present the setup, intermediate results, and experiments leading to our main contribution.

2.1.1 Motivation

In order to present our mathematical model, we first introduce the characteristics and mechanism for SHSs admissions in NYC. SHSs admit students uniquely based on the student's score on the highly competitive SHSAT. The NYC Department of Education (DOE) acknowledges that there is a disparity in students' abilities to prepare for the test, and so classifies some students as disadvantaged. This classification uses criteria such as their household income and the middle school they attended, which together constitute a proxy for socioeconomic status [48]. Following the DOE's definition, we divide students who take the SHSAT into two groups: non-disadvantaged (G_1) and disadvantaged (G_2). We find that the distribution of SHSAT scores[1] of the two groups (Figure 2.1a) exhibit a significant distributional shift, but match closely (as measured by Wasserstein distance) if the scores of disadvantaged students are adjusted by a multiplicative factor of $\frac{1}{\beta} \approx \frac{1}{0.88} \approx 1.13$ (Figure 2.1b), or an additive factor of $\gamma = 49$ points (Figure 2.1c).



(a) raw SHSAT scores (b) multiplicative shift (c) additive shift

Figure 2.1: Distribution of SHSAT scores for students in groups G_1 and G_2 for the 2016–17 academic year. Scores between the groups align closely after multiplicative shift of $\beta \approx 0.88$, or under an additive shift of 49 points.

While the exact mechanism of action and its causal factors are both unknown and

[1]We thank the NYC DOE for providing us this data under a non-disclosure agreement.

hotly debated in the literature[2], the consensus stands that performance gaps between socioeconomic groups stem not from differences in innate ability, but from disadvantages that hinder students' potential [49]. It is therefore natural to postulate that the distribution of innate ability ought to coincide when students are arbitrarily partitioned into groups based on their socioeconomic status.

Motivated by these observations and the literature on performance gaps between socioeconomics groups, we consider the following model. The *true potential* (unobserved innate ability) Z of a student is sampled from the Pareto distribution[3], while the *perceived potential* (observed performance) $\hat{Z}$ is equal to Z for G_1 students and to βZ for G_2, students where $\beta \in (0, 1)$ is some *bias factor*. We choose this bias model for its computational tractability and because it gives a good approximation to the SHSAT score distribution of admitted students (see Figure 2.2); we discuss the implications of using different models of bias in more detail in Section 2.6.

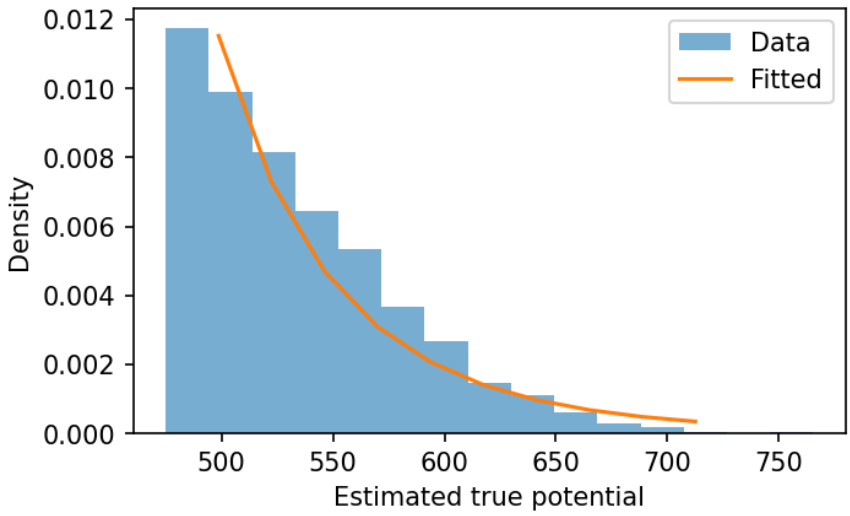


Figure 2.2: Distribution of estimated true potentials of students who score high enough to receive an offer from a SHS under the multiplicative model. The best fitting Pareto distribution has parameter $\alpha = 8.9$.

[2]Various mechanisms for the existence of this disparity have been debated and studied, and it is difficult to establish causality. We appreciate that this is a contentious issue and do not take a strong stance on it, but do note that implementers may adjust the bias factor and other model parameters to better fit their data and model of bias.

[3]The choice of the Pareto distribution to model potentials is inspired by a body of empirical work, see for example [50] on the achievements of individuals in many professions.

In our model, schools rank students based on their perceived potentials $\hat{Z}$. To be able to find tractable policies, we assume in our theoretical analysis that all students share the same ranking of schools (for example based on US News Rankings). This assumption abstracts out considerations that may be important for students such as proximity of a school [51], limits on the length of preference lists [9], or strong preferences of students for certain high schools[4]. However, we later argue experimentally, by dropping this assumption and applying it to the case of stable matching, that our qualitative results are robust to relaxations of our stylized model.

We then study the impact of interventions that we call *vouchers*, in the form additional resources (e.g. tutoring, test prep, or a scholarship) allocated to certain disadvantaged students. In our model, such a voucher enables a disadvantaged student to unlock their innate ability and perform at their true potential. We therefore refer to this process as *debiasing*. This model of intervention by additional resources is motivated by real world examples where supplemental instruction such as tutoring has been shown to be effective to close the performance gap, see for example [52]. However, such interventions are costly, and so they cannot be offered to everyone. This leads to a natural resource allocation question.

2.1.2 Our Contributions

We investigate the impact that bias has on both disadvantaged and non-disadvantaged students on two opposing measures of aggregate mistreatment, then quantify the effect that these interventions have on the mistreatment, and identify the best populations they should be targeted to. We discuss qualitative results in the real-world context of New York City Specialized High Schools by first estimating the parameters of our model from the data, and then evaluating the effect of various interventions on the actual stable matching outcomes. Our key findings are as follows:

[4]For instance in the 2016–17 SHSAT cohort, 56% of students indicated Stuyvesant or Brooklyn Tech as their first preference, with 76% naming at least one of the two in their top two preferences.

1. **Asymmetric impact and minority effect:** We observe that under reasonable assumptions on the parameters (such as disadvantaged students constituting a minority), the impact of bias on G_2 (disadvantaged) students is much bigger than the slight advantage that G_1 students obtain (see Figure 2.3 and Figure 2.7). Moreover, at a societal level, the presence of bias excludes most disadvantaged students from top schools (see Appendix B.3 and Figure B.1 therein), which explains a phenomenon often seen in the real world [53].

2. **Deterministic Centralized Interventions:** We next study the impact of deterministic voucher allocation. To measure the impact of bias on disadvantaged students, we define the *mistreatment* of a student as the difference in the ranking of the school the disadvantaged student gets matched to under bias compared to the unbiased setting. We study mechanisms to allocate vouchers to reduce aggregate measures of mistreatment, which we interpret as measures of (group) fairness. We first show that under two such very different measures, maximum benefit is achieved by providing vouchers to average-performing (rather than top-performing) disadvantaged students, assuming that their abilities are Pareto-distributed. These findings challenge existing scholarship/aid allocation mechanisms, addressing one of the key questions facing policy makers on how to distribute resources.

3. **Incentive Compatible Voucher Distribution:** We next observe that the deterministic allocation of vouchers to average performers creates an incentive for some top students to underperform. More generally, we show that the only deterministic policy that is incentive compatible distributes vouchers to top students. However, this policy has a small impact on reducing aggregate mistreatment. We therefore discuss two classes of mechanisms for randomly allocating vouchers that are incentive compatible. In particular, one of them (that we term *Proportional to Mistreatment*) still favors middle-performing students while guaranteeing that the maximum ex-

pected mistreatment is lower than under any deterministic policy. The other class of random policies we study is incentive compatible under general potential distributions. These policies have the additional benefit of being *individually fair* (in the sense of a Lipschitz condition) so that the probability of receiving a voucher for students with similar potentials is similar.

4. **Alternate Models of Bias:** We study variations of our simple multiplicative bias model, and its limitations in the context of the literature on discrimination. We quantify the impact of applying our model to contexts where bias arises due to a more sophisticated process, and discuss experimental results quantifying efficiency loss from such model misspecification. We then extend our results from the case of uniform multiplicative bias to the case of uniform additive bias. We show that our main takeaways continue to hold under an additive model or when the simple multiplicative model is applied under moderate levels of model misspecification.

5. **Experimental validation:** We then validate our theoretical results on admissions data to the SHSs for the academic year 2016–17. Although our model assumes that students have homogeneous preferences, we compute the stable matching using their SHSAT scores and reported heterogeneous preferences. We find that our key theoretical takeaways are still valid under relaxation of the homogeneous preference assumption: for instance, the shape of students' mistreatment resembles the prediction of our model (including the fact that average performers are the most mistreated) and that our voucher distribution program improves the mistreatment across the board. We further show that the best ranges of students to give vouchers to obtained via our theory, are qualitatively similar to the best empirically found ranges under the real data with heterogeneous preferences. This leads to our policy insights, which we discuss next.

6. **Policy Insights:** Motivated by the goal of maximizing the impact of limited resources, in this work we propose that additional training and resources should be offered to average performers rather than top performers. At a high level, the two assumptions that lead to this result are that (1) the concentration of students who perform around the average, compared to a much smaller cohort who make up the top performers, and (2) that given enough opportunities and support, the performance of the two cohorts of students ought to be indistinguishable. The key phenomenon that arises due to these assumptions is that a small deviation in an average disadvantaged student's perceived performance leads to a significant change in the rank of the school they are matched to.

   (1) is a key characteristic of many common distributions, including the Pareto distribution investigated in this paper. (2) is supported by education policy research and discrimination literature which shows that additional resources can positively impact low achieving student groups. Our work complements this line of work through mathematical analysis in order to optimally target limited resources.

   The current rationale behind most scholarship programs is to reward top performers, driven by a desire for meritocracy, and to drive better top performance by creating competition. Our analysis, on the other hand, suggests that if the goal is to maximize positive impact or more equitable outcomes for disadvantaged students, more support should be given to average performers. Moreover, this can be achieved with an incentive compatible randomized policy tailored to the distribution of potentials. However, in the case that truly nothing is known about the distribution of student potentials (so, in particular, condition (1) cannot be assumed), we show that the only policies guaranteed to be incentive compatible are those where the probability of receiving vouchers increases with the perceived performance.

The rest of the chapteris organized as follows. In Section 2.1.3 we review the prior liter-

ature related to this work. In Section 2.2, we formally introduce our mathematical model for a continuous matching market with bias. In Section 2.3 we analyze the effects of bias on both disadvantaged and non-disadvantaged students, introducing the key concepts of *displacement* and *mistreatment*. We then consider deterministic policies for reducing such bias via a centralized approach in Section 2.4, quantifying their impact on students and discussing various notions of (group) fairness. In particular, we present two theorems that quantify the optimal deterministic debiasing sets under different measures of fairness. In Section 2.5, we show that such deterministic policies fail to be *incentive compatible* and *individually fair*, and introduce the randomized assignment of vouchers to satisfy these fairness conditions, and show that such randomized policies achieve a lower maximum mistreatment than their deterministic counterparts. In Section 2.6 we discuss variations on models of bias, quantifying the impact of applying our stylized model to contexts with more intricate forms of bias, and extending our results on deterministic debiasing to an additive model. In Section 2.7 we apply our policies to the real-world dataset of SHSs admissions for the 2016–17 cohort, then close with a discussion in Section 2.8.

2.1.3 Related Work

The most common way to model admissions to schools is through a two-sided market, consisting of schools and students respectively, where each agent has an ordered preference over agents acceptable to them on the other side of the market. This model has been used to match doctors to hospitals by the National Residency Matching Program since the 1960s, and it has since gained widespread notoriety when [3] used it to reform the admissions process for New York City public high schools in 2003. Since then, admission decisions in NYC have been centralized and are (essentially) governed by the classical Gale-Shapley Deferred Acceptance algorithm [1]. The simplicity of the algorithm, as well as the drastic improvement in the quality of the matching it provides when compared to the pre-2003 method have led to academic and public acclaim, and spurred applica-

tions in many other systems (see, e.g. [54]). However, this mechanism does not naturally address problems like school segregation and class diversity, which have worsened and become more and more of a concern in recent years [55, 56, 57, 58]. The scientific community and policy-makers have reacted in various ways such as by modifying the mathematical model to incorporate group-specific quotas or proportionality constraints [59, 60, 61]. However, there is evidence that adding such constraints may even hurt the very students they were meant to help [62, 63, 64], and the question of legal challenges abounds.

There is a long line of work on affirmative action policies in theory and in practice [65, 66, 67, 68, 64, 38]; and alternatives such as the "top 10%" admissions criteria implemented in Texas [69]. Substitute mechanisms such as the top 10% criteria deviate significantly from current practice bar esoteric implementations, and it is unclear whether such criteria improve the status quo or worsen it. For instance [70] found a significant negative impact on the admissions rates of minorities if affirmative action policies for college admissions were replaced by top x% rules. We take a completely different approach to improving the outcomes for disadvantaged students by voucher distribution, which will naturally help with the downstream impacts in the education pipeline towards economic opportunities [71, 72].

Our work is somewhat related to an increasing body of work on test-taking itself, in particular policies that make tests optional or allow applicants to take tests multiple times. Test-optional policies are typically studied in the context of college admissions (see e.g. [73, 74]), but are largely not adaptable to public high school admissions due to significant differences in admissions dynamics[5]. A recent study in [75] shows the impact of being able to retake SAT exams and that reporting all the scores leads to more equi-

[5]For better or worse, admissions to the most sought after colleges are administered by large admissions offices staffed to evaluate nuanced applications, with significant autonomy and shielding from scrutiny. Almost all public school admissions processes on the other hand are deeply scrutinized by the public, and choose simple mechanisms with high explainability, up to the extreme case of SHS admissions being dictated by law. Further, colleges tend to be selective and have decentralized admissions processes, whereas public schools must provide education for all, through a centralized assignment process.

table outcomes as well as a more accurate signal for colleges[6]. Further, a recent study [77] focused on the design of a fair admissions process by identifying conditions where standardized tests should be dropped, while our paper mostly focuses on pre-admissions policies. We finally remark that our proposed interventions are politically and practically palpable as they do not require changing the admissions criteria (a significant hurdle to implementation [78]). In New York for instance, this would itself require changing state law (Hecht-Calandra Act).

Further, any admissions policy is susceptible to manipulation by applicants. Recent work by [79] has considered strategic behavior of students in a classification setting, where each student can expend some bounded amount of resources to improve their test-score performance and convert a “reject" decision to an “accept" decision. The school can provide subsidies to students to reveal their true potential. They further identify cases where providing a subsidy can make the group receiving the subsidy worse-off. Though our work considers a completely different model, we also find theoretical conditions under which voucher distribution can in fact worsen some fairness metrics over the disadvantaged groups, and investigate strategic behavior of students, see Section 2.4 and Section 2.5.1.

Various selection problems have been investigated in models with a multiplicative bias introduced by [80] (including [81, 82, 83, 84]), but to the best of our knowledge, this paper is the first to investigate it in the role of school admissions. There exists some literature such as [85, 86] on understanding the impact of family backgrounds on student preferences, but this is orthogonal to the questions we study here. Our work complements existing work in the education and policy literature which shows additional resources can positively impact low achieving student groups [87, 88], in particular, [52] studies the impact of supplemental instruction on disadvantaged students.

Lastly, our work is related to the modeling of bias and discrimination itself, which is

[6]The DOE allows students to take the SHSAT both in their 8th grade and in their 9th grade with slightly different tests, but only about 6% of test takers are in 9th grade [76].

an active area of research in economics [44, 89, 90]. The DOE is required by law to use only the SHSAT score to decide admissions to the SHSs, and so does not directly discriminate against any student based on group membership (since it does not take this information into account in the decision). Instead, the discrepancies in performance that arise in this setting are caused by pre-existing bias and discrimination at earlier stages in the educational pipeline, and so specifying the process by which such bias arises is relevant. Our model of uniform multiplicative bias bears some resemblance to what is known as *taste-based discrimination* in the literature: models where agents hold uniform dis-taste towards members of some group. In the modern world, many guardrails exist to disbar and discourage direct taste-based discrimination in consequential decisions such as in school admissions (e.g. based on "protected attributes"). Therefore, while taste-based discrimination models make some attempt to describe why bias or prejudice come to be, more modern work commonly takes bias to be primarily of the *statistical discrimination* kind [91]. While these models do not explain why initial bias has arisen, they posit that the processes that beget discrimination can be explained via economically rational behavior of individual agents [92]. In particular, such phenomena may arise even when agents do not directly imbue dis-taste towards some group, but due to distributional differences between groups. A common feature in such models is that the decision maker relies on some signal that is noisier and hence a less accurate predictor of the true characteristic of interest (such as innate ability) for discriminated populations [93]. For instance, if the variance of test scores of students in the disadvantaged group are higher than those of the non-disadvantaged group, then a decision maker facing two students with the same above average score but different group membership may be inclined to choose in favor of the non-disadvantaged student, rationally believing their score to be a more accurate predictor of their performance (and the disadvantaged student's score more likely to be a fluke). We refer the reader to the survey in [94]. We address such bias processes with higher noise in Section 2.6 and Appendix B.7.

2.2 A continuous matching market

We now introduce our stylized matching model for school choice. For tractability, we follow a recent trend in the literature and assume both schools and students to be continuous sets (see Appendix B.1 for a discussion on this choice). We denote the set of students by Θ and for each student $\theta \in \Theta$, endow them with a *true potential* $Z(\theta)$ sampled from some probability distribution. We interpret this true potential as the innate ability of student θ. For the rest of this section, we assume $Z(\theta) \sim \text{Pareto}(1, \alpha)$ (note that all students then have true potentials exceeding 1). This assumption is relaxed in Section 2.5.1 where we consider randomized voucher programs under different distributional assumptions. We occasionally identify a student's true potential $Z(\theta)$ with θ, the student itself.

We denoted the mass of schools by the unit interval, $[0, 1]$, and assume that all students rank schools in the same order with school 0 the best, and school 1 the worst. Schools on the other hand rank students uniquely based on their potential. This conveniently lets us identify the matching with the complementary cumulative distribution function (ccdf) of the students' potentials. That is, let μ denote the matching when schools rank students according to their true potentials, then student θ gets matched to school $\mu(\theta) = 1 - F(Z(\theta))$ where $F(t)$ is the cumulative distribution function (cdf) of the distribution of student potentials. For convenience, we denote the ccdf by $\bar{F} = 1 - F$, so $\mu(\theta) = \bar{F}(Z(\theta))$. In the Pareto case, we get $\mu(\theta) = Z(\theta)^{-\alpha}$.

In our model, however, not all students are perceived at their true potential. Instead, we consider the case where a proportion $p \in [0, 1]$ of students is disadvantaged and they perform at a level lower than their innate ability due to some kind of *bias*. Formally, we consider the student body to be composed of two groups: a proportion $1 - p$ of *non-disadvantaged* students G_1, and a proportion p of *disadvantaged* students G_2. We let $\hat{Z}(\theta)$ denote the *perceived potential* of student $\theta \in \Theta$. While students in G_1 are perceived at their true potential, we now assume that the perceived potential of disadvantaged students are

biased by a constant multiplicative factor $\beta \in (0, 1]$. That is, if $\theta \in G_2$ then $\hat{Z}(\theta) = \beta Z(\theta)$; otherwise, if $\theta \in G_1$ then $\hat{Z}(\theta) = Z(\theta)$.

Let F_1 and F_2 be the cdfs for the perceived potentials of students in G_1 and G_2, respectively, then

$$F_1(t) = 1 - t^{-\alpha}; \qquad F_2(t) = 1 - \beta^{\alpha} t^{-\alpha}.$$

Note that the domain of F_1 is $[1, \infty)$, whereas the domain of F_2 is $[\beta, \infty)$. We now consider the matching of students to schools under this biased regime (when schools use perceived potentials to rank students), which we denote by $\hat{\mu}$. For a student $\theta \in \Theta$, $\hat{\mu}(\theta)$ is equal to the mass of students whose perceived potentials exceed $\hat{Z}(\theta)$, and one can compute

$$\hat{\mu}(\theta) = \begin{cases} (1-p)\bar{F}_1(\hat{Z}(\theta)) + p\bar{F}_2(\hat{Z}(\theta)) & \text{if } \theta \in G_1, \\ (1-p)\bar{F}_1(\hat{Z}(\theta) \vee 1) + p\bar{F}_2(\hat{Z}(\theta)) & \text{if } \theta \in G_2, \end{cases} \tag{2.1}$$

where $\vee$ is the maximum operator. Note that when $\beta = 1$ (no bias), (2.1) equals $\mu(\theta)$ for all $\theta \in \Theta$.

Formally, we define a *matching* in this market to be a surjective measurable function $\gamma : \Theta \to [0, 1]$, such that the mass of students mapped to a measurable set of schools $S \subseteq [0, 1]$ coincides with the standard Lebesgue measure ν of S. That is, any surjective function γ from Θ to $[0, 1]$ is a matching if

$$\nu(\gamma^{-1}(S)) := (1-p) \int_{\theta \in \gamma^{-1}(S) \cap G_1} dF_1(\hat{Z}(\theta)) + p \int_{\theta \in \gamma^{-1}(S) \cap G_2} dF_2(\hat{Z}(\theta))$$

is equal to the standard Lebesgue measure of S for all measurable $S \subseteq [0, 1]$. One can easily check that μ and $\hat{\mu}$ defined above are matchings.

Example 20. *Suppose $\alpha = 3$ so student scores are sampled from Pareto$(1, 3)$. Suppose Maya $\in G_2$ scores $Z(\text{“Maya”}) = 1.4$, and Lisa $\in G_1$ scores $Z(\text{“Lisa”}) = 1.3$. In the unbiased setting, Maya*

gets matched to schoool $\bar{F}(Z(\text{“}Maya\text{”})) = 1 - (1 - 1/(1.4^3)) \approx 0.3644$ *while Lisa gets matched to* $\bar{F}(Z(\text{“}Lisa\text{”})) = 1 - (1 - 1/(1.3^3)) \approx 0.4552$. *On the other hand, in the biased case with* $\beta = 0.9$, *we have* $\hat{Z}(\text{“}Maya\text{”}) = 1.26$, *while* $\hat{Z}(\text{“}Lisa\text{”}) = 1.3$. *If* $p = 0.2$, *we have that Maya and Lisa are matched to schools*

$$\hat{\mu}(\text{“}Maya\text{”}) = 0.4729 \quad \textit{and} \quad \hat{\mu}(\text{“}Lisa\text{”}) = 0.4305, \tag{2.2}$$

respectively to a significantly worse (slightly better) school than they used to in the setting without bias. Note that Lisa has a smaller true potential than Maya but is assigned to a better school in the biased setting.

2.3 Impact of Bias on Students

Our first goal is to understand how much bias affects agents in the market. In particular, we would like to quantify the loss of efficiency for students[7] when students are matched to schools under $\hat{\mu}$ instead of μ. Formally, we define the displacement and mistreatment of a student as follows.

Definition 21 (Displacement and mistreatment)**.** *Let* $\theta \in \Theta$, *let* μ *be the matching in the absence of bias (using* $Z(\theta)$*), and let* γ *be some other matching. We define*

1. *the* displacement *of* θ *under* γ *as* $\text{disp}_\gamma(\theta) = \gamma(\theta) - \mu(\theta)$*; and*
2. *the* mistreatment *of* θ *under* γ *as* $m_\gamma(\theta) = \max(0, \gamma(\theta) - \mu(\theta))$.

We often drop the subscript when the matching at hand is clear from context.

Note that if $\theta \in G_1$, the displacement under $\hat{\mu}$ is non-positive, and if $\theta \in G_2$, it is non-negative. The displacement for $\hat{\mu}$ can be calculated using the formulae for μ and $\hat{\mu}$ in (2.1).

[7] In Appendix B.3, we take the schools' perspective and show that there is effectively no loss of efficiency for schools under this model, creating little incentive for them to intervene at the individual school level. We also measure there the diversity of the admitted cohort in our model.

Proposition 22. *For any student $\theta \in G_2$, the displacement under $\hat{\mu}$ is given by:*

$$\text{disp}_{\hat{\mu}}(\theta) = \begin{cases} (1-p)\,(Z(\theta))^{-\alpha}\,(\beta^{-\alpha}-1) & \text{if } Z(\theta) \geq \frac{1}{\beta}, \\ (1-p)\,(1-(Z(\theta))^{-\alpha}) & \text{if } Z(\theta) \leq \frac{1}{\beta}. \end{cases}$$

For any student $\theta \in G_1$, we have $\text{disp}_{\hat{\mu}}(\theta) = -p\,(1-\beta^{\alpha})\,(Z(\theta))^{-\alpha}$. Thus, the maximum displacement of $(1-p)(1-\beta^{\alpha})$ is experienced by a G_2 student with true potential $1/\beta$; and the most significant negative displacement of $-p(1-\beta^{\alpha})$ is experienced by a G_1 student with true potential 1.

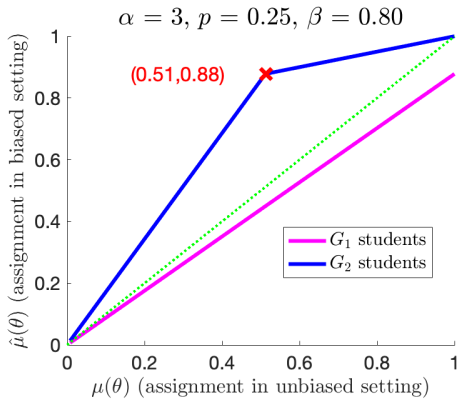


Figure 2.3: Schools students are matched to under $\hat{\mu}$ and μ. The green dotted line is a line of slope one, representing the place a student should be placed if *there were no bias in the system*.

One can interpret this result intuitively as follows. Starting from the top school, G_1 students gradually take up more seats than they would without bias, and thus gradually push G_2 students to worse schools. This process stops once all G_1 students are matched to schools, and the only students that remain to be matched are G_2 students. As a result, in the lowest ranked schools, all students are G_2 students. Hence, the difference in ranks of the schools G_2 students are matched to decreases towards the end. Figure 2.3 gives a pictorial illustration of Proposition 22. From there, one can clearly see how the most mistreated students are average performers. This intuition will be fundamental in devising policies to counter the effect of bias.

2.4 Deterministic Centralized Interventions

In this section, we discuss deterministic interventions, where the central planner is able to assign vouchers to some targeted subset of disadvantaged students to *debias* them in order to reduce mistreatment. When a disadvantaged student is chosen for such a voucher (in the form of supplemental instruction, test prep, a scholarship, or similar), we assume that they receive the support they need to realize their innate ability and perform at their true potential by the time their performance is measured (e.g. when they take the SHSAT). Such interventions are costly, so deciding which set of students to assign these vouchers to in order to mitigate bias as much as possible subject to some budget constraint becomes a key question. These policies are deterministic, and the decision of whether to assign a voucher to a student (hence, debias them) depend only on the potential of the students. In the next section, we will discuss randomized policies where the decisions will also depend on the outcome of a random coin flip.

Deterministic Debiasing Sets: Formally, the planner chooses some measurable subset $T \subseteq [1, \infty)$ of disadvantaged students to debias, which we call a *deterministic debiasing set* (DDS). We let $\hat{c} \in [0, 1]$ be the budget of the central planner, then define $\mathcal{T}(\hat{c})$ to be the set of all DDSs respecting this budget, that is the set of all measurable $T \subseteq [1, \infty)$ with $\int_T dF_1 \leq \hat{c}$. Additionally, let $\mathcal{T}^c(\hat{c}) \subseteq \mathcal{T}(\hat{c})$ be the set of such sets that are also closed and connected (i.e. they are intervals $[a, b] \subseteq [1, \infty)$ that satisfy $F_1(b) - F_1(a) \leq \hat{c}$). If $\theta \in T$, then we set $\hat{Z}(\theta) = Z(\theta)$, so that student performs at their true potential. Let $\mu_T : \Theta \to [0, 1]$ be the matching after G_2 students whose true potentials lie in T have been debiased[8]. Write m_T for the mistreatment under μ_T as in Definition 21. The mistreatment is the drop in the rank of the school the student is matched to (if this drop is positive): a student θ has mistreatment equal to 0 if they are assigned to a school at least as good as $\mu(\theta)$. In

[8]In this section we assume the debiasing decision is based on *true potentials*. In the case of deterministic one-to-one bias, the distinction is not important, but we later discuss debiasing on perceived potentials when it becomes relevant.

the following, we evaluate a voucher distribution by its effect on the mistreatment of G_2 students, since only G_2 students may experience strictly positive mistreatment. It is easy to see that after the interventions, no student $\theta \in G_1$ will be matched to a school worse than $\mu(\theta)$. This is because our interventions focus on helping (certain) G_2 students reveal their true potentials, hence for any G_1 student θ, no student with potential lower than $Z(\theta)$ can have a perceived potential higher than $\hat{Z}(\theta) = Z(\theta)$.

Fairness Considerations: Finding a set of students to allocate vouchers to is a resource allocation problem with natural fairness considerations that guide the choice of the measures to be optimized[9]. For the cohort of disadvantaged students, we take the view of finding a distribution of vouchers so that the mistreatment across G_2 students is as balanced or equitable as possible. We analyze two representative fairness measures in this regard: (1) the *positive area under the mistreatment curve* over all disadvantaged students, and (2) the *maximum mistreatment* experienced in this cohort. The former is the continuous L^1 norm (under the F_1 measure) of the mistreatment after voucher allocation, or the *positive area under the curve* (PAUC), which we denote by σ. The latter is the continuous L^∞ norm of mistreatment and we denote it by mm. Formally, for a matching γ, we define

$$\sigma(\gamma) := \int_\Theta m_\gamma \, dF_1 = \|m_\gamma(\theta)\|_1, \tag{2.3}$$

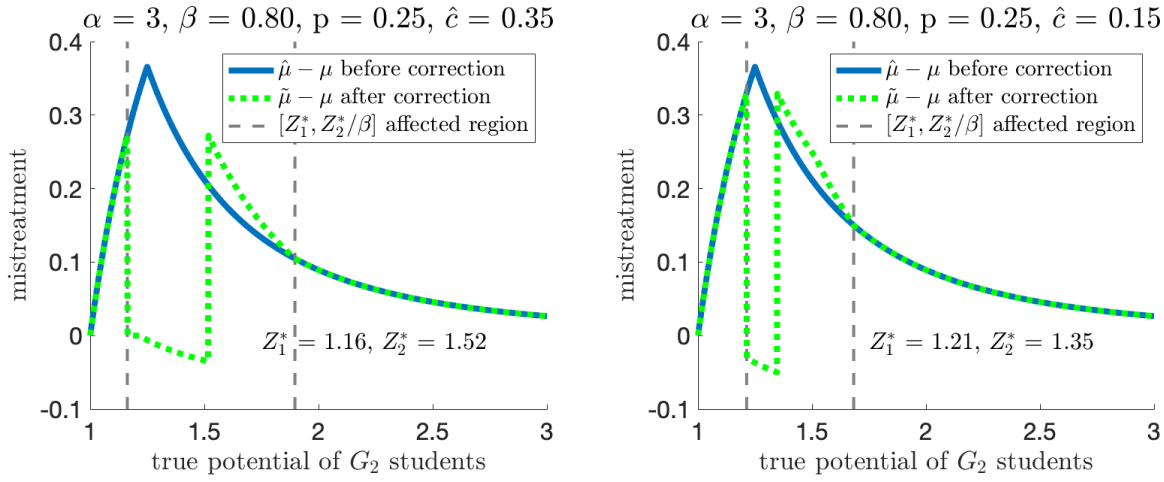
$$mm(\gamma) := \sup_{\theta \in \Theta} m_\gamma(\theta) = \lim_{p \to \infty} \left(\int_\Theta \left| m_\gamma \right|^p \, dF_1 \right)^{1/p} = \|m_\gamma(\theta)\|_\infty. \tag{2.4}$$

These notions of fairness have been axiomatically established and are well-studied in the literature. For example, the *min-max* notion of fairness has been considered in [96], whereas the positive area under the curve corresponds to average mistreatment of group G_2: it is a group notion of fairness consider in many fairness related studies [97, 98, 99]. Since we will show that the optimal interventions at the two extremes L^1 and L^∞ target

[9]To read a more detailed philosophical discussion on relevant philosophies of equality and decision-making, we refer the interested reader to the 1979 Tanner Lectures on Human Values ([95]).

qualitatively similar sets of students, we expect the solution for any other L^p norm to also behave similarly[10], and so restrict our analysis to L^1 and L^∞.

Optimal Deterministic Strategies: We now proceed to proving our main results in the deterministic settings, which fully describe the optimal debias intervals in our model. We show in Figure 2.5 how much the two fairness measures can be improved as a function of the budget $\hat{c}$. We define $\mathcal{T}_{mm}(\hat{c}) = \arg\min_{T \in \mathcal{T}(\hat{c})} mm(\mu_T)$, in other words, $\mathcal{T}_{mm}(\hat{c})$ is the collection of sets T that minimize $mm(\mu_T)$ among sets with $\int_T dF_1 \leq \hat{c}$. The next result gives an exact characterization of $\mathcal{T}_{mm}(\hat{c})$, assuming[11] $p < 1 - \beta^\alpha$.



(a) optimal debiasing when $\hat{c}$ is large

(b) optimal debiasing when $\hat{c}$ is small

Figure 2.4: Maximum mistreatment before and after optimal voucher allocation.

Theorem 23. *Assume $p < 1 - \beta^\alpha$. Then there exists a set $T = [Z_1^*, Z_2^*] \in \mathcal{T}_{mm}(\hat{c})$ such that all other sets in $\mathcal{T}_{mm}(\hat{c})$ differ from T on a set of measure zero. If $\hat{c} \geq \frac{(1-p)(1-\beta^\alpha)}{1-p+1-\beta^\alpha}$, then*

$$Z_1^* = \left(\frac{(1-p) + (\frac{1}{\beta^\alpha} - 1)\hat{c}}{\frac{1}{\beta^\alpha} - p} \right)^{-\frac{1}{\alpha}} \quad \text{and} \quad Z_2^* = \left(\frac{(1-p)(1-\hat{c})}{\frac{1}{\beta^\alpha} - p} \right)^{-\frac{1}{\alpha}},$$

[10] An L^p norm on a probability space with p small generally measures the average of a function, whereas a large p measures its "peakiness", with $p = \infty$ equaling the essential supremum, and values in between trading off between these properties (for a further discussion on the relationship between L^p spaces, see [100]).

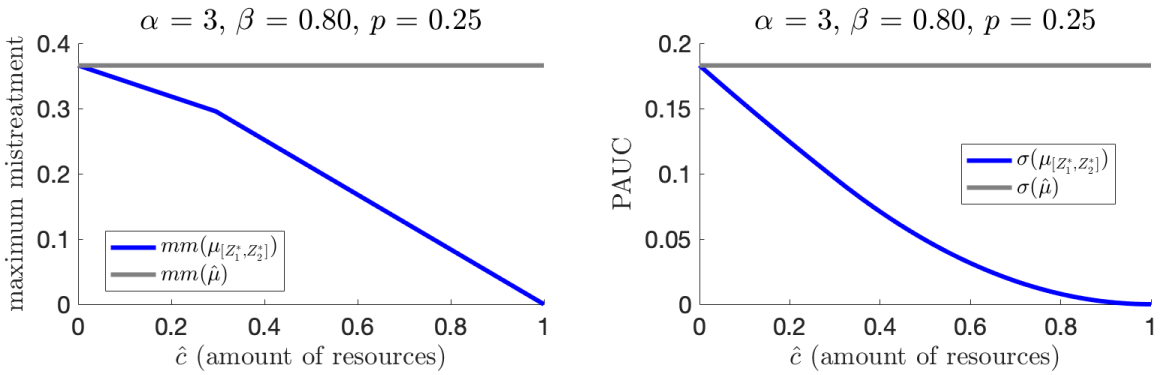
[11] We refer to the end of Section 2.5.1 for a discussion on the various technical assumptions on data from Section 2.4 and Section 2.5.1

and $mm(\mu_{[Z_1^*,Z_2^*]}) = (1-p)(1-\beta^\alpha)\frac{1-\hat{c}}{1-p\beta^\alpha}$, *reduced from* $mm(\hat{\mu}) = (1-p)(1-\beta^\alpha)$. *Conversely, if* $\hat{c} \leq \frac{(1-p)(1-\beta^\alpha)}{1-p+1-\beta^\alpha}$, *then:*

$$Z_1^* = \left(\frac{(1-p-\hat{c})\beta^\alpha}{1-p} + \hat{c}\right)^{-\frac{1}{\alpha}} \quad \text{and} \quad Z_2^* = \left(\frac{(1-p-\hat{c})\beta^\alpha}{1-p}\right)^{-\frac{1}{\alpha}},$$

and $mm(\mu_{[Z_1^*,Z_2^*]}) = (1-p-\hat{c})(1-\beta^\alpha) + p\hat{c}$.

We include the proof of Theorem 23 in Appendix B.4. Interestingly, our proof also shows that if vouchers are not distributed carefully, one may actually *increase* the maximum mistreatment and, more generally, shows which debiasing sets lead to an improvement over the status quo. A pictorial representation of Theorem 23 is given in Figure 2.4. The two sub-figures correspond to two choices of $\hat{c}$. Moreover, Figure 2.5a shows how much $mm(\mu_{[Z_1^*,Z_2^*]})$ decreases as the budget, $\hat{c}$, increases.



(a) maximum mistreatment

(b) PAUC

Figure 2.5: Effect of bias after debiasing the optimal set of G_2 students given budget $\hat{c}$.

We next consider the positive area under the curve (PAUC): this is the aggregate amount of mistreatment experienced by all G_2 students. In this case, we restrict our attention to debiasing G_2 students whose potentials are in a connected set—this is a justifiable implementation in practice (otherwise a student might feel fairly treated given that someone with a better potential as well as someone with a worse potential receives the voucher). This assumption also makes our analysis more tractable. In particular, let

$\mathcal{T}^c_{auc}(\hat{c}) = \arg\min_{T\in\mathcal{T}^c(\hat{c})} \sigma(\mu_T)$ be the set of $T \in \mathcal{T}^c(\hat{c})$ that minimize $\sigma(\mu_T)$. The next result gives an explicit description of these sets when assuming, again that $p < 1 - \beta^\alpha$ and additionally that $p < 0.5$.

Theorem 24. *Assume $p < 1 - \beta^\alpha$ and $p < 0.5$. Then $\mathcal{T}^c_{auc}(\hat{c})$ is made up of a unique set $T = [Z_1^*, Z_2^*]$. If $\hat{c} \geq \frac{(1-p)(1-\beta^\alpha)}{2-p-\beta^\alpha-p\beta^\alpha+p\beta^{2\alpha}}$, then:*

$$Z_2^* = \left(\frac{(1-p)(1-\hat{c})}{p\beta^\alpha + \frac{1}{\beta^\alpha} - 2p}\right)^{-\frac{1}{\alpha}} \quad \textit{and} \quad Z_1^* = \left(\frac{(1-p)(1-\hat{c})}{p\beta^\alpha + \frac{1}{\beta^\alpha} - 2p} + \hat{c}\right)^{-\frac{1}{\alpha}},$$

and $\sigma(\mu_{[Z_1^*,Z_2^*]}) = \frac{1}{2}(1-p)(1-\beta^\alpha)\left(\frac{(\frac{1}{\beta^\alpha}-p)(1-\hat{c})^2}{p\beta^\alpha+\frac{1}{\beta^\alpha}-2p}\right)$, down from $\sigma(\hat{\mu}) = \frac{1}{2}(1-p)(1-\beta^\alpha)$. Otherwise, if $\hat{c} \leq \frac{(1-p)(1-\beta^\alpha)}{2-p-\beta^\alpha-p\beta^\alpha+p\beta^{2\alpha}}$, then:

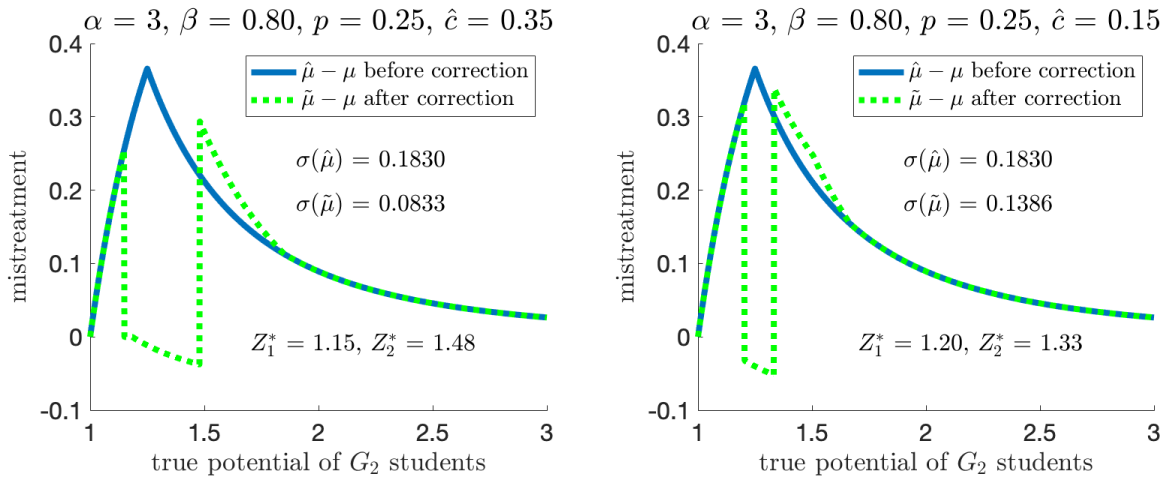
$$Z_2^* = \left(\frac{(p\beta^\alpha - 1)\hat{c} + (1-p)}{(1-p)\frac{1}{\beta^\alpha}}\right)^{-\frac{1}{\alpha}} \quad \textit{and} \quad Z_1^* = \left(\frac{(p\beta^\alpha - 1)\hat{c} + (1-p)}{(1-p)\frac{1}{\beta^\alpha}} + \hat{c}\right)^{-\frac{1}{\alpha}},$$

and $\sigma(\mu_{[Z_1^*,Z_2^*]}) = \frac{1}{2}(1-p)(1-\hat{c})^2 - \frac{1}{2}\beta^\alpha\left(\frac{[(p\beta^\alpha-1)\hat{c}+(1-p)]^2}{1-p} + p\hat{c}^2\right)$.

The proof of Theorem 24 is given in Appendix B.5, a pictorial representation is presented in Figure 2.6. The two sub-figures again show two different choices of $\hat{c}$. Figure 2.5b shows how much $\sigma(\mu_{[Z_1^*,Z_2^*]})$ decreases as $\hat{c}$ increases.

In Table B.2 of the e-Companion, we compare the optimal ranges disadvantaged students to debias under the two measures of fairness, with parameters $\alpha = 3$, $\beta = 0.8$, and $p = 1/4$. We find on average a 95% overlap of the optimal intervals under the two measures of fairness. In particular, both measures suggest that vouchers should be given to the average (middle performing) students.

Although these results highlight an important deviation from the current practice of prioritizing top-performing students for scholarships, we highlight in the next section two fundamental problems with such policies.



(a) optimal debiasing when $\hat{c}$ is large

(b) optimal debiasing when $\hat{c}$ is small

Figure 2.6: PAUC before and after optimal voucher allocation.

2.5 Incentive Compatible and Individually Fair Voucher Distribution

In the last section, we characterized the *deterministic* intervals to distribute vouchers to in order to minimize either the maximum mistreatment or the positive area under the mistreatment curve for the disadvantaged student group. In this section, we introduce two natural and desirable properties that the policies developed in Section 2.4 fail to have. Once we recognize these flaws, we show in Section 2.5.1 how we can shift from a deterministic voucher distribution policy to a randomized one in order to satisfy them.

Our first property is *individual fairness*, which requires that similar individuals be treated similarly [101]. While a formal definition of this concept is postponed to Section 2.5.1, we observe here that the policies developed in Section 2.4 fail to be individually fair as individuals close to the boundary of the debiasing interval are treated very differently depending on whether they are inside or outside of it.

Our second property is *incentive compatibility* (see, e.g., [102]). In general, it requires that no individual can benefit from misrepresenting their features. In our setting, we assume that a student is able to misrepresent themselves as appearing to have lower perceived potential (e.g., intentionally achieving a lower score on a test) in order to be part

of the set of students that get allocated vouchers. Recall that a DDS is a measurable set $T \subseteq [1, \infty)$. A DDS is *incentive compatible* if no student is assigned to a better school if they misreport their performance. Formally, assume that a voucher given to a disadvantaged student with reported perceived potential $\beta Z(\theta)$, will improve their performance up to $Z(\theta)$[12]; then, a DDS T is incentive compatible if for each $x \in [1, \infty) \setminus T$ and $x' \in T$ with $x > x'$, we have $\beta x \geq x'$.

Lemma 25. *Assume $\beta \in [0, 1)$ and let $T \neq \emptyset$ be an incentive compatible DDS. Then T is of the form $\{\theta \in \Theta : Z(\theta) \geq \delta\}$ or $\{\theta \in \Theta : Z(\theta) > \delta\}$ for some value $\delta \in [1, \infty)$.*

We defer the proof to Appendix B.2. This lemma shows that, if we care about incentive compatibility and require that vouchers are distributed deterministically, then the only feasible mechanism is to debias all students that have potential above some cutoff δ (i.e. the top students). However, we showed in the last section that such policies are not optimal. To overcome these flaws in deterministic policies, we next turn to randomization.

2.5.1 Randomized assignment of vouchers

In this section (and in its proofs in Appendix B.6) we abuse notation and identify a student θ with their true potential $Z(\theta)$ for simplicity.

A *Randomized Voucher Program* (RVP) is a measurable function $\rho : \Theta \to [0, 1]$ that gives, for each $\theta \in \Theta$, the probability that a G_2 student with true potential θ is assigned a voucher. Observe that if $\rho(\theta) \in \{0, 1\}$ for all $\theta \in \Theta$, then $\rho^{-1}(1)$ is a measurable set and therefore also a deterministic debiasing set (DDS) as in the definition in Section 2.4; likewise, given a DDS T we can construct the RVP $\rho_T(\theta) = \mathbb{1}_{\{\theta \in T\}}$ that coincides with a given DDS.

[12]This assumption is justified by the fact that additional training is usually commensurate with the (perceived) level of a student.

The main class of RVPs investigated in this section are those we call *Proportional-to-Mistreatment* (PropM), denoted by ρ_m and defined as

$$\rho_m(\theta) := \frac{2\hat{c}}{(1-\beta^\alpha)(1-p)} m_{\hat{\mu}}(\theta), \tag{2.5}$$

for some $\hat{c} \in (0, 1/2]$ (recall that $m_{\hat{\mu}}(\theta)$ is the mistreatment of a student with real potential θ when no vouchers are distributed). It is easy to see that $\hat{c}$ is the expected proportion of disadvantaged students that will get a voucher, that is $\hat{c} = \int_\Theta \rho_m \, dF$. Intuitively, ρ_m assigns a larger probability of being debiased to students with a higher mistreatment.

As we show next, under broadly applicable technical hypotheses on the parameters, PropMs satisfy many of the properties that deterministic voucher allocations fail to have. Moreover, we will show that they can lower the maximum expected mistreatment. To state these results formally, we first extend concepts from deterministic DDSs to RVPs. We let $\mu_\rho(\theta)$ be the expected school that a student with true potential $\theta \in \Theta$ is assigned to under ρ. An explicit computation of μ_ρ for arbitrary ρ can be found in Lemma 112 of Appendix B.6. Now all prior definitions and notation carry over, including the mistreatment m_ρ, and maximum mistreatment mm_ρ.

An RVP ρ is *incentive compatible* if $\mu_\rho(\theta') \geq \mu_\rho(\theta)$ for all $\theta' < \theta$. That is, an RVP is incentive compatible if a student with true potential θ is not better off by manipulating themselves to appear as having a true potential $\theta' < \theta$.

We define individual fairness as a Lipschitz condition on ρ. We say an RVP ρ is k-*individually fair* if, for each $\theta, \theta' \in [1, \infty)$, $|\rho(\theta) - \rho(\theta')| \leq k|\theta - \theta'|$ (note that under this definition, no non-trivial DDS is k-individually fair for any k). We can now state the main result from this section, whose proof is deferred to Appendix B.6.3. Observe that $mm^*(\hat{c})$ is the maximum mistreatment achieved by the best deterministic policy minimizing this metric, as computed in Theorem 23.

Theorem 26. *Let ρ_m be a PropM defined as in* (2.5) *for some $\hat{c} \in (0, 1/2]$ and assume $p \leq 0.5$.*

Let $mm^*(\hat{c}) = \min_{T \in \mathcal{T}(\hat{c})} mm(\mu_T)$. *Then:*

1. ρ_m *is* $\frac{2\hat{c}\alpha}{1-\beta^\alpha}$*-individually fair.*

2. ρ_m *is incentive compatible for*

$$\hat{c} \leq \frac{1-p}{2\left[p(1-\beta^\alpha) + (1-p)(\beta^{-\alpha}-1)\right]}. \tag{2.6}$$

3. *Suppose* $p < 1-\beta^\alpha$ *and* $\hat{c} \leq \frac{(1-p)(1-\beta^\alpha)}{1-p+1-\beta^\alpha}$. *Then* $mm_{\rho_m} \leq mm^*(\hat{c})$ *if*

$$\hat{c} \geq 1 - \frac{p+1-\beta^\alpha}{4p(1-\beta^\alpha)}. \tag{2.7}$$

(2.6) and (2.7) give complementary conditions on the amount of vouchers that can be given out. On one hand, (2.6) suggests that distributing too many vouchers prevents incentive compatibility of the PropM. In fact, a $\hat{c}$ too large causes students performing just above the most mistreated student to be incentivized to artificially lower their score, as the absolute value of the derivative of the PropM becomes large around its maximum. On the other hand, (2.7) suggests that we need to distribute enough vouchers to see the maximum expected mistreatment mm_{ρ_m} drop below the optimal deterministic one $mm^*(\hat{c})$. This is because the optimal deterministic policy debiases the most mistreated student straight away whereas the PropM distributes vouchers more widely, and so the maximum expected mistreatment does not immediately drop as significantly. As we discuss at the end of the section, both conditions are satisfied for a large range of parameters.

PropMs represent therefore a more robust and theoretically satisfying alternative to reducing the maximum mistreatment that is at least as effective as the deterministic voucher assignments developed in Section 2.4.

We observe however, that to design a non-trivial incentive compatible RVP, it is essential to have knowledge of the distribution of student potentials. We say an RVP ρ is *Increasing-with-Potential* (IwP) if $\rho(\theta) \geq \rho(\theta')$ for all $\theta > \theta'$. An IwP assigns a higher prob-

ability of being debiased to students with higher potential. It can therefore be interpreted as a randomized counterpart of the DDS from Lemma 25 that allocated vouchers to the top performing students (in particular, the DDS from Lemma 25 is IwP).

General distributions of potentials: For the rest of this section, we relax the Pareto assumption and consider the more general version of the model defined in Section 2.2 where the true potentials of students are allowed to be drawn from any continuously integrable distribution F. All definitions naturally extend to this setting. We first show that under mild technical conditions, IwPs are incentive compatible with respect to any F. We prove this fact in Appendix B.6.4.

Lemma 27. *Suppose ρ is IwP and such that it is everywhere continuously differentiable except a countable set of isolated points where it has right-continuous jump discontinuities. Then, for any distribution of true potentials F, ρ is incentive compatible.*

The following theorem gives a converse to the previous statement and is also proved in Appendix B.6.4.

Theorem 28. *Suppose ρ is an RVP. Let $\theta \in [1, \infty)$ such that ρ is continuously differentiable in some neighborhood of θ but $\rho'(\theta) < 0$. Then, for any $\beta \in (0, 1)$ and $p \in (0, 1)$, there exists a continuous distribution F such that if true potentials are distributed according to F, ρ is not incentive compatible.*

Theorem 28 implies that without any information on the distribution of student potentials, the only voucher distribution policies that are guaranteed to be incentive compatible are those that allocate vouchers with higher probability to higher performing students. Examples of such policies are lotteries for students whose potential is above some certain threshold. Hence, if no information on the distribution of students potential can be assumed, it may be reasonable for policy-makers to stick to a more conservative distribution of vouchers which rewards top-performing students.

Discussion on technical assumptions: We now discuss the technical assumptions on the parameters of the model in Section 2.4 and Section 2.5. In Theorem 23, Theorem 24, and Theorem 26 we assume $p < 1 - \beta^\alpha$. Note that the right hand side is equal to $\mathbb{P}(F_2 \in [\beta, 1])$, that is, the proportion of disadvantaged students whose perceived potential is less than 1 (the smallest perceived potential of any non-disadvantaged student). The condition $p < 1 - \beta^\alpha$ therefore requires that the proportion of disadvantaged students out of the whole student population is no more than the proportion of disadvantaged students that are perceived as being worse than any non-disadvantaged student. In Theorem 24, we further assume $p < 0.5$, and in Theorem 26 we assume both an upper and a lower bound on $\hat{c}$. The conditions need to be checked and do not always hold, but we note that all conditions hold for many reasonable choices of $(\alpha, \beta, p, \hat{c})$. For instance, they hold if $\beta = .8$, $\alpha = 3$, $p < .4$ (as in Figure 2.5 and the Figure 2.4(b)), and $\hat{c} \leq 1/4$; or if $\beta = .9$, $\alpha = 8.9$, $p \leq 1/3$, and $\hat{c} \leq 1/4$ (as in our numerical experiments in Section 2.7).

2.6 Alternate Models of Bias

Until now we have studied a relatively simple model of uniform multiplicative bias that posits that a disadvantaged student θ with true potential $Z(\theta)$ is perceived at potential $\hat{Z}(\theta) = \beta Z(\theta)$; where $\beta \in (0, 1]$ is fixed. We now investigate deviations from this model.

Model Misspecification and Bias Processes: As we discuss in Section 2.1.3, our work touches on the questions of how bias and prejudice arises and propagates. While our model is analytically tractable, it does not necessarily capture the realities of bias. In Appendix B.7, we study the impact of various types of model misspecification on the predictions of our model. In particular, we look at two types of models. One class is those where the perceived potentials of disadvantaged students are noisier than those of the non-disadvantaged students (therefore resembling the statistical discrimination models

discussed in Section 2.1.3), either by adding noise to the bias factor, or by adding noise to the resultant perceived potential. The second class is additive models, where bias takes the form of an additive shift in perceived potentials. In both of these cases (and the case of a mixture), we empirically show that applying our simple multiplicative model yields close to optimal debiasing ranges, and conclude that our model is robust to such model misspecification.

An additive bias model: We next turn our attention to an additive model of bias, and extend some of the results in Section 2.4 to the case of additive (in place of multiplicative) bias. We consider the same setting as Section 2.2 where the multiplicative bias model was introduced, but now the perceived potential of a student is given by $\hat{Z}(\theta) = Z(\theta) - \gamma$ if $\theta \in G_2$ (and θ does not receive a voucher), with $\hat{Z}(\theta) = Z(\theta)$ otherwise. This then gives

$$F_1(t) = 1 - t^{-\alpha}, \qquad F_2(t) = 1 - (t+\gamma)^{-\alpha}, \tag{2.8}$$

with domains $[1, \infty)$ and $[1-\gamma, \infty)$, respectively. The expression given in (2.1) for the biased matching $\hat{\mu}(\theta)$ continues to hold under the additive definition of $\hat{Z}$, and this fact yields the following result, which is an analogue of Proposition 22.

Proposition 29. *Under additive bias of $\gamma \geq 0$, for any student $\theta \in G_2$, the displacement under $\hat{\mu}$ is given by:*

$$\mathrm{disp}_{\hat{\mu}}(\theta) = \begin{cases} (1-p)\left((Z(\theta)-\gamma)^{-\alpha} - (Z(\theta))^{-\alpha}\right), & \text{if } Z(\theta) \geq 1+\gamma, \\ (1-p)\left(1 - (Z(\theta))^{-\alpha}\right), & \text{if } Z(\theta) < 1+\gamma. \end{cases} \tag{2.9}$$

For any student $\theta \in G_1$, we have $\mathrm{disp}_{\hat{\mu}}(\theta) = -p((Z(\theta))^{-\alpha} - (Z(\theta)+\gamma)^{-\alpha})$. Thus, the maximum displacement of $(1-p)(1-(1+\gamma)^{-\alpha})$ is experienced by a G_2 student with potential $1+\gamma$; and the most significant negative displacement of $-p(1-(1+\gamma)^{-\alpha})$ is experienced by a G_1 student with potential 1.

Optimal Deterministic Debiasing: We next extend Theorem 23 which establishes the optimal debiasing interval under maximum mistreatment to the additive model. Let $\mathcal{S}^c(\hat{c})$ be the set of closed and connected subsets of $[1, \infty)$ such that for $S \in \mathcal{S}^c(\hat{c})$, $\int_S dF_1 \leq \hat{c}$. Similarly to the multiplicative case, we define μ_S to be the matching under the additive model when students in S receive vouchers. Let then $\mathcal{S}^c_{mm}(\hat{c}) = \arg\min_{S \in \mathcal{S}^c(\hat{c})} mm(\mu_S)$ be the collection of sets in $\mathcal{S}^c(\hat{c})$ that minimize maximum mistreatment. The following result, which we prove in Appendix B.8, is analogous to Theorem 23.

Theorem 30. *The set $\mathcal{S}^c_{mm}(\hat{c})$ consists of a unique set $S = [Y_1^*, Y_2^*]$ where Y_1^* and Y_2^* are computed as follows. $Y_2^* = \min\{U_1, U_2\}$ where U_1 is the positive solution to*

$$(1-p)(1-\hat{c}) = \mathbb{P}(F \geq U_1 - \gamma) - p\mathbb{P}(F \geq U_1), \tag{2.10}$$

and U_2 is the positive solution to

$$(1-p)(1-\hat{c}) = (1-p)\mathbb{P}(F \geq U_2 - \gamma) + p\hat{c}, \tag{2.11}$$

and $Y_1^* = (\hat{c} + (Y_2^*)^{-\alpha})^{-1/\alpha}$.

We remark that it is possible to similarly compute the optimal DDS under the PAUC measure. In the proof of Theorem 30, we give an expression for the mistreatment for a given student, and one can then integrate this against F_1 to compute PAUC. One easily argues that this expression has a global minimum, but the resultant expressions are not amenable to analytical computations, so we do not include this result.

Multiplicative vs Additive Models: We close by noting that we fitted both the multiplicative as well as the additive model to our real data, and as measured by both Wasserstein distance and KL-divergence, the multiplicative model did fit it better, which is why we focus heavily on that model as our main object of study in this paper.

2.7 Experimental Case Study

Our theoretical analysis has shown that mistreatment under various metrics can be substantially reduced via a targeted intervention tailored to the distribution of student potential. We now use data from NYC with real test scores and a student population with heterogeneous preferences over schools to compute optimal policies for reducing student mistreatment. We show that our theoretical model provides a reasonable approximation despite some deviations from the data, and importantly, that our qualitative results continue to hold. Our theoretical analysis is therefore instrumental in identifying effective debiasing policies for real-world applications, and can be optimized empirically for real data.

Dataset: There are eight Specialized High Schools (SHSs) in NYC which are consistently ranked among the best schools in the city. Admission to these schools is highly competitive, and is determined solely by the score a student achieves on the SHS Admissions Test (SHSAT). An intake of only about $5,000$ students gets selected every year from a pool of $29,000$ applicants who take the test. We apply our model to the dataset of 2016–17 academic year SHS admissions which include for each student their SHSAT score, their preference list over the SHSs, and whether the DOE deems them disadvantaged or not.

Pre-processing & Model Fitting: As in Section 2.1.1, we estimate the distributional shift in scores between the two groups (see Figure 2.1), then assume that reversing this shift gives the innate ability for each student, and use this scaled score as the (in reality unobservable) true potential. In our dataset, we fit $\beta = 0.882$ for a multiplicative model and $\gamma = 49$ points for an additive model[13]. We take the original scores to be the perceived potentials, and the scaled scores to be the true potentials: that is, for G_1 students, we always use the raw SHSAT score, and for G_2 students with raw score s in the dataset, we

[13]We choose parameters that minimize the Wasserstein distance between the two distributions. The additive shift becomes $49/475 = 0.103$ after normalization.

use s/β (or $s + \gamma$) as their true potential if they receive a voucher (i.e. are debiased) and s otherwise.

In this section, all matchings on real data are computed as stable matchings with the student-proposing deferred-acceptance algorithm, using the true preferences of students, and with schools choosing students based solely on their perceived score. This mimics closely the real SHS admissions process. We extend the definition of displacement in the natural way: as the difference in the ranking of the school a student is assigned to (in their own preference list) between the matching at hand and the matching that uses the estimated true potentials. Due to the heterogeneity of student preferences, the displacement for a given student may be positive, negative, or zero regardless of if they receive a voucher or not. Because of this variability, we use the positive area under the mistreatment[14] curve (PAUC) measure exclusively to compare interventions as it averages out this effect.

Our theoretical model assumes a balanced market, whereas only a small number of those who apply to SHSs are admitted. We therefore discard those students whose score is lower than a cutoff of 475 points, whom we compute to not receive admissions in any case[15]. This right tail of the student scores now closely matches the Pareto distribution (after dividing by the minimum score) with $\alpha = 8.856$ in the multiplicative case (see Figure 2.2) and $\alpha = 9.315$ in the additive case. Of these students, we compute the proportion that are disadvantaged as $p = 0.319$ for the multiplicative case and $p = 0.300$ for the additive case[16]. These normalizations yield a subset of students that form a balanced market and whose distributional properties approximate our model well.

Model Characteristics: We first observe empirically that without intervention, all G_1 students (magenta dots in Figure 2.7) have non-positive displacement and all G_2 students

[14]Recall mistreatment is the non-negative part of displacement.

[15]We compute this cutoff by performing a stable matching using the true potentials and rounding down to the nearest 5 points the score of the last student that gets admitted in this matching.

[16]Because of the different distributional assumption, a different number of G_2 students end up above the cutoff in the two different models.

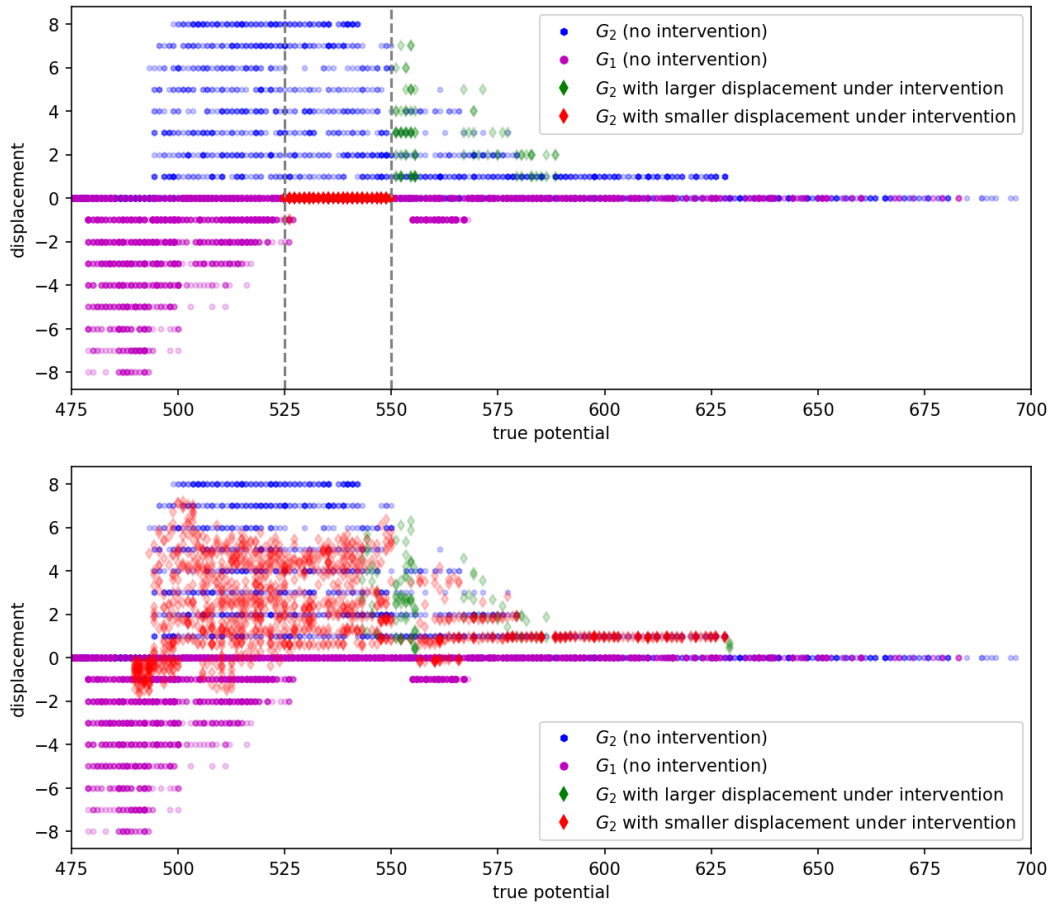


Figure 2.7: The displacement of G_1 and G_2 students in the SHSAT dataset from NYC DOE. Blue and magenta dots respectively show the displacement of disadvantaged and non-disadvantaged students when there is no intervention. In the top figure, the students in the debiased range (dashed lines) are offered vouchers, in the bottom figure vouchers are offered with probability given in Figure 2.8. In the bottom figure with the randomized voucher program, we plot the average displacement over 100 repetitions. In both figures, we plot the displacement of disadvantaged students whose assigned schools change, with red dots representing those going to more preferred schools and green dots for those going to less preferred schools.

(blue dots) have non-negative displacement, as predicted by our analysis in Section 2.3. Furthermore, we consider deterministic and randomized interventions with $\hat{c} = 0.17$. In Figure 2.7 (top), deterministic vouchers are offered to students between the two dashed lines. All G_2 students that receive vouchers (red dots) have a displacement of at most zero, but some G_2 students might (green dots) fare worse, particularly the ones who are scoring slightly higher than the range to which vouchers are offered, as they are overtaken by some other G_2 students just below them. This highlights the non-incentive compatible nature of such deterministic policies (such students have incentive to underperform).

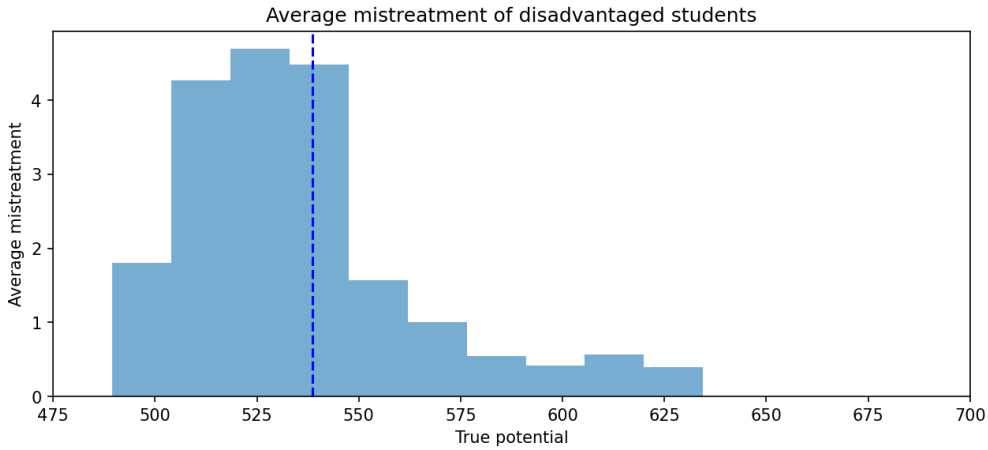


Figure 2.8: The average bucketed mistreatment of students as computed from empirical data. Note the peak at 525, which represents an average student (cf. Figure 2.4, the curve representing mistreatment before intervention). The vertical line indicates the value of $1/\beta$ where the theoretical PropM debiasing policy would have its maximum probability of assigning a voucher.

Applying the randomized voucher program to this dataset requires further modifications. As observed earlier, since students have heterogeneous preferences, their mistreatment is also heterogeneous. In the worst case two students with the same score may have different mistreatment. To produce an empirical PropM (see Figure 2.8), we divide the potentials of admitted students into 20 equally sized buckets and compute an average mistreatment within each such bucket. The PropM is then normalized to be proportional to this average mistreatment, with magnitude determined by the budget of vouchers available. Since PropM is a randomized voucher program (such that a given student might get a voucher in one realization but not another), we run this experiment 100 times with different seeds and take the average displacement (see Figure 2.7, bottom). Our experiments show that the maximum mistreatment is reduced compared to the deterministic allocation, and more generally, mistreatment improves across the board indicating a more equitable outcome. Note that due to the heterogeneity of preferences and binned averaging, PropM is not in fact incentive compatible, as indicated by the G_2 students with larger displacement under intervention (green diamonds). However, unlike the deterministic debiasing procedure, students with an incentive to underperform are interspersed with students with no incentive to underperform, making it harder for students to game the

system ex-ante. Moreover, it is not uncommon that some theoretically incentive compatible mechanisms exhibits in practice some lack of incentive compatibility. For instance, the NYC School Match mechanism curtails the preference lists of students to at most 12 schools [3], incentivizing students to be at least partially strategic.

Theoretical vs. empirical intervals: We next compare theoretically optimal intervals with those found to be empirically optimal. The basic primitive in our analysis is the routine that is given an interval of students to debias and computes the PAUC. By computing actual stable matchings and their PAUC, we produce a reliable benchmark to compare both theoretical and empirical intervals. We use this routine in a grid search to find the empirically optimal debiasing intervals.

We consider two cases: a low budget regime ($\hat{c} = 0.1$) and an abundant budget regime ($\hat{c} = 0.4$). For both cases, we compute the theoretically optimal intervals by applying our theorems to the parameter values fitted earlier. Table 2.1 shows the optimal ranges found both via our theory and via empirical grid search. The differences between the theoretical and empirical models are minor, with the biggest difference being under the additive assumption. As earlier remarked, this is also the less-well fitting model, but even then, the intervals are qualitatively close as opposed to debiasing top students.

Model	Range	$\hat{c} = 0.1$	$\hat{c} = 0.4$
multiplicative	theoretical	[526.73, 547.19]	[506.84, 582.99]
multiplicative	empirical	[527, 543]	[505, 561]
additive	theoretical	[516.94, 530.91]	[499.82, 558.24]
additive	empirical	[529, 544]	[508, 567]

Table 2.1: Comparison of optimal ranges of students to offer vouchers to, obtained empirically and theoretically (based on our formulas), under two different budgets.

Overall, we find further evidence of reasonability of our assumptions since the empirical results on real-world data match the optimal target distribution of students predicted by our assortative model.

2.8 Discussion

The qualitative takeaways from our work speak to a much ingrained systemic problem that limits access to opportunities—how can one understand the impact of bias on societal practices and systematically account for biases in the real world? Indeed, resources available for meaningful interventions in an existing system are limited, and there is resistance to change in the form of lawsuits and pushback to changes in admissions policies that attempt at a more equitable process. Thus, our focus is on understanding the impact of minimally invasive use of targeted resources, as opposed to changing the matching mechanism itself.

From our analysis, we are able to highlight several qualitative properties using simple models of bias applied to matching mechanisms:

1. **Disparate Impact:** The disparity in admissions is experienced much more by the disadvantaged group of students, compared to the marginal advantage for the rest.

2. **Interventions:** A carefully-designed randomized voucher distribution program can counter some of the effects of bias, while also being incentive compatible and individually fair. We further showed empirically that our qualitative results remain unchanged when applied to a real-world dataset (the SHS admissions process in New York).

3. **Resources:** Additional resources centrally distributed to slightly above-average students overall in the system (such as top performers in schools with high economic need index) would maximally impact measures of group fairness. Targeting resources at students based on their performance provides an important lever for policy makers to improve fairness.

These takeaways are a first step, and in no way address all the systemic problems in school admissions process—such as access to counselors, transport to schools or fa-

milial support towards education. But they do help us understand the most impacted student groups, and provide a mathematical basis to policymakers to make changes to allocation of public funds. We have shared the results of this work and are in discussions with the Department of Education of New York City. Further, our analysis leads to open questions such as theoretically optimal interventions under other structured student preferences and qualitative analyses when the distribution of student potentials is not Pareto distributed.

Chapter 3: On Quota-Filling and Kuhn Choice Functions

Joint work with Yuri Faenza, and Benjamin Rubio.

3.1 Introduction

In classical stable matching theory, the preferences of every agent are specified by a strict ordering over their acceptable partners on the other side of the market[1]. When the matching is *one-to-one*—meaning that each agent gets matched to at most one partner on the other side—such a list fully specifies the preferences an agent may have. Due to the simplicity and elegance of this model, strict orderings have also been extensively studied in the context of *one-to-many*, and *many-to-many* matching, where agents may be matched to multiple partners on the other side. In these cases, however, one can gain significantly in generality and expressive power by allowing agents to specify more granular preference via *choice functions*. A choice function is a function that maps each set of offered partners to the subset of those that the agent would prefer being matched to. We discuss the formal definition in Section 3.2.

Choice functions need only satisfy the conditions of *substitutability* and *consistency* for a stable matching to exist and be found via Roth's generalization of the Deferred Acceptance algorithm [103]. While choice functions represent a formally elegant extension in theory, practical use is significantly hindered by various limitations on information exchange, a phenomenon that is particularly prevalent in large markets. For this reason, few real-world applications explicitly utilize choice functions in stable matching. For instance, many school districts—such as the one in New York City—only allow schools to

[1]This means that the agent compiles an ordering of the agents on the other side, omitting any that they do not find acceptable. While results are most elegant when the list is taken to be strict, there is also a robust body of work on ties and tie-breaking.

express their preferences over students via strict preference lists. This is in contrast to the increasing desire of schools to assemble balanced cohorts of students from diverse backgrounds, which has been shown to produce positive effects on student outcomes [104, 105]. Strict preference lists are then a poor substitute for expressing preference in such markets. For instance, we show that there exists a choice function such that when approximated by any preference list, the chosen set of students will coincide in no more than one student in the worst case. Finding tractable ways for schools to indicate their preference beyond preference lists is therefore an important research direction.

In this chapter, we investigate choice functions in the context of school choice, in particular those that satisfy the property of being *quota-filling*. Such a condition is natural when schools may not waste resources by rejecting a student unless they are able to admit a more desirable one. We show that quota-filling choice functions cannot be used in practice under the common *offline model* of stable matching because the Deferred Acceptance algorithm fails to terminate in polynomial time. We then propose a class of *Kuhn choice functions* that are amenable to real world use and possess many desirable properties. We present several results on the approximability of different types of choice functions, establishing a hierarchy of quota-filling choice. These theoretical results are complemented by computational results where we explore Kuhn and non-Kuhn choice functions over small cohorts further.

3.1.1 Contributions

We investigate the gap between the elegant theory of choice functions present in the literature and the practical constraints of large real-world matching markets like school choice. Our work focuses on quota-filling choice functions, which are natural in this setting. We present results on the following topics:

1. The limitations of choice functions in the offline matching model. We begin by showing how general quota-filling choice functions are not viable in practice under what we

call the *offline model* of stable matching, used by most real-world matching markets. More precisely, we show that the number of quota-filling choice functions is doubly exponential in the number of partners on the other side of the market, which shows that representing or communicating such choice functions is computationally infeasible for any reasonably-sized market. This may explain why many school districts resort to the simpler, less expressive models that use strict preference lists.

2. Kuhn choice Functions as a practical alternative. The unwieldiness of general choice functions motivates the search for a more compactly representable, yet expressive subclass of choice functions. We propose *Kuhn choice functions* for this purpose, which arise from maximum-weight matchings in an auxiliary bipartite graph, a construction that lends itself to several desirable properties. We show that Kuhn choice functions are highly interpretable and can be represented compactly, making them ideal for markets where these properties are paramount, such as school choice.

3. Results on the Kuhn recognition problem. We show that while Kuhn choice functions possess desirable properties, recognizing whether an arbitrary quota-filling choice function is Kuhn or not, is itself a highly non-trivial task. We present results for the case of a school with two seats, and in general show that deciding whether a given choice function is Kuhn has high query complexity.

4. Approximation and the hierarchy of choice functions. Motivated by the fact that determining whether a given preference system belongs to a specific class can be difficult, we turn to the question of approximation. We analyze the relationships between Kuhn choice functions and other natural classes, including responsive preferences and the entire class of quota-filling choice functions. We establish a formal hierarchy of these function classes by proving a series of approximability and inapproximability results. Our results formalize the expressive power lost or gained when moving between models.

5. Computational results. We discuss practical aspects of using choice functions for stable matching via a number of computational results. We present a Mixed-Integer Pro-

gram (MIP) to decide whether a given quota-filling choice function is in the class of Kuhn choice, discuss various representations of choice functions, then close by investigating the number of choice functions of various types for small parameter sizes, and discussing a few particular cases of non-Kuhn choice functions.

3.1.2 Organization of the Chapter

The rest of this chapter is organized as follows. We begin in Section 3.2 by introducing the theory of stable matching with choice functions, and defining many of the standard classes of choice functions considered in the literature. In Section 3.3.1 we discuss the offline model of stable matching as used by many school choice programs, and highlight the practical constraints with communication complexity in such settings. This motivates our introduction of Kuhn choice functions in Section 3.3.2 where we discuss their properties and the problem of recognizing whether a path-independent, quota-filling choice function belongs in this class. In Section 3.3.3 we introduce our notion of approximation for choice functions, and construct the hierarchy of choice via various results on the approximability and inapproximability of choice functions. Proofs of our main results appear in Section 3.4 with short commentary. In Section 3.5, we complement our theoretical results with a discussion on the computational aspects on applying choice functions in stable matching. We close the chapter in Section 3.6 by arguing that markets such as the New York City public school choice market should adopt the more expressive class of preference systems captured by Kuhn choice functions due to their many desirable properties.

3.2 Preliminaries of Choice Functions

In this section, we introduce choice functions as an extension of preference lists in the stable matching problem. Here we take X to be a set of students, and consider the choice function describing the preferences of a single school in a school choice model. We will

discuss the market as a whole later.

Definition 31 (Choice function)**.** *A choice function C on a ground set X is a set function $C : 2^X \to 2^X$ such that $C(S) \subseteq S$ for all $S \subseteq X$. Given a set of offered partners S, the choice function outputs $C(S)$, the subset of partners that the agent prefers.*

Two important regularity properties of choice functions are *substitutability* and *consistency* [106].

Definition 32 (Substitutability)**.** *C is substitutable if for all $S \subseteq X$ and $T \subseteq S$,*

$$C(S) \cap T \subseteq C(T).$$

Equivalently, C is substitutable if for all $S \subseteq X$, if $b \in C(S)$ and $T \subseteq S$, then $b \in C(T \cup \{b\})$.

In other words, if b is chosen out of S, then it will be chosen out of any subset of S where it is available. If this condition is met, we say that students are *substitutes* to the school. It rules out *complementarity*, where a school might prefer two students, say Jesse (j) and Nicole (n), together but not individually. In such a case we would have $C(\{j, n\}) = \{j, n\}$ but $C(\{j\}) = \emptyset$, which violates substitutability. In school matching students rarely exhibit this type of coupling, but it is more common in other contexts such as assembling a team of athletes. Jesse and Nicole in that case might have a long history of playing together and make a great pair on the team by complementing each other, yet may each be lackluster independently.

Definition 33 (Consistency)**.** *C is consistent if for all $S \subseteq X$, and $T \subseteq S$,*

$$C(S) \subseteq T \implies C(S) = C(T).$$

Consistency guarantees that removing elements that were not chosen from S does not change the choice $C(S)$. This condition is also known as the property of *irrelevance of rejected contracts* in the literature. Another key property is *path-independence* [106].

Definition 34 (Path-independence)**.** *C is path-independent if for all $S, T \subseteq X$,*

$$C(S \cup T) = C(C(S) \cup T).$$

Path-independence is also referred to in the literature as being *Plottian* due early work in [107]. The next lemma is well known in the literature, stating that a choice function is substitutable and consistent if and only if it is *path-independent*. We provide a proof in Appendix C.1 for completeness.

Lemma 35. *A choice function C is path-independent if and only if it is substitutable and consistent.*

In the context of public school choice, schools are considered a public resource that ought to not be wasted. This means that they should accept any student that applies as long as they have seats to spare. This motivates the definition of a *quota-filling* choice function [108].

Definition 36 (Quota-filling)**.** *A choice function C is* quota-filling *with* quota *q if for all $S \subseteq X$,*

$$|C(S)| = \min\{|S|, q\},$$

that is, $|C(S)| = q$ if $|S| \geq q$, and $C(S) = S$ otherwise.

Quota-filling choice functions with quota q are also often referred to as q-accepting or q-acceptant choice functions in the literature. Another well known result is that consistency always holds for a quota-filling choice function that satisfies substitutability. Again, a proof appears in Appendix C.1.

Lemma 37. *Suppose C is quota-filling and substitutable. Then it is consistent.*

From now on, we assume all choice functions are path-independent and quota-filling unless otherwise stated.

Matching and Stability: We next introduce the concept of stability in the setting of one-to-many matching for school choice, where schools have preferences given by choice functions [103]. We consider a set of students X and a set of schools S. We assume each school $s \in S$ has a capacity of q *seats*. For simplicity of exposition, we assume all schools have equal capacities, but all results continue to hold in the case of heterogeneous capacities. Students are assumed to have a (strict) preference list over acceptable schools S, and the preference of schools is assumed to be described by a q-quota-filling and path-independent choice function. For convenience, we formalize this structure and call it a matching market.

Definition 38 (Matching market)**.** *We say $\mathcal{M} = (X, S, \{\succ_x\}_{x \in X}, \{C_s\}_{s \in S})$ is a matching market, where X is a set of students, S is a set of schools, every student $x \in X$ is endowed with a strict preference order $\succ_x$ over a subset of schools (those the student deems acceptable), and every school $s \in S$ has a q-quota-filling choice function C_s over students in X for some $q \in \mathbb{N}$.*

A matching for a matching market is an assignment of students to schools.

Definition 39 (One-to-many matching)**.** *A function $\mu : X \cup S \to 2^{X \cup S}$ is called a (one-to-many)* matching *for the matching market $\mathcal{M} = (X, S, \{\succ_x\}_{x \in X}, \{C_s\}_{s \in S})$ if for all $x \in X$ and $s \in S$,*

1. *$\mu(x) \in S \cup \{\emptyset\}$,*
2. *$\mu(s) \subseteq X$, and*
3. *$\mu(x) = s$ if and only if $x \in \mu(s)$.*

We abuse notation and write $\mu(x)$ for the sole element when $x \in X$. We say an agent $a \in X \cup S$ remains unmatched under μ if $\mu(a) = \emptyset$.

In order to state the definition of a stable matching, we first define a *blocking pair*, that blocks a matching from being stable.

Definition 40 (Blocking pair)**.** *Given a matching μ for market $\mathcal{M} = (X, S, \{\succ_x\}_{x \in X}, \{C_s\}_{s \in S})$, we say a student-school pair (s, x) is a* blocking pair *for μ if both the student and the school would prefer being matched to each other than the partners they were assigned to in the matching. Formally (s, x) is a blocking pair if*

1. $s \succ_x \mu(x)$*, and*
2. $x \in C_s(\mu(s) \cup \{x\})$.

If a blocking pair exists for μ in $\mathcal{M}$, we say μ admits *a blocking pair, and the pair* blocks *the matching.*

Finally we define a stable matching as one that is individually rational for all agents and is absent of blocking pairs.

Definition 41 (Stable matching)**.** *A matching μ is stable for the market $\mathcal{M} = (X, S, \{\succ_x\}_{x \in X}, \{C_s\}_{s \in S})$ if the following hold:*

1. *(Individual rationality for students.) For all $x \in X$, either $\mu(x) = \emptyset$ or $\mu(x)$ is an acceptable school for x.*
2. *(Individual rationality for schools.) For all $s \in S$, $C_s(\mu(s)) = \mu(s)$.*
3. *(No blocking pairs.) μ admits no blocking pairs in $\mathcal{M}$.*

Intuitively, for agents to participate in a given matching market, they must have a guarantee that the resultant matching produced by the mechanism respects their preferences. In particular, the first two conditions of stability require that the matching respects the preferences of agents by not matching them to undesirable partners. The last condition, requiring the absence of blocking pairs, guarantees that agents are not incentivized to deviate from the matching, which also motivates calling such a matching stable.

Stability is a stringent condition, and in particular, is not defined by who is matched to whom, but rather by who is not matched to whom. It is therefore not obvious that such

matchings ought to exist in the first place. A seminal result in the field is the guaranteed existence of stable matchings via the *Deferred Acceptance* algorithm.

The Deferred Acceptance Algorithm: The following algorithm is an extension of the classical Gale-Shapley [1] algorithm that allows one to find a stable matching in one-to-many markets defined with choice functions [109].

Algorithm 42 (Deferred Acceptance)**.**

Input:

A matching market $\mathcal{M} = (X, S, \{\succ_x\}_{x \in X}, \{C_s\}_{s \in S})$.

Procedure:

1. *(Initialization.) Set every student* $x \in X$ *to be unassigned.*

2. *(Proposal.) Each unassigned student* $x \in X$ *applies to their most preferred school that has not yet rejected them.*

3. *(Deferred Acceptance.) Each school* $s \in S$ *considers the* T *of students that have currently applied to the school. It tentatively accepts the set of students* $C_s(T)$ *given by its choice function, and rejects all other students in* $T \setminus C_s(T)$.

4. *(Loop Condition.) If any student was rejected in the last step, such a student becomes unassigned, and the algorithm returns to the proposal stage.*

5. *(Termination.) If no student was rejected, the algorithm terminates. Every school* $s \in S$ *accepts the students currently tentatively accepted, and we set* $\mu(s)$ *equal to this set. For student* $x \in X$*, we set* $\mu(x)$ *equal to the school they were accepted to, or* $\emptyset$ *if they are unassigned.*

Output:

A matching μ *for* $\mathcal{M}$.

The next theorem states that when applied to a market with path-independent choice functions, the algorithm produces a stable matching. For a proof, we refer the reader to [103, 110, 111].

Theorem 43 (Deferred Acceptance)**.** *Applying Algorithm 42 to a matching market* $\mathcal{M} = (X, S, \{\succ_x\}_{x \in X}, \{C_s\}_{s \in S})$ *where* C_s *is path-independent for every* $s \in S$ *yields a stable matching* μ *for* $\mathcal{M}$*. The algorithm terminates in time polynomial in the number of agents and time required to evaluate a choice function. Furthermore, for every other matching* μ' *for* $\mathcal{M}$*, every student* $x \in X$ *weakly prefers* μ *to* μ'*.*

We make a few remarks on this algorithm and its implications. First note that if the choice function of a school were not substitutable, then that school may reject some student that they would later prefer when offered in conjunction with another one. It is therefore clear why substitutability is required.

Note also that the result itself is quite remarkable: not only does it guarantee that a stable matching can always be found in polynomial time when choice functions are readily available, but the last property implies that this matching is preferred by all students.

It is no coincidence that when the students propose to schools at each round, the output is optimal for students. Indeed, if the schools proposed to students[2], we would find a (possible different) matching such that it would be optimal for schools. Stable matchings have rich structure and the sets of stable matchings between these two extrema can be efficiently enumerated. This rich structure may be exploited to implement many efficient algorithms over stable matchings. While many such results are well known in the case that strict preference lists are used, recent work in [112] shows that such structure exists also when preferences are defined by quota-filling choice functions.

[2]Although in this work we are mostly interested in the one-to-many student-proposing version of Deferred Acceptance, both the algorithm and the proof may be naturally extended to the case where schools propose to students, and the case of many-to-many matching where every agent has preferences given by choice functions.

Elementary classes of choice functions: We now define two simple classes of choice functions that are built from preference lists.

Definition 44 (q-responsive choice function)**.** *Given a strict preference list $\succ$ over X and a quota q, we define the* q-responsive *choice function $C_\succ$ over X as $C_\succ(S) = \max_\succ(S, q)$ for all $S \subseteq X$ (where the notation $\max_\succ(S, q)$ denotes the top-q elements of S according to $\succ$).*

Applying Algorithm 42 to a market $\mathcal{M}$ where all choice functions are responsive yields exactly the classic Deferred Acceptance algorithm of [1]. Choice functions therefore generalize strict preference lists. Another simple class of choice functions that is strictly more general than the responsive class is that of lexicographic choice [113].

Definition 45 (Lexicographic choice function)**.** *Given q strict preference lists over X, denoted $\{\succ_i\}_{i=1,\dots,q}$, the* lexicographic choice function *C is defined as follows. For $S \subseteq X$, we let $C_1 = \max_{\succ_1} S$, and let $C_i = \max_{\succ_i} S \setminus C_{i-1}$ for $i = 2, \dots, q$, finally, we let $C(S) = C_q$. In other words, we think of dividing the school into q ordered seats, each seat at its turn using a preference list to choose its favorite remaining student.*

It is straightforward to show that both q-responsive and lexicographic choice functions are path-independent and quota-filling. By choosing all the preference lists for a lexicographic choice function to coincide, we recover exactly a q-responsive choice functions. Every responsive choice functions is therefore also lexicographic.

3.3 Models and Results

In this section, we present our main theoretical results.

3.3.1 The Offline Model of Stable Matching

As per Theorem 43, the Deferred Acceptance algorithm runs in time polynomial in the number of agents and the time taken to evaluate a choice function. Most works in the literature assume what we call the *oracle model*, where the central planner may query

any choice functions at any point during the algorithm in $O(1)$ time. In this model, the Deferred Acceptance procedure clearly terminates in polynomial time.

However, in the real world, such an assumption is often much too strong. In the context of school choice for instance, school districts generally utilize an *offline model*. In this model, all agents first compile their preferences, then communicate them to the central planner before the algorithm is executed. The central planner then proceeds to compute the stable matching offline with no further interaction with the agents during the algorithm execution. For the Deferred Acceptance algorithm to terminate in polynomial time under this offline model, we therefore require that each choice function can be written down and communicated to the central planner in polynomial time.

This model is very common in the real world, especially in the case where the matching occurs rarely, the market is large, or it takes significant time for agents to compile their preferences. For instance, the New York City public high school matching currently uses the offline model, with each school submitting a q-responsive choice function (where q is the number of seats per school) to the Department of Education (DOE). It would be impossible in practice for the DOE to coordinate with schools to solicit their preferences during the matching process (which in practice is run on a computer in a matter of seconds).

The following somewhat informal lemma shows how the size of the respective class of preference is intimately tied to the space and time required to communicate them, and therefore the running time of the Deferred Acceptance algorithm in the offline model.

Lemma 46. *Let $\mathcal{X}$ be a finite set with cardinality $|\mathcal{X}|$. Any deterministic communication scheme that can uniquely identify any element $x \in \mathcal{X}$ must use $\Theta(\log |\mathcal{X}|)$ bits in the worst case, and this bound is achievable.*

Counting arguments of this nature are well known in theoretical computer science, we provide a brief proof in Appendix C.1. It is not hard to see that the number of responsive

choice functions over any set of students X is at most[3] $O(|X+1|!)$, which means that they can be communicated in time $O(|X|\log|X|)$, and therefore the Deferred Acceptance algorithm using preference lists terminates in polynomial time in the offline model.

In [114] the authors count the size of various classes of choice functions, including the class of path-independent choice functions, which they show to be doubly exponential in size. This implies that an arbitrary path-independent choice function cannot be communicated to the central planner in polynomial time, and so cannot be used to produce an efficient Deferred Acceptance algorithm. In [112], the authors show that a further subset of these choice functions that satisfy the property of *cardinal monotonicity* (see Section 3.4.2) is still doubly exponential.

We prove the following theorem, that shows that even in the case of quota-filling, path-independent choice functions, the class is too large.

Theorem 47. *The number of substitutable and quota-filling choice functions on ground set X with $|X| = n$ is $2^{\Omega\left(\frac{2^{\lfloor n/2\rfloor-1}}{\sqrt{\lfloor n/2\rfloor-1}}\right)}$.*

The proof is deferred to Section 3.4.1. The following corollary summarizes the implications of this fact.

Corollary 48. *Substitutable and quota-filling choice functions cannot be communicated efficiently, in particular, the Deferred Acceptance algorithm in the offline model using such choice function does not terminate in time polynomial in the number of agents.*

Proof. This follows from Theorem 47 and Lemma 46. □

A key obstacle to the use of many practical choice functions in the real world is therefore the communication cost of choice functions. In the next section, we propose a class of choice functions that lead to a polynomial time algorithm when used in the offline model for stable matching.

[3] If preference lists are complete, they number $|X|!$, but if they may be incomplete, we can add an extra "tombstone" element and discard any students after it, which gives an upper bound of $|X+1|!$.

3.3.2 Kuhn Choice Functions

The lack of efficient representation for arbitrary path-independent and quota-filling choice functions shown in the last section motivates our search for a more tractable subclass. We now propose the class of so-called *Kuhn choice functions* as a practical choice.

Definition 49 (Kuhn choice function)**.** *Consider some ground set X and a quota q. Let G be the complete bipartite graph on $[q] \cup X$. For each edge $(i, j) \in [q] \times X$, define a strictly positive weight $w(i, j) > 0$.*

The Kuhn choice function *$C(S)$ with weights w is defined for all $S \subseteq X$ by the set of nodes of S that are matched in the maximum-weight matching in the subgraph $[q] \times S$. That is*

$$C(S) = S \cap \left(\arg\max_{M \in \mathcal{M}_{[q] \cup S}} \{w(M)\} \right), \tag{3.1}$$

where $\mathcal{M}_{[q] \cup S}$ is the set of all matchings in the subgraph of G restricted to nodes in $[q] \cup S$. We require that the maximizer in the definition of $C(S)$ is unique.

The naming is after Harold Kuhn, one of the early pioneers of the Hungarian method for finding maximum-weight matchings in bipartite graphs [115], work that also shows that Kuhn choice functions can be evaluated efficiently for any $S \subseteq X$.

Kuhn choice functions have the following natural interpretation: divide the school into q seats (some of which may be identical), then for every seat i and every student j, assign a positive value $w(i, j)$ to that seat-student pair. The Kuhn choice function given by these weights now chooses from any offered set of students, the subset that maximizes total value. Such values may represent the "aptitude" of students for the various seats, or for instance some monetary utility of assigning that student to that given seat. This simple interpretation and easy construction is a key merit of Kuhn choice functions. We define the value of the maximum-weight matching itself as the *valuation* of the Kuhn choice function.

Definition 50 (Valuation)**.** *Let C be a Kuhn choice function given by weights w, then define the* valuation *of C as*

$$w_C(S) = \max_{M \in \mathcal{M}_{[q] \cup S}} \{w(M)\}.$$

Similar objects based on maximum-weight matchings have been studied widely in the literature in different areas. Shapley may have been the first to show they satisfy a certain complementarity condition [116]. The valuation function on the other hand, is widely studied in the context of the Kelso-Crawford market of indivisible goods [117] where it is known by various names, such as an OXS valuation. These are a strict subclass of all gross substitutes valuation functions (a sufficient condition for an equilibrium to exist in such markets). The gross substitutes condition is known to be equivalent to the concept of $M^\natural$-concavity [118] in the field of discrete convex analysis. Similarly for S with $|S| \leq q$, w_C is known to be a valuated matroid. We refer the reader to [119] for a comprehensive survey on gross substitutes valuation functions.

The requirement that all maximizers be unique—the *unique-maximizer* property—is what sets Kuhn choice functions apart from related objects. If all maximizers were not distinct, then C would not necessarily produce a well-defined choice function (but would produce a choice correspondence instead [120]). To illustrate, observe that without the unique-maximizer property, one could set all weights equal, whereby every matching would be maximal. Then the task of deciding which set of students to choose would reduce to some kind of tie-breaking rule which would need to encapsulate all the structure of the choice function.

We next turn to properties of Kuhn choice functions.

Lemma 51. *Kuhn choice functions are quota-filling and path-independent.*

The proof is deferred to Section 3.4.2. Furthermore, Kuhn choice functions can indeed be efficiently communicated.

Theorem 52. *The encoding length of any Kuhn choice function is polynomial in the size of the ground set.*

The proof appears in Section 3.4.2. In particular, Theorem 52 guarantees that when used in the offline model, Kuhn choice functions yield a polynomial time algorithm for stable matching.

Recognition of Kuhn choice functions: While Kuhn choice functions are simple to construct, it is natural to ask whether one can easily tell if some arbitrary choice function is Kuhn or not. The authors in [121] show that given a choice function, verifying whether it satisfies substitutability itself takes an expected number of queries exponential in $|X|$, the size of the ground set. We therefore obviously require that choice functions be path-independent and quota-filling. In contrast, there may be some elegant structure of certain subclasses of choice function that would allow us to infer how they act on many sets without querying each one directly.

We prove the following theorem which shows that for $q \geq 3$, determining whether a given choice function is Kuhn or not cannot be done in a polynomial number of queries in q.

Theorem 53. *For any $q \geq 3$, there exists $n = O(q)$, a Kuhn choice function C over $X = [n]$, and a family $\mathcal{H}$ of non-Kuhn q-quota-filling path-independent choice functions over X such that an oracle must query C on $2^{\Omega(q)}$ subsets of X in order to distinguish C from $\mathcal{H}$.*

We defer the proof to Section 3.4.3. While this result implies that there is no hope in deciding whether a given choice function is Kuhn or not efficiently, it may still be possible that given a Kuhn choice function, one could efficiently find find weights that realize it via some form of oracle access. The difficulty in finding such weights is largely in the assignment problem: for every $S \subseteq X$ one must find an assignment of the q seats to the q chosen students in $C(S)$, and these assignments must furthermore be somehow consistent across every subset $S \subseteq X$. However, once one has chosen an assignment, finding the

exact weights themselves (or whether they exist) can be done in polynomial time with a simple linear program. See Section 3.5 for a discussion on computational aspects.

Before we explore this path further, we discuss some existing results on rationalizability of choice functions.

Rationalizability and the Value Oracle: In the recent work [122], the authors show that a choice function is path-independent if and only if it is *rationalizable* by a utility function that satisfies ordinal concavity. This result is summarized in the next definition and theorem.

Definition 54 (Rationalizability)**.** *A choice function C on X is* rationalizable *by a utility function $u : 2^X \to \mathbb{R}$ if for all $S \subseteq X$,*

$$C(S) = \arg\max_{T \subseteq S} \{u(T)\},$$

where the maximizer is unique.

Clearly all utility functions give rise to a choice function of some form, so the question is then what must be demanded of the utility function to guarantee path-independence. This question is answered by the following result.

Theorem 55. *A choice function C on X is path-independent, if and only if there exists a utility function $u : 2^X \to \mathbb{R}$ that rationalizes C such that for all $S, S' \subseteq X$ and $x \in S \setminus S'$, there exists $x' \in (S' \setminus S) \cup \{\emptyset\}$ such that either $u(S) < u(S - x + x')$ or $u(S') < u(S' - x' + x)$.*

See [122] for a proof, which proceeds by constructing an appropriate utility function. Rationalizability of arbitrary choice functions is very similar to the definition of Kuhn choice functions, with the difference being that the Kuhn utility function is the maximum-weight matching in the appropriate subgraph. In other words, a choice function C is Kuhn if and only if the valuation w_C rationalizes C.

As remarked earlier, finding weights itself is not difficult once an assignment from $[q]$ to $C(S)$ has been found for every $S \subseteq X$. A slightly easier question is then whether given the actual valuation w_C of a Kuhn choice function C allows one to find the weights themselves easily. We call this the *valuation-oracle* model. We show that in the case that $q = 2$, one can find structure in the weights given the valuation.

Theorem 56. *Under the valuation-oracle model, given a q-quota-filling choice function C with $q = 2$, one can efficiently verify whether C is Kuhn or not and if it is, find weights that rationalize it.*

The proof is deferred to Section 3.4.3. The case for $q \geq 3$ is an open problem.

In this section we have introduced Kuhn choice functions and shown that they satisfy several desirable properties. We have argued that they are readily usable if specified directly via weights, but deciding if a given choice function is Kuhn or finding the weights with only oracle access is difficult. We next turn our attention to understanding the hierarchy of choice functions.

3.3.3 Approximability and the Hierarchy of Choice

Up to this point, we have studied various classes of choice functions. To more granularly understand the relationship between these different classes, we now construct a *hierarchy of choice functions*, and discuss approximability between them.

Definition 57 (Classes of choice functions)**.** *Let X be a ground set of students, with $|X| = n$, and let $q = 1, 2, \ldots, n$ be some quota. We define the following classes of choice functions over X:*

- $\mathcal{P}^n$*: all choice functions that are path-independent;*
- $\mathcal{Q}^n_q$*: all choice functions that are path-independent and q-quota-filling;*
- $\mathcal{K}^n_q$*: all choice functions that are Kuhn with q seats;*
- $\mathcal{L}^n_q$*: all choice functions that are lexicographic with q seats; and*

- $\mathcal{R}_q^n$: *all choice functions that are q-responsive.*

From Section 3.2, we know for all q, n,

$$\mathcal{R}_q^n \subseteq \mathcal{L}_q^n \subseteq \mathcal{K}_q^n \subseteq \mathcal{Q}_q^n \subseteq \mathcal{P}^n.$$

To go beyond simple statements of inclusion, we define approximability for quota-filling choice functions as follows.

Definition 58 (α-approximability)**.** *Let $C, C' \in \mathcal{Q}_q^n$, then we say C α-approximates C' if for all $S \subseteq X$, and $|S| > q$,*

$$|C(S) \cap C'(S)| \geq \alpha q.$$

Note that if $C, C' \in \mathcal{Q}_q^n$, then $C(S) = C'(S)$ for all $S \subseteq X$ with $|S| \leq q$.

Our first result in this realm is that for any path-independent quota-filling choice function, one can always find a preference list that matches at least one seat correctly.

Lemma 59 ($1/q$-approximability of $\mathcal{Q}_q^n$ with $\mathcal{R}_q^n$)**.** *Let $C \in \mathcal{Q}_q^n$ be arbitrary, then there exists some $C' \in \mathcal{R}_q^n$ such that C' $1/q$-approximates C.*

The next result, however, shows that in the worst case responsive choice functions fail to approximate even lexicographic choice better than this.

Lemma 60 ($2/q$-inapproximability of $\mathcal{L}_q^n$ with $\mathcal{R}_q^n$)**.** *Given $q \geq 3$, there exists $n = O(q^2)$, and $C \in \mathcal{L}_q^n$ such that for all $C' \in \mathcal{R}_q^n$, C' does not $2/q$-approximate C.*

Further, a similar gap exists between quota-filling choice functions and Kuhn choice functions.

Lemma 61 ($2/q$-inapproximability of $\mathcal{Q}_q^n$ with $\mathcal{K}_q^n$)**.** *There exists $n, q \in \mathbb{N}$ and $C \in \mathcal{Q}_q^n$ such that for all $C' \in \mathcal{K}_q^n$, C' does not $2/q$-approximate C.*

Our final approximability result shows that the approximation neighborhood of responsive choice functions is doubly exponential within quota-filling choice functions.

Lemma 62. *For $n = O(q)$, there exists $C \in \mathcal{R}_q^n$ whose $1/q$-approximation neighborhood is of size $2^{\Omega\left(\frac{2^q-1}{\sqrt{q-1}}\right)}$ in $\mathcal{Q}_q^n$.*

These results together show that while responsive choice can approximate quota-filling choice up to one seat, a big gap exists in what can be represented between responsive and lexicographic choice, and between Kuhn and quota-filling choice.

3.4 Proofs of Main Results

3.4.1 Number of quota-filling choice functions

In this section we prove Theorem 47. We begin by defining cardinal monotonicity, a requirement on the size of choice that is weaker than that of being quota-filling.

Definition 63 (Cardinal monotonicity)**.** *A choice function C is* cardinal monotone *if for all $S \subseteq X$ and all $T \subseteq S$, $|C(T)| \leq |C(S)|$.*

We will proceed by embedding a large class of cardinal monotone choice functions into the set of quota-filling choice functions. To do so, we make use of the completion of a choice function with a preference list, which allows us to produce quota-filling choice functions from non-quota-filling ones.

Definition 64. *Let C be a choice function on ground set X and let $\succ$ be a strict preference list over X. Suppose there is some q such that for all $S \subseteq X$, $|C(S)| \leq q$. Define the $(q, \succ)$-completion of C to be the set function C' defined by*

$$C'(S) = C(S) \cup M(S),$$

for all $S \subseteq X$, where

$$M(S) = \max_{\succ}(S \setminus C(S), q - |C(S)|).$$

Here $\max_{\succ}(T, k)$ are the first k elements of T according to $\succ$. In other words $C'(S)$ picks $C(S)$ then selects the $\succ$-top elements left in S to fill it up to q elements.

The next lemma shows that the completion is also substitutable and now quota-filling.

Lemma 65. *Let C be a substitutable and cardinal monotone choice function on X. Suppose $\succ$ is a strict preference list over X and $q \geq \max_{S \subseteq X} |C(S)|$. Then the $(q, \succ)$-completion of C is substitutable and quota-filling with quota q.*

Proof. Let C' be the $(q, \succ)$-completion. C' is clearly quota-filling. For substitutability we must show that for $T \subseteq S$, we have $C'(S) \cap T \subseteq C'(T)$. By the substitutability of C, we have (in the notation of Definition 64)

$$C'(S) \cap T = (C(S) \cap T) \cup (M(S) \cap T) \subseteq C(T) \cup (M(S) \cap T)$$

By construction, $C(T) \subseteq C'(T)$. Now suppose $x \in M(S) \cap T$ and $x \notin C(T)$. By substitutability of C, $T \setminus C(T) \subseteq S \setminus C(S)$, so $x \in S \setminus C(S)$. Therefore, since x is among the $q - |C(S)|$ largest elements (by $\succ$) in $S \setminus C(S)$, it must also be among the $q - |C(S)|$ largest in the subset $T \setminus C(T)$. Further $q - |C(T)| \geq q - |C(S)|$ by cardinal monotonicity, so $x \in M(T) \subseteq C'(T)$. □

The key to Theorem 47 is now that one can expand the ground set appropriately and choose the right preference list to injectively embed cardinal monotone choice functions in the class of quota-filling ones.

We now show that one can choose the preference list $\succ$ in such a way that a set X and a preference list $\succ$ over X such that the completion yields an injective map from the set of cardinal monotone choice functions to quota-filling ones.

Theorem 66. *For each $q \geq 1$, there is an injective map between substitutable, cardinal monotone choice functions on q elements and substitutable q-quota-filling choice functions on $2q$ elements.*

Proof. Let X be the ground set of the cardinal monotone choice functions with $|X| = q$, and let X' be a "copy" of X such that $X \cap X' = \emptyset$. Let $\succ$ be any strict preference list over $X \cup X'$ such that for all $x' \in X'$ and $x \in X$, $x \succ x'$. We now describe the injective mapping of cardinal monotone choice functions over X to quota-filling choice functions on $X \cup X'$.

Let C be any substitutable, cardinal monotone choice function with ground set X. Now define C' to be the $(q, \succ)$-completion of C on $X \cup X'$ (with C extended to $X \cup X'$ in such a way that it never accepts any elements in X'). C' is then q-quota-filling and substitutable by Proposition 65. Furthermore for $S \subseteq X$, we get $C'(S \cup X') \cap X = C(S)$, so the mapping is injective. □

In [112], the authors prove the following theorem that shows that the number of substitutable and cardinal monotone choice functions is doubly exponential.

Theorem 67. *The number of substitutable and cardinal monotone choice functions on ground set X with $|X| = n$ is $2^{\Omega\left(\frac{2^{n-1}}{\sqrt{n-1}}\right)}$.*

Now the proof of Theorem 47 follows easily.

Theorem 47. *The number of substitutable and quota-filling choice functions on ground set X with $|X| = n$ is $2^{\Omega\left(\frac{2^{\lfloor n/2 \rfloor - 1}}{\sqrt{\lfloor n/2 \rfloor - 1}}\right)}$.*

Proof. Combine Theorem 66 and Theorem 67. □

3.4.2 Properties of Kuhn choice functions

In this section we prove Lemma 51 and Theorem 52. We begin with the former.

Lemma 68. *Kuhn choice functions are quota-filling and path-independent.*

Proof. Kuhn choice functions are clearly quota-filling: since all weights are strictly positive and the graph is complete, the maximum weight matching will always contain exactly q elements or as many as possible. It remains to show that such choice functions are substitutable.

Suppose $T \subseteq S \subseteq X$ with $b \in C(S)$, and assume for a contradiction that $b \notin C(T \cup \{b\})$. Let M be the maximum-weight matching on $[q] \cup S$, and M' the one on $[q] \cup T \cup \{b\}$. If both matchings are not $[q]$-perfect, the result follows trivially, so assume this is not the case.

Since b is only matched in M, there exists some maximal M-alternating path P in $M \cup M'$ starting at b. Note that the first edge is in M and goes from b to $[q]$, and since both matchings are $[q]$-perfect, we can continue the path back to X. But this edge is in M', so we are back to T. By similar reasoning, we conclude that P is of even length and contained in $[q] \cup T \cup \{b\}$.

$M \triangle P$ is therefore a maximal matching in $[q] \cup S$, and $M' \triangle P$ is a maximal matching in $[q] \cup T \cup \{b\}$. If the weight of P were positive, then M would not be of maximum weight, and if it were negative, M' would not be of maximum weight (and the weight cannot be 0 as the edge weights are strictly positive). This is the required contradiction. □

We now prove Theorem 52. We need the following result, which is Theorem 10.2 of [123].

Theorem 69 (Bounds on facet and vertex complexity)**.** *Let P be a rational polyhedron in $\mathbb{R}^n$ of facet complexity φ and vertex complexity ν, then $\nu \leq 4n^2\varphi$ and $\varphi \leq 4n^2\nu$.*

The facet complexity of a polyhedron $P \subseteq \mathbb{R}^n$ is the smallest integer $\varphi \geq n$ such that there is a set of rational inequalities $Ax \leq b$ with the size of each inequality being at most φ. The vertex complexity is similarly the smallest integer $\nu \geq n$ such that there are vertices $x_1, \ldots, x_k$ and vectors $y_1, \ldots, y_t$ with size at most ν such that P is the convex hull of the x_i plus the cone formed by the y_j.

The following is a direct consequence of this theorem.

Corollary 70. *If the rational system $Ax \leq b$ has each inequality with size at most ℓ and it has a vertex, then each of its vertices has size at most $4n^2\ell$.*

We now use this corollary to prove Theorem 52.

Theorem 52. *The encoding length of any Kuhn choice function is polynomial in the size of the ground set.*

Proof. Let C be any Kuhn choice function with q seats on ground set X with $|X| = n$. There are nq weights that need to be written down, and so it remains to show that the encoding size of each weight is polynomial in nq.

We will apply the last corollary to an appropriately chosen polyhedron. Let $w(i, j)$ for $(i, j) \in [q] \times X$ be the weights for C. Since we know the weights, we also know for every $S \subseteq X$ what the maximum-weight matching on $[q] \times S$ is. Denote this by M_S. It therefore holds that any set of weights, say $w'(i, j)$ for $(i, j) \in [q] \times X$, yield the same choice function if for all $S \subseteq X$,

$$\sum_{i \in [q]} w'(i, M_S(i)) > \sum_{i \in [q]} w'(i, M(i)), \qquad \forall\, M \in \mathcal{M}_{[q] \times S} \setminus \{M_S\},$$

where again $\mathcal{M}_{[q] \times S}$ is the set of all matchings in the complete bipartite graph $[q] \times S$. It is not hard to see that non-negative affine scaling of weights does not change the choice function they represent[4], which means we can replace the strict inequality by a non-strict inequality if we add +1 to the right hand side.

Each of these inequalities now has $2q + 1$ terms, and each term has a coefficient of ± 1, so the facet complexity is $O(q)$. Note that the system has nq variables. By construction, this rational system is non-empty and since $w' \geq 0$ holds, it must have a vertex. Applying Corollary 70 directly gives us an upper bound of $O(n^2q^3)$ on the size of a weight function realizing the Kuhn choice function, which completes the proof. □

[4]That is, if $w(i, j)$ are weights for C, then so are $a \cdot w(i, j) + b$ for all $a > 0$, $b \geq 0$.

We note that an alternative proof can be obtained via the Frank-Tardos preprocessing scheme [124], as pointed out to us by László Végh [125].

3.4.3 Complexity of Kuhn recognition

We now prove Theorem 53 and Theorem 56.

Notation: In this section we use the following notation. For a family of sets $\mathcal{A}$ and a set B with $B \cap \bigcup \mathcal{A} = \emptyset$, we denote $\mathcal{A} \oplus B = \{A \cup B : A \in \mathcal{A}\}$. For a set A and $k \in \mathbb{N}$, denote $\binom{A}{k} = \{T \subseteq A : |T| = k\}$. For $l, r \in \mathbb{Z}$ with $l \leq r$, denote $[l..r] = \{i \in \mathbb{Z} : l \leq i \leq r\}$. We denote by $>$ the natural order on $\mathbb{N}$, and define $C_>$ as the canonical responsive choice function on $[n]$.

Proof of THeorem 53: We first show how to construct a family of almost-lexicographic choice functions over $X = [n]$.

Definition 71. *Let $q \geq 3$ and $n \geq q + 3$, and let $A \subseteq [7..n]$ with $|A| = q - 3$ be arbitrary. We define the* A-modification *of the canonical q-responsive choice function C_A as follows. For all $S \subseteq X$ we set*

$$C_A(S) = \begin{cases} C_>(S) \setminus \{2\} \cup \{1\}, & \text{if } S \in \mathcal{F}(A), \\ C_>(S), & \text{otherwise}, \end{cases}$$

where $C_>$ is the canonical q-responsive choice function (that chooses the largest available elements in S), and

$$\mathcal{F}(A) = \{\{1,2,4,5\}, \{1,2,3,6\}, \{1,2,5,6\}\} \oplus A.$$

The A-modification is also quota-filling and path-independent.

Lemma 72. *For any A, C_A is quota-filling and path-independent.*

Proof. The fact that C_A is quota-filling follows from the definition. It remains to show C_A is substitutable, that is, for all $T \subseteq S \subseteq X$, if $z \in C_A(S)$ then $z \in C_A(T \cup \{z\})$. From the definition of C_A and the fact that $C_>$ is substitutable, we only need to check the case when $z = 1$ and $z = 2$ because on other elements, C_A acts identically to $C_>$. In both cases if $|S| \leq q + 1$ (this includes the case $S \in \mathcal{F}(A)$), the result follows because C_A is quota-filling; and for any $|S| \geq q + 2$, we have $\{1, 2\} \cap C_A(S) = \emptyset$ since they are the two least elements. This completes the proof. □

To show C_A is not Kuhn, we need the following fact about Kuhn choice functions interpreted as gross substitutes valuations.

Lemma 73. *Let C be a Kuhn choice function, then for all $S, T \subseteq X$ with $|S| = |T| = q$, and for all $s \in S \setminus T$, we have*

$$w_C(S) + w_C(T) \leq \max_{t \in T \setminus S} \{w_C(S \setminus \{s\} \cup \{t\}) + w_C(T \setminus \{t\} \cup \{s\})\}.$$

Proof. This is the valuated matroid exchange property, but w_C is a valuated matroid, see Lemma 8.5 of [119]. □

Lemma 74. *C_A is not a Kuhn choice function.*

Proof. Assume for contradiction that C_A were Kuhn and there were some weights $w(i, j)$, $i \in [q]$, $j \in [n]$ that realize C_A. In this proof, we write $a = 3$, $b = 4$, $c = 5$ and $d = 6$, and $w = w_{C_A}$ for clarity. Observe that since $|A| = q - 3$, we may apply Lemma 73 to the union of A with 3 distinct elements from $\{1, 2, a, b, c, d\}$.

By the definition of C_A, it holds that

$$w(A \cup \{1,2,a,b\}) = w(A \cup \{2,a,b\}) > w(A \cup \{1,a,b\}),$$
$$w(A \cup \{1,2,a,c\}) = w(A \cup \{2,a,c\}) > w(A \cup \{1,a,c\}),$$
$$w(A \cup \{1,2,a,d\}) = w(A \cup \{1,a,d\}) > w(A \cup \{2,a,d\}),$$
$$w(A \cup \{1,2,b,c\}) = w(A \cup \{1,b,c\}) > w(A \cup \{2,b,c\}),$$
$$w(A \cup \{1,2,b,d\}) = w(A \cup \{2,b,d\}) > w(A \cup \{1,b,d\}),$$
$$w(A \cup \{1,2,c,d\}) = w(A \cup \{1,c,d\}) > w(A \cup \{2,c,d\}).$$

Now define

$$X = w(A \cup \{2,a,b\}) + w(A \cup \{1,c,d\}),$$
$$Y = w(A \cup \{2,a,c\}) + w(A \cup \{2,b,d\}),$$
$$Z = w(A \cup \{1,a,d\}) + w(A \cup \{1,b,c\}).$$

Then applying Lemma 73 to Y and Z yields

$$Y < \max\{X,Z\}, \qquad Z < \max\{X,Y\}.$$

But this means that $Y < X$ and $Z < X$. We now argue that in fact $X < Z$ or $X < Y$, which leads to a contradiction. Consider the multigraph that contains both matchings that define X, that is, $A \cup \{2,a,b\}$ and $A \cup \{1,c,d\}$. Since every element in A has degree 2, no path in the multigraph terminates in A. The multigraph must therefore contain three disjoint alternating paths from $\{2,a,b\}$ to $\{1,c,d\}$. Observe that there is no path from 1 to 2 because otherwise we could swap along that path in the two matchings and get the

same total weight, yielding a contradiction because

$$X = w(A \cup \{1, a, b\}) + w(A \cup \{2, c, d\}) < w(A \cup \{2, a, b\}) + w(A \cup \{1, c, d\}) = X.$$

Instead, the path starting from 2 must terminate at c or d, and likewise the path starting from 1 must terminate at a or b. Suppose there is an alternating path from 2 to c and from 1 to a, then we may swap along it without modifying the total weight, so

$$X = w(A \cup \{c, 1, b\}) + w(A \cup \{a, 2, d\}) < w(A \cup \{1, b, c\}) + w(A \cup \{1, a, d\}) = Z.$$

One can verify that every other possible combination of alternating paths yields either $X < Z$ or $X < Y$, which contradicts that both $Y < X$ and $Z < X$ hold. □

We are now ready to prove Theorem 53.

Theorem 53. *For any $q \geq 3$, there exists $n = O(q)$, a Kuhn choice function C over $X = [n]$, and a family $\mathcal{H}$ of non-Kuhn q-quota-filling path-independent choice functions over X such that an oracle must query C on $2^{\Omega(q)}$ subsets of X in order to distinguish C from $\mathcal{H}$.*

Proof. Let $q \geq 3$ and $n \geq q+3$ be arbitrary. Note that C_A is q-quota-filling, path-independent, and not Kuhn for any choice of $A \subseteq [7..n]$ with $|A| = q - 3$. Furthermore, observe that if A and A' both satisfy these conditions then $\mathcal{F}(A) \cap \mathcal{F}(A') = \emptyset$ if $A \neq A'$. Define $\mathcal{H} = \{C_>\} \cup \{C_A : A \subseteq [7..n], |A| = q - 3\}$. For the oracle to discard some C_A from $\mathcal{H}$, it must query one of $\mathcal{F}(A)$, as C_A acts exactly like $C_>$ on all other sets. In other words, the oracle must query a set S where $S \cap [7..n] = A$ in order to discard C_A. Therefore the oracle must make at least $|\{A \subseteq [7..n] : |A| = q - 3\}| = \binom{n-6}{q-3}$ queries in order to rule out every C_A from $\mathcal{H}$. Choosing $n = 2q$ completes the proof. □

Proof of Theorem 56: For the remainder of this section, let $q = 2$ and let n be arbitrary. Denote the seats $[q]$ by $A = \{a, b\}$ for brevity. We first state two auxiliary lemmas.

Lemma 75 (Local 0-edge lemma)**.** *For $q = 2$ and all $s \in [q]$, there exists $i \in [n]$ such that for all $S \subseteq [n]$ with $|S| \geq q$, the edge (s, i) is not present in the maximum-weight matching for $C(S)$.*

Proof. Note that if $|S| > 2$, then that maximum-weight matching is also the same one as for $S' = C(S)$, so it suffices to show the claim for $|S| = 2$. Suppose for a contradiction that this were false. Then take without loss of generality $a \in [q]$ to be the seat it fails for, and suppose that for all $i \in [n]$, (a, i) were in the maximum-weight matching for at least one $S \subseteq [n]$. Consider one such matching that contains (a, i), and note that there is some $j \in [n] \setminus \{i\}$ such that (b, j) is also in that matching. This by definition implies that

$$w(a, i) + w(b, j) > w(a, j) + w(b, i).$$

Define the mapping $p : i \mapsto j$, which is well-defined for all $i \in [n]$ (one can arbitrarily pick j if there are multiple choices). Observe that there must therefore be some $\ell \in [n]$ and $m \geq 1$ such that $p^m(\ell) = \ell$. That is, p must contain a cycle. Now write $p^0(i) = i$ and this gives

$$\begin{aligned}
&\sum_{k=1}^{m} \left(w(a, p^{k-1}(\ell)) + w(b, p^k(\ell)) \right) \\
&\quad > \sum_{k=1}^{m} \left(w(a, p^k(\ell)) + w(b, p^{k-1}(\ell)) \right) \\
&\quad = w(a, p^m(\ell)) + \sum_{k=2}^{m} w(a, p^{k-1}(\ell)) + w(b, p^0(\ell)) + \sum_{k=1}^{m-1} w(b, p^k(\ell)) \\
&\quad = w(a, p^0(\ell)) + \sum_{k=2}^{m} w(a, p^{k-1}(\ell)) + w(b, p^m(\ell)) + \sum_{k=1}^{m-1} w(b, p^k(\ell)) \\
&\quad = \sum_{k=1}^{m} \left(w(a, p^{k-1}(\ell)) + w(b, p^k(\ell)) \right).
\end{aligned}$$

Which is a contradiction. □

Lemma 76 (Global 0-edge lemma)**.** *For $q = 2$, there exists $(s, i) \in [q] \times [n]$ such that for all $S \subseteq [n]$, the edge (s, i) is not in the maximum-weight matching for $C(S)$.*

Proof. Let (a, i_a) and (b, i_b) be the two edges that never appear in a maximum-weight matching for any S with $|S| \geq q$ as in Lemma 75, for $a, b \in [q]$ respectively. It suffices to show that one of (a, i_a) or (b, i_b) is never present in the maximum-weight matching for any sets $S \subseteq X$ with $|S| = 1$. But the only cases where (a, i_a) or (b, i_b) may appear are exactly when $S = \{i_a\}$ or $S = \{i_b\}$.

Suppose for a contradiction that neither (a, i_a) nor (b, i_b) satisfy the conditions of the premise. Then for $S = \{i_a\}$, the maximum-weight matching must be the edge (a, i_a) so that $w(a, i_a) > w(b, i_a)$, and likewise for $S = \{i_b\}$ it must be (b, i_b) so $w(b, i_b) > w(b, i_b)$. This implies $w(a, i_a) + w(b, i_b) > w(a, i_b) + w(b, i_a)$ so when $S = \{i_a, i_b\}$, the maximum-weight matching must contain both (a, i_a) and (b, i_b), but now this contradicts the claim that these edges never appear in such an S as per Lemma 75, which completes the proof. □

Using these two lemmas, we are can prove Theorem 56.

Theorem 56. *Under the valuation-oracle model, given a q-quota-filling choice function C with $q = 2$, one can efficiently verify whether C is Kuhn or not and if it is, find weights that rationalize it.*

Proof. Fix $q = 2$ and $n \geq 2$ and suppose we are given a Kuhn choice function C via its valuation function w_C for all $S \subseteq [n]$.

We describe an algorithm. First, guess an edge that satisfies Lemma 76, let this without loss of generality be $(a, 1)$. We then guess an edge that satisfies Lemma 75 for the remaining seat, let this now be $(b, 2)$ without loss of generality.

Now observe that if these choices of global and local 0-edges are correct, then $w(b, 1) = w_C(\{1\})$ because we know that the matching consists of the edge $(b, 1)$. Similarly, when $S = \{1, i\}$ for $i \in [2..n]$, the matching must consist of (a, i) and $(b, 1)$, so $w(a, i) = w_C(\{1, i\}) - w(b, 1)$. Similar reasoning allows us to deduce that for $i = [3..n]$, $w(b, i) = w_C(\{2, i\}) - w(a, 2)$.

Assuming our choices of the 0-edges were correct, we have now recovered all weights

except for $w(a,1)$ and $w(b,2)$. But it is not hard to see that we can without loss of generality set $w(a,1) = 0$ and $w(b,2) = 0$, since the first edge is never chosen and the second edge may be chosen only in the singleton $S = \{2\}$.

If the guesses for the 0-edges were correct, one can now verify in polynomial time for all $S \subseteq X$ with $|S| = 2$ that the found weights rationalizes C.

But note that the number of possible guesses for the two 0-edges is n^2, and the verification procedure given a guess runs in polynomial time, so one can enumerate every possibility for the 0-edges and check whether the found weights rationalize C in time polynomial in n. □

3.4.4 Proofs for the approximation hierarchy

In this section, we present the missing proofs on approximability.

Lemma 77 ($1/q$-approximability of $\mathcal{Q}_q^n$ with $\mathcal{R}_q^n$)**.** *Let $C \in \mathcal{Q}_q^n$ be arbitrary, then there exists some $C' \in \mathcal{R}_q^n$ such that C' $1/q$-approximates C.*

Proof. Construct sets C_i as follows; let $C_1 = C(X)$, and $C_i = C(X \setminus \cup_{j=1}^{i-1} C_j)$ for $i = 2, \ldots, \lceil n/q \rceil$. Let $\succ$ be any total order on X such that $x \succ y$ if $x \in C_i$, $y \in C_j$ and $i < j$. Let C' be the q-responsive choice function constructed from $\succ$.

We argue that C' $1/q$-approximates C. To see this, let $S \subseteq X$ be arbitrary. Set also $\ell = \min\{i : S \cap C_i \neq \emptyset\}$. Then since ℓ is minimal, C' prefers students in C_ℓ over any other student in $S \setminus C_\ell$, so we must have $S \cap C_\ell \subseteq C'(S)$, and in particular $S \cap C_\ell = C'(S) \cap C_\ell$. Further, by construction, $S \subseteq \cup_{j=\ell}^{\lceil n/q \rceil} C_j$, but $C(\cup_{j=\ell}^{\lceil n/q \rceil} C_j) = C_\ell$ and so by substitutability, $S \cap C_\ell = S \cap C(\cup_{j=\ell}^{\lceil n/q \rceil} C_j) \subseteq C(S)$. We have shown $C'(S) \cap C_\ell = S \cap C_\ell \subseteq C(S)$, but ℓ was chosen so that this intersection is non-empty, which proves the claim. □

Lemma 78 ($2/q$-inapproximability of $\mathcal{L}_q^n$ with $\mathcal{R}_q^n$)**.** *Given $q \geq 3$, there exists $n = O(q^2)$, and $C \in \mathcal{L}_q^n$ such that for all $C' \in \mathcal{R}_q^n$, C' does not $2/q$-approximate C.*

Proof. Fix $q \geq 3$, and construct a Kuhn choice function C with q schools and $n = q^2$ students as follows. Let seat $i = 1, \ldots, q$ prefer students $i, q+i, 2q+i, \ldots, (q-1)q+i$ in that order, and find others unacceptable.

Suppose there were some $C' \in \mathcal{R}_q^n$ such that it $2/q$-approximated C and let $\succ$ be its preference list. For each school $i = 1, \ldots, q$, let m_i be the student that $\succ$ ranks last, and let $M = \{m_1, \ldots, m_q\}$. Now let $j = \min_i \{m_i\}$ and let $T = \{q+j, 2q+j, \ldots, (q-1)q+j\}$ and $S = T \cup M$. Observe that $M \cap T = \{m_i\}$ since otherwise i is not minimal, so $|S| = 2q-1$. Further, note that for all $x_1 \in T$ and $x_2 \in M$, $x_1 \succ x_2$ by the definition of M and T. In particular, this implies $C'(S) = T$, but $C(S) = M$, which completes the proof since $|M \cap T| = 1$. □

Lemma 79 ($2/q$-inapproximability of $\mathcal{Q}_q^n$ with $\mathcal{K}_q^n$)**.** *There exists $n, q \in \mathbb{N}$ and $C \in \mathcal{Q}_q^n$ such that for all $C' \in \mathcal{K}_q^n$, C' does not $2/q$-approximate C.*

Proof. In Definition 82 of Section 3.5, we discuss the Devil's choice function, which is a non-Kuhn choice function with $n = 6$, $q = 2$. The result follows since $C \notin \mathcal{K}_q^n$. □

Lemma 80. *For $n = O(q)$, there exists $C \in \mathcal{R}_q^n$ whose $1/q$-approximation neighborhood is of size $2^{\Omega\left(\frac{2^{q-1}}{\sqrt{q-1}}\right)}$ in $\mathcal{Q}_q^n$.*

Proof. Fix q and let $n = 2q$. In Section 3.4.1 we construct $2^{\Omega\left(\frac{2^{q-1}}{\sqrt{q-1}}\right)}$ quota-filling and path-independent choice functions on $[n]$. Call this family $\mathcal{H}$. In the construction, we divide $[n]$ into two sets of size q: X and X', and let $\succ$ be any strict preference list over $X \cup X'$ such that for $x \in X, x' \in X'$ we have $x \succ x'$. We now argue that each $C' \in \mathcal{H}$ is in the $1/q$-approximation neighborhood of $C_\succ$. Fix $C' \in \mathcal{H}$. We need to show that for any $S \subseteq X \cup X'$ with $|S| \geq q+1$,

$$|C'(S) \cap C_\succ(S)| \geq 1.$$

Observe that $|S \cap X| \leq |X| = q$, so $S \cap X' \neq \emptyset$. Because both $C_\succ$ and C' (by construction as a

$(\succ, q)$-completion) follow the order $\succ$ on X', both must also pick the maximal element in $S \cap X'$ according to $\succ$, so $\max_{\succ}(S \cap X', 1) \in C'(S) \cap C_{\succ}(S)$, which completes the proof. □

3.5 Computational results

We now discuss various computational aspects of using quota-filling choice functions in stable matching.

Notation: When it is clear from context, we concatenate sets and write them without curly braces, for instance writing $C(123) = 12$ for $C(\{1,2,3\}) = \{1,2\}$.

Recognizing Kuhn choice functions: We first describe how to decide whether a given set of inputs and outputs of a choice function can be extended to a Kuhn choice function. This takes the form of the following theorem.

Theorem 81 (MIP for Kuhn recognition)**.** *Let C be a q-quota-filling choice function over $X = [n]$. Suppose that we are given $C(S)$ for all $S \in \mathcal{S}$ where $\mathcal{S} \subseteq 2^X$ (and all $S \in \mathcal{S}$ satisfy $|S| \geq q$). Then the following mixed-integer program will have an optimal solution with objective value $\varepsilon = 1$ if and only if there exists a Kuhn choice function that coincides exactly with the given values of C on $\mathcal{S}$.*

$$
\begin{aligned}
\max \quad & \varepsilon \\
s.t. \quad & W(S,\sigma) = \sum_{i \in [q]} w\left(i, \sigma(i)\right) && \forall\, S \in \mathcal{S},\ \forall\, \sigma \in \Pi([q], S), \\
& M(S) = \max_{\sigma \in \Pi([q], C(S))} v(S,\sigma) && \forall\, S \in \mathcal{S}, \\
& M(S) \geq \varepsilon + W(S,\sigma) && \forall\, S \in \mathcal{S},\ \forall\, \sigma \in \Pi([q], S) \setminus \Pi([q], C(S)), \\
& w(i,j) \geq 1 && \forall\, i \in [q], \forall\, j \in [n], \\
& 0 \leq \varepsilon \leq 1.
\end{aligned}
$$

Here we denote by $\Pi(U, V)$ the set of all injections $\sigma : U \to V$. In particular $\Pi([q], C(S))$ is the set of permutations of $C(S)$ across the q positions.

Furthermore, if such a solution exists, the values of the decision variables $w(i, j)$ with $(i, j) \in [q] \times [n]$ are weights that rationalize this Kuhn choice function.

Proof. Suppose the MIP has a solution with $\varepsilon = 1$. By its definition, $W(S, \sigma)$ is the weight of the matching that matches the students S to seats $[q]$ according to σ. $M(S)$ is then the maximum-weight matching in $[q] \cup C(S)$. Note that this constraint is where the non-linearity appears (making this formulation a MIP). We require that every matching that does not assign students in the right way (that is, $\sigma(S) \neq C(S)$) has a strictly lower weight, in particular it has at least a unit gap. The weights further satisfy $w(i, j) > 0$. Finally note that the formulation does not necessitate that sets $S' \notin \mathcal{S}$ have unique maximizers. However, since the maximizers are unique on the prescribed sets $\mathcal{S}$, one can simply perturb the weights by some small amount and the maximizers on the prescribed sets remain, and all other maximizers also become unique. Therefore the produced weights certainly rationalize a Kuhn choice function that coincides with C on $\mathcal{S}$.

Suppose now that C were Kuhn and there existed some weights $w(i, j) > 0$ that rationalize it, we must show the MIP has a solution. Define

$$\delta = \min \left\{ \min_{\substack{S \in 2^X, \\ \sigma, \sigma' \in \Pi([q], S) \setminus \Pi([q], C(S))}} \left\{ \left| \sum_{i \in [q]} w(i, \sigma(i)) - w(i, \sigma'(i)) \right| \right\}, \min_{(i,j) \in [q] \times [n]} \{ w(i, j) \} \right\}.$$

If we multiply all weights by δ^{-1}, then clearly $w(i, j) \geq 1$ for all i, j, and each maximizer is separated by at least 1. These weights $\delta^{-1} w(i, j)$ therefore satisfy the MIP with $\varepsilon = 1$, as required. □

We note that this formulation is presented for exposition but is not ideal in real-world use. It contains many redundant inequalities and can be further tightened when used in practice. For instance for $T \subset S$, one can add the constraint $M(S) \geq M(T) + \varepsilon$. That said,

modern MIP solvers will already remove many of the redundancies in the presolve stage.

The lattice representation: Consider a q-quota-filling path-independent choice function C on some ground set X. One convenient way to visualize C is the *lattice representation* of a choice function. Such a representation is constructed by observing that for each $T \subseteq X$ with $|T| = q$, $C^{-1}(T)$ has a unique inclusion-wise maximal set, say S, such that $C(S) = T$ (this is a straightforward corollary of path-independence). Further, $C(S \setminus \{x\}) = T$ for all $x \notin C(S)$, and for each of the q elements $y \in C(S)$, $C(S \setminus \{y\})$ yields a different choice, so S has exactly q descendant nodes. One can show that this in fact gives rise to a distributive lattice [126] (when one adds sets $T \subseteq X$ with $|T| < q$). Figure 3.1 shows an example of two quota-filling and path-independent (but non-Kuhn) choice functions through their lattice representation.

We now discuss the two smallest examples of non-Kuhn choice functions. We call these the *Devil's choice function* and the *Devil's cousin*.

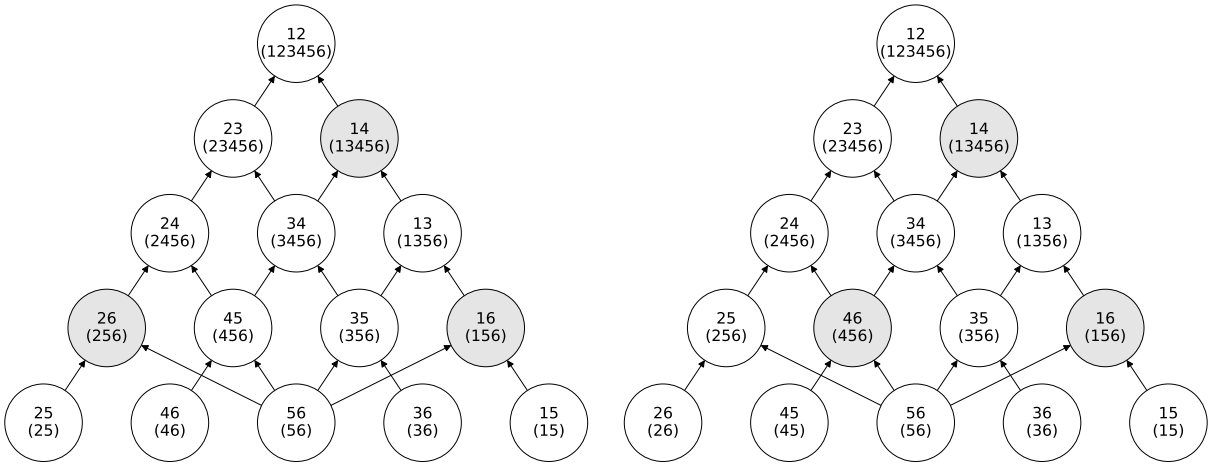
Definition 82 (Devil and Devil's cousin)**.** *Let $n = 6$ and $q = 2$, and define the* Devil's choice function *and the* Devil's cousin *as per the lattice representations in Figure 3.1. In particular, the Devil is identical to the responsive choice function choosing the least elements except for the choices $C(13456) = 14$, $C(256) = 26$, and $C(156) = 16$. Likewise the Devil's cousin is identical to the responsive choice function except for the choices $C(13456) = 14$, $C(456) = 46$, and $C(156) = 16$.*

Lemma 83. *For $n = 6$ and $q = 2$, the only non-Kuhn choice functions are the Devil's choice function and the Devil's cousin.*

This lemma can be shown computationally. There does not appear to be a trivial proof that these two choice functions are non-Kuhn, instead, the proofs are tedious and follow along the same lines as that of Lemma 74.

Counting Kuhn choice functions: For small q and n, one can compute exactly the number of path-independent and Kuhn choice functions. Table 3.1 show these counts for

$q = 2, 3, 4$, respectively, for small values of n. Here we consider choice functions equal if there exists some permutation $\sigma : [n] \to [n]$ such that $C(S) = C'(\sigma(S))$ for all $S \subseteq [n]$. That is, if the choice functions are equal after some relabeling of the ground set X. The MIP in Theorem 81 was used along with an exhaustive search to compute these tables. This method breaks down for larger values of n. The combinatorial structure in the MIP grows increasingly complex, in particular choosing the right assignment that arises in the maximum constraint. Further, simply deduplicating the choice functions becomes computationally infeasible for large n.



(a) Devil's choice function

(b) Devil's cousin

Figure 3.1: Lattice representations of the Devil's choice function and Devil's cousin. Nodes where the choice function deviates from the responsive choice function are shaded. Each node represents an (inclusion-wise maximal) $S \subseteq X$ such that $C(S) = T$, where T is on the first line and S on the second (inside braces). Observe in particular that each edge represents adding (or removing) an element to (from) S. In this particular case the lattice representations are very regular (and indeed have identical structure), but we note that this is not the case in general.

3.6 Conclusion

In this chapter we have investigated various aspects of quota-filling, path-independent choice functions. We have argued that this general class is too large to be used in the offline model for stable matching, and then proposed Kuhn choice functions as a practical yet powerful subclass. We have shown that while Kuhn choice functions are hard to recognize if not specified directly via weights, they have many desirable properties such as

	$q = 2$		$q = 3$		$q = 4$	
n	**All**	**Non-Kuhn**	**All**	**Non-Kuhn**	**All**	**Non-Kuhn**
2	1	0	–	–	–	–
3	1	0	1	0	–	–
4	2	0	1	0	1	0
5	6	0	2	0	1	0
6	40	2	24	1	3	0
7	560	116	2954	1037	119	21

Table 3.1: Number of path-independent quota-filling choice functions that are Kuhn and non-Kuhn for small values of n and q.

being efficiently representable and having high interpretability, making them ideal for use in the offline model. We have additionally studied the hierarchy of choice functions and shown that Kuhn choice functions are capable of representing a much larger class of preference systems than strict preference lists. These results support the use of Kuhn choice functions in markets such as New York City school choice, which would give schools richer ability to describe their preferences and to assemble classes of diverse students.

Epilogue

In this thesis, we have presented three strands of research exploring tradeoffs between information, tractability, and fairness within large matching markets. In Chapter 1 we studied the Serial Dictatorship under random markets where student preferences were limited in length, and showed that under our hypotheses students in large balanced markets prefer longer lists. In Chapter 2 we studied the impact of the presence of bias on admissions to the New York City Specialized High Schools, and showed that while disadvantaged students experience outsized mistreatment due to bias, it can be effectively mitigated (as measured by two forms of aggregate mistreatment) by targeting the average-top performing students with additional resources. Finally in Chapter 3 we proposed Kuhn choice functions as a practical alternative to strict preference lists in the context of school choice, arguing for their good fit and versatility under the offline model, and placing them within the hierarchy of quota-filling path-independent choice functions.

References

[1] D. Gale and L. S. Shapley, "College admissions and the stability of marriage," *The American Mathematical Monthly*, vol. 69, no. 1, pp. 9–15, 1962.

[2] D. M. Herszenhorn, "Revised admission for high schools," *The New York Times*, Oct. 2003.

[3] A. Abdulkadiroğlu, P. A. Pathak, and A. E. Roth, "The New York City high school match," *American Economic Review*, vol. 95, no. 2, pp. 364–367, 2005.

[4] I. Ashlagi, A. Graur, I. Lo, and K. Mentzer, "Overbooking with priority-respecting reassignment," *Presentation at the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO'23)*, 2023.

[5] E. Bampis, B. Escoffier, and P. Youssef, "Online 2-stage stable matching," *Discrete Applied Mathematics*, vol. 341, pp. 394–405, 2023.

[6] Y. Faenza, A. Foussoul, and C. He, "Two-stage stochastic stable matching," in *International Conference on Integer Programming and Combinatorial Optimization*, Springer, 2024, pp. 154–167.

[7] Y. Faenza, A. Foussoul, and C. He, "Minimum cut representability of stable matching problems," *arXiv preprint arXiv:2504.04577*, 2025.

[8] S. Miyazaki and K. Okamoto, "Jointly stable matchings," *Journal of Combinatorial Optimization*, vol. 38, no. 2, pp. 646–665, 2019.

[9] C. Calsamiglia, G. Haeringer, and F. Klijn, "Constrained school choice: An experimental study," *American Economic Review*, vol. 100, no. 4, pp. 1860–74, 2010.

[10] G. Artemov, Y.-K. Che, and Y. He, "Strategic 'mistakes': Implications for market design research," *NBER working paper*, 2017.

[11] N. DOE. "The pre-k application is now open!" Accessed: 2025-02-09. (Jan. 2023).

[12] N. R. M. P. NRMP, "National resident matching program, results and data: 2023 main residency match," 2023.

[13] N. Immorlica and M. Mahdian, "Incentives in large random two-sided markets," *ACM Transactions on Economics and Computation (TEAC)*, vol. 3, no. 3, pp. 1–25, 2015.

[14] Y. Kanoria, S. Min, and P. Qian, "In which matching markets does the short side enjoy an advantage?" In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, SIAM, 2021, pp. 1374–1386.

[15] B. Pittel, "The average number of stable matchings," *SIAM Journal on Discrete Mathematics*, vol. 2, no. 4, pp. 530–549, 1989.

[16] B. Pittel, "On likely solutions of a stable marriage problem," *The Annals of Applied Probability*, vol. 2, no. 2, pp. 358–401, 1992.

[17] I. Ashlagi, Y. Kanoria, and J. D. Leshno, "Unbalanced random matching markets: The stark effect of competition," *Journal of Political Economy*, vol. 125, no. 1, pp. 69–98, 2017.

[18] B. Pittel, "On likely solutions of the stable matching problem with unequal numbers of men and women," *Mathematics of Operations Research*, vol. 44, no. 1, pp. 122–146, 2019.

[19] I. Bó and R. Hakimov, "Iterative versus standard deferred acceptance: Experimental evidence," *The Economic Journal*, vol. 130, no. 626, pp. 356–392, 2020.

[20] A. Abdulkadiroğlu and T. Sönmez, "Random serial dictatorship and the core from random endowments in house allocation problems," *Econometrica*, vol. 66, no. 3, pp. 689–701, 1998.

[21] A. Abdulkadiroğlu and T. Sönmez, "School choice: A mechanism design approach," *American economic review*, vol. 93, no. 3, pp. 729–747, 2003.

[22] S. Bade, "Random serial dictatorship: The one and only," *Mathematics of Operations Research*, vol. 45, no. 1, pp. 353–368, 2020.

[23] A. Bogomolnaia and H. Moulin, "A new solution to the random assignment problem," *Journal of Economic theory*, vol. 100, no. 2, pp. 295–328, 2001.

[24] L. H. Ehlers and B. Klaus, "Normative properties for object allocation problems: Characterizations and trade-offs," in *Online and Matching-based Market Design*, N. Immorlica, F. Echenique, and V. V. Vazirani, Eds., Cambridge University Press, 2023.

[25] N. Arnosti, "A continuum model of stable matching with finite capacities," *arXiv preprint arXiv:2205.12881*, 2022.

[26] N. Arnosti, "Lottery design for school choice," *Management Science*, vol. 69, no. 1, pp. 244–259, 2023.

[27] E. M. Azevedo and J. D. Leshno, "A supply and demand framework for two-sided matching markets," *Journal of Political Economy*, vol. 124, no. 5, pp. 1235–1268, 2016.

[28] N. Arnosti, "Short lists in centralized clearinghouses," in *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, 2015, pp. 751–751.

[29] R. M. Karp, U. V. Vazirani, and V. V. Vazirani, "An optimal algorithm for on-line bipartite matching," in *Proceedings of the twenty-second annual ACM symposium on Theory of computing*, 1990, pp. 352–358.

[30] N. Devanur and A. Mehta, "Online matching in advertisement auctions," in *Online and Matching-based Market Design*, N. Immorlica, F. Echenique, and V. V. Vazirani, Eds., Cambridge University Press, 2023.

[31] Z. Huang and T. Tröbst, "Applications of online matching," N. Immorlica, F. Echenique, and V. V. Vazirani, Eds., 2023.

[32] E. Kamke, *Differentialgleichungen lösungsmethoden und lösungen*. Chelsea Publishing Company, 1959.

[33] K. Henk and N. Peyman, "Series solution of high order abel, bernoulli, chini and riccati equations," *Kyungpook Mathematical Journal*, vol. 62, no. 4, pp. 729–736, 2022.

[34] A. Abdulkadiroğlu and T. Sönmez, "House allocation with existing tenants," *Journal of Economic Theory*, vol. 88, no. 2, pp. 233–260, 1999.

[35] Y. Faenza, S. Gupta, and X. Zhang, "Discovering opportunities in new york city's discovery program: Disadvantaged students in highly competitive markets," in *EC'23: Proceedings of the 24th ACM Conference on Economics and Computation*, ACM, 2023.

[36] D. P. Kroese, T. Taimre, and Z. I. Botev, *Handbook of Monte Carlo methods*. John Wiley & Sons, 2013.

[37] S. N. Ethier and T. G. Kurtz, *Markov processes: characterization and convergence*. John Wiley & Sons, 2009.

[38] I. Quinn Capers, D. Clinchot, L. McDougle, and A. G. Greenwald, "Implicit racial bias in medical school admissions," *Academic Medicine*, vol. 92, no. 3, pp. 365–369, 2017.

[39] S. P. Corcoran and E. C. Baker-Smith, "Pathways to an elite education: Application, admission, and matriculation to New York City's specialized high schools," *Education Finance and Policy*, vol. 13, no. 2, pp. 256–279, 2018.

[40] E. Shapiro, "Should a single test decide a 4-year-old's educational future?" *New York Times*, 2019.

[41] NYC DOE, *2019 NYC High School Directory*, https://bigappleacademy.com/wp-content/uploads/2018/06/HSD_2019_ENGLISH_Web.pdf, 2019.

[42] M. J. Lovaglia, J. W. Lucas, J. A. Houser, S. R. Thye, and B. Markovsky, "Status processes and mental ability test scores," *American Journal of Sociology*, vol. 104, no. 1, pp. 195–228, 1998.

[43] E. Shapiro, "Racist? fair? biased? Asian-American alumni debate elite high school admissions," *The New York Times Magazine*, 2019.

[44] J. Boschma and R. Brownstein, "The concentration of poverty in American schools," *The Atlantic*, vol. 29, 2016.

[45] J. Ashkenas, H. Park, and A. Pearce, "Even with affirmative action, Blacks and Hispanics are more underrepresented at top colleges than 35 years ago," *New York Times*, pp. 1–18, 2017.

[46] Gratz v. Bollinger, "Gratz v. Bollinger, 539 U.S. 244 (2003).," 2003.

[47] E. Shapiro and V. Wang, "Amid racial divisions, mayor's plan to scrap elite school exam fails," *New York Times*, 2019.

[48] NYC DOE, *Specialized high schools proposal*, https://www.schools.nyc.gov/docs/default-source/default-document-library/specialized-high-schools-proposal, 2018.

[49] G. Considine and G. Zappalà, "The influence of social and economic disadvantage in the academic performance of school students in Australia," *Journal of sociology*, vol. 38, no. 2, pp. 129–148, 2002.

[50] A. Clauset, C. R. Shalizi, and M. E. Newman, "Power-law distributions in empirical data," *SIAM review*, vol. 51, no. 4, pp. 661–703, 2009.

[51] S. Burgess, E. Greaves, A. Vignoles, and D. Wilson, "What parents want: School preferences and school choice," *The Economic Journal*, vol. 125, no. 587, pp. 1262–1289, 2015.

[52] H. Yue, R. S. Rico, M. K. Vang, and T. A. Giuffrida, "Supplemental instruction: Helping disadvantaged students reduce performance gap," *Journal of Developmental Education*, pp. 18–25, 2018.

[53] E. Shapiro, "Only 8 Black students are admitted to Stuyvesant High School," *New York Times*, 2021.

[54] P. Biró, "Student admissions in Hungary as Gale and Shapley envisaged," *University of Glasgow Technical Report TR-2008-291*, 2008.

[55] Y. Kamada and F. Kojima, "Fair matching under constraints: Theory and applications," *Review of Economic Studies*, vol. 91, no. 2, pp. 1162–1199, 2024.

[56] J. Kucsera and G. Orfield, "New York State's extreme school segregation: Inequality, inaction and a damaged future," 2014.

[57] E. Shapiro, "Segregation has been the story of New York City's schools for 50 years," *The New York Times Magazine*, March 26, 2019.

[58] E. Shapiro and K. K. R. Lai, "How New York's elite public schools lost their Black and Hispanic students," *The New York Times Magazine*, June 03, 2019.

[59] P. Biró, T. Fleiner, R. W. Irving, and D. F. Manlove, "The college admissions problem with lower and common quotas," *Theoretical Computer Science*, vol. 411, no. 34-36, pp. 3136–3153, 2010.

[60] T. Nguyen and R. Vohra, "Stable matching with proportionality constraints," *Operations Research*, 2019.

[61] K. Tomoeda, "Finding a stable matching under type-specific minimum quotas," *Journal of Economic Theory*, vol. 176, pp. 81–117, 2018.

[62] B. Backes, "Do affirmative action bans lower minority college enrollment and attainment?: Evidence from statewide bans," *Journal of Human resources*, vol. 47, no. 2, pp. 435–455, 2012.

[63] D. Fershtman and A. Pavan, ""soft"' affirmative action and minority recruitment," *American Economic Review: Insights*, vol. 3, no. 1, pp. 1–18, 2021.

[64] I. E. Hafalir, M. B. Yenmez, and M. A. Yildirim, "Effective affirmative action in school choice," *Theoretical Economics*, vol. 8, no. 2, pp. 325–363, 2013.

[65] A. Abdulkadiroğlu, "College admissions with affirmative action," *International Journal of Game Theory*, vol. 33, pp. 535–549, 2005.

[66] P. Arcidiacono, E. M. Aucejo, H. Fang, and K. I. Spenner, "Does affirmative action lead to mismatch? a new test and evidence," *Quantitative Economics*, vol. 2, no. 3, pp. 303–333, 2011.

[67] H. Chade, G. Lewis, and L. Smith, "Student portfolios and the college admissions problem," *Review of Economic Studies*, vol. 81, no. 3, pp. 971–1002, 2014.

[68] J. Chan and E. Eyster, "Does banning affirmative action lower college student quality?" *American Economic Review*, vol. 93, no. 3, pp. 858–872, 2003.

[69] Texas Comptroller of Public Accounts, *Top 10% rule*, Accessed: 2024-07-01, 2024.

[70] M. C. Long, "Race and college admissions: An alternative to affirmative action?" *review of Economics and Statistics*, vol. 86, no. 4, pp. 1020–1033, 2004.

[71] S. Kannan, A. Roth, and J. Ziani, "Downstream effects of affirmative action," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 2019, pp. 240–248.

[72] S. Coate and G. C. Loury, "Will affirmative-action policies eliminate negative stereotypes?" *The American Economic Review*, pp. 1220–1240, 1993.

[73] Z. Liu and N. Garg, "Test-optional policies: Overcoming strategic behavior and informational gaps," in *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, 2021, pp. 1–13.

[74] W. Dessein, A. Frankel, and N. Kartik, "Test-optional admissions," *arXiv preprint arXiv:2304.07551*, 2023.

[75] M. Niu, S. Kannan, A. Roth, and R. Vohra, "Best vs. all: Equity and accuracy of standardized test score reporting," in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 2022, pp. 574–586.

[76] New York City Independent Budget Office, *The Specialized High School Admissions pipeline*, Accessed: 2025-07-29, 2024.

[77] N. Garg, H. Li, and F. Monachou, "Standardized tests and affirmative action: The role of bias and variance," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021, pp. 261–261.

[78] W. W. Chin, "Equity and excellence, four years later," *City Journal*, Dec. 2022, Accessed: 2024-07-01.

[79] L. Hu, N. Immorlica, and J. W. Vaughan, "The disparate effects of strategic manipulation," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 2019, pp. 259–268.

[80] J. Kleinberg and M. Raghavan, "Selection problems in the presence of implicit bias," in *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[81] L. E. Celis, A. Mehrotra, and N. K. Vishnoi, "Interventions for ranking in the presence of implicit bias," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020, pp. 369–380.

[82] L. E. Celis, C. Hays, A. Mehrotra, and N. K. Vishnoi, "The effect of the Rooney Rule on implicit bias in the long term," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021, pp. 678–689.

[83] J. Salem and S. Gupta, "Secretary problems with biased evaluations using partial ordinal information," *Management Science*, 2023.

[84] V. Emelianov, N. Gast, K. P. Gummadi, and P. Loiseau, "On fair selection in the presence of implicit variance," in *Proceedings of the 21st ACM Conference on Economics and Computation*, 2020, pp. 649–675.

[85] J. Hastings, T. J. Kane, and D. O. Staiger, "Heterogeneous preferences and the efficacy of public school choice," *NBER Working Paper*, vol. 2145, pp. 1–46, 2009.

[86] M. Laverde, "Unequal assignments to public schools and the limits of school choice," *Unpublished working paper*, 2020.

[87] T. S. Dee and B. Jacob, "The impact of No Child Left Behind on student achievement," *Journal of Policy Analysis and management*, vol. 30, no. 3, pp. 418–446, 2011.

[88] R. Greenwald, L. V. Hedges, and R. D. Laine, "The effect of school resources on student achievement," *Review of educational research*, vol. 66, no. 3, pp. 361–396, 1996.

[89] K. Lang and J.-Y. K. Lehmann, "Racial discrimination in the labor market: Theory and empirics," *Journal of Economic Literature*, vol. 50, no. 4, pp. 959–1006, 2012.

[90] R. G. Fryer Jr, "The importance of segregation, discrimination, peer dynamics, and identity in explaining trends in the racial achievement gap," in *Handbook of Social Economics*, vol. 1, Elsevier, 2011, pp. 1165–1191.

[91] E. S. Phelps, "The statistical theory of racism and sexism," *The american economic review*, vol. 62, no. 4, pp. 659–661, 1972.

[92] K. Arrow, "The theory of discrimination," *Discrimination in Labor Markets*, 1973.

[93] D. J. Aigner and G. G. Cain, "Statistical theories of discrimination in labor markets," *Ilr Review*, vol. 30, no. 2, pp. 175–187, 1977.

[94] H. Fang and A. Moro, "Theories of statistical discrimination and affirmative action: A survey," *Handbook of social economics*, vol. 1, pp. 133–200, 2011.

[95] A. Sen, "Equality of what?" *The Tanner lecture on human values*, vol. 1, 1979.

[96] A. Kumar and J. Kleinberg, "Fairness measures for resource allocation," in *Proceedings 41st Annual Symposium on Foundations of Computer Science*, IEEE, 2000, pp. 75–85.

[97] V. Conitzer, R. Freeman, N. Shah, and J. W. Vaughan, "Group fairness for the allocation of indivisible goods," in *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI)*, 2019.

[98] C. Dwork and C. Ilvento, *Group fairness under composition*, 2018.

[99] M. T. Marsh and D. A. Schilling, "Equity measurement in facility location analysis: A review and framework," *European Journal of Operational Research*, vol. 74, no. 1, pp. 1–17, 1994.

[100] G. B. Folland, *Real analysis: modern techniques and their applications*. John Wiley & Sons, 1999, vol. 40.

[101] 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*, ACM, 2012, pp. 214–226.

[102] T. Roughgarden, "Algorithmic game theory," *Communications of the ACM*, vol. 53, no. 7, pp. 78–86, 2010.

[103] A. E. Roth, "Stability and polarization of interests in job matching," *Econometrica: Journal of the Econometric Society*, pp. 47–57, 1984.

[104] N. C. Gottfredson, A. T. Panter, C. E. Daye, W. A. Allen, L. F. Wightman, and M. E. Deo, "Does diversity at undergraduate institutions influence student outcomes?" *Journal of Diversity in Higher Education*, vol. 1, no. 2, p. 80, 2008.

[105] P. Gurin, E. Dey, S. Hurtado, and G. Gurin, "Diversity and higher education: Theory and impact on educational outcomes," *Harvard educational review*, vol. 72, no. 3, pp. 330–367, 2002.

[106] A. M.A. and M. A.V., "General theory of best variants choice: Some aspects," *IEEE Transactions on Automatic Control*, vol. 26, no. 5, pp. 1030–1040, 1981.

[107] C. R. Plott, "Path independence, rationality, and social choice," *Econometrica: Journal of the Econometric Society*, pp. 1075–1091, 1973.

[108] A. Alkan, "On preferences over subsets and the lattice structure of stable matchings," *Review of Economic Design*, vol. 6, no. 1, pp. 99–111, 2001.

[109] A. E. Roth, "Stability and polarization of interests in job matching," *Econometrica*, vol. 52, no. 1, pp. 47–57, 1984.

[110] J. W. Hatfield and P. R. Milgrom, "Matching with contracts," *American Economic Review*, vol. 95, no. 4, pp. 913–935, 2005.

[111] O. Aygün and T. Sönmez, "Matching with contracts: Comment," *American Economic Review*, vol. 103, no. 5, pp. 2050–2051, 2013.

[112] Y. Faenza and X. Zhang, "Affinely representable lattices, stable matchings, and choice functions," *Mathematical Programming*, vol. 197, no. 2, pp. 721–760, 2023.

[113] C. P. Chambers and M. B. Yenmez, "On lexicographic choice," *Economics Letters*, vol. 171, pp. 222–224, 2018.

[114] F. Echenique, "Counting combinatorial choice rules," *Games and Economic Behavior*, vol. 58, no. 2, pp. 231–245, 2007.

[115] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.

[116] L. S. Shapley, "Complements and substitutes in the opttmal assignment problem," *Naval Research Logistics Quarterly*, vol. 9, no. 1, pp. 45–48, 1962.

[117] A. S. Kelso Jr and V. P. Crawford, "Job matching, coalition formation, and gross substitutes," *Econometrica: Journal of the Econometric Society*, pp. 1483–1504, 1982.

[118] S. Fujishige and Z. Yang, "A note on kelso and crawford's gross substitutes condition," *Mathematics of Operations Research*, vol. 28, no. 3, pp. 463–469, 2003.

[119] R. P. Leme, "Gross substitutability: An algorithmic survey," *Games and Economic Behavior*, vol. 106, pp. 294–316, 2017.

[120] K. Bando, K. Imamura, and Y. Kawase, "Properties of path-independent choice correspondences and their applications to efficient and stable matchings," in *Proceedings of the 26th ACM Conference on Economics and Computation*, 2025, pp. 4–4.

[121] J. W. Hatfield, N. Immorlica, and S. D. Kominers, "Testing substitutability," *Games and Economic Behavior*, vol. 75, no. 2, pp. 639–645, 2012.

[122] K. Yokote, I. E. Hafalir, F. Kojima, and M. B. Yenmez, "Rationalizing path-independent choice rules," *arXiv preprint arXiv:2303.00892*, 2023.

[123] A. Schrijver, *Theory of linear and integer programming*. John Wiley & Sons, 1998.

[124] A. Frank and É. Tardos, "An application of simultaneous diophantine approximation in combinatorial optimization," *Combinatorica*, vol. 7, no. 1, pp. 49–65, 1987.

[125] L. Végh, Personal communication, 2022.

[126] M. R. Johnson and R. A. Dean, "Locally complete path independent choice functions and their lattices," *Mathematical social sciences*, vol. 42, no. 1, pp. 53–87, 2001.

[127] G. Teschl, *Ordinary differential equations and dynamical systems*. American Mathematical Society, 2024, vol. 140.

[128] G. Grimmett and D. Stirzaker, *Probability and random processes*. Oxford university press, 2020.

[129] A. E. Roth and M. Sotomayor, "Two-sided matching," *Handbook of game theory with economic applications*, vol. 1, pp. 485–541, 1992.

[130] N. Arnosti, "A continuum model of stable matchings with finite capacities," Talk at Simons Institute for the Theory of Computing, 2019.

Appendix A: Additional Details for Chapter 1

A.1 Background on Differential Equations and Stochastic Processes

This appendix contains some preliminaries that are required to follow the proofs in the main text, in particular on the theory of ordinary differential equations (ODEs) and initival value problems (IVPs), as well as probability and Markov theory.

Ordinary Differential Equations

An ordinary differential equation is a description of the local dynamics of a system in terms of its current state and derivatives; in particular, a first order ordinary differential equation describes the rate of change of a quantity x as a function of the current time t and the current state $x(t)$. That is, $x'(t) = f(t, x(t))$ for some f. A problem with such a description along with boundary data (e.g., $x(t_0) = x_0$) is called an *initial value problem*. A main theme in differential equations is to understand when such local descriptions give rise to solutions on a large domain, such as globally. Lipschitz continuity is an important property in differential equations, as it is a sufficient condition for the existence of such solutions in a neighborhood of where it holds as we will argue next.

Definition 84 (Lipschitz continuity)**.** *A function $f : \mathbb{R}^m \to \mathbb{R}^n$ is* Lipschitz continuous *under norm $\|\cdot\|$ on $U \subseteq \mathbb{R}^m$ if there exists some $L \geq 0$ such that*

$$\|f(x) - f(y)\| \leq L \|x - y\|,$$

for all $x, y \in U$. We call L the Lipschitz constant of f on U.

The standard result for the existence and uniqueness of solutions to initial value prob-

lems of ordinary differential equations comes from variations of the Picard-Lindelöf Theorem, which roughly state that if f is Lipschitz at a point, a solution exists and is unique in some neighborhood of that point. The following is a version that extends this fact to the positive reals for the one dimensional case, adapted from [127, Theorem 2.2 and Corollary 2.6].

Theorem 85. *Consider the initial value problem*

$$x' = f(t, x), \quad x(t_0) = x_0, \tag{A.1}$$

where $(t_0, x_0) \in U$ for some open set $U \subseteq \mathbb{R}^2$ and $f : U \to \mathbb{R}$ is a Lipschitz continuous function. Then if $[t_0, \infty) \times \mathbb{R} \in U$ there exists a unique solution $x(t)$ to (A.1) *for all $t \geq t_0$.*

Probability and Markov theory

In this section we briefly cover some fundamentals of general probability and Markov theory. The treatment is not fully formal as we omit certain regularity conditions and assumptions—however such conditions are immediately satisfied by the objects we study in this chapter. We refer the reader to [128] for a more thorough discussion of these topics.

Recall that a random variable X on a probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ is a measurable real-valued function from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{L})$.

Definition 86 (Stochastic process)**.** *A stochastic process on a state space $\mathcal{S} \subseteq \mathbb{R}$ is a set of random variables $\{X_t\}_{t \in \mathcal{T}}$ on the same probability triple that are indexed by some $\mathcal{T}$ (thought of as time) and each takes values in $\mathcal{S}$ (that is, $X_t \in \mathcal{S}$ for all $t \in \mathcal{T}$).*

Common cases are $\mathcal{T} = \{0, 1, 2, 3, \dots\}$ or $\mathcal{T} = \{0, \Delta, 2\Delta, 3\Delta, \dots\}$ for some $\Delta > 0$ for *discrete time* processes and $\mathcal{T} = [0, \infty)$ for *continuous time* processes.

Convergence. Consider a sequence $\{X^n\}_{n=1,2,\dots}$ of random variables on the same probability triple $(\Omega, \mathcal{F}, \mathbb{P})$. We recall some ways in which X^n may converge to some other

random variable X.

Definition 87 (Convergence in Probability)**.** *The sequence of random variables* $\{X^n\}_{n=1,2,\ldots}$ *converges to the random variable X in probability, denoted* $X^n \xrightarrow{\mathcal{P}} X$ *if for all* $\varepsilon > 0$*, as* $n \to \infty$

$$\lim_{n\to\infty} \mathbb{P}\left(|X^n - X| \geq \varepsilon\right) \to 0.$$

A stochastic process can converge pointwise (for all $t \in \mathcal{T}$) in probability, but clearly a stronger condition is that of uniform convergence, where this convergence is uniform through $\mathcal{T}$.

Definition 88 (Uniform convergence in probability)**.** *The sequence of stochastic processes* $\left\{X_t^n\right\}_{t\in\mathcal{T},n=1,2,\ldots}$ *on* $\mathcal{T}$ *converges uniformly in probability to* $\{X_t\}_{t\in\mathcal{T}}$ *if for all* $\varepsilon > 0$*, as* $n \to \infty$*,*

$$\mathbb{P}\left(\sup_{t\in\mathcal{T}} \left|X_t^n - X_t\right| \geq \varepsilon\right) \to 0.$$

That is, the maximum deviation over the index set between the X_t^n and X_t is itself bounded and vanishes in probability.

Definition 89 (Convergence in r-mean)**.** *The sequence of random variables* $\{X^n\}_{n=1,2,\ldots}$ *converges to a random variable X in r-mean, denoted* $X^n \xrightarrow{L^r} X$ *if for all* $\varepsilon > 0$*,* $\lim_{n\to\infty} \mathbb{E}\left(|X^n - X|^r\right) \to 0$*, as* $n \to \infty$*.*

Note in particular that for $r = 1$, this is convergence in mean, that is, $\mathbb{E}\left(X^n\right) \to \mathbb{E}\left(X\right)$. Convergence in probability does not automatically imply convergence in r-mean, though the converse is true. One special case where the converse holds is when X^n and X are bounded.

Lemma 90. *Suppose* $X^n \xrightarrow{\mathcal{P}} X$ *and* $|X| \leq M$ *and* $|X^n| \leq M$ *for some* $M \geq 0$*. Then* $X^n \xrightarrow{L^r} X$ *for all* $r \geq 1$*.*

Markov processes. A Markov process is a stochastic process which additionally satisfies a property of being *memoryless*, meaning that its future evolution is entirely dictated by its current state independent of the past. We restrict ourselves to time-homogeneous Markov chains with finite state spaces.

Definition 91 (Discrete time Markov chain)**.** *A* discrete time Markov chain *is a stochastic process on a finite state space $\mathcal{S} \subseteq \mathbb{R}$ and index set $\mathcal{T} = \{0, 1, 2, 3, \dots\}$ that satisfies the property*

$$\mathbb{P}\left(X_{t+1} = j \mid X_t = i_t, X_{t-1} = i_{t-1}, \dots, X_1 = i_1\right) = \mathbb{P}\left(X_{t+1} = j \mid X_t = i_t\right),$$

for all states $j, i. \in \mathcal{S}$ and all $t \in \mathcal{T}$. The evolution of such a Markov chain is therefore determined by its (one step) transition probabilities *$p(i, j)$ for $i, j \in \mathcal{S}$, arranged in a transition matrix P defined by*

$$p(i, j) = \mathbb{P}\left(X_{t+1} = j \mid X_t = i\right).$$

Definition 92 (Continuous time Markov chain)**.** *A* continuous time Markov chain *is a stochastic process on a finite state space $\mathcal{S} \subseteq \mathbb{R}$ and index set $\mathcal{T} = [0, \infty)$ that satisfies the property*

$$\mathbb{P}\left(X_{t+s} = j \mid X_t = i, X_{t_l} = i_l, X_{t_{l-1}} = i_{l-1}, \dots, X_0 = i_0\right) = \mathbb{P}\left(X_s = j \mid X_0 = i\right), \tag{A.2}$$

for all $s \geq 0$, $t > t_l > t_{l-1} > \cdots > t_1 > t_0 \geq 0$ and all $j, i, i. \in \mathcal{S}$.

Note that (A.2) implies that the transition times must also be random and memoryless. It turns out that this property embodies continuous time Markov chains with a large amount of structure. In particular, the time between transitions (when the chain moves from one state to another) must be exponentially distributed (this is the only memoryless continuous distribution, satisfying $\mathbb{P}(X \leq x \mid X \geq y) = \mathbb{P}(X \leq x - y)$).

In the discrete case the evolution was defined by the one step transition probabilities P. In the continuous case, these are replaced by a Matrix-valued function of time $P(t)$. It

turns out there is a compact way to represent the possible transitions via *transition rates* $q(i,j)$ that describe the instantaneous rate of moving from state i to state j as formalized by the following lemma.

Theorem 93 (Transition rates)**.** *For a continuous time Markov chain on a finite state space, let* $p(i,j;t) = \mathbb{P}(X_t = j \mid X_0 = i)$ *for* $t \geq 0$*. Denote by* $P(t)$ *the matrix with entries* $p(i,j;t)$*. Then there exists a* transition rate *matrix* Q *with entries* $q(i,j)$ *that is the unique solution to*

$$Q = \lim_{t \to 0} \frac{P(t) - I}{t},$$

note that $q(i,i) = -\sum_{j \neq i} q(i,j)$*.*

We next introduce the concept of a Poisson point process in order to build continuous time Markov chains from discrete time ones. Intuitively, a homogeneous Poisson point process simply counts the number of events that have happened until time t, where the time between successive events is iid and exponentially distributed.

Definition 94 (Homogeneous Poisson point process)**.** *Let* $\lambda > 0$*. Define the stochastic process* $\{H_t\}_{t \geq 0}$ *with state space* $\{0,1,2,\dots\}$ *as follows. For* $i = 1,2,\dots$ *let* $E_i \sim \text{Exp}(\lambda)$ *(an exponentially distributed random variable with rate* λ*), and let* $H_t = \sup_{i=0,1,2,\dots} \left\{ \sum_{j=1}^{i} E_j \leq t \right\}$*. We call* $\{H_t\}_{t \geq 0}$ *the homogeneous Poisson process with rate* λ *on* $t \geq 0$*. Note that* $H_t \sim \text{Pois}(\lambda t)$ *(a Poisson distributed random variable with mean* λt*), and in particular* $\mathbb{E}(H_t) = \lambda t$ *for all* $t \geq 0$*.*

A Homogeneous Poisson point process is an example of a very simple Markov chain.

Lemma 95. *A homogeneous Poisson point process with rate* λ *is a continuous time Markov chain on the countable state space* $\mathcal{S} = \{0,1,2,\dots\}$ *with transition rates*

$$q(i,j) = \begin{cases} \lambda, & j = i+1, \\ -\lambda, & j = i, \\ 0, & \textit{otherwise}. \end{cases}$$

Given a continuous time Markov chain, one can decompose it into a discrete time Markov chain that describes the transition probabilities at jump times, and a per-state rate of an Exponential distribution that describes the duration that the chain stays in that state before jumping to the next.

The following theorem gives one variation of a theorem that allows easily constructing continuous time Markov chains from discrete time Markov chains.

Theorem 96 (Constant-rate Markov Chain Embedding)**.** *Let $\{X_n\}_{n=0,1,2,\ldots}$ be a discrete time Markov chain on finite state space $\mathcal{S}$ with $X_0 = s_0$ for some initial state $s_0 \in \mathcal{S}$. Let $\lambda > 0$ be some rate and let $\{H_t\}_{t\geq 0}$ be the homogeneous Poisson point process with rate λ on $[0, \infty)$.*

Define a stochastic process $\{Z_t\}_{t\geq 0}$ by $Z_t = X_{H_t}$. Then $\{Z_t\}_{t\geq 0}$ is a continuous time Markov chain whose state space is $\mathcal{S}$, initial state is s_0, and that has rates $q(i, j) = \lambda p(i, j)$ for $i \neq j$ and $q(i, i) = -\lambda(1 - p(i, i))$. This continuous time Markov chain is called an embedded *chain because its transition rates follow the transition probabilities of the discrete chain, that is, $\mathbb{P}(Z_s = j \mid Z_t = i) = p(i, j)$ if $H_t = H_s + 1$.*

Martingales. A martingale is a stochastic process with the property that its expected value in the future is the current value.

Definition 97 (Martingale)**.** *A stochastic process $\{X_t\}_{t\geq 0}$ is a* martingale *if $\mathbb{E}(X_s \mid X_t) = X_t$ for all $s \geq t$.*

The homogeneous Poisson process corrected for its mean is a martingale. That is, $H_t - \lambda t$ is a martingale. The following is a standard result in martingale theory.

Theorem 98 (Doob's martingale inequality)**.** *If $\{X_t\}_{t\geq 0}$ is a martingale, then for all $T \geq 0$ and $C > 0$,*

$$\mathbb{P}\left(\sup_{t\in[0,T]} X_t \geq C\right) \leq \frac{\mathbb{E}(\max(X_T, 0))}{C}.$$

Further, for $r \geq 1$, we have

$$\mathbb{P}\left(\sup_{t\in[0,T]} |X_t| \geq C\right) \leq \frac{\mathbb{E}\left(|X_T|^r\right)}{C^r}.$$

A.2 Extension to Schools Having Multiple Seats

In Section 1.2.4, we discussed an extension of our model for the case that each school has $q = 1, 2, 3, \ldots$ seats. In this appendix, we trace the steps of Section 1.2 for the case of arbitrary q in more detail and discuss in more depth the results stated in the main body.

The discrete model with multiple seats. We fix $q \in \mathbb{N}$ and do not include it in superscripts for brevity. We let $n \in \mathbb{N}$ be the number of schools and $d \geq 1$ be the list length. It is now not enough to keep track just of $T_i^{n,d}$, now defined to be the number of schools with no remaining seats after students $\{1, 2, \ldots, i-1\}$ have had their turn. Instead, we must additionally keep track of the vector $S_i = (S_i^0, S_i^1, S_i^2, \ldots, S_i^{q-1})$ counting the number of schools with $k = 0, 1, \ldots, q-1$ seats taken at this point. Given this information, we can now see the market dynamics.

When it is the turn of student i, the probability that they get matched to any school in their list (or equivalent the complement of the probability that they pick exactly d schools with no seats remaining) is still the same one given in (1.3),

$$\mathbb{P}\left(M_i^{n,d} = 1 \mid T_i^{n,d} = k\right) = \begin{cases} 1 - \frac{\binom{k}{d}}{\binom{n}{d}}, & k \geq d, \\ 1, & \text{otherwise.} \end{cases}$$

It is also approximated by the same expression as Lemma 12, derived from sampling with replacement. On the other hand, if a student picks any school with a remaining seat, then that school will have k seats occupied with probability $S_i^k / \sum_{j=0}^{q-1} S_i^j$ (that is, conditional on the school having a remaining seat, we simply randomly pick any such school). The update rule for the counts is then as follows. If the chosen school has k seats taken, we

have $S_{i+1}^k = S_i^k - 1$, and if $k < q-1$ then $S_{i+1}^{k+1} = S_i^{k+1} + 1$ whereas if $k = q-1$ then $T_{i+1}^{n,d} = T_i^{n,d} + 1$. Note that we always have $T_i^{n,d} + \sum_{j=0}^{q-1} S_i^j = n$.

The continuous model with multiple seats. We now construct continuous analogues to $T_i^{n,d}$ and S_i for $i = \lfloor tn \rfloor$, letting $x_d(t)$ be the proportion of schools with all seats taken, and $y_d^k(t)$ be the proportion of schools with k seats taken for $k = 0, \ldots, q-1$. Initially no schools have any seats taken, so

$$\begin{aligned} y_d^0(0) &= 1, \\ y_d^k(0) &= 0, \qquad k = 1, 2, \ldots, q-1, \\ x_d(0) &= 0. \end{aligned}$$

The probability that student t is matched to any school is $1 - x_d(t)^d$, and otherwise, the proportion of time they get matched to a school with k seats taken is proportional to S_i^k. We therefore get that the rate at which schools with k seats taken become schools with $k+1$ seats taken is

$$(1 - x_d(t)^d) \frac{y_d^k(t)}{\sum_{m=0}^{q-1} y_d^m(t)} = \frac{1 - x_d(t)^d}{1 - x_d(t)} y_d^k(t),$$

since $x_d(t) + \sum_{k=0}^{q-1} y_d^k(t) = 1$. Note further that this is a positive flow into y_d^{k+1} and negative for y_d^k. We have each of these flows from y_d^k to y_d^{k+1} for $k = 0, \ldots, q-1$ and one from y_d^{q-1} to x_d. For convenience, define

$$\gamma_d(t) = \frac{1 - x_d(t)^d}{1 - x_d(t)}.$$

Denoting derivative with respect to time with a dot for clarity to avoid multiple superscripts, this gives us the following differential equation

$$\begin{aligned}
\dot{y}_d^0(t) &= -\gamma_d(t) y_d^0(t), \\
\dot{y}_d^k(t) &= \gamma_d(t)(y_d^{k-1}(t) - y_d^k(t)), \qquad k = 1, \dots, q-1, \\
\dot{x}_d(t) &= \gamma_d(t) y_d^{q-1}(t).
\end{aligned}$$

Putting these together, we have the initial value problem (1.8) stated in the main body. Note in particular that for $q = 1$, we get

$$\begin{aligned}
y_d^0(0) &= 1, \qquad \dot{y}_d^0(t) = -\frac{1 - x_d(t)^d}{1 - x_d(t)} y_d^0(t), \\
x_d(0) &= 0, \qquad \dot{x}_d(t) = \frac{1 - x_d(t)^d}{1 - x_d(t)} y_d^0(t).
\end{aligned}$$

This directly gives us $\dot{y}_d^0(t) + \dot{x}_d(t) = 0$, so with the initial condition, we have $y_d^0(t) + x_d(t) = 1$ for all $t \geq 0$. Rearranging since $x_d(t) < 1$, we write $\frac{y_d^0(t)}{1 - x_d(t)} = 1$, and the last equation reduces to the familiar $x_d(0) = 0$, $\dot{x}_d(t) = 1 - x_d(t)^d$, showing how this is an extension of the $q = 1$ case in (1.4).

Connection with the discrete model. Following the steps in Section 1.2.3, we can apply a slightly more general version of Theorem 6 for this multi-dimensional case, again from [36, Theorem 17.3.1] to prove the following lemma.

Lemma 99. *Fix $d \in \mathbb{N}$, $q = 1, 2, \dots$ and let $T_i^{n,d}$ and S_i^k for $k = 0, \dots, q-1$ be defined as in this section. Then $n^{-1} T_{\lfloor tn \rfloor}^{n,d} \to x_d(t)$ and $n^{-1} S_{\lfloor tn \rfloor}^k \to y_d^k(t)$ for $k = 0, 1, \dots, q-1$ uniformly in probability as $n \to \infty$, where $\{x_d(t), y_d^k(t)\}$ for $k = 0, 1, \dots, q-1$ is the unique solution satisfying the initial value problem* (1.8) *for $t \geq 0$.*

Proof. The proof follows identically to the proof of Theorem 6 in Section 1.3, adapted to the case of multiple dimensions. □

We also have the following lemma connecting the match probability for $q > 1$. Note that the difference to Lemma 7 is simply that this time $\dot{x}_d(t)$ is not necessarily equal to $1 - x_d(t)^d$.

Lemma 100. *For all $d \in \mathbb{N}$, $t \geq 0$, we have $\mathbb{P}(M^{n,d,q}_{\lfloor tn \rfloor} = 1) \to 1 - x_d(t)^d$ as $n \to \infty$.*

Proof. The proof follows similarly to the proof of Lemma 7. □

Dynamics of the multiple seats model for $d = 1$. Observe that for $d = 1$, we have $\gamma_d \equiv 1$. As in the earlier continuous model, we can again solve this explicitly. It's easy to see we get $y_1^0(t) = \exp(-t)$, and for $k = 1$, we now have $\dot{y}_1^1 = \exp(-t) - y_1^1$, which yields $y_1^1(t) = t\exp(-t)$. Following this pattern, we have

$$y_1^k(t) = \frac{t^k}{k!}e^{-t}, \qquad k = 0, \ldots, q-1,$$
$$x_1(t) = 1 - e^{-t}\sum_{k=0}^{q-1}\frac{t^k}{k!}. \tag{A.3}$$

Readers familiar with probability or phase-type distributions will note that $x_1(t)$ is the cumulative distribution function of an Erlang distribution with parameters $k = q$ and $\lambda = 1$. Figure A.1 shows the solution for $d = 1$, $q = 4$.

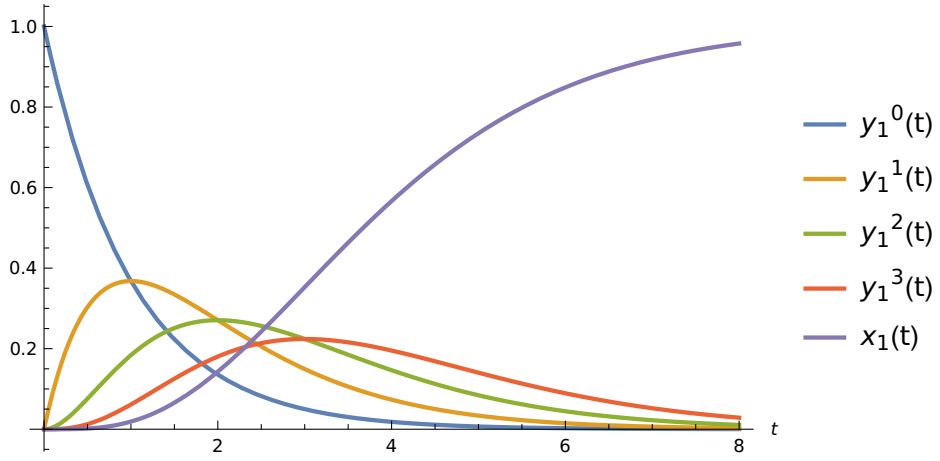


Figure A.1: Solution to the multiple-seat continuous market for $d = 1$ and $q = 4$.

Discussion of dynamics for $q > 1$. The fact that γ_d multiplies every equation in (1.8) suggests the following idea. Define an initial value problem for τ_d as

$$\tau_d(0) = 0, \qquad \dot{\tau}_d(t) = \frac{1 - x_1(\tau_d(t))^d}{1 - x_1(\tau_d(t))}. \tag{A.4}$$

It is then not hard to show that the following holds

$$\begin{aligned} y_d^k(t) &= y_1^k(\tau_d(t)), \qquad k = 0, 1, \ldots, q-1 \\ x_d(t) &= x_1(\tau_d(t)). \end{aligned} \tag{A.5}$$

This suggests a procedure for computing solutions to the initial value problem defined by (1.8): first compute a solution to (A.3), then solve (A.4) to compute $\tau_d(t)$ and plug it into (A.5).

We further remark that one can interpret $\tau_d(x)$ as describing the ratio of how much faster students are matched to schools when they get $d > 1$ tries compared to a single try. $1 - x_1^d$ is the probability of getting matched to any school given a list of length d and $1 - x_1$ is that probability with lists of length 1. The interpretation via (A.5) then tells us that the case for general d is exactly the same as for $d = 1$ except now the rate at which students are matched to schools is rescaled by a time-dependent factor of τ_d.

A.3 Missing Proofs

This appendix is divided into two sections in order to reduce nesting and allow the reader to follow the high-level techniques in our technical theorems. This section contains the main technical theorems preceded by a list of intricate inequalities needed in their proofs. The proofs of those inequalities are placed in Section A.3.2.

A.3.1 Main technical proofs

Notation. In this appendix, $x_d(t)$ always refers to the (unique) solution of the initial value problem (1.4) in Section 1.2.2 given by

$$x_d(0) = 0, \quad x_d'(t) = 1 - x_d(t)^d.$$

We often rely on its integral representation introduced in Lemma 15, which states that

$$t = \int_0^{x_d(t)} \frac{1}{1-u^d}\, du.$$

We use the notation $x_d(t)$, $x(d,t)$ or simply x interchangeably.

A useful identity. We begin by proving a useful technical lemma that we use heavily.

Lemma 101. *Let $d \geq 1$ and $z \in [0,1)$, then*

$$\int_0^z \frac{du}{1-u^d} = z\left(1 + \sum_{r=1}^{\infty} \frac{(z^d)^r}{rd+1}\right). \tag{A.6}$$

Let $q \in (0,1]$ and $A \in [0,1)$, then

$$\int_0^{A^q} \frac{1+\log(u)}{1-u^{1/q}}\, du = A^q\left((1+\log(A^q))\sum_{n=0}^{\infty}\frac{qA^n}{n+q} - \sum_{n=0}^{\infty}\frac{q^2A^n}{(n+q)^2}\right). \tag{A.7}$$

Proof. $(1-u^d)^{-1}$ is the sum of a geometric series in $u^d \in (0,1)$, so with $z \in [0,1)$ write

$$\int_0^z \frac{du}{1-u^d} = \int_0^z \sum_{r=0}^{\infty} u^{rd}\, du = \sum_{r=0}^{\infty}\int_0^z u^{rd}\, du = \sum_{r=0}^{\infty}\frac{z^{rd+1}}{rd+1} = z\left(1 + \sum_{r=1}^{\infty}\frac{(z^d)^r}{rd+1}\right).$$

The validity of exchanging the order of summation and integration follows by the Tonelli theorem since the integrand is non-negative.

For the second expression, we similarly write

$$\begin{aligned}\int_0^{A^q} \frac{1+\log(u)}{1-u^{1/q}}\,du &= \int_0^{A^q} \sum_{n=0}^{\infty} u^{n/q}(1+\log(u))\,du \\ &= \sum_{n=0}^{\infty} \int_0^{A^q} u^{n/q}(1+\log(u))\,du \\ &= \sum_{n=0}^{\infty} \left(\frac{qA^{n+q}}{n+q} + \frac{qA^{n+q}\log(A^q)}{n+q} - \frac{q^2A^{n+q}}{(n+q)^2}\right) \\ &= A^q\left((1+\log(A^q))\sum_{n=0}^{\infty}\frac{qA^n}{n+q} - \sum_{n=0}^{\infty}\frac{q^2A^n}{(n+q)^2}\right).\end{aligned}$$

Here again, one may apply the Tonelli theorem: even though the integrand is itself not non-negative, one can split $1+\log(u)$ into two parts getting a non-negative and a non-positive integral. □

Proof of Lemma 18. We next prove Lemma 18 which our main theorems hinge one, and which states bounds on the match probability of the last student in a balanced market. Before we proceed to prove the lemma, we state some inequalities that we will need, whose proofs are deferred to the next section.

Lemma 102. *For $q \in [0,1]$, we have*

$$\left(\frac{q+1}{q+2}\right)^q \left(1+\left(\frac{q}{q+1}\right)\log(q+2)\right) \geq 1, \tag{A.8}$$

$$\left(\frac{q+2}{q+4}\right)^q \left(1-\frac{q^2(q+2)}{(q+1)(q+4)} - q\log\left(\frac{2}{q+4}\right)\right) \leq 1. \tag{A.9}$$

Lemma 103. *For $d \geq 1$, we have $\left(\frac{2d+1}{4d+1}\right)^{1/d} \leq x(d,1) \leq \left(\frac{d+1}{2d+1}\right)^{1/d}$.*

Proof. We first show the upper bound. Recall $x(d,1)$ is defined via

$$1 = \int_0^{x(d,1)} \frac{du}{1-u^d}.$$

$(1-u^d)^{-1}$ is strictly positive, so the integral is increasing in the upper limit and it suffices to show

$$\int_0^{x(d,1)} \frac{du}{1-u^d} \leq \int_0^{\left(\frac{d+1}{2d+1}\right)^{1/d}} \frac{du}{1-u^d},$$

or equivalently

$$1 \leq \int_0^{\left(\frac{d+1}{2d+1}\right)^{1/d}} \frac{du}{1-u^d}.$$

Now apply the identity from (A.6) in Lemma 101 to write

$$\int_0^z \frac{du}{1-u^d} = z\left(1+\sum_{r=1}^{\infty}\frac{(z^d)^r}{rd+1}\right) \geq z\left(1+\sum_{r=1}^{\infty}\frac{(z^d)^r}{r(d+1)}\right) = z\left(1-\frac{1}{d+1}\log(1-z^d)\right).$$

The last line follows from the Maclaurin series of $\log(1-y)$ for $y \leq 1$. We need to show that this is lower bounded by 1. Substitute $q = 1/d$, so $q \in (0,1]$ and $z = \left(\frac{q+1}{q+2}\right)^q$, to get

$$\begin{aligned} z\left(1-\frac{1}{d+1}\log(1-z^d)\right) &= \left(\frac{q+1}{q+2}\right)^q\left(1-\frac{1}{1/q+1}\log\left(1-\frac{q+1}{q+2}\right)\right) \\ &= \left(\frac{q+1}{q+2}\right)^q\left(1+\left(\frac{q}{q+1}\right)\log(q+2)\right). \end{aligned}$$

We complete the proof by showing this is lower bounded by 1 in inequality (A.8) of Lemma 102.

For the lower bound on $x(d,1)$, we apply identical reasoning up to the application of the identity in (A.6) of Lemma 101 and must now show

$$1 \geq z\left(1+\sum_{r=1}^{\infty}\frac{(z^d)^r}{rd+1}\right).$$

We now substitute $z = \left(\frac{2d+1}{4d+1}\right)^{1/d}$ with $q = 1/d$, and write

$$\begin{aligned} z\left(1 + \sum_{r=1}^{\infty} \frac{(z^d)^r}{rd+1}\right) &= z\left(1 + \frac{z^d}{d+1} + \sum_{r=2}^{\infty} \frac{(z^d)^r}{rd+1}\right) \\ &\le z\left(1 + \frac{z^d}{d+1} + \sum_{r=2}^{\infty} \frac{(z^d)^r}{rd}\right) \\ &= z\left(1 + \frac{z^d}{d+1} - \frac{z^d}{d} - \frac{1}{d}\log(1-z^d)\right) \\ &= z\left(1 - \frac{z^d}{d(d+1)} - \frac{1}{d}\log(1-z^d)\right) \\ &= \left(\frac{q+2}{q+4}\right)^q \left(1 - \frac{q^2(q+2)}{(q+1)(q+4)} - q\log\left(\frac{2}{q+4}\right)\right). \end{aligned}$$

We again defer the inequality to (A.9) of Lemma 102 where we show this is upper bounded by 1. □

Proof of Theorem 8 We next prove Theorem 8, by bounding an appropriate integral. Again, we state some inequalities up front (whose proofs appear in the next section) before proceeding to the main proof.

Lemma 104. *For $q \in [0,1]$ and $A = \frac{q+1}{q+2}$, we have*

$$-\frac{\log(2)}{1+q\log(2)} \le \log(A) \le -\frac{2\log^2(2)}{q+2\log(2)}, \tag{A.10}$$

$$2A(1 + qA(\log(4) - 1)) \le 1 + q\log(2), \tag{A.11}$$

$$\sum_{n=2}^{\infty} \frac{A^{n-2}}{n+q} \le 2(\log(4) - 1). \tag{A.12}$$

Theorem 8. *For all $\ell > d \ge 1$ and for all $t \in (0,1]$, $x'_\ell(t) > x'_d(t)$.*

Proof. Applying Lemma 16, it suffices to show that for $t \in (0,1]$, $d \ge 1$, we have

$$\int_0^{x(d,t)} \frac{1+\log u}{1-u^d}\, du < 0. \tag{A.13}$$

This is because if $x'_d(t)$ is strictly increasing in d, then for any $\ell \geq d$ it must also hold that $x'_\ell(t) > x'_d(t)$ for $t \in (0,1]$.

Observe that $x(d,t)$ is strictly increasing in t and that $x(d,t) \leq t$. The integrand in (A.13) is negative on $u \in [0,1/e]$ and vanishes at $u = 1/e$ after which it remains non-negative. The result therefore follows immediately for $x(d,t) \in [0,1/e]$, and for $x(d,t) > 1/e$, one may write

$$\int_0^{x(d,t)} \frac{1+\log u}{1-u^d}\,du < \int_0^{x(d,1)} \frac{1+\log u}{1-u^d}\,du.$$

In Lemma 18 we showed that for $d \geq 1$,

$$x(d,1) \leq \left(\frac{d+1}{2d+1}\right)^{1/d}.$$

To prove the theorem, it therefore remains to show that for $d \geq 1$,

$$\int_0^{\left(\frac{d+1}{2d+1}\right)^{1/d}} \frac{1+\log(u)}{1-u^d}\,du \leq 0.$$

To proceed, change variables with $q = 1/d$ (so $q \in (0,1]$) and define $A = \frac{d+1}{2d+1} = \frac{q+1}{q+2}$, then apply the identity (A.7) of Lemma 101 to write

$$\begin{aligned}
\int_0^{A^q} \frac{1+\log(u)}{1-u^{1/q}}\,du &= A^q\left((1+\log(A^q))\sum_{n=0}^{\infty}\frac{qA^n}{n+q} - \sum_{n=0}^{\infty}\frac{q^2A^n}{(n+q)^2}\right)\\
&= A^q\left((1+q\log(A))\left(1+\frac{qA}{q+1}+qA^2\sum_{n=2}^{\infty}\frac{A^{n-2}}{n+q}\right) - 1 - \frac{q^2A}{(q+1)^2} - \sum_{n=2}^{\infty}\frac{q^2A^n}{(n+q)^2}\right)\\
&\leq A^q\left((1+q\log(A))\left(1+\frac{qA}{q+1}+2qA^2(\log(4)-1)\right) - 1 - \frac{q^2A}{(q+1)^2}\right)\\
&\leq A^q\left((1+q\log(A))2A(1+qA(\log(4)-1)) - 1 - \frac{q^2}{(q+2)^2}\right)\\
&\leq A^q\left(\left(1-\frac{2q\log^2(2)}{q+\log(4)}\right)(1+q\log(2)) - 1 - \frac{q^2}{(q+2)^2}\right).
\end{aligned}$$

In the first inequality we removed non-positive terms and applied (A.12) of Lemma 104. In the second inequality we applied the facts that $1 + \frac{qA}{q+1} = 2A$ and $\frac{q^2}{(q+2)^2} \leq \frac{q^2 A}{(q+1)^2}$. The last line follows from (A.10) and (A.11) of Lemma 104.

Since A^q is positive, we need to show the rest of the expression is non-positive. This is true at equality for $q = 0$, so it suffices to show that the derivative is non-positive, which is the case since

$$\begin{aligned}
&\frac{\partial}{\partial q}\left(\left(1 - \frac{2q\log^2(2)}{q+\log(4)}\right)(1+q\log(2)) - 1 - \frac{q^2}{(q+2)^2}\right) \\
&\quad = q\left(\frac{\log(2)\left(1-2\log^2(2)\right)(q+\log(16))}{(q+\log(4))^2} - \frac{4}{(q+2)^3}\right) \\
&\quad \leq q\left(\frac{\log(2)\left(1-2\log^2(2)\right)(1+\log(16))}{(0+\log(4))^2} - \frac{4}{(1+2)^3}\right) \\
&\quad \approx -0.09q \\
&\quad \leq 0,
\end{aligned}$$

as required. □

Proof of Theorem 9 We close this section with a proof of Theorem 9. The following inequalities are used in the proof, they will be verified in the next section.

Lemma 105. *For $q \in [0,1]$ and $A = \frac{1}{2} + \frac{q}{q+1}$, the following inequalities hold*

$$\log\left(\frac{1}{2} + \frac{q}{q+1}\right) \geq \frac{q}{q+1} - \log(2) \geq \frac{q}{2} - \log(2), \tag{A.14}$$

$$\log(A^q) \geq -\frac{\log(2)}{4}. \tag{A.15}$$

Theorem 9. *For all $d \geq 1$, there exists $c(d) > 1$ such that for all $\ell > d$ and $t \geq c(d)$, $x'_\ell(t) < x'_d(t)$. Furthermore, $c(d) \to 1$ as $d \to \infty$.*

Proof. Via an application of Lemma 16, it suffices to show that for all $d \geq 1$, there exists $c(d)$, such that

1. $c(d) \geq 1$,
2. $c(d) \to 1$ as $d \to \infty$, and
3. for all $t \geq c(d)$,

$$\int_0^{x(d,t)} \frac{1 + \log u}{1 - u^d}\, du > 0. \tag{A.16}$$

This is because $c(d)$ is decreasing, and so for any $\ell \geq d$ and $t \geq c(d)$, it must hold that $x'_\ell(t) < x'_d(t)$, which would complete the proof.

For $d = 1$, the result follows from the discussion in Section 1.2.3, so we take $d > 1$ from now on. Define $\varphi(d)$ and $c(d)$ as follows

$$\varphi(d) = \left(\frac{1}{2} + \frac{1}{d+1}\right)^{1/d},$$
$$c(d) = \int_0^{\varphi(d)} \frac{1}{1 - v^d}\, dv.$$

That is, we define $c(d)$ implicitly via $\varphi(d) = x(d, c(d))$. Observe that $(1 - v^d)^{-1}$ is strictly positive for $d > 1$, $v \in [0, 1)$, so for $x(d,t) \in [0,1)$, such an implicit equation is well defined, and in particular, $x(d,t)$ is strictly increasing in t.

We now show that this choice of c satisfies the three conditions of the premise.

For $d > 1$, we have $1 > \varphi(d) > \left(\frac{d+1}{2d+1}\right)^{1/d}$, and by Lemma 18, $\left(\frac{d+1}{2d+1}\right)^{1/d} \geq x(d,1)$, so $c(d) \geq 1$. This shows the first condition.

Note that c is continuous, so to show $\lim_{d\to\infty} c(d) = 1$, it suffices to show $\lim_{q\to 0^+} c(1/q) =$

1. To do so, fix $q \in (0,1)$, then apply identity (A.6) of Lemma 101,

$$\begin{aligned} c(1/q) &= \int_0^{\varphi(1/q)} \frac{1}{1-v^{1/q}}\,dv \\ &= \varphi(1/q)\left(1+\sum_{r=1}^{\infty}\frac{\varphi(1/q)^{r/q}}{r/q+1}\right) \\ &= \left(\frac{1}{2}+\frac{q}{q+1}\right)^q\left(1+q\sum_{r=1}^{\infty}\frac{\left(\frac{1}{2}+\frac{q}{q+1}\right)^r}{r+q}\right), \end{aligned}$$

and now it's easy to see that $\lim_{q\to 0^+} c(1/q) = 1$. This shows the second condition.

We finally need to show that for $t \geq c(d)$,

$$\int_0^{x(d,t)} \frac{1+\log u}{1-u^d}\,du \geq 0. \tag{A.17}$$

We have $t \geq c(d)$ if and only if $x(d,t) \geq x(d,c(d)) = \varphi(d)$. Since the integrand of (A.17) is positive for $u > 1/e$, and $\varphi(d) > 1/e$, we have that for $t \geq c(d)$,

$$\int_0^{x(d,t)} \frac{1+\log u}{1-u^d}\,du > \int_0^{\varphi(d)} \frac{1+\log u}{1-u^d}\,du, \tag{A.18}$$

and so it suffices to prove the right hand side is non-negative.

To do so, apply again identity (A.7) of Lemma 101 with $q = 1/d$ and $A = \frac{1}{2}+\frac{1}{d+1} = \frac{1}{2}+\frac{q}{q+1}$

to get

$$\begin{aligned}
\int_0^{\varphi(d)} \frac{1+\log u}{1-u^d}\,du &= \int_0^{A^q} \frac{1+\log u}{1-u^d}\,du \\
&= A^q\left((1+\log(A^q))\sum_{n=0}^{\infty}\frac{qA^n}{n+q} - \sum_{n=0}^{\infty}\frac{q^2A^n}{(n+q)^2}\right) \\
&= qA^q\left(\log(A) + (1+\log(A^q))\sum_{n=1}^{\infty}\frac{A^n}{n+q} - \sum_{n=1}^{\infty}\frac{qA^n}{(n+q)^2}\right) \\
&= qA^q\left(\sum_{n=1}^{\infty}\frac{(-1)^{n+1}(A-1)^n}{n} + (1+\log(A^q))\sum_{n=1}^{\infty}\frac{A^n}{n+q} - \sum_{n=1}^{\infty}\frac{qA^n}{(n+q)^2}\right) \\
&= \sum_{n=1}^{\infty} qA^{n+q}\left[-\frac{1}{n}\left(\frac{1-A}{A}\right)^n + (1+\log(A^q))\frac{1}{n+q} - \frac{q}{(n+q)^2}\right].
\end{aligned}$$

We have $qA^{n+q} \geq 0$, so defining

$$K_n(q) = -\frac{1}{n}\left(\frac{1-A}{A}\right)^n + (1+\log(A^q))\frac{1}{n+q} - \frac{q}{(n+q)^2},$$

it suffices to show $K_n(q) \geq 0$ for all $n = 1, 2, 3, \ldots$ in order to complete the proof. We first show $K_1(q) \geq 0$ and $K_2(q) \geq 0$, then prove $K_n(q) \geq 0$ for $n \geq 3$.

To show $K_1(q) \geq 0$, use (A.14) of Lemma 105 to write

$$\begin{aligned}
K_1(q) &= -\frac{1-A}{A} + (1+q\log(A))\frac{1}{q+1} - \frac{q}{(q+1)^2} && \text{(A.19)} \\
&= \frac{q}{(q+1)^2(3q+1)}\left[q^2+q+2+(q+1)(3q+1)\log\left(\frac{1}{2}+\frac{q}{q+1}\right)\right] && \text{(A.20)} \\
&\geq \frac{q}{(q+1)^2(3q+1)}\left[q^2+q+2+(q+1)(3q+1)\left(\frac{q}{q+1}-\log(2)\right)\right] && \text{(A.21)} \\
&= \frac{q}{(q+1)^2(3q+1)}\left[2-\log(2)+2q(1-2\log(2))+q^2(4-3\log(2))\right] && \text{(A.22)} \\
&\geq \frac{q}{(q+1)^2(3q+1)}\left[2-\log(2)-q\right] && \text{(A.23)} \\
&\geq 0. && \text{(A.24)}
\end{aligned}$$

Here we used $4 - 3\log(2) \geq 0$ and $2 - 4\log(2) \geq -1$.

For $K_2(q) \geq 0$, write similarly using (A.14) of Lemma 105 again

$$K_2(q) = -\frac{1}{2}\left(\frac{1-A}{A}\right)^2 + (1+q\log(A))\frac{1}{q+2} - \frac{q}{(q+2)^2} \tag{A.25}$$

$$= \frac{1}{2(3q+1)^2(q+2)^2}\left[-(q-1)^2(q+2)^2 + 4(3q+1)^2 + 2q(3q+1)^2(q+2)\log\left(\frac{1}{2}+\frac{q}{q+1}\right)\right] \tag{A.26}$$

$$\geq \frac{1}{2(3q+1)^2(q+2)^2}\left[-(q-1)^2(q+2)^2 + 4(3q+1)^2 + 2q(3q+1)^2(q+2)\left(\frac{q}{2}-\log(2)\right)\right] \tag{A.27}$$

$$\geq \frac{1}{2(3q+1)^2(q+2)^2}\left[-(q-1)^2(q+2)^2 + 4(3q+1)^2 - 3q(3q+1)^2\right] \tag{A.28}$$

$$= \frac{q}{2(3q+1)^2(q+2)^2}\left[25 - q(q^2+29q-21)\right] \tag{A.29}$$

$$\geq 0. \tag{A.30}$$

Here we used the fact that $2(q+2)\left(\frac{q}{2}-\log(2)\right) \geq -3$ (this is an increasing quadratic on $q \in [0,1]$, and at $q = 0$ it takes value $-4\log(2) > -3$). The last line follows because $q^2+29q-21 \in [-21, 9]$.

We now proceed to show $K_n(q) \geq 0$, first on $q \in [0, 1/2)$, then on $q \in [1/2, 1]$. For the former, observe that $A \geq 1/2$ so $\log(A) \geq \log(1/2) = -\log(2)$, and write

$$\begin{aligned}
K_n(q) &= -\frac{1}{n}\left(\frac{1-A}{A}\right)^n + (1+\log(A^q))\frac{1}{n+q} - \frac{q}{(n+q)^2} \\
&= -\frac{1}{n}\left(\frac{1-q}{3q+1}\right)^n + \frac{n}{(n+q)^2} + \frac{q\log(A)}{n+q} \\
&\geq -\frac{1}{n}\left(\frac{1-q}{3q+1}\right)^n + \frac{n}{(n+q)^2} - \frac{q\log(2)}{n+q} \\
&\geq -\frac{1}{n}(1-q)^n + \frac{n}{(n+q)^2} - \frac{q\log(2)}{n} \\
&= \frac{1}{n}\left(\frac{n^2}{(n+q)^2} - (1-q)^n - q\log(2)\right).
\end{aligned}$$

In the second inequality we use $-(3q+1)^{-n} \geq -1$ and $-(n+q)^{-1} \geq -n^{-1}$. Note that

$\frac{n^2}{(n+q)^2} \geq 1 - \frac{2q}{n}$. Further, $1-(1-q)^n$ is concave in q and equals 0 at $q=0$ and $1-2^{-n}$ at $q=1/2$, so then $1-(1-q)^n \geq 2q(1-2^{-n})$ on $q \in [0,1/2]$. This yields

$$\begin{aligned} nK_n(q) &\geq \frac{n^2}{(n+q)^2} - (1-q)^n - q\log(2) \\ &\geq 1 - \frac{2q}{n} - (1-q)^n - q\log(2) \\ &\geq 2q(1-2^{-n}) - \frac{2q}{n} - q\log(2) \\ &= q\left(2(1-2^{-n}) - \frac{2}{n} - \log(2)\right). \end{aligned}$$

The multiplier for q is increasing in n and positive for $n=3$, so $K_n(q) \geq 0$ for $n \geq 3$ and $q \leq 1/2$.

It remains to show that $K_3(q) \geq 0$ for $n \geq 3$ and $q \in [1/2,1]$. To do so, define $\alpha = \frac{\log(2)}{4} \approx 0.173$, then apply (A.15) of Lemma 105 which states $\log(A^q) \geq -\alpha$ to get

$$\begin{aligned} K_n(q) &= -\frac{1}{n}\left(\frac{1-A}{A}\right)^n + (1+\log(A^q))\frac{1}{n+q} - \frac{q}{(n+q)^2} \\ &\geq -\frac{1}{n}\left(\frac{1-A}{A}\right)^n + (1-\alpha)\frac{1}{n+q} - \frac{q}{(n+q)^2} \\ &\geq -\frac{1}{n}\left(\frac{1-A}{A}\right)^n + (1-\alpha)\frac{1}{n+1} - \frac{1}{(n+1)^2} \\ &= \frac{1}{n}\left(\frac{n((1-\alpha)n-\alpha)}{(n+1)^2} - \left(\frac{1-A}{A}\right)^n\right). \end{aligned}$$

Everything inside the brace is increasing in both q and n, as we now argue. Only the second term depends on q, and writing $\frac{1-A}{A} = \frac{1-q}{1+3q}$, it is clear that this is decreasing in q, and so $-\left(\frac{1-A}{A}\right)^n$ is also increasing in q for fixed n. Also note that $\frac{1-A}{A} \in [0,1]$, so $-\left(\frac{1-A}{A}\right)^n$ is increasing in n. Finally we argue that the first term is increasing in n, by taking a derivative and noting that it is non-negative for $n \geq 3$,

$$\frac{\partial}{\partial n}\left[\frac{n((1-\alpha)n-\alpha)}{(n+1)^2}\right] = \frac{n(2-\alpha)-\alpha}{(n+1)^3} \geq 0.$$

We have shown that $K_n(q) \geq 0$ for $q \in [0,1]$ for all $n = 1, 2, 3, \ldots$, and so

$$\int_0^{x(d,t)} \frac{1+\log u}{1-u^d}\, du \geq \int_0^{\varphi(d)} \frac{1+\log u}{1-u^d}\, du = \sum_{n=1}^{\infty} qA^{n+q} K_n(q) \geq 0. \tag{A.31}$$

This shows the last point, and concludes the proof. □

A.3.2 Various useful inequalities

In this section, we prove Lemmas 102, 104 and 105, which amount to exercises in calculus and do not hold much value for exposition or understanding the nature of our main results. We prove Lemma 104 before Lemma 102 as the latter depends on the former.

Lemma 106. *For $q \in [0,1]$ and $A = \frac{q+1}{q+2}$, we have*

$$-\frac{\log(2)}{1+q\log(2)} \leq \log(A) \leq -\frac{2\log^2(2)}{q+2\log(2)}, \tag{A.10}$$

$$2A(1+qA(\log(4)-1)) \leq 1+q\log(2), \tag{A.11}$$

$$\sum_{n=2}^{\infty} \frac{A^{n-2}}{n+q} \leq 2(\log(4)-1). \tag{A.12}$$

Proof. For (A.10), begin by noting that the inequality holds at equality for $q = 0$. Now multiplying the first half by $1 + q\log(2)$, it is equivalent to

$$0 \leq \big(1+q\log(2)\big)\log\left(\frac{q+1}{q+2}\right) + \log(2).$$

We take the second derivative with respect to q

$$\frac{\partial}{\partial q}\left[\big(1+q\log(2)\big)\log\left(\frac{q+1}{q+2}\right)\right] = \log(2)\log\left(\frac{q+1}{q+2}\right) + \frac{1+q\log(2)}{(q+1)(q+2)},$$
$$\frac{\partial^2}{\partial^2 q}\left[\big(1+q\log(2)\big)\log\left(\frac{q+1}{q+2}\right)\right] = \frac{q(3\log(2)-2)+4\log(2)-3}{(q+1)^2(q+2)^2}.$$

This is strictly negative for $q \in [0,1]$, and so the original function is concave, which means

that

$$(1+q\log(2))\log\left(\frac{q+1}{q+2}\right)+\log(2)\geq(1-q)\cdot 0+q\left[(1+\log(2))\log\left(\frac{2}{3}\right)+\log(2)\right]\geq 0. \tag{A.32}$$

The second inequality of (A.10) holds at equality for $q=0$. Further, one can show that the derivative of the difference is negative for $q\in[0,1)$ since

$$\frac{\partial}{\partial q}\left(\log(A)+\frac{2\log^2(2)}{q+\log(4)}\right)=\frac{1}{(q+1)(q+2)}-\frac{2\log^2(2)}{(q+\log(4))^2}.$$

The fact that the right hand side is negative is simple to verify, for example by checking that

$$(q+1)(q+2)\geq\frac{(q+\log(4))^2}{2\log^2(2)}.$$

For (A.11), we need the following to be non-positive

$$2A(1+qA(\log(4)-1))-(1+q\log(2))=\frac{q^2}{(q+2)^2}\left[q(\log(8)-2)-3+\log(16)\right].$$

Now the inner term is clearly increasing in q, and negative at $q=1$, so (A.11) holds.

For (A.12) we start with the following elementary series identity for $y<1$,

$$-\log(1-y)=\sum_{n=1}^{\infty}\frac{y^n}{n}. \tag{A.33}$$

In particular, this now gives us

$$\sum_{n=2}^{\infty}\frac{2^{2-n}}{n}=-2+4\sum_{n=1}^{\infty}\frac{2^{-n}}{n}=2(\log(4)-1).$$

In order to verify (A.12), it therefore remains to show that the following difference is non-

positive,

$$\begin{aligned}\sum_{n=2}^{\infty}\frac{A^{n-2}}{n+q}-\sum_{n=2}^{\infty}\frac{2^{2-n}}{n} &= \sum_{n=2}^{\infty}\left(\frac{A^{n-2}}{n+q}-\frac{2^{2-n}}{n}\right)\\ &= -\frac{2q}{3(q+3)}+\sum_{n=4}^{\infty}\left(\frac{A^{n-2}}{n+q}-\frac{2^{2-n}}{n}\right)\\ &\le -\frac{2q}{3(q+3)}+\sum_{n=4}^{\infty}\left(\frac{A^{n-2}-2^{2-n}}{n}\right)\\ &= -\frac{2q}{3(q+3)}+\sum_{n=4}^{5}\left(\frac{A^{n-2}-2^{2-n}}{n}\right)+\left(\frac{A^4-2^{-4}}{6}\right)+\sum_{n=7}^{\infty}\left(\frac{A^{n-2}-2^{2-n}}{n}\right).\end{aligned}$$

We now proceed to bound each of the last three terms. Note that if $g(q)$ is a smooth function with $g(0) = 0$ and $g'(q) \le M$ on $q \in [0,1]$, then via the fundamental theorem of calculus, one has

$$g(q) = g(q) - g(0) = \int_0^q g'(y)\,dy \le \int_0^q M\,dy = q \cdot M. \tag{A.34}$$

We apply this identity to $g_n(q) = \frac{A^{n-2}-2^{2-n}}{n}$ for various values of n: observe that it is smooth and satisfies $g_n(0) = 0$. We compute

$$g_n'(q) = \frac{A^{n-1}(n-2)}{n(q+1)^2}, \qquad g_n''(q) = \frac{A^n(n-2)(n-2q-5)}{n(q+1)^4}. \tag{A.35}$$

In particular, we have $g_n' \ge 0$ for $n \ge 4$. For $n = 4$ and $n = 5$ we further have $g_n'' \le 0$, which means g_n' is decreasing in q, and so $g_n'(q) \le g_n'(0)$, which establishes an upper bound we will use with (A.34),

$$\sum_{n=4}^{5}\left(\frac{A^{n-2}-2^{2-n}}{n}\right) = \sum_{n=4}^{5} g_n(q) \le \sum_{n=4}^{5} q \cdot g_n'(0) = q\sum_{n=4}^{5}\left(\frac{2^{n-1}(n-2)}{n}\right) = \frac{q}{10}.$$

For $n = 6$, one can verify via (A.35) that $g_6'(q)$ is maximized at $q = 1/2$, so $g_6'(q) \le g_6'(1/2)$,

and

$$\frac{A^4 - 2^{-4}}{6} = g_6(q) \le q \cdot g_6'(1/2) = \frac{72q}{3125}.$$

Finally for $n \ge 7$, we have $g_n'' \ge 0$, so g_n itself is convex, but note again $g_n(0) = 0$, so

$$\sum_{n=7}^{\infty} \left(\frac{A^{n-2} - 2^{2-n}}{n} \right) = \sum_{n=7}^{\infty} g_n(q) \le \sum_{n=7}^{\infty} q \cdot g_n(1) = q \sum_{n=7}^{\infty} \left(\frac{(2/3)^{n-2} - 2^{2-n}}{n} \right) \le \frac{q}{25}.$$

The bound in the last inequality can be computed using (A.34).

Combining the last three bounds on $g_n(q)$ for various n, we complete the proof with

$$\begin{aligned}
\sum_{n=2}^{\infty} \frac{A^{n-2}}{n+q} - \sum_{n=2}^{\infty} \frac{2^{2-n}}{n} &\le -\frac{2q}{3(q+3)} + \sum_{n=4}^{5} \left(\frac{A^{n-2} - 2^{2-n}}{n} \right) + \left(\frac{A^4 - 2^{-4}}{6} \right) + \sum_{n=7}^{\infty} \left(\frac{A^{n-2} - 2^{2-n}}{n} \right) \\
&\le -\frac{2q}{3(q+3)} + \frac{q}{10} + \frac{72q}{3125} + \frac{q}{25} \\
&\le -\frac{2q}{3(q+3)} + \frac{q}{6} \\
&= -\frac{q(1-q)}{6(3+q)} \\
&\le 0,
\end{aligned}$$

which finally establishes (A.12). □

Lemma 107. *For $q \in [0,1]$, we have*

$$\left(\frac{q+1}{q+2} \right)^q \left(1 + \left(\frac{q}{q+1} \right) \log(q+2) \right) \ge 1, \tag{A.8}$$

$$\left(\frac{q+2}{q+4} \right)^q \left(1 - \frac{q^2(q+2)}{(q+1)(q+4)} - q \log\left(\frac{2}{q+4} \right) \right) \le 1. \tag{A.9}$$

Proof. For (A.8), we move $\left(\frac{q+1}{q+2} \right)^q$ to the other side and will operate on the logarithm of the

inequality to equivalently show that

$$\log\left(1+\left(\frac{q}{q+1}\right)\log\left(q+2\right)\right) \geq q\log\left(\frac{q+2}{q+1}\right). \tag{A.36}$$

We start by stating two bounds for the logarithm. For $y \in [0,1]$, we have

$$\log(1+y) \geq \frac{2y}{2+y}, \tag{A.37}$$

$$\log(2+y) \geq \log(2) + \frac{y}{2} - \frac{y^2}{4(1+y)}. \tag{A.38}$$

Both are easy to verify as they hold for $y = 0$ and the difference of the left and the right side has positive derivative on $y \in [0,1]$. In fact, the first holds for all $y \geq 0$ (note $2y/(2+y) = y - y^2/(2+y)$),

$$\frac{\partial}{\partial y}\left(\log(1+y) - y + \frac{y^2}{2+y}\right) = \frac{y^2}{(1+y)(2+y)^2} \geq 0.$$

The second for $y \in [0,\sqrt{2}]$,

$$\frac{\partial}{\partial y}\left(\log(2+y) - \log(2) - \frac{y}{2} + \frac{y^2}{4(1+y)}\right) = \frac{y(2-y^2)}{4(1+y)^2(2+y)} \geq 0.$$

Applying (A.37) and (A.38) to the left hand side of (A.36), we get

$$\begin{aligned}
&\log\left(1+\left(\frac{q}{q+1}\right)\log\left(q+2\right)\right) \\
&\quad\geq \log\left(1+\left(\frac{q}{q+1}\right)\left(\log(2)+\frac{q}{2}-\frac{q^2}{4(q+1)}\right)\right) \\
&\quad= \log\left(1+\frac{4q(q+1)\log(2)+q^2(q+2)}{4(q+1)^2}\right) \\
&\quad\geq 2\times\frac{4q(q+1)\log(2)+q^2(q+2)}{4(q+1)^2}\times\frac{4(q+1)^2}{4q(q+1)\log(2)+q^2(q+2)+8(q+1)^2} \\
&\quad= \frac{8q(q+1)\log(2)+2q^2(q+2)}{4q(q+1)\log(2)+q^2(q+2)+8(q+1)^2}.
\end{aligned}$$

Next apply (A.10) of Lemma 104, to the right hand side of (A.36),

$$\frac{q\log(2)}{1+q\log(2)} \geq q\log\left(\frac{q+2}{q+1}\right).$$

The proof now hinges on showing that

$$\frac{8q(q+1)\log(2)+2q^2(q+2)}{4q(q+1)\log(2)+q^2(q+2)+8(q+1)^2} \geq \frac{q\log(2)}{1+q\log(2)}.$$

By collecting terms and simplifying, this holds if and only if

$$q^2\left[q^2\log(2)-2(1-\log(2))(2\log(2)-1)q+4(\log(2)-1)^2\right] \geq 0.$$

The inner term is now a quadratic in q, and in fact it has negative discriminant, so it has no real roots and is always either positive or negative. One verifies that it is positive at $q=0$, so the quadratic is positive for all $q \in [0,1]$, as required.

For (A.9), we first show the following three inequalities hold

$$\left(\frac{q+4}{q+2}\right)^q \geq \frac{(q+4)^2}{16}, \tag{A.39}$$

$$\log\left(2+\frac{q}{2}\right) \leq \log(2)+\frac{1}{4}\left(1-\frac{q}{10}\right)q, \tag{A.40}$$

$$\left(\frac{q+4}{q+2}\right)^{q-1} \geq \frac{1}{2}+\frac{q}{8}(1+4\log(2)). \tag{A.41}$$

Observe that each of these holds for $q=0$ and $q=1$ and all expressions are smooth, so it suffices to verify they hold for $q \in (0,1)$, for instance by checking the sign of the derivative.

For (A.39), take the logarithm of both sides and rearrange, after which we need

$$\log\left(\frac{q+4}{q+2}\right) \geq \frac{1}{q}\log\left(\frac{(q+4)^2}{16}\right) = \frac{2}{q}\log\left(1+\frac{q}{4}\right).$$

But now the left hand side always exceeds $\log(5/3) > 1/2$, and the right hand side never exceeds $1/2$ because $\log(1 + q/4) \leq q/4$.

For (A.40), take the derivative of the difference of the left and right terms, which is clearly negative

$$\frac{\partial}{\partial q}\left(\log\left(2+\frac{q}{2}\right)-\frac{1}{4}\left(1-\frac{q}{10}\right)q-\log(2)\right)=\frac{1}{q+4}+\frac{q-5}{20}=\frac{(q-1)q}{20(q+4)}\leq 0.$$

For (A.41), apply (A.39) to the derivative of the difference to get

$$\begin{aligned}
&\frac{\partial}{\partial q}\left[\left(\frac{q+4}{q+2}\right)^{q-1}-\frac{1}{2}-\frac{q}{8}(1+4\log(2))\right]\\
&\quad=\left(\frac{q+2}{q+4}\right)\left(\frac{q+4}{q+2}\right)^{q}\left(\frac{2(1-q)}{(2+q)(4+q)}+\log\left(\frac{q+4}{q+2}\right)\right)-\frac{1}{8}-\frac{1}{2}\log(2)\\
&\quad\geq\left(\frac{q+2}{q+4}\right)\frac{(q+4)^2}{16}\left(\frac{2(1-q)}{(2+q)(4+q)}+\log\left(\frac{q+4}{q+2}\right)\right)-\frac{1}{8}-\frac{1}{2}\log(2)\\
&\quad=\frac{(q+2)(q+4)}{8}\left(\frac{1}{2}\log\left(\frac{q+4}{q+2}\right)-\frac{q+4\log(2)}{(q+2)(q+4)}\right)\\
&\quad\geq\frac{1}{2}\log\left(\frac{q+4}{q+2}\right)-\frac{q+4\log(2)}{(q+2)(q+4)}.
\end{aligned}$$

This last expression vanishes at $q = 0$, so we take one further derivative and show it is positive,

$$\frac{\partial}{\partial q}\left[\frac{1}{2}\log\left(\frac{q+4}{q+2}\right)-\frac{q+4\log(2)}{(q+2)(q+4)}\right]=\frac{2(-3q+4q\log(2)-8+12\log(2))}{(q+2)^2(q+4)^2}\geq 0,$$

since the denominator is positive and the numerator is linear and positive for $q = 0$ and $q = 1$.

Now to finally prove (A.9), observe it holds if only if

$$\left(\frac{q+4}{q+2}\right)^{q}+\frac{q^2(q+2)}{(q+1)(q+4)}-q\log\left(2+\frac{q}{2}\right)-1\geq 0. \tag{A.42}$$

Now apply (A.40) and (A.41) to the left hand side and write

$$\begin{aligned}
&\left(\frac{q+4}{q+2}\right)^q + \frac{q^2(q+2)}{(q+1)(q+4)} - q\log\left(2+\frac{q}{2}\right) - 1 \\
&\quad = \left(\frac{q+4}{q+2}\right)\left(\frac{q+4}{q+2}\right)^{q-1} + \frac{q^2(q+2)}{(q+1)(q+4)} - q\log\left(2+\frac{q}{2}\right) - 1 \\
&\quad \geq \left(\frac{q+4}{q+2}\right)\left(\frac{1}{2}+\frac{q}{8}(1+4\log(2))\right) + \frac{q^2(q+2)}{(q+1)(q+4)} - q\left(\log(2)+\frac{1}{4}\left(1-\frac{q}{10}\right)q\right) - 1 \\
&\quad = \frac{q^2\left(q^4-3q^3-11q^2-20q^2\log(2)+53q-100q\log(2)+100-80\log(2)\right)}{40(q+1)(q+2)(q+4)}
\end{aligned}$$

This is positive if $q^4 - 3q^3 - 11q^2 - 20q^2\log(2) + 53q - 100q\log(2) + 100 - 80\log(2) \geq 0$, but this is a polynomial with two real roots, both at $q > 1$, so since this equation is positive for $q = 0$, it must be non-negative for all $q \in [0,1]$. □

Lemma 108. *For $q \in [0,1]$ and $A = \frac{1}{2} + \frac{q}{q+1}$, the following inequalities hold*

$$\log\left(\frac{1}{2}+\frac{q}{q+1}\right) \geq \frac{q}{q+1} - \log(2) \geq \frac{q}{2} - \log(2), \tag{A.14}$$

$$\log(A^q) \geq -\frac{\log(2)}{4}. \tag{A.15}$$

Proof. For (A.14), write the Maclaurin series of $\log(1-y)$

$$\begin{aligned}
\log\left(\frac{1}{2}+\frac{q}{q+1}\right) &= -\sum_{n=1}^{\infty}\frac{1}{n}\left(\frac{1}{q+1}-\frac{1}{2}\right)^n \\
&= \frac{1}{2} - \frac{1}{q+1} - \sum_{n=2}^{\infty}\frac{1}{n}\left(\frac{1}{q+1}-\frac{1}{2}\right)^n \\
&\geq \frac{1}{2} - \frac{1}{q+1} - \sum_{n=2}^{\infty}\frac{1}{n}\left(\frac{1}{2}\right)^n \\
&= \frac{1}{2} - \frac{1}{q+1} + \frac{1}{2} - \log(2) \\
&= \frac{q}{q+1} - \log(2) \\
&\geq \frac{q}{2} - \log(2).
\end{aligned}$$

For (A.15), observe that A is concave increasing in q, and $\log(\cdot)$ is concave increasing, so $\log(A)$ is concave increasing in q. One can also compute that $\log(A) = \log(1/2)$ at $q = 0$ and it vanishes at $q = 1$, so since $\log(A)$ is now concave, we must have $\log(A) \geq (1-q)\log(1/2)$ for $q \in [0,1]$. This yields $\log(A^q) = q\log(A) \geq q(1-q)\log(1/2)$, with the right hand side being minimized at $q = 1/2$ with a value of $\log(1/2)/4$, as required. □

A.4 Plots for Numerical Experiments

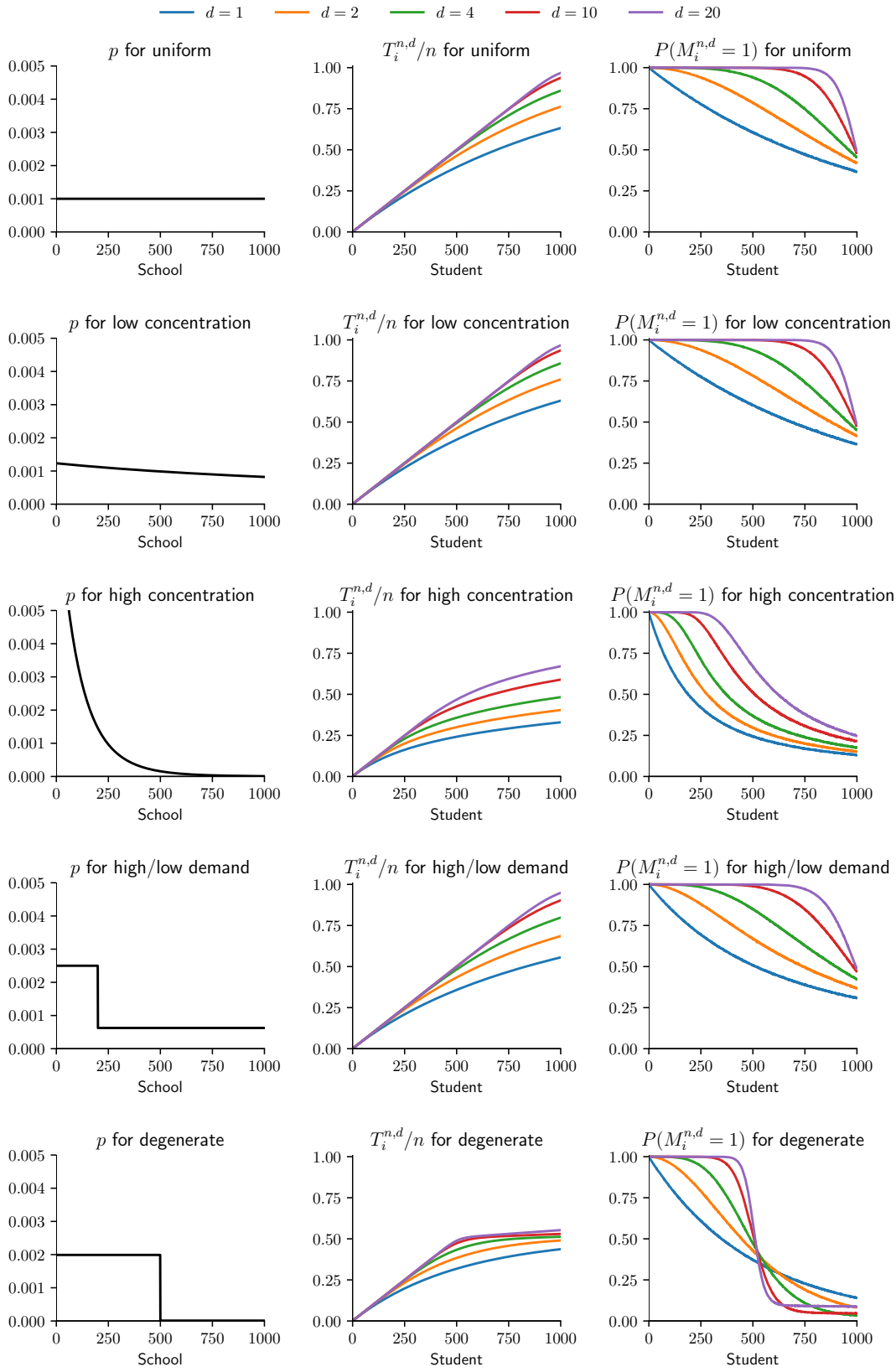


Figure A.2: Numerics for non-uniform sampling of schools.

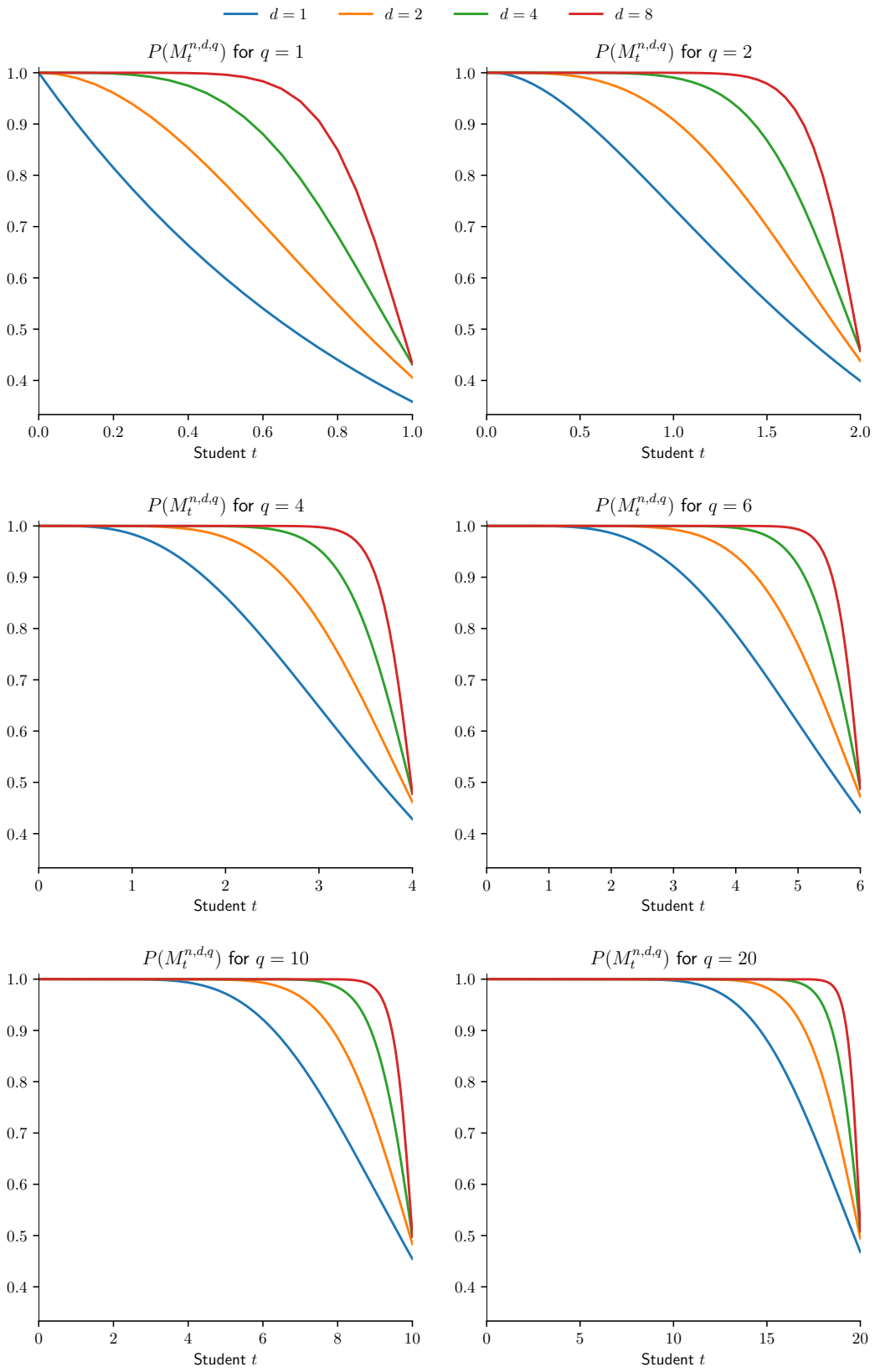


Figure A.3: Solutions to the multi-seat initial value problem for various values of d and q.

Appendix B: Additional Details for Chapter 2

B.1 Discussion on discrete versus continuous models

Traditionally, matching markets are assumed to be discrete [1, 129]. There has been however, in recent years, an interest for models where one or both sides of the markets are continuous [130, 27]. This is justified by the fact that, in many applications, markets are large, hence predictions in continuous markets often translate with a good degree of accuracy to discrete ones. Moreover, continuous markets are often analytically more tractable than discrete ones (see, again, [130, 27]). Our case is no exception: the continuous model allows us to deduce precise mathematical formulae, while we show through experiments that those formulae are a good approximation to the discrete case. We also provide additional experiments evaluating the robustness of our results under relaxation of assumptions, such as that of a unique bias factor for all students in G_2. We remark that the goal of this study is not to provide a mechanism to admits students to schools, for which the assumption of all rankings of schools as well as of students being the same would be too simplistic. On the contrary, as we want to understand the impact of bias at a macroscopic level, we believe our approximation to be meaningful and useful since as in our model any reasonable mechanism would output the same assignment. Note that in the classical discrete model, when schools and students have unique ranking, there is only one stable assignment, which is also Pareto-optimal for students. A similar statement holds for the appropriate translations of those concepts to our model.

B.2 Proof of Lemma 25

Proof. Assume T is not of the form $\{\theta \in \Theta : Z(\theta) \geq \delta\}$ or $\{\theta \in \Theta : Z(\theta) > \delta\}$ for some δ and let U be a connected and inclusionwise maximal subset of T that is bounded. Take the smallest number $\delta_1 \in [1, \infty)$ so that $Z(\theta) \leq \delta_1$ for all $\theta \in U$. Let $\overline{\theta} \in \theta$ such that $Z(\overline{\theta}) = \delta_1$.

Assume first that $\overline{\theta} \in U$. Then, for each $\epsilon_1 > 0$, there exists $\epsilon \in (0, \epsilon_1]$ such that $Z^{-1}(\delta_1 + \epsilon) \notin T$. Since $\beta < 1$ and by continuity of $Z(\cdot)$, there exists $\epsilon_2 > 0$ such that $\beta(\delta_1 + \epsilon) < \delta_1$ for all $\epsilon < \epsilon_2$. We can then take an appropriate $x \in [\delta_1, \delta_1 + \epsilon_2] \setminus T$ and $x' = \delta_1$ to show that T is not incentive compatible.

Next assume $\overline{\theta} \notin U$. In particular, we have $\overline{\theta} \notin T$. Similarly to the case above, we can find $\epsilon > 0$ such that $x' = \delta_1 - \epsilon$ satisfies $Z^{-1}(x') \in U \subseteq T$ and $\beta\delta_1 < x'$. Setting $x = \delta_1$, we deduce that T is not incentive compatible. □

B.3 Impact on Schools

This appendix explores the schools' perspective: the impact of bias on *utility* (quality of accepted students) and *diversity* for schools, as well as school-driven interventions such as interviews. In the notation of the two-group model in Section 2.2, we define the *utility* $u_\gamma(s)$ of a school s under matching $\gamma \in \{\mu, \hat{\mu}\}$, as

$$u_\gamma(s) \coloneqq \int_{\theta \in \gamma^{-1}(s) \cap G_1} Z(\theta) dF_1(\hat{Z}(\theta)) + \int_{\theta \in \gamma^{-1}(s) \cap G_2} Z(\theta) dF_2(\hat{Z}(\theta)). \tag{B.1}$$

That is, the utility of a school is the average true potential of admitted students. Continuing from Example 20, let s_M (resp. s_L) be the school Maya (resp. Lisa) is assigned to in the biased setting. Following Proposition 109, the utilities of s_M, s_L in the unbiased setting are $u_\mu(s_M) = 1.283$ and $u_\mu(s_L) = 1.324$, while in the biased setting, they are $u_{\hat{\mu}}(s_M) = 1.280$ and $u_{\hat{\mu}}(s_L) = 1.320$. Hence, the change in the utilities of the two school between the two settings is negligible. We develop the theory to validate these observations in this appendix.

We discuss first the impact of bias on the average true potential of students accepted by a school. Let $s \in [0,1]$ denote the school that is ranked in the $s \times 100\%$ position among the continuous range of schools. As the next proposition shows, the impact on the utilities of schools is negligible for all schools other than the lowest ranked schools. This is because for each school, although the average potential of assigned G_1 students is lower than it should be, its assigned G_2 students have much higher true potentials. And thus, the toll on the utility due to unqualified G_1 students is partially canceled out by the overqualified G_2 students and the net effect is minimal. On the other hand, some lower ranked schools that only admit G_2 students fare better in the biased setting (since they admit over-qualified G_2 candidates).

Proposition 109. *For school s, its utility under the unbiased (resp. biased) models are respectively*

$$u_\mu(s) = s^{-\frac{1}{\alpha}} \quad \text{and} \quad u_{\hat{\mu}}(s) = \begin{cases} \dfrac{1-p+p\beta^\alpha}{1-p+p\beta^{\alpha+1}} \left(\dfrac{s}{1-p+p\beta^\alpha}\right)^{-\frac{1}{\alpha}} & \text{if } s \le 1-p+p\beta^\alpha, \\ \left(\dfrac{s-(1-p)}{p}\right)^{-\frac{1}{\alpha}} & \text{if } s > 1-p+p\beta^\alpha. \end{cases}$$

The key idea in the proof is to first compute the *cutoffs* at each school for each of the two groups, that is, the minimum *perceived* potential needed for a student to be matched to a given school. Once these are known, using Bayes' rule, we deduce the minimum *real* potential needed by students of each group to attend the school. From the latter, we can immediately compute the average utility of each school.

Proof. In order for a student θ to be assigned to a school that is at least as good as s, their perceived potential $\hat{Z}(\theta)$ needs to be high enough to satisfy $(1-p)\bar{F}_1(1 \vee \hat{Z}(\theta)) + p\bar{F}_2(\hat{Z}(\theta)) \le s$. That is, we need

$$\hat{Z}(\theta) \ge d(s) := \begin{cases} \left(\dfrac{s}{1-p+p\beta^\alpha}\right)^{-\frac{1}{\alpha}} & \text{if } s \le 1-p+p\beta^\alpha, \\ \beta\left(\dfrac{s-(1-p)}{p}\right)^{-\frac{1}{\alpha}} & \text{if } s > 1-p+p\beta^\alpha. \end{cases}$$

We call $d(s)$ the *cutoff* for school s. With the cutoffs, we can compute the utilities of schools. We start with the formula for $u_{\hat{\mu}}(s)$. First note that by Bayes rule, the probability that a given student with perceived potential $\hat{Z}(\theta) \geq 1$ belongs to G_1 is $\frac{1-p}{1-p+p\beta^{\alpha+1}}$. Using Equation (2.1), observe that the G_2 student whose perceived potential is 1 (i.e., true potential is $\frac{1}{\beta}$) is matched to school $1 - p + p\beta^{\alpha}$. Thus, if $s \geq 1 - p + p\beta^{\alpha}$, s is only assigned with G_2 students. Therefore, when $s \leq 1 - p + p\beta^{\alpha}$,

$$u_{\hat{\mu}}(s) = \frac{1-p}{1-p+p\beta^{\alpha+1}} d(s) + \frac{p\beta^{\alpha+1}}{1-p+p\beta^{\alpha+1}} \frac{d(s)}{\beta} = \frac{1-p+p\beta^{\alpha}}{1-p+p\beta^{\alpha+1}} \left(\frac{s}{1-p+p\beta^{\alpha}} \right)^{-\frac{1}{\alpha}}.$$

And when $s > 1 - p + p\beta^{\alpha}$, we have

$$u_{\hat{\mu}}(s) = d(s)/\beta = \left(\frac{s-(1-p)}{p} \right)^{-\frac{1}{\alpha}}.$$

One the other hand, when there is no bias against G_2 students, we simply have $u_{\mu}(s) = s^{-\frac{1}{\alpha}}$. □

As one readily observes from Proposition 109, the negative impact of bias on schools' utility is negligible. Hence, from an operational perspective, it may be hard to convince schools to autonomously put in place mechanisms to alleviate the effect of bias given the limited impact on them.

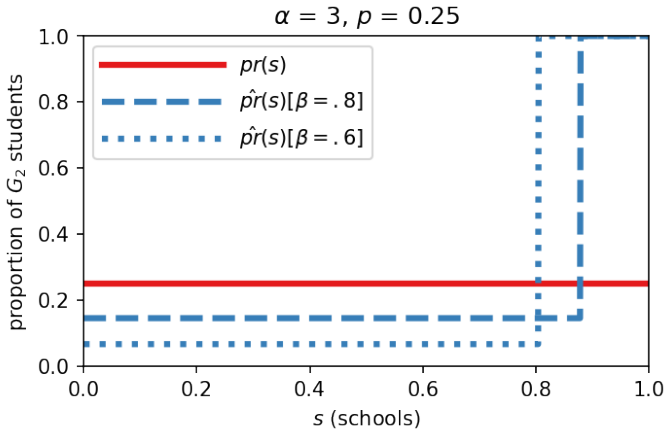


Figure B.1: Proportion of G_2 students in higher ranked schools decreases significantly in the biased setting.

Let $pr(s)$ (resp. $\hat{pr}(s)$) be the proportion of G_2 students assigned to school s when there is no bias (resp. there is bias) against G_2 students. Since the distribution of potentials is the same for both G_1 and G_2 students, it is immediate that $pr(s) = p$ when there is no bias.

Proposition 110. *Without bias, we have $pr(s) = p$. Under the biased setting, we have*

$$\hat{pr}(s) = \begin{cases} \dfrac{p\beta^\alpha}{1-p+p\beta^\alpha} & \text{if } s \leq 1-p+p\beta^\alpha, \\ 1 & \text{if } s > 1-p+p\beta^\alpha. \end{cases}$$

Proof. The formula for $\hat{pr}(s)$ follows from the analysis of utility of schools in Proposition 109. □

A visual comparison of $pr(s)$ and $\hat{pr}(s)$ can be found in Figure B.1 for different values of β and p. In particular, we show that the proportion of G_2 students in higher ranked schools decreases significantly in the biased setting.

B.4 Proof of Theorem 23 and related facts

B.4.1 Technical discussion

The main idea of the proof is to first assume that the set T forms a connected set (i.e., a closed interval). Then, we can express $mm(\mu_T)$ as a function of the endpoints of T and work out the minimizing interval. We next drop the assumption that T is connected and show that the optimal set of students to debias remains the same. The analysis we give is actually more general, and presents results under which vouchers improve the mistreatment of students *lexicographically*. Interestingly, it also shows that, if vouchers are not distributed carefully, one may actually *worsen the most mistreated students*.

B.4.2 A more general approach

The analysis we give is actually leads to a more general statement than Theorem 23, and has the goal of investigating conditions under which giving vouchers can improve over the status quo. More formally, for bounded functions $f, g : G_2 \to \mathbb{R}$, we write $f \succ g$ if we can partition G_2 in two sets S, S' (with possibly $S' = \emptyset$) so that $f(\theta) = g(\theta)$ for $\theta \in S'$ and $\sup_{\theta \in S} f(\theta) > \sup_{\theta \in S} g(\theta)$. Note that $\succ$ is transitive and antisymmetric, and can be interpreted as a continuous equivalent of the classical lexicographic ordering for discrete vectors. In particular, if we let $f = \gamma - \mu$ and $g = \gamma' - \mu$ for matchings γ, γ', then $\sup_{\theta \in G_2}(\gamma - \mu)(\theta) > \sup_{\theta \in G_2}(\gamma' - \mu)(\theta)$ implies $f \succ g$ (taking $S = G_2$). Now suppose we debias student in $T = [Z_1, Z_2]$ for some $T \in \mathcal{T}(\hat{c})$, and let $f := \hat{\mu} - \mu$, $g := \mu_T - \mu$. Table B.1 provides conditions under which $f \succ g$ (i.e., intervention reduces the maximum mistreated experienced by G_2 students). In particular it shows that for certain combinations of the data and the choice of Z_1 and Z_2, giving vouchers may actually lead to worse (according to $\succ$) matchings. One can check that under assumption $p < 1 - \beta^\alpha$, all conditions given in Table B.1 for different cases are satisfied.

CASE	subcase	condition for $\hat{\mu} - \mu \succ \mu_T - \mu$
I. $\beta Z_2 \geq Z_1$	1. $1 \leq \beta Z_1$	$p < 1 - \left(\frac{Z_1}{Z_2}\right)^\alpha$
	2. $\beta Z_1 \leq 1 \leq \beta Z_2$	$p < 1 - \left(\frac{1}{\beta Z_2}\right)^\alpha$
II. $\beta Z_2 \leq Z_1$	1. $1 \leq \beta Z_1$	$p < 1 - \beta^\alpha$
	2. $\beta Z_1 \leq 1 \leq \beta Z_2$	$p\left(\left(\frac{1}{Z_1}\right)^\alpha - \left(\frac{1}{Z_2}\right)^\alpha\right) < (1-p)\left(\frac{1}{\beta^\alpha} - 1\right)\left(\beta^\alpha - \left(\frac{1}{Z_2}\right)^\alpha\right)$
	3. $\beta Z_2 \leq 1$	Not possible: $g \succ f$ in this case.

Table B.1: Sufficient conditions for $\hat{\mu} - \mu \succ \mu_T - \mu$ by cases, where $T = [Z_1, Z_2]$. Each strict inequality, when replaced with its non-strict counterpart, gives instead a necessary condition.

In this first part of the proof, we proceed as follows. First, we assume that $T \in \mathcal{T}^c(\hat{c})$. That is, we assume $T = [Z_1, Z_2]$ with extreme points $Z_1 < Z_2$. For simplicity, we let $\tilde{\mu}$

denote μ_T. We then compare $f := \hat{\mu} - \mu$ and $g := \tilde{\mu} - \mu$ using the relation $\succ$.

Note that, if we let S be the set of students in G_2 with potential in $[Z_1, Z_2/\beta]$ and $S' := G_2 \setminus S$, we have $f(\theta) = g(\theta)$ for $\theta \in S'$. That is, only G_2 students whose true potential lies in interval $[Z_1, Z_2/\beta]$ are affected by the intervention. Hence, $\sup_{\theta \in S} f > \sup_{\theta \in S} g$ if and only if $f \succ g$. We divide the analysis in the following two major cases: the first case is when $\beta Z_2 \geq Z_1$ (i.e., when $[\beta Z_1, \beta Z_2]$ and $[Z_1, Z_2]$ overlap) and the second case is when $\beta Z_2 \leq Z_1$. For both major cases, we will consider two subcases: $\beta Z_1 \geq 1$, $\beta Z_1 \leq 1 \leq \beta Z_2$. And for the second major case, we also need to consider the subcase where $\beta Z_2 \leq 1$. The results for all cases are summarized in the Table B.1.

Observation 111. *If there is an interval $[Z_1, Z_2]$ that is of either case I.2 or case II.2 such that $\mu_{[Z_1,Z_2]} - \mu \prec \hat{\mu} - \mu$ with $S = G_2$, then the optimal range must be of case I.2 or case II.2. This is because for any interval $[Z_1', Z_2']$ that is not of case I.2 or case II.2, we have*

$$\sup_{\theta \in \Theta}\{(\mu_{[Z_1',Z_2']} - \mu)\} \geq \sup_{\theta \in \Theta}\{\hat{\mu} - \mu\} > \sup_{\theta \in \Theta}\{\mu_{[Z_1,Z_2]} - \mu\}.$$

As it turns out, indeed, the optimal range will be either case I.2 or case II.2, and exactly which one the optimal solution is depends on the amount of resources, i.e., the value of $\hat{c}$.

We now show the first half of Theorem 23, i.e., we assume $\hat{c} \geq \frac{(1-p)(1-\beta^\alpha)}{1-p+1-\beta^\alpha}$. The proof steps are outlined below. Each step can be shown by simple algebra and is thus omitted.

(1). We first show that $[Z_1^*, Z_2^*]$ is of case I.2. That is, we show $\beta Z_2^* \geq Z_1^*$ and $Z_1^* \leq \frac{1}{\beta} \leq Z_2^*$.

By writing out the formula for $\mu_{[Z_1,Z_2]} - \mu$, one can see that for an interval $[Z_1, Z_2]$ of case I.2 or case II.2, $\mu_{[Z_1,Z_2]} - \mu$ increases on $[1, Z_1]$, deceases on $[Z_2, \infty]$, and it is non-positive on $[Z_1, Z_2]$. This means $\sup_{\theta \in \Theta}\{\mu_{[Z_1,Z_2]} - \mu\}$ is achieved either at Z_1 or Z_2.

(2). Next, we show that $[Z_1^*, Z_2^*]$ is an *exact* range, that is, $(\frac{1}{Z_1^*})^\alpha - (\frac{1}{Z_2^*})^\alpha = \hat{c}$. Moreover, let θ_1^* and θ_2^* be the G_2 students whose potentials are Z_1^* and Z_2^* respectively. Then, $(\mu_{[Z_1,Z_2]} - \mu)(\theta_1^*) = (\mu_{[Z_1,Z_2]} - \mu)(\theta_2^*)$ and thus, they are both equal to $\sup_{\theta \in \Theta}\{\mu_{[Z_1,Z_2]} - \mu\}$.

Together with the assumption $p < 1 - \beta^\alpha$, we have $\sup_{\theta\in\Theta}\{\mu_{[Z_1^*,Z_2^*]} - \mu\} \leq \sup_{\theta\in\Theta}\{\hat{\mu} - \mu\}$. Thus, due to Observation 111, it is sufficient to compare $[Z_1^*, Z_2^*]$ only with intervals $[Z_1, Z_2]$ of case I.2 and case II.2 (i.e, when $\beta Z_1 \leq 1 \leq \beta Z_2$). Since $[Z_1^*, Z_2^*]$ is exact, we must either have $Z_1 > Z_1^*$ or $Z_2 < Z_2^*$.

(3). Lastly, we show that for any other feasible range $[Z_1, Z_2]$ of case I.2 or case II.2, we must have $\sup_{\theta\in\Theta}\{\mu_{[Z_1,Z_2]}-\mu\} > \sup_{\theta\in\Theta}\{\mu_{[Z_1^*,Z_2^*]}-\mu\}$. Let θ_1 and θ_2 be the G_2 students whose potentials are Z_1 and Z_2. It suffices to show

i). if $Z_1 > Z_1^*$, then $(\mu_{[Z_1,Z_2]} - \mu)(\theta_1) > (\mu_{[Z_1^*,Z_2^*]} - \mu)(\theta_1^*)$;

ii). if $Z_2 < Z_2^*$, then $(\mu_{[Z_1,Z_2]} - \mu)(\theta_2) > (\mu_{[Z_1^*,Z_2^*]} - \mu)(\theta_2^*)$.

For the second half of the theorem, we will follow similar steps and reasoning, outlined below.

(1). We first show that $[Z_1^*, Z_2^*]$ is of case II.2. That is to show $\beta Z_2^* \leq Z_1^*$ and $Z_1^* \leq \frac{1}{\beta} \leq Z_2^*$.

(2). We check that $[Z_1^*, Z_2^*]$ is an exact range. And let θ_1^* and θ_2^* be the G_2 students whose potentials are Z_1^* and Z_2^* respectively, we want to show that $(\mu_{[Z_1,Z_2]}-\mu)(\theta_1^*) = (\mu_{[Z_1,Z_2]} - \mu)(\theta_2^*)$, which implies that both are $\sup_{\theta\in\Theta}\{\mu_{[Z_1^*,Z_2^*]} - \mu\}$.

(3). We show $\mu_{[Z_1^*,Z_2^*]} - \mu \prec \hat{\mu} - \mu$, which, unlike in the previous case, is not immediate from the assumption $p < 1 - \beta^\alpha$.

Again, due to Observation 111, it is sufficient to compare $[Z_1^*, Z_2^*]$ only with regions $[Z_1, Z_2]$ of case I.2 and case II.2 (i.e, when $\beta Z_1 \leq 1 \leq \beta Z_2$).

(4). As before, we will show two cases, which is enough because $[Z_1^*, Z_2^*]$ is exact and one of the two cases is bound to happen. Again, let θ_1 and θ_2 be the G_2 students whose potentials are Z_1 and Z_2 respectively. We want to show

i). if $Z_1 > Z_1^*$, then $(\mu_{[Z_1,Z_2]} - \mu)(\theta_1) > (\mu_{[Z_1^*,Z_2^*]} - \mu)(\theta_1^*)$,

ii). otherwise, we must have $Z_2 < Z_2^*$, and then $(\mu_{[Z_1,Z_2]} - \mu)(\theta_2) > (\mu_{[Z_1^*,Z_2^*]} - \mu)(\theta_2^*)$.

Now let $T^* \in \mathcal{T}(\hat{c})$ be the optimal solution without the restriction that sets in $\mathcal{T}(\hat{c})$ are connected. We will show that T^* differs from $[Z_1^*, Z_2^*]$ in a set of measure zero. First, in order to have $\sup(\mu_{T^*} - \mu) \leq \sup(\mu_{[Z_1^*,Z_2^*]} - \mu) =: s$, in T^*, we must debias all students θ whose mistreatment $(\hat{\mu} - \mu)(\theta)$ is greater than s. That is, we must have $T_1^* := [Z_1^*, Z^{(1)}] \subseteq T^*$, where $Z^{(1)} := Z(\theta^{(1)}) \geq 1/\beta$ and $(\hat{\mu} - \mu)(\theta^{(1)}) = s$. There is a G_2 student $\theta^{(2)}$ such that $Z^{(2)} := Z(\theta^{(2)}) > Z^{(1)}$ and $(\mu_{T_1^*} - \mu)(\theta^{(2)}) = s$. We have moreover that $(\mu_{T_1^*} - \mu)(\theta) \geq s$ for all $\theta \in G_2$ such that $Z(\theta) \in [Z^{(1)}, Z^{(2)}]$. Thus, we must also have $[Z^{(1)}, Z^{(2)}] \in T^*$. Let $T_2^* := [Z_1^*, Z^{(2)}]$. We can repeat the argument and observe that there is a G_2 student $\theta^{(3)}$ such that $Z^{(3)} := Z(\theta^{(3)}) > Z^{(2)}$ and $(\mu_{T_2^*} - \mu)(\theta) \geq s$ for $\theta \in G_2$ such that $Z(\theta) \in [Z^{(2)}, Z^{(3)}]$ and conclude that $T_3^* := [Z_1^*, Z^{(3)}]$ must be contained in T^*. Continuously applying the same argument, we have $\lim_{n\to\infty} Z(\theta^{(n)}) = Z_2^*$ and thus the claim follows.

B.5 Proof of Theorem 24 and related facts

Assume $T = [Z_1, Z_2]$ is the range of true potentials of G_2 students we want to debias. For simplicity, as in previous sections, let $\tilde{\mu}$ denote μ_T. In order to obtain the minimizer of $\sigma(\tilde{\mu} - \mu)$, first, we want to compute $\sigma(\tilde{\mu} - \mu)$ for each of the cases in Table B.1.

For $1 \leq t_1 \leq t_2 \in \mathbb{R} \cup \{+\infty\}$, let $\sigma_{t_1}^{t_2}(f) := \int_{t_1}^{t_2} \max(f(t), 0) dF_1(t)$ for any function $f : [1, \infty] \to [0, 1]$. When $t_1 = 1$ and $t_2 = \infty$, we simply write $\sigma(f)$, which is consistent with previous notations. Note that with $\sigma(\hat{\mu} - \mu)$ as a reference, it actually suffices to compute only $\sigma_{Z_1}^{Z_2/\beta}(\tilde{\mu} - \mu)$, because minimizing $\sigma(\tilde{\mu} - \mu)$ is equivalent to maximizing $\sigma_{Z_1}^{Z_2/\beta}(\hat{\mu} - \mu) - \sigma_{Z_1}^{Z_2/\beta}(\tilde{\mu} - \mu)$ since $(\hat{\mu} - \mu)(\theta) = (\tilde{\mu} - \mu)(\theta)$ for all $\theta \in G_2$ with $Z(\theta) \notin [Z_1, Z_2/\beta]$.

For each case, we give an explicit formula for $\sigma_{Z_1}^{Z_2/\beta}(\hat{\mu} - \mu) - \sigma_{Z_1}^{Z_2/\beta}(\tilde{\mu} - \mu)$. These formulae can be computed via simply integration, and are thus omitted. In addition, we analyze how this value changes (increase or decrease) with respect to Z_1 and Z_2.

CASE I – Subcase 1. After integrating, we have

$$\sigma_{Z_1}^{Z_2/\beta}(\hat{\mu}-\mu)-\sigma_{Z_1}^{Z_2/\beta}(\tilde{\mu}-\mu)=\frac{(1-p)\left(\frac{1}{\beta^\alpha}-1\right)}{2}\left(\frac{1}{Z_1}\right)^{2\alpha}+\frac{p-p\beta^\alpha-\frac{1}{\beta^\alpha}+1}{2}\left(\frac{1}{Z_2}\right)^{2\alpha}.$$

Now, to analyze how this quantity changes with Z_1 and Z_2, we first simplify some of the terms, which will also be used in later sections. Let $x=(\frac{1}{Z_2})^\alpha\in[0,1]$ and let $(\frac{1}{Z_1})^\alpha=c+x\in[0,1]$ for some $c\le\hat{c}$. Also, let $g(x,c):=\sigma_{Z_1}^{Z_2/\beta}(\hat{\mu}-\mu)-\sigma_{Z_1}^{Z_2/\beta}(\tilde{\mu}-\mu)$. Then,

$$g(x,c)=\frac{(1-p)\left(\frac{1}{\beta^\alpha}-1\right)}{2}(c+x)^2+\frac{p-p\beta^\alpha-\frac{1}{\beta^\alpha}+1}{2}x^2.$$

First order conditions (FOC) show that $g(x,c)$ increases as x increases (or equivalently, as Z_2 decreases) and as c increases (meaning that the constraint $(\frac{1}{Z_1})^\alpha-(\frac{1}{Z_2})^\alpha\le\hat{c}$ is effectively $(\frac{1}{Z_1})^\alpha-(\frac{1}{Z_2})^\alpha=\hat{c}$).

CASE I – Subcase 2. In this case, we have

$$\sigma_{Z_1}^{Z_2/\beta}(\hat{\mu}-\mu)-\sigma_{Z_1}^{Z_2/\beta}(\tilde{\mu}-\mu)=-\frac{1}{2}(1-p)\beta^\alpha+(1-p)\left(\left(\frac{1}{Z_1}\right)^\alpha-\frac{1}{2}\left(\frac{1}{Z_1}\right)^{2\alpha}\right)+\frac{p-p\beta^\alpha-\frac{1}{\beta^\alpha}+1}{2}\left(\frac{1}{Z_2}\right)^{2\alpha}.$$

Now, for the analysis, similarly, write

$$g(x,c)=\text{const}+(1-p)\left((c+x)-\frac{1}{2}(c+x)^2\right)+\frac{p-p\beta^\alpha-\frac{1}{\beta^\alpha}+1}{2}x^2.$$

Then, FOC shows that $g(x,c)$ is an increasing function w.r.t. c, and it is an increasing function w.r.t. x on $[0,h_{\mathrm{I}}(c)]$ and is a decreasing function on $[h_{\mathrm{I}}(c),1]$, where $h_{\mathrm{I}}(c)=\frac{(1-p)(1-c)}{p\beta^\alpha+\frac{1}{\beta^\alpha}-2p}$.

CASE II – Subcase 1. In this case, $\sigma_{Z_1}^{Z_2/\beta}(\hat{\mu}-\mu) - \sigma_{Z_1}^{Z_2/\beta}(\tilde{\mu}-\mu)$ equals to

$$\left(\frac{1}{Z_1}\right)^{2\alpha}\left(\frac{(1-p)\left(\frac{1}{\beta^\alpha}-1\right)+p\beta^\alpha}{2}\right) - \left(\frac{1}{Z_2}\right)^{2\alpha}\left(\frac{(1-p)\left(\frac{1}{\beta^\alpha}-1\right)+p\beta^\alpha}{2}\right) + p\left(\frac{1}{Z_2}\right)^{2\alpha} - p\left(\frac{1}{Z_1}\right)^{\alpha}\left(\frac{1}{Z_2}\right)^{\alpha}.$$

Now, for the analysis, let $A = [(1-p)(\frac{1}{\beta^\alpha}-1)+p\beta^\alpha]/2 \geq 0$. Then, $g(x,c) = A(c+x)^2 - Ax^2 + px^2 - p(c+x)(x)$. Checking the FOCs, we have that $g(x,c)$ is an increasing function w.r.t. c and w.r.t. x.

CASE II – Subcase 2. We have

$$\begin{aligned}\sigma_{Z_1}^{Z_2/\beta}(\hat{\mu}-\mu) - \sigma_{Z_1}^{Z_2/\beta}(\tilde{\mu}-\mu) = &-\frac{1}{2}(1-p)\beta^\alpha + \left(\frac{1}{Z_1}\right)^{2\alpha}\left(\frac{-(1-p)+p\beta^\alpha}{2}\right) + (1-p)\left(\frac{1}{Z_1}\right)^{\alpha} \\ &- \left(\frac{1}{Z_2}\right)^{2\alpha}\left(\frac{(1-p)\left(\frac{1}{\beta^\alpha}-1\right)+p\beta^\alpha}{2}\right) + p\left(\frac{1}{Z_2}\right)^{2\alpha} - p\left(\frac{1}{Z_1}\right)^{\alpha}\left(\frac{1}{Z_2}\right)^{\alpha}.\end{aligned}$$

For the analysis, again let $A = [(1-p)(\frac{1}{\beta^\alpha}-1)+p\beta^\alpha]/2 \geq 0$ and $B = [-(1-p)+p\beta^\alpha]/2 < 0$. Then,

$$g(x,c) = \text{const} + B(c+x)^2 + (1-p)(c+x) - Ax^2 + px^2 - p(c+x)(x),$$

and for c, it is an increasing function; and for x, it is an increasing function on $[0, h_{\text{II}}(c)]$ and is a decreasing function on $[h_{\text{II}}(c), 1]$, where

$$h_{\text{II}}(c) = \frac{(p\beta^\alpha - 1)c + (1-p)}{(1-p)\frac{1}{\beta^\alpha}}.$$

CASE II – Subcase 3. Lastly, we have that $\sigma_{Z_1}^{Z_2/\beta}(\hat{\mu}-\mu) - \sigma_{Z_1}^{Z_2/\beta}(\tilde{\mu}-\mu)$ equals to

$$\left(\frac{1}{Z_1}\right)^{2\alpha} B + (1-p)\left(\frac{1}{Z_1}\right)^{\alpha} - \left(\frac{1}{Z_2}\right)^{2\alpha} B - (1-p)\left(\frac{1}{Z_2}\right)^{\alpha} + p\left(\frac{1}{Z_2}\right)^{2\alpha} - p\left(\frac{1}{Z_1}\right)^{\alpha}\left(\frac{1}{Z_2}\right)^{\alpha}.$$

For the analysis, write $g(x,c) = B(c+x)^2 + (1-p)(c+x) - Bx^2 - (1-p)x + px^2 - p(c+x)x$. $g(x,c)$ is a decreasing function in x. The sign of $\frac{\partial g(x,c)}{\partial c}$ is actually not clear in this subcase. But for the purpose of finding the minimizer of $\sigma(\tilde{\mu}-\mu)$, this is not important because for

a fixed value of c, $g(x, c)$ achieves its maximum when x is of the value such that $[Z_1, Z_2]$ is of subcase 2, of either case I or case II.

Not that for a fixed value of $\hat{c}$, as Z_1 gets larger (or equivalently as Z_2 gets larger, or as $x := (\frac{1}{Z_2})^\alpha$ gets smaller), the range $[Z_1, Z_2]$ goes from case II to case I. In particular, for each value of $\hat{c}$, such transition happens exactly when $\beta Z_2 = Z_1$. That is, when

$$\hat{c} = \left(\frac{1}{\beta^\alpha} - 1\right)\left(\frac{1}{Z_2}\right)^\alpha \quad \Leftrightarrow \quad \left(\frac{1}{Z_2}\right)^\alpha = \frac{\hat{c}\beta^\alpha}{1-\beta^\alpha}.$$

With simple algebra, one can easily check that

$$\hat{c} = \frac{(1-p)(1-\beta^\alpha)}{2-p-\beta^\alpha-p\beta^\alpha+p\beta^{2\alpha}} \quad \Rightarrow \quad h_{\mathrm{I}}(\hat{c}) = h_{\mathrm{II}}(\hat{c}) = \frac{\hat{c}\beta^\alpha}{1-\beta^\alpha}.$$

Therefore, when

$$\hat{c} \geq \frac{(1-p)(1-\beta^\alpha)}{2-p-\beta^\alpha-p\beta^\alpha+p\beta^{2\alpha}},$$

both $h_{\mathrm{I}}(\hat{c})$ and $h_{\mathrm{II}(\hat{c})}$ are no more than $\dfrac{\hat{c}\beta^\alpha}{1-\beta^\alpha}$. Thus, the maximum value of $\sigma(\hat{\mu}-\mu) - \sigma(\tilde{\mu}-\mu)$ is achieved when $x = h_{\mathrm{I}}(\hat{c})$; and when $\hat{c} \leq \frac{(1-p)(1-\beta^\alpha)}{2-p-\beta^\alpha-p\beta^\alpha+p\beta^{2\alpha}}$, both $h_{\mathrm{I}}(\hat{c})$ and $h_{\mathrm{II}(\hat{c})}$ are no less than $\dfrac{\hat{c}\beta^\alpha}{1-\beta^\alpha}$. Thus, the maximum value of $\sigma(\hat{\mu}-\mu)-\sigma(\tilde{\mu}-\mu)$ is achieved when $x = h_{\mathrm{II}}(\hat{c})$.

B.6 Proofs from Section 2.5.1

B.6.1 Auxiliary results for Section 2.5.1

Recall that, in this section, we consider the generalization of the model from Section 2.2 where students' true potential follow a generic continuous, integrable cdf F. Moreover, we write $[x]^+ := \max(0, x)$ for a number or a function x. Recall that, similarly to Section 2.5.1, we abuse notation and identify a student θ with their potential $Z(\theta)$.

Budget	Multiplicative			Additive			Difference	
$\hat{c}$	PAUC	MM	Difference	PAUC	MM	Difference	PAUC	MM
0.1	44.4%	45.2%	0.9%	45.4%	46.1%	0.7%	1.1%	0.9%
0.2	37.6%	39.3%	1.7%	39.7%	41.1%	1.4%	2.1%	1.8%
0.3	30.8%	33.3%	2.5%	34.1%	35.9%	1.8%	3.3%	2.6%
0.4	26.4%	28.5%	2.0%	30.2%	31.8%	1.6%	3.7%	3.3%
0.5	22.0%	23.7%	1.7%	26.0%	27.4%	1.4%	4.0%	3.7%
0.6	17.6%	19.0%	1.4%	21.6%	22.8%	1.2%	4.0%	3.9%
0.7	13.2%	14.2%	1.0%	17.0%	18.0%	0.9%	3.8%	3.7%
0.8	8.8%	9.5%	0.7%	12.1%	12.7%	0.7%	3.3%	3.3%

Table B.2: Proportion of disadvantaged students above theoretically optimal debiasing ranges under multiplicative/additive models and PAUC/maximum mistreatment aggregate mistreatment measures for $\alpha = 3$ and $p = 1/4$, with $\beta = 0.8$ for multiplicative models and $\gamma = 0.252$ for additive. For instance, the first value, 44.4% indicates that under budget $\hat{c} = 0.1$ and the multiplicative model, the top-44.4% to 54.4% of students were debiased.

Lemma 112. *Let ρ be an RVP. Under any continuous distribution of potentials F, we have*

$$\mu_\rho(\theta) = \rho(\theta)\left((1-p)\int_\theta^\infty dF + p\left[\int_\theta^{\theta/\beta}\rho\, dF + \int_{\theta/\beta}^\infty dF\right]\right) + (1-\rho(\theta))\cdot \left((1-p)\int_{\beta\theta}^\infty dF + p\left[\int_{\beta\theta}^{\theta}\rho\, dF + \int_\theta^\infty dF\right]\right), \tag{B.2}$$

$$m_\rho(\theta) = [0, (1-\rho(\theta))(1-p)\int_{\beta\theta}^{\theta} dF + p\left[(1-\rho(\theta))\int_{\beta\theta}^{\theta}\rho\, dF - \rho(\theta)\int_\theta^{\theta/\beta}(1-\rho)\, dF\right]]^+. \tag{B.3}$$

Proof. Suppose a student appears to have potential τ, possibly after having been debiased. Then under μ_ρ, they will be matched to school $s(\tau)$ given by

$$s(\tau) = (1-p)\int_\tau^\infty dF + p\left[\int_\tau^{\tau/\beta}\rho\, dF + \int_{\tau/\beta}^\infty dF\right],$$

that is, they will appear after all non-disadvantaged students with true potential exceeding τ; those disadvantaged students with potential exceeding τ/β; and those disadvantaged students who receive a voucher and have potential in the interval $(\tau, \tau/\beta)$.

A student with true potential θ now receives a voucher with probability $\rho(\theta)$, so by

the law of total expectation, we have $\mu_\rho(\theta) = \rho(\theta)s(\theta) + (1-\rho(\theta))s(\beta\theta)$, which is exactly (B.2). (B.3) follows from (B.2) and the definitions of displacement and $\mu(\theta)$. □

We next report more useful expressions for μ_ρ and μ'_ρ.

Proposition 113. *Let ρ be an RVP. For all $\theta \in \Theta$, we have*

$$\begin{aligned}\mu_\rho(\theta) = -\rho(\theta)\left((1-p)\int_{\beta\theta}^{\theta} dF + p\left[\int_{\theta}^{\theta/\beta}(1-\rho)\,dF + \int_{\beta\theta}^{\theta}\rho\,dF\right]\right)\\ + \left((1-p)\int_{\beta\theta}^{\infty} dF + p\left[\int_{\beta\theta}^{\theta}\rho\,dF + \int_{\theta}^{\infty} dF\right]\right).\end{aligned} \tag{B.4}$$

Moreover, if μ_ρ is differentiable at θ, we have

$$\begin{aligned}\mu'_\rho(\theta) = -f(\theta) - \rho'(\theta)\left[p\int_{\theta}^{\theta/\beta}(1-\rho)\,dF + (1-p)\int_{\beta\theta}^{\theta} dF + p\int_{\beta\theta}^{\theta}\rho\,dF\right]\\ - p\rho(\theta)\left[\frac{1}{\beta}(1-\rho(\theta/\beta))f(\theta/\beta) - (1-\rho(\theta))f(\theta)\right]\\ + (1-\rho(\theta))\left[(1-p)(f(\theta)-\beta f(\beta\theta)) + p(f(\theta)\rho(\theta) - \beta f(\beta\theta)\rho(\beta\theta))\right].\end{aligned} \tag{B.5}$$

Proof. (B.4) follows by simple rearrangement of (B.2), and (B.5) follows by standard mechanics of derivative computation. □

Definition 114. *The RVP that assigns no vouchers, denoted ρ_0 is defined by $\rho_0(\theta) := 0$ for all $\theta \in [1,\infty)$. Note that $\mu_{\rho_0}(\theta) = (1-p)\int_{\beta\theta}^{\infty} dF + p\int_{\theta}^{\infty} dF$ and $m_{\rho_0}(\theta) = m_{\hat{\mu}}(\theta) = (1-p)\int_{\beta\theta}^{\theta} dF$.*

B.6.2 Necessary and sufficient conditions for incentive compatibility

In this section we develop necessary and sufficient conditions for incentive compatibility through the concept of well-behavedness and prove an important technical lemma.

Definition 115 (Well-behaved RVP)**.** *We call an RVP ρ* well-behaved *if it is everywhere continuously differentiable except for a set of isolated points where it has non-negative, right-continuous jump discontinuities.*

Lemma 116 (Necessary and sufficient conditions for incentive compatibility)**.** *Let ρ be a well-behaved RVP and F be an arbitrary continuous distribution of potentials. ρ is incentive compatible with respect to F if and only if, for all θ such that ρ is continuously differentiable at θ, we have $\rho'(\theta) \geq 0$ or $\mu'_\rho(\theta) \leq 0$.*

Proof. Recall that ρ is incentive compatible if μ_ρ is everywhere non-increasing. Observe from (B.4) in Proposition 113 that μ_ρ is continuous at θ if and only if ρ is continuous at θ. On the other hand, if μ_ρ is not continuous at θ then it must have a negative jump-discontinuity caused by a positive jump-discontinuity of ρ (since all other terms of (B.4) are positive). Further note that if μ_ρ is not continuously differentiable at $\theta \in \Theta$, then ρ is not continuously differentiable at θ, $\beta\theta$ or θ/β; so the set of points where μ_ρ is not continuously differentiable also forms an isolated set.

Let $\theta \in \Theta$ where μ_ρ is continuously differentiable. Then, μ_ρ is non-increasing if and only if $\mu'_\rho(\theta) \leq 0$. From (B.5), one can see that $\mu'_\rho(\theta) \leq 0$ if $\rho'(\theta) \geq 0$.

We have established that μ_ρ is continuous at all but an isolated set of negative jump-discontinuities, and that μ_ρ is continuously differentiable and non-increasing at all but an isolated set of points. μ_ρ is therefore everywhere non-increasing, as required. □

Lemma 117. *Suppose ρ is a well-behaved RVP such that for all θ where ρ is continuously differentiable, we have $\rho'(\theta) \geq -\phi(\theta)$, with $\phi(\theta) := \frac{\alpha(1-p)}{\theta[p(1-\beta^\alpha)+(1-p)(\beta^{-\alpha}-1)]}$. Then, ρ is incentive compatible.*

Proof. By Lemma 116, it suffices to show that $\mu'_\rho(\theta) \leq 0$ for θ such that $\rho'(\theta)$ exists and is continuous, and $\rho'(\theta) < 0$. Now define

$$\mathcal{L} = p\int_\theta^{\theta/\beta}(1-\rho)\,dF + (1-p)\int_{\beta\theta}^{\theta} dF + p\int_{\beta\theta}^{\theta}\rho\,dF \text{ and } W = -\rho(\theta)(1-p)f(\theta) - (1-\rho(\theta))(1-p)\beta f(\beta\theta),$$

and note that $\mathcal{L} \geq 0$, and $W \leq 0$. Simple calculations based on (B.5) in Proposition 113

shows that $\mu'_\rho(\theta) \leq -\rho'(\theta)\mathcal{L} + W$. It is therefore enough to prove $-\rho'(\theta) \leq \frac{-W}{\mathcal{L}}$. Compute

$$\mathcal{L} \leq p\int_\theta^{\theta/\beta} dF + (1-p)\int_{\beta\theta}^{\theta} dF \leq \theta^{-\alpha}\left[p(1-\beta^\alpha) + (1-p)(\beta^{-\alpha}-1)\right] \text{ and } -W \geq (1-p)\min\{f(\theta), \beta f(\beta\theta)\} = \frac{\alpha(1-}{\theta^{1+\alpha}}$$

This yields $\frac{-W}{\mathcal{L}} \geq \frac{\alpha(1-p)}{\theta[p(1-\beta^\alpha)+(1-p)(\beta^{-\alpha}-1)]} = \phi(\theta)$.We have shown $\frac{-W}{\mathcal{L}} \geq \phi(\theta)$, which combined with the assumption that $\phi(\theta) \geq -\rho'(\theta)$ completes the proof. □

B.6.3 Proof of Theorem 26: properties of PropMs

We next prove Lemmas 118, 119, and 122, which together constitute Theorem 26.

Lemma 118. *The proportional-to-mistreatment RVP ρ_m is $\frac{2\hat{c}\alpha}{1-\beta^\alpha}$-individually fair.*

Proof. ρ_m is everywhere continuous and continuously differentiable on Θ, except at $\theta = 1/\beta$. ρ_m is therefore Lipschitz for a constant given by the supremum of the absolute value of the derivative, which occurs at $\theta = 1$ where $\rho'_m(\theta) = \frac{2\hat{c}\alpha}{1-\beta^\alpha}$. □

Lemma 119. *The proportional-to-mistreatment RVP ρ_m is incentive compatible for*

$$\hat{c} \leq \frac{1-p}{2\left[p(1-\beta^\alpha) + (1-p)(\beta^{-\alpha}-1)\right]}.$$

Proof. Applying Lemma 117, it suffices to show

$$-\rho'_m(\theta) = 2\alpha\hat{c}\beta^{-\alpha}\theta^{-\alpha-1} \leq \phi(\theta) = \frac{\alpha(1-p)}{\theta\left[p(1-\beta^\alpha) + (1-p)(\beta^{-\alpha}-1)\right]}. \tag{B.6}$$

for $\theta \geq 1/\beta$ (since $\rho'_m(\theta) \geq 0$ for $\theta < 1/\beta$). Note that this is tightest when $\theta = 1/\beta$, which gives the condition for $\hat{c}$. □

Lemma 120. *For the proportional-to-mistreatment RVP ρ_m, we have*

$$\sup_{\theta\in\Theta}(1-\rho_m(\theta))\int_{\beta\theta}^{\theta} dF = (1-\beta^\alpha)\xi(\hat{c}), \quad \textit{where} \quad \xi(\hat{c}) := \begin{cases} 1-2\hat{c}, \hat{c} \leq 1/4, \\ \frac{1}{8\hat{c}}, \hat{c} > 1/4. \end{cases}$$

Proof. With $F(a,b) := \int_a^b dF$, define $q := \frac{2\hat{c}}{1-\beta^\alpha}$ and $y(\theta) := F(\beta\theta, \theta)$. Now write

$$(1-\rho_m(\theta))\int_{\beta\theta}^{\theta} dF = \left(1 - \frac{2\hat{c}}{1-\beta^\alpha}F(\beta\theta,\theta)\right)F(\beta\theta,\theta) = (1-qy(\theta))y(\theta).$$

This is a quadratic in y that increases from $y = 0$ to its maximum at $y = 1/(2q)$. Observe that $\max_\theta F(\beta\theta,\theta) = 1-\beta^\alpha$ which is attained at $\theta = 1/\beta$. This means that if $\hat{c} \le 1/4$, then

$$y(\theta) = F(\beta\theta,\theta) \le 1-\beta^\alpha \le \frac{1-\beta^\alpha}{4\hat{c}} = \frac{1}{2q},$$

and the maximum of the quadratic over y is realized at the maximum value of y. Thus,

$$\sup_{\theta\in\Theta,\hat{c}\le 1/4}(1-qy(\theta))y(\theta) = \left(1-\frac{2\hat{c}}{1-\beta^\alpha}(1-\beta^\alpha)\right)(1-\beta^\alpha) = (1-\beta^\alpha)(1-2\hat{c}).$$

On the other hand if $\hat{c} > 1/4$, then this expression reaches its maximum when the quadratic does, at $y = 1/(2q)$, giving

$$\sup_{\theta\in\Theta,\hat{c}>1/4}(1-qy(\theta))y(\theta) = \left(1-q\frac{1}{2q}\right)\frac{1}{2q} = \left(1-\frac{1}{2}\right)\frac{1}{2q} = \frac{1-\beta^\alpha}{8\hat{c}}.$$

Combining the two completes the proof. □

Lemma 121. *For the proportional-to-mistreament RVP, ρ_m, the maximum mistreament mm_{ρ_m} satisfies $mm_{\rho_m} \le (1-p(1-2\hat{c}))(1-\beta^\alpha)\xi(\hat{c})$, where $\xi(\cdot)$ is defined as in Lemma 120.*

Proof. Abbreviate $\rho = \rho_m$. Apply $\rho(\theta) \le \rho(1/\beta)$ to (B.3) and simplify to get

$$\begin{aligned} m_{\rho_m}(\theta) &\le \left[(1-p)(1-\rho(\theta))\int_{\beta\theta}^{\theta} dF + p(1-\rho(\theta))\rho(1/\beta)\int_{\beta\theta}^{\theta} dF\right]^+ \\ &= \left[(1+p(\rho(1/\beta)-1))(1-\rho(\theta))\int_{\beta\theta}^{\theta} dF\right]^+. \end{aligned}$$

The thesis follows by taking the supremum over $\theta \in \Theta$, applying Lemma 120, and substituting $\rho(1/\beta) = 2\hat{c}$. □

Recall that we let $mm^*(\hat{c})$ be the maximum mistreatment achieved by the optimal policy from Theorem 23 with the amount of resources being $\hat{c}$. We have

$$mm^*(\hat{c}) = \begin{cases} (1-p-\hat{c})(1-\beta^\alpha) + \hat{c}p & \text{if } \hat{c} \leq \frac{(1-p)(1-\beta^\alpha)}{1-p+1-\beta^\alpha}, \\ (1-p)(1-\beta^\alpha)\frac{1-\hat{c}}{1-p\beta^\alpha} & \text{otherwise.} \end{cases} \tag{B.7}$$

Lemma 122. *Let* $p < \min\{1-\beta^\alpha, 1/2\}$ *and* $1 - \frac{p+1-\beta^\alpha}{4p(1-\beta^\alpha)} \leq \hat{c} \leq \frac{(1-p)(1-\beta^\alpha)}{1-p+1-\beta^\alpha}$. *Then* $mm_{\rho_m} \leq mm^*(\hat{c})$.

Proof. Let $Q := mm^*(\hat{c}) - mm_{\rho_m}$. We need to show $Q \geq 0$. Using Theorem 23 and Lemma 121, compute

$$Q = mm^*(\hat{c}) - mm_{\rho_m} \geq (1-p)(1-\beta^\alpha) - \hat{c}(1-\beta^\alpha - p) - (1+p(2\hat{c}-1))(1-\beta^\alpha)\xi(\hat{c}).$$

For $\hat{c} \leq 1/4$, we now have

$$Q \geq \hat{c}\left[(1-4p(1-\hat{c}))(1-\beta^\alpha) + p\right]. \tag{B.8}$$

If $p \leq \frac{1}{4}$, then the right-hand side of (B.8) is nonnegative, concluding the proof. Thus, assume $p > \frac{1}{4}$. Since $\hat{c} > 0$, we can drop the leading $\hat{c}$, so for $Q \geq 0$, we need $p(1-4(1-\beta^\alpha)(1-\hat{c})) \geq -(1-\beta^\alpha)$. Rearranging leads to the thesis. Consider next the case where $\hat{c} \geq 1/4$. We want to show

$$(1-p)(1-\beta^\alpha) - \hat{c}(1-\beta^\alpha - p) - (1+p(2\hat{c}-1))(1-\beta^\alpha)\frac{1}{8\hat{c}} \geq 0.$$

Since $\hat{c} > 0$, we can multiply by $\hat{c}$ to get a quadratic in $\hat{c}$; call the resulting expression $W(\hat{c})$:

$$W(\hat{c}) = \hat{c}(1-p)(1-\beta^\alpha) - \hat{c}^2(1-\beta^\alpha - p) - (1+p(2\hat{c}-1))(1-\beta^\alpha)\frac{1}{8}.$$

Since $p < 1 - \beta^\alpha$, $W''(\hat{c}) \leq 0$, hence this is a concave quadratic. One can verify that if $p \leq 1/2$ then $W(1/4) \geq 0$ and $W(1/2) \geq 0$, which means that W must also be non-negative for $\hat{c} \in [1/4, 1/2]$, as required. □

B.6.4 Increasing-with-Potential RVPs

Proof. Directly from Lemma 116. □

Proof. We claim that there exists $\delta > 0$ with $\delta < \theta(1 - \beta)$ such that on $I := (\theta - \delta, \theta + \delta)$, the following properties hold for all $t \in I$: ρ is continuous and differentiable at t; ρ is monotonically decreasing at t; and $0 < \rho(t) \leq (1 + \rho(\theta))/2$. The existence of an interval that satisfies the first and second properties follows since ρ is continuously differentiable in some neighborhood of θ and has strictly negative derivative. The third follows since ρ has a strictly negative derivative at θ, so it must be strictly bounded away from 0 and 1 itself. Then, one can restrict δ to guarantee the same for t close to θ. Note also that $I \subset (\beta\theta, \theta/\beta)$.

Next, fix $\varepsilon > 0$, then one can construct a distribution f that satisfies the following conditions: f is continuous and differentiable everywhere; $f(\theta) = \varepsilon$; $f(t) = 0$ for $t \notin I$; and $\int_\theta^{\theta+\delta} f(t)\, dt \geq \frac{1}{2}$. This can be done for instance by constructing a piece-wise constant function that satisfies all but the first condition, then smoothing it out with an appropriate bump function via standard techniques.

From (B.5) and $p(1-\rho(\theta))(1-2\rho(\theta))+\rho(\theta) \in [0,1]$, we compute

$$\begin{aligned}
\mu'_\rho(\theta) &= -\rho'(\theta)\left(p\int_\theta^{\theta/\beta}(1-\rho)\,dF + (1-p)\int_{\beta\theta}^{\theta} dF + p\int_{\beta\theta}^{\theta}\rho\,dF\right) \\
&\quad - \frac{1}{\beta}p\rho(\theta)f\left(\frac{\theta}{\beta}\right)\left(1-\rho\left(\frac{\theta}{\beta}\right)\right) - \beta(1-\rho(\theta))f(\beta\theta)(1-p(1-\rho(\beta\theta))) \\
&\quad - f(\theta)(p(1-\rho(\theta))(1-2\rho(\theta))+\rho(\theta)) \\
&= (-\rho'(\theta))\left(p\int_\theta^{\theta+\delta}(1-\rho)\,dF + (1-p)\int_{\theta-\delta}^{\theta} dF + p\int_{\theta-\delta}^{\theta}\rho\,dF\right) - \varepsilon(p(1-\rho(\theta))(1-2\rho(\theta))+\rho(\theta)) \\
&\geq (-\rho'(\theta))p\int_\theta^{\theta+\delta}(1-\rho)\,dF - \varepsilon \;\geq\; \frac{1}{2}(-\rho'(\theta))p(1-\rho(\theta))\int_\theta^{\theta+\delta} dF - \varepsilon \\
&\geq \frac{1}{4}(-\rho'(\theta))p(1-\rho(\theta)) - \varepsilon. \qquad \text{(B.9)}
\end{aligned}$$

Now the first term in (B.9) is strictly positive, and we can freely choose ε strictly smaller in magnitude to get $\mu'_\rho(\theta) > 0$. Note that although ρ might not be well-behaved everywhere, it is well behaved on I, and we can apply Lemma 116 to this point to get that ρ is not incentive compatible for θ, completing the proof. □

B.7 Impact of model misspecification

In this appendix, we study the robustness of our framework under model misspecification. That is, we investigate the impact of applying our simple model of constant multiplicative bias when the true process by which bias arises is more complicated. In particular, we study additive models and models where idiosyncratic randomness exists within the bias factor or potentials of disadvantaged students. Based on computational experiments we show that our main takeaway holds, and that applying our results would lead to little efficiency loss except in the case of very high randomness.

Setup: We generate simulated data with parameters chosen to match those we fit to our real data. In the language of Section 2.2, all students $\theta \in \Theta$ have a true potential $Z(\theta)$ sampled i.i.d. from a Pareto$(1,\alpha)$ distribution with $\alpha = 9$, and we identify students with

their potential, so we write θ for their true potential. A proportion $p = 0.3$ of students is disadvantaged and they appear at a perceived potential $\hat{Z}(\theta)$, where $\hat{Z}$ is some random variable with $\hat{Z}(\theta) \leq \theta$. We study various models for $\hat{Z}(\cdot)$.

The central planner has a budget $\hat{c}$ and applies our model with $\hat{Z}(\theta) = \beta\theta$ to choose an interval[1] of disadvantaged students to debias as instructed by Theorem 24. We call this the *theoretical* debias interval. Due to randomness within the model, it does not make sense to measure the maximum mistreatment, so we concentrate solely on the positive area under the mistreatment curve (PAUC) introduced in Section 2.4.

We then compare the theoretical interval to the optimal empirical interval if the full bias process were known to the central planner a priori. We compute such an interval using grid search, which we call the *empirical* debias interval.

Models: For each model, we let η be some fixed parameter. We report results for the cases where the true bias process takes each of the following forms:

1. $\hat{Z}(\theta) = \theta - \eta$, a deterministic additive model;
2. $\hat{Z}(\theta) = (\eta + \varepsilon)\theta$ for $\varepsilon \sim \text{Normal}(0, .02)$, minor Gaussian noise in bias factor;
3. $\hat{Z}(\theta) = (\eta + \varepsilon)\theta$ for $\varepsilon \sim \text{Uniform}(-.05, .05)$, minor uniform noise in bias factor;
4. $\hat{Z}(\theta) = \theta - \eta + \varepsilon$ for $\varepsilon \sim \text{Normal}(0, .1)$, additive, medium Gaussian noise in bias factor;
5. $\hat{Z}(\theta) = \eta\theta + \varepsilon$ for $\varepsilon \sim \text{Normal}(0, .1)$, medium Gaussian noise in potential;
6. $\hat{Z}(\theta) = (\eta + \varepsilon)\theta$ for $\varepsilon \sim \text{Uniform}(-.15, .15)$, medium uniform noise in bias factor;
7. $\hat{Z}(\theta) = \eta\theta + \varepsilon$ for $\varepsilon \sim \text{Uniform}(-.3, .3)$, large uniform noise in potential; and
8. $\hat{Z}(\theta) = (\eta + \varepsilon)\theta$ for $\varepsilon \sim \text{Uniform}(-.3, .3)$, large uniform noise in bias factor.

[1]We assume the central planner cannot observe the true potentials and must therefore debias disadvantaged students chosen based on perceived potentials. For the case of uniform multiplicative bias, this makes no difference, but when the bias process has randomness, this is an important detail.

The first model in particular is the additive model studied in Section 2.6, and the fourth model is exactly the statistical discrimination model of [91]. All models except the first (which is deterministic) can be interpreted as adaptations of a statistical discrimination model, as they contain the key feature of increased variance of the disadvantaged group. We choose for each model the fixed parameter η in such a way that if the central planner would apply our theoretical model of constant multiplicative bias, they would fit exactly $\beta = 0.88$ as the Wasserstein metric minimizing parameter. This yields a set of experiments that can be readily compared. The best fit for the η parameter for each model is shown in Table B.3.

Model	1	2	3	4	5	6	7	8
Parameter	0.130	0.878	0.876	0.151	0.860	0.857	0.843	0.848

Table B.3: Best-fits for the model parameter η.

Simulations: We perform experiments with two budgets, $\hat{c} = 0.1$ and $\hat{c} = 0.4$, and run simulations with 1 million students in order to adequately approximate the continuous market. Many of these models increase the variance of the distribution of disadvantaged student scores, so the theoretical interval would often debias a significantly smaller proportion of students than allowed by the budget, naturally leading to a lower PAUC reduction and making comparison difficult. Because of this phenomenon, we fix the upper endpoint of the theoretical interval, but choose the lower endpoint such that it fills up the budget.

Table B.4 and Table B.5 summarize the results for $\hat{c} = 0.1$ and $\hat{c} = 0.4$ respectively. For each model, we report the aggregate mistreatment as measured by the PAUC metric (introduced in Section 2.4) for three cases: no debiasing, under the empirically optimal debiasing, and under the theoretically optimal debiasing. We report the reduction in PAUC given by both debiasing methods as well as their difference.

Model	PAUC			PAUC Reduction		Difference
	No debiasing	Empirical	Theoretical	Empirical	Theoretical	
1	0.2261	0.1870	0.1870	17.28%	17.28%	0.00%
2	0.2398	0.2009	0.2009	16.21%	16.20%	0.00%
3	0.2406	0.2029	0.2029	15.67%	15.65%	0.01%
4	0.2276	0.1978	0.1980	13.12%	13.00%	0.12%
5	0.2406	0.2105	0.2108	12.50%	12.39%	0.11%
6	0.2395	0.2138	0.2149	10.72%	10.27%	0.45%
7	0.2360	0.2048	0.2057	13.23%	12.82%	0.41%
8	0.2337	0.2033	0.2042	13.03%	12.64%	0.38%

Table B.4: Comparison of PAUC reductions between theoretically and empirically optimal intervals for $\hat{c} = 0.1$.

Low Budget: In the small budget case ($\hat{c} = 0.1$, see Table B.4), the difference in using the empirically optimal and the theoretically optimal debiasing intervals is minuscule in every case. In the case of largest difference in PAUC reduction, the empirical interval is able to reduce PAUC by 10.72%, whereas the theoretically optimal interval would have reduced it by 10.27%, a difference of only 0.45%.

Model	PAUC			PAUC Reduction		Difference
	No debiasing	Empirical	Theoretical	Empirical	Theoretical	
1	0.2261	0.0850	0.0850	62.39%	62.38%	0.00%
2	0.2398	0.0994	0.0994	58.56%	58.55%	0.01%
3	0.2406	0.1062	0.1064	55.86%	55.80%	0.06%
4	0.2276	0.1146	0.1194	49.66%	47.54%	2.12%
5	0.2406	0.1254	0.1311	47.87%	45.51%	2.36%
6	0.2395	0.1284	0.1402	46.37%	41.47%	4.90%
7	0.2360	0.1015	0.1202	56.99%	49.08%	7.92%
8	0.2337	0.0991	0.1187	57.62%	49.22%	8.40%

Table B.5: Comparison of PAUC reductions between theoretically and empirically optimal intervals for $\hat{c} = 0.4$.

High Budget: For the large budget case ($\hat{c} = 0.4$, see Table B.5), the difference is more pronounced, yet still limited. The difference is negligible for models 1–3. Medium Gaussian noise in either potential or bias factor causes a small difference in PAUC reduction,

in the order of 2–2.5%. In the case of medium uniform noise in bias factor, the difference starts being more noticeable at 4.9%, and with the cases of large uniform noise in potential or bias, the difference goes to 7.92% and 8.4% respectively.

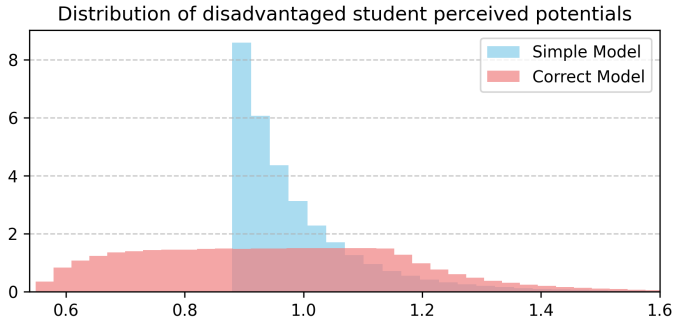


Figure B.2: Difference in disadvantaged student perceived potentials for model 8.

Most Misspecified Case: In the worst case, the absolute difference in PAUC is 0.0196, meaning that on average, disadvantaged students in the correctly specified model achieve a ranking of approximately 2 percentage points higher than if the central planner were to apply the misspecified simpler model. We note that to achieve such a difference, the model has to be highly misspecified: for example, our model would predict the perceived potentials of disadvantaged students to have mean 0.99 and standard deviation 0.125, whereas the correct model in this case has mean 0.95 and standard deviation 0.231. One can see this difference in Figure B.2, one would clearly observe that the model is not appropriate.

Additive Case: We finally remark that our simple model has excellent performance under the case that the true model is any of models 1–3. In particular, in the case of the true model being additive, or the bias factor being slightly idiosyncratic, our model predictions apply virtually unchanged.

We conclude that our results are highly robust to model misspecification.

B.8 Proof of Theorem 30

To simplify notation, we define the following. Recall $F \sim \text{Pareto}(1, \alpha)$ is the distribution of true potentials of all students, and let ϕ be a *bias map* that gives the perceived potential of a disadvantaged student given their true potential. ϕ therefore encodes information both about the nature of bias, as well as any debiasing steps taken. In this case we consider an additive bias model where we debias an interval $S = [Y_1, Y_2]$, so that $\phi(x) = x$ if $x \in S$ and $\phi(x) = x - \gamma$ otherwise. Let $H \sim \phi(F)$ be the distribution of perceived potentials of disadvantaged students, and then note that $(1-p)F + pH$ is the distribution of perceived potentials of all students in aggregate in the presence of bias and any debiasing. In the fair matching (where ϕ is the identity map), a disadvantaged student is matched to school with rank $\mu(x) = \mathbb{P}(F \geq x)$, and in the ranking with bias and debiasing of S, they are matched to $\mu_\phi(x) = \mathbb{P}((1-p)F + pH \geq \phi(x))$. We therefore have under S the displacement

$$\begin{aligned}
\text{disp}_S(x) &= \mu_\phi(x) - \mu(x) \\
&= 1 - \mathbb{P}((1-p)F + pH \leq \phi(x)) - (1 - \mathbb{P}(F \leq x)) \\
&= \mathbb{P}(F \leq x) - [p\mathbb{P}(\phi(F) \leq \phi(x)) + (1-p)\mathbb{P}(F \leq \phi(x))] \\
&= \begin{cases} p\left[\mathbb{P}(F \leq x) - \mathbb{P}(\phi(F) \leq x)\right], & x \in S, \\ \mathbb{P}(F \leq x) - [p\mathbb{P}(\phi(F) \leq x - \gamma) + (1-p)\mathbb{P}(F \leq x - \gamma)], & x \notin S. \end{cases}
\end{aligned}$$

By carefully dividing into the cases where $Y_2 - Y_1 < \gamma$ and $Y_2 - Y_1 \geq \gamma$, one can show

$$\begin{aligned}
\text{disp}_S(x) &= \begin{cases} (1-p)\mathbb{P}\left(F \in [x-\gamma, x]\right), & x \leq Y_1 \text{ or } x \geq Y_2 + \gamma, \\ -p\mathbb{P}\left(F \in [Y_2, \max\{x+\gamma, Y_2\}]\right), & x \in (Y_1, Y_2), \\ (1-p)\mathbb{P}\left(F \in [x-\gamma, x]\right) + p\mathbb{P}\left(F \in [\max\{x-\gamma, Y_1\}, Y_2]\right), & x \in [Y_2, Y_2+\gamma), \end{cases} \\
&= (1-p)\mathbb{P}\left(F \in [x-\gamma, x]\right) \mathbb{1}_{\{x \notin S\}} + \begin{cases} -p\mathbb{P}\left(F \in [Y_2, \max\{x+\gamma, Y_2\}]\right), & x \in (Y_1, Y_2), \\ p\mathbb{P}\left(F \in [\max\{x-\gamma, Y_1\}, Y_2]\right), & x \in [Y_2, Y_2+\gamma). \end{cases}
\end{aligned} \tag{B.10}$$

We are now ready to complete the proof.

Proof. Let $S = [Y_1, Y_2]$ be some interval to debias, then (B.10) implies:

1. $m_\emptyset(x) = m_S(x)$ for $x \notin (Y_1, Y_2 + \gamma)$,
2. $m_S(x) = 0$ for $x \in (Y_1, Y_2)$,
3. $m_S(x) \geq m_\emptyset(x)$ for $x \in [Y_2, Y_2 + \gamma)$, and
4. $m_S(x)$ is decreasing for $x \geq \max\{1 + \gamma, Y_2\}$.

Further, if $1 + \gamma \notin S$ then $\max_x m_S(x) \geq \max_x m_\emptyset(x)$. To see this, suppose $Y_1 > 1 + \gamma$, then the result follows because $\max_x m_\emptyset(x) = m_\emptyset(1 + \gamma)$. Otherwise if $Y_2 \in [1, 1 + \gamma]$, we must have $1 + \gamma \in [Y_2, Y_2 + \gamma]$, so $\max_x m_S(x) \geq m_S(1 + \gamma) \geq m_\emptyset(1 + \gamma) = \max_x m_\emptyset(x)$.

We therefore know that the optimal $S \in \mathcal{S}^c(\hat{c})$ contains $1 + \gamma$ and that it minimizes $\max\{m_S(Y_1), m_S(Y_2)\}$. Now write

$$\begin{aligned}
m_S(Y_1) &= (1-p)\mathbb{P}\left(F \in [Y_1 - \gamma, Y_1]\right) = (1-p)\mathbb{P}\left(F \leq Y_1\right), \\
m_S(Y_2) &= (1-p)\mathbb{P}\left(F \in [Y_2 - \gamma, Y_2]\right) + p\mathbb{P}\left(F \in [\max\{Y_2 - \gamma, Y_1\}, Y_2]\right).
\end{aligned}$$

Recall that $\mathbb{P}(F \in [Y_1, Y_2]) = \hat{c}$. We need $m_S(Y_1) = m_S(Y_2)$, so write

$$
\begin{aligned}
(1-p)\mathbb{P}(F \leq Y_1) &= (1-p)\mathbb{P}(F \in [Y_2 - \gamma, Y_2]) + p\mathbb{P}(F \in [\max\{Y_2 - \gamma, Y_1\}, Y_2]) \\
\iff (1-p)(1-\hat{c}) &= (1-p)\mathbb{P}(F \geq Y_2 - \gamma) + p\mathbb{P}(F \in [\max\{Y_2 - \gamma, Y_1\}, Y_2]) \\
&= (1-p)\mathbb{P}(F \geq Y_2 - \gamma) + p(\mathbb{P}(F \geq \max\{Y_2 - \gamma, Y_1\}) - \mathbb{P}(F \geq Y_2)) \\
&= (1-p)\mathbb{P}(F \geq Y_2 - \gamma) + p(\min\{\mathbb{P}(F \geq Y_2 - \gamma), \mathbb{P}(F \geq Y_1)\} - \mathbb{P}(F \geq Y_2)) \\
&= \min\{\mathbb{P}(F \geq Y_2 - \gamma) - p\mathbb{P}(F \geq Y_2), (1-p)\mathbb{P}(F \geq Y_2 - \gamma) + p\hat{c}\}.
\end{aligned}
$$

Since both terms inside the minimum are decreasing in Y_2, the Y_2 that solves this equation is given by $\min\{U_1, U_2\}$ where U_1 solves $(1-p)(1-\hat{c}) = \mathbb{P}(F \geq U_1 - \gamma) - p\mathbb{P}(F \geq U_1)$, and U_2 solves $(1-p)(1-\hat{c}) = (1-p)\mathbb{P}(F \geq U_2 - \gamma) + p\hat{c}$. These are exactly the expressions sought for in the theorem. □

Appendix C: Additional Details for Chapter 3

C.1 Proofs of some well known results

In this appendix we provide for completeness proofs of some results that are known in the literature.

Lemma 123. *A choice function C is path-independent if and only if it is substitutable and consistent.*

Proof. Suppose C is path-independent. We first show substitutability. Let $S \subseteq X$, $b \in C(S)$, and $T \subseteq S$. For a contradiction, suppose $b \notin C(T \cup \{b\})$. Let $Q = T \cup \{b\}$ so that $b \notin C(Q)$, then by path-independence, we have that

$$C(S) = C(Q \cup (S \setminus Q)) = C(C(Q) \cup (S \setminus Q)).$$

But now observe that $b \notin C(Q)$, and $b \notin S \setminus Q$, so $b \notin C(Q) \cup (S \setminus Q)$ and $b \notin C(S)$, a contradiction.

To show consistency, suppose $T \subseteq S \subseteq X$ with $C(S) \subseteq T$, then we must have

$$C(S) = C(S \cup T) = C(\mathcal{S}(T) \cup T) = C(T),$$

as required.

For the opposite direction, let C be substitutable and consistent, and let $S, T \subseteq X$. We first apply substitutability to $S' = S \cup T$, $T' = S$. We have $T' \subseteq S'$, so therefore

$$C(S \cup T) \cap S \subseteq C(S).$$

Now write

$$\begin{aligned} C(S \cup T) &= C(S \cup T) \cap (S \cup T) \\ &= (C(S \cup T) \cap S) \cup (C(S \cup T) \cap T) \\ &\subseteq C(S) \cup T. \end{aligned}$$

Finally, apply the definition of consistency to $S' = S \cup T$ and $T' = C(S) \cup T$. Since $C(S \cup T) \subseteq C(S) \cup T$, therefore it must hold that $C(S \cup T) = C(C(S) \cup T)$, as required. □

Lemma 124. *Suppose C is quota-filling and substitutable. Then it is consistent.*

Proof. Let $S \subseteq X$ and $T \subseteq S$ such that $C(S) \subseteq T$. We wish to show $C(S) = C(T)$. By substitutability we have $C(S) \cap T \subseteq C(T)$ and since $C(S) \subseteq T$, this yields $C(S) \subseteq C(T)$. Since $T \subseteq S$, we must have $|C(T)| \leq |C(S)|$ by the quota-filling property, but this immediately implies that $C(S) = C(T)$, as required. □

Lemma 125. *Let $\mathcal{X}$ be a finite set with cardinality $|\mathcal{X}|$. Any deterministic communication scheme that can uniquely identify any element $x \in \mathcal{X}$ must use $\Theta(\log |\mathcal{X}|)$ bits in the worst case, and this bound is achievable.*

Proof. We define a deterministic communication scheme as a function $f : \mathcal{X} \to \{0,1\}^*$ where $\{0,1\}^*$ denotes the set of all binary strings. For the encoding scheme to identify elements uniquely, we therefore require that f is injective, so $|f(\mathcal{X})| = |\mathcal{X}|$. Let L be the maximum length of any binary string in $f(\mathcal{X})$, then counting the number of binary strings with size at most L, we have $2^{L+1} - 1 \geq |f(\mathcal{X})| = |\mathcal{X}|$, which directly implies the bound. The bound can be achieved by enumerating $f(\mathcal{X})$ and encoding each element by a fixed length binary string of length $\lceil \log_2(|\mathcal{X}|) \rceil$. □