How More Equitable Assignment Mechanisms Can Increase School-level Segregation

by

Orrie B. Page

B.S. Computer Science, Economics, and Data Science
Massachusetts Institute of Technology, 2022

Submitted to the Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degrees of

MASTER OF ENGINEERING IN COMPUTER SCIENCE, ECONOMICS, AND DATA SCIENCE

at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2023

Authored by: Orrie B. Page
Department of Electrical Engineering and Computer Science
May 19, 2023

Certified by: Parag A. Pathak
Class of 1922 Professor of Economics

Accepted by: Katrina LaCurts
Department of Electrical Engineering and Computer Science
Chair, Master of Engineering Thesis Committee

How More Equitable Assignment Mechanisms Can Increase School-level Segregation

by

Orrie B. Page

Submitted to the Department of Electrical Engineering and Computer Science
on May 19, 2023, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Computer Science, Economics, and Data Science

Abstract

I investigate how implementing a school assignment mechanisms that gives families influence over their child's placement can lead to an increase in school level segregation when compared to a more typical "neighborhood school" mechanism. Using data from a major U.S. school district and a theoretical model I show that under certain conditions, this surprising outcome can occur even when housing within the city is geographically segregated by race.

Thesis Supervisor: Parag A. Pathak
Title: Class of 1922 Professor of Economics

Acknowledgments

I would first like to thank my supervisor, Parag Pathak, for welcoming me into Blueprint Labs as a summer research assistant back in 2022. His help in my quest to begin graduate school has been invaluable, and I would not be where I am today without him.

I also want to thank my research team at Blueprint Labs, Eryn Heying, Jim Shen, and Carol Gao, without whom the research that went into this thesis would not have been possible.

Finally, and most importantly, I want to thank my family: my wonderful parents, Jenna Bednar and Scott Page, for always being there to support me and for telling me what I need to hear, even when I don't want to hear it; my little brother, Cooper Page, for his unwavering kindness; and of course, our two dogs, Oda and Roland, for reminding me to come back home to Michigan every once in a while.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The School Assignment Problem

In small towns the question of school assignment is relatively straightforward–students enroll in their local elementary school, move on to their local middle school, and eventually graduate from their local high school. There are of course always cases of families choosing to send their child to schools that are further away for special programming, but by-and-large, students who stay in the public school system attend the closest school to their front door. In major U.S. cities this process is much more complicated. School boards in major cities are faced with the daunting task of assigning thousands of students to potentially over 100 different schools. Each student comes from a different home with a different family, each of whom might have entirely different preferences over which schools they would like their child to attend. In addition, some students might have disabilities or require special programming. Others may seek specific bilingual opportunities. With each child having 10 to 20 schools that are within a reasonable radius of their home, school boards turn to complicated assignment mechanisms to place students in classrooms.

School districts have their own set of broad goals as well, such as seeking to establish equal access to high quality schools for all subsets of students, to minimize racial isolation in schools, or to shorten the distance that students must travel to reach their assigned school. Financial costs are also always a concern, which add pressure to minimizing busing loads. In many cities in the U.S. these appear to be is conflicting goals. Geographical racial segregation is still present in many of today's cities and in these racially divided cities it can be challenging to integrate schools without drastically increasing commute times. In order to make meaningful impacts on diversity in classrooms, districts would need to bus minority students into schools in neighborhoods with few minority house holds and vice-versa. Similarly, it can be difficult to decrease commute times without reducing the equality of access to high quality schools. The unfortunate reality in America is that many if not most of the high quality public schools are already found in wealthier–and whiter–regions of the city.

1.2 School Assignment Mechanisms

The problem of assigning students to schools can be modeled as a two sided matching problem. Each district is presented with its own version of this problem and in order to solve these, school districts across the country implement complicated matching mechanisms. The mechanisms themselves differ from city to city and we are yet to reach academic or industry consensus on which mechanism, or set of mechanisms, are most effective. The fact that there is so much diversity in the types of mechanisms used shouldn't come as a surprise; each city is faced with a unique set of challenges, a different student population, and its own set of long term goals. Because of this, we should expect that a mechanism which works particularly well for one district, has no guarantee of working at all for another.

In this section I will lay out two mechanisms. One is the mechanism currently being used by one of Americas major public school districts. We will refer to it as the "choice" mechanism. The other is what we call the "neighborhood" mechanism, which is similar to what you would see implemented in most smaller cities. The American school district whose mechanism I will detail has graciously granted me access to over 15 years of their assignment records, but has requested to remain anonymous.

1.2.1 The Choice Mechanism

The choice mechanism, as mentioned, is what is currently implemented in my sample city's school district. The underlying idea behind the choice mechanism is that families are given the opportunity to submit preferences across a set of schools that are relatively near to their own home–and that those preferences will be taken into account when assigning children to schools. A naive version of this mechanism would simply ask each family to submit preferences over all schools within a fixed radius of their home but there are two main problems with that implementation. The first problem is that it is unfair to students who live on the periphery of the district, they could conceivably be left with only a very small selection of schools to pick between while students in the center of the district would be overwhelmed with options. The other issue is that it is distinctly unfair to students who live in under-served sections of the city. The unfortunate reality, in my sample city and in most other cities across America, is that good schools tend to be near other good schools, and bad schools tend to be near other bad schools. This means that with the naive implementation some students might only have 1 or 2 high quality school options while others might have 20.

To solve this issue, the district had to find a way of ensuring that all students have the ability to rank schools that the district deemed high quality. They did this by creating a tier system for all of their schools. Each school was ranked on the scale of tier 1, the best schools, to tier 4, the lower performing schools.[1]

[1]Note that the process of assigning schools to tiers involves some quantitative cutoffs, like student growth scores, and also some qualitative measures as decided by the district. For my purposes I have taken these underlying tiers to be a relatively accurate measure of school quality.

Once tiers have been made (a process that is redone every school year) students are given a "basket" of schools to rank. Their basket will first include all schools within a 1 mile radius, as well as any schools that their siblings attend. Then, to ensure all students have the opportunity to rank several high quality schools, additional schools are added to a students basket to ensure that every student's basket meets the following requirements:

1. Basket contains at least 2 tier 1 schools.
2. Basket contains at least 4 schools that are at least tier 2.
3. Basket contains at least 6 schools that are at least tier 3.

When schools are added to a students basket to satisfy one of the above requirements, the nearest school that satisfies the requirement is the one added. For example, if a student only had 5 schools that were at least tier 3, and the nearest school to them that wasn't already in their basket was a tier 1 school, that tier 1 school would be added.

Once baskets have been made, families submit their preferences as a ranking across the schools in their basket.[2] After families have submitted their preferences they are each given a random lottery number (which represents the schools preferences over the students) and then the Gale-Shapley algorithm is run to determine the final school assignments.

The goal of this mechanism when implemented was two fold. First it wanted to bring students closer to home, which is the reason that students are only allowed to rank schools that are near to their home. The second goal was to increase the equity of access to high quality schools. This is why more schools are added to students choice menus to ensure every student is offered a comparable number of high quality schools.

[2]families do not need to rank all of the schools in their basket. In some cases the baskets are very long and so it is very common to see families only rank a subset of the possible schools.

1.2.2 The Neighborhood Mechanism

The default mechanism is what I will refer to as the “neighborhood" mechanism. Under the neighborhood mechanism, the Gale-Shapley algorithm is still used, but instead of allowing families to submit preferences, each student’s “preferences” are defaulted to be purely proximity based. So for example their top preference would be for the school closest to them, second preference would be for the second nearest school, and so forth. Similarly, instead of using a lottery, schools now prefer the closest student to them first, the next closest second and so on.

When Gale-Shapley is used over this set of preferences we are left with a set of assignments in which every student is attending the closest possible school to them–subject to capacity constraints. What this means in practice is that if a student was assigned to school B but lives closer to school A, then it must be that every student who was assigned to school A lives closer to school A than our example student.

This mechanism is much more similar to the mechanisms used in smaller cities, where students primarily attend whichever school is closest to their home. It is the solution to the matching problem that minimizes the bussing cost for the city, but its big drawback is that it doesn’t give families the opportunity to choose to attend a higher quality school that is farther away. Another pitfall of this system is that because students are attending their nearest school, the racial and economic demographics in a school are heavily dependent on the demographics of the area immediately surrounding it. In a city like our example city, where housing is racially clustered, this could lead to racial segregation between schools.

1.3 The Gale-Shapley Deferred Acceptance Algorithm

In the above subsections I reference the Gale-Shapley deferred acceptance algorithm a few times. It is used both in the assignment of students to schools in the choice mechanism, and in our approximation of the neighborhood mechanism. Because it is such a key mechanic for this analysis I include here a brief section in which I explain how the Gale-Shapley algorithm works and explain a few of its convenient properties.

The Gale-Shapley deferred acceptance algorithm, which is often referred to as either "the Gale-Shapley algorithm" or just "deferred acceptance", first appeared in D. Gale and L. Shapley's 1962 paper on college admissions and the stability of marriage[16]. The algorithm is a method of finding a stable equilibrium in a two sided matching problem. A stable equilibrium as defined as an equilibrium in which there is no blocking pair, a pair that would be better off if they had not participated in the algorithm.

The easiest way to explain how the algorithm works is to consider the marriage application. In this example there is a set of men and a set of women. Each man has a hierarchical ranking over all women, including the option to not be paired with any woman–meaning there may be some women he would prefer to be alone than to be paired with. Similarly each woman has a hierarchical ranking over all men, again including the option to not be paired with any man.

The algorithm is then run in a series of rounds. In round 1, every man proposes to the woman who he most prefers. Each woman then evaluates her set of proposers and selects the one she most prefers from them to be her fiance, and rejects all of the others. In round 2 all men who were rejected in round 1 now propose to their second choice woman. Again each woman evaluates her set of proposers–including her current fiance if she has one–and selects the one she most prefers, rejecting the rest. The algorithm continues like this where in round i, all of the men who were rejected in round $i - 1$, but who still prefer a woman who they have not yet proposed to yet over having no partner, will propose to the most preferred woman who they have no yet proposed to. The algorithm ends when either every man is engaged, or when all men who are not engaged prefer to be alone over being with any of the women they haven't proposed to yet.

In the context of our school assignment problems, we can think of the women as being the schools, and the men as being the students. In this case the school would not be "engaged" to a single student but instead to a set of students. For example if 100 students "proposed" to a given school in round 1, but the school only had a capacity of 50, the school would select the 50 students it most prefers and reject the rest. In the mechanisms we look at in the paper, the school preferences are based on either a random lottery number, or on proximity to the school, but you could also imagine cases where it is based on test results or something correlated with the quality of the student.

The Gale-Shapley algorithm, in addition to its simplicity, has two handy properties. The first is that in every final outcome, every man will do as well as he could have expected to do. What this means is that for a given man, if he is in the end married to his ith choice woman, it must be that the $i - 1$ women who he would prefer over his wife, are all married to men who they prefer over him. The other handy property is that its outcomes are always stable. As I mentioned before, a stable equilibrium as defined as an equilibrium in which there is no blocking pair. In the marriage example, a blocking pair would be if there were some man m_i who is left married to woman w_i, and some woman w_j, who is left married to man m_j, where man m_i prefers w_j to w_i, and woman w_j prefers m_i to m_j. In that scenario man m_i and woman w_j would be better off if they had not participated in the mechanism at all and had just married each other–this type of outcome is not possible under the Gale-Shapley algorithm.

Chapter 2

Existing Literature

Thinking about school assignment from a matching perspective has a long tradition. Over the last 20 odd years, substantial effort has gone into trying to better understand these school assignment mechanisms, especially ones that take into account family preferences through some form of choice mechanism. In cities, these choice mechanisms typically use either the Deferred Acceptance algorithm [16] or the Top Trading Cycles algorithm [17] to create the student-school pairings, but in some smaller towns, the process is less rigid and is sometimes even done manually. In all cases, the efficacy of choice mechanisms is the topic of much debate. Some papers (and books) have been written advocating for choice based school assignment mechanisms [8] [11]. Arguing that they are necessary for us to reach a more balanced society. While others have warned of the potential pitfalls and unintended consequences of these mechanisms, like the segregation issue my thesis focuses on. [1] [6].

Research on this topic can be broadly split into three categories. The first category of research looks at how families make decisions when submiting preferences over schools. The second category looks at the impacts that school placement has on future success for given students. And the final category looks at the impact that different assignment mechanisms have on the demographics of the schools in the system.

My thesis primarily falls into this last category, as I am interested in the degree to which choice mechanisms can increase segregation between schools, but I also look briefly at the first category, as I consider how differences in preferences can drive this segregation. In this chapter I give a brief overview of some of the existing literature on this subject, and explain how my thesis fits into the landscape.

2.1 How Families Select School Preferences

A common through-line in arguments over how choice mechanisms have the potential to increase segregation is the idea that families of different demographic groups have different preferences over schools. There are those that disagree with this [14], but by-and-large there is a agreement that this is an important factor. In this section, I give some academic context for how scholars believe families are making these decisions.

There is a substantial amount of evidence that families are more likely to rank a school highly if the current students at that school are of the same race as them. A clear-cut example of this is in a 2005 study conducted by Denessen, Driessena and Sleegers [10]. The group was studying the school choice mechanism in the Dutch educational system and found that Muslim families had a very strong preference towards Islamic education, in schools where most of the other students were also Muslim. The result of this, somewhat trivially, is that the system ended up with a few schools containing almost all of the Muslim students and virtually nobody else.

This raises an interesting question though about the goal of our school choice mechanisms. If an uninitiated observer were to be given data on a school system with this form of segregation–where the Muslim population was almost completely isolated–they would likely point to it as an issue. But all of the Muslim families who chose to send their children to those schools prefer their child to be there over the rest of the educational options, so the question then becomes: who is this an issue for? It is difficult for a school board to tell a subset of the population that they should be making different school selections, if that subset of the population is pleased with the assignments they are receiving.

In some cases, the issue with these types of preferences are more obvious. For example Saporito's 2014 paper investigates a very similar issue [15]. It looks at magnet schools in a large city and finds that white families tend to avoid applying to schools that have large non-white populations. He goes on to demonstrate that this tendency cannot be explained purely by other non-race related characteristic, such as exam scores or poverty. He then demonstrates how this has led to further racial segregation within the school system. Most people would probably agree that this is an issue–white families intentionally avoiding schools with minorities–but in the case of minorities selecting to be with other minorities like the Muslim example, the question becomes less clear.

Other studies have disagreed entirely with the idea that families from different demographics have different preferences over schools. To see an example of this we can look to Courtney A. Bell's 2009 study on the role of choice sets in the process [5]. Bell was interested to see how the set of schools parents had the opportunity to rank influenced the outcome of the eventual match. To study this question she repeatedly interviewed 48 families over a 9 month period leading up to their children enrolling in school in a Midwestern city. She found that the primary cause for quality differences in school selections across demographic groups was the range of options that families had in their choice basket, not the families choice processes. The implication of this is that all families have somewhat similar sets of preferences when it comes to selecting schools. A result that directly contradicts the studies done by Saporito and by Denessen, Driessena and Sleegers.

Other studies have looked at how parents evaluate the quality of a school when making selections. For example a 2020 paper by Atila Abdulkadiroglu, Parag Pathak, Jonathan Schellenberg, and Christopher R. Walters studied the relationship between family preferences, school quality, and peer quality [3]. Their paper looked at the high school assignment mechanism in New York City and concluded that preferences were highly correlated with the number of high-achieving students who enrolled in the school, but that after controlling for this peer quality effect, preference was uncorrelated with the effectiveness of the school itself. This is in some ways a similar result to Stephen Ball and Carol Vincent's 1998 paper on parental gossip [4]. In this paper, they compared two different types of information parents receive on schools. The first was grapevine gossip which they referred to as 'hot' knowledge. The other was more formal information, like that published by schools or by districts, they referred to this as 'cold' knowledge. In the paper they find that parents actually place more weight on hot knowledge than on cold knowledge. This could explain Abdulkadiroglu, Pathak, Schellenberg, and Walters' result, if families with high achieving students are in some sense coordinating.

While my thesis is not focused on taking a direct stance on the relationship between families demographic background and their preferences over schools, it is an important aspect of my analysis. In Chapter 5, I show a substantial difference in the relative preference for school quality between White/Asian students and Black-/Hispanic students, even when controlling for proximity. My results support those of Saporito and Denessen, Driessena and Sleegers.

2.2 The Impact of Choice Mechanisms

My thesis explores the phenomenon of choice based mechanisms resulting in more segregation than neighborhood school mechanisms. This phenomenon is surprising because it directly contradicts the idea that low-income students and minorities should stand to gain from these choice mechanisms. The idea that these mechanisms should help disadvantaged students has strong support and rational. Disadvantaged students are less likely to live close to high quality schools, and so by implementing choice mechanisms where they can choose to attend a school that is not their closest option, it would seem as though they should be strictly better off. Many people, including Chubb and Moe in 1990 [8] and Viteritti in 1999 [11] have made arguments along these lines, claiming that choice based mechanisms are the key to achieving a higher level of educational equality in our cities.

Interestingly many scholars who have had the data to study this empirically have actually found the opposite to be true–meaning they have found that choice mechanisms tend to further segregate schools, hurting disadvantaged students even more than neighborhood mechanisms–just like in my sample city.

Consider for example Söderström and Uusitalo's 2010 paper which studied the effect that a school choice mechanisms had on school-level racial segregation in Stockholm [13]. Their paper looked at Stockholm's switch from a nearest school model–which is for all intensive purposes the same as the neighborhood mechanism described in Chapter 1–to a school choice mechanism in 2000. The paper is similar to this thesis in that it compares closely a school system under two different possible mechanisms; a neighborhood mechanisms and a choice mechanisms. Their paper finds that the transition to a choice mechanism lead to an increase in segregation by family background. This is of course in line with my findings and contrary to the idea that choice mechanisms should help disadvantaged students.

A similar study was conducted in England by Rebecca Allen in 2007 [1]. In this study, Allen looked at elementary schools in England, a system with unclear central organization that has allowed about half of its students to attend schools that would not have been assigned to them through a neighborhood mechanism. Using similar methodology to my own, Allen is able to simulate a counterfactual world in which every student had attended their nearest school. In comparing the real outcome with her simulation, Allen finds that there is a significantly lower amount of segregation in the nearest school simulation than in the true outcome. The main differences between my thesis and Allen's results are that Allen focuses primarily on ability level segregation, and to some degree wealth segregation, but not race. She also does not model the situation to explain underlying forces.

A much larger nationwide study was also conducted in Sweden in 2015 by Anders Böhlmarkb, Helena Holmlundc and Mikael Lindahld [6] [7]. This study explored Sweden's slow adoption of choice based mechanisms across the country to see how school level segregation changed in regions where school choice was more or less prevalent. The results of this experiment were somewhat inconclusive. On the one hand the group found that choice mechanisms did correlate with a substantially higher level of segregation. But their research did not include any simulations of counterfactual neighborhood mechanism scenarios. This meant that they were unable to say whether the correlation existed because of choice mechanisms increasing segregation, or whether it existed because regions with particularly bad segregation were more likely to adopt choice mechanisms. In their concluding discussion, they mention that political pressures in Sweden are more likely to cause segregated regions to adopt choice mechanisms than homogeneous regions, which they theorize could have caused the observed effect.

It should be noted that not all empirical studies into the effects that choice mechanisms have on segregation have found the same result. For example a 2004 paper by Douglas Archbald studied how segregation changed in magnet schools that used choice mechanisms for assignment, and did not find any difference between segregation in the magnet schools when compared to surrounding schools that had been filled using a neighborhood mechanism [2]. A 2014 study conducted by Tomeka Davis found similar results [9]. In Davis' study, they found that magnet schools populated through choice mechanisms did not have statistically different segregation levels when compared to neighborhood mechanism schools in the same area. One interesting result from Davis' paper was that they found that classrooms in the choice schools were more mixed than classrooms in neighborhood schools, particularly honors classrooms. This is out of the scope of my thesis though, as I did not have access to classroom level data. I also focus on elementary school students, where classroom differentiation is not yet as serious of an issue.

The final paper I want to mention was actually a PhD dissertation out of Stockholm University by Dany Kessel [12]. In Kessel's dissertation, they looked at modifications of choice mechanisms that imposed soft quotas for schools. By using choice basket data they simulated the mechanism in action under various modifications and found that by using these quotas, choice mechanisms could actually lead to more integrated schools.

There are two ways my thesis adds to this existing literature. The first, and most obvious, is that it gives a new case study of a city where implementing a choice mechanism had the effect of increasing segregation when compared to a neighborhood mechanism. The second, and perhaps more interesting addition my thesis makes, is that it offers a model to help explain why this phenomenon occurs. All of the above papers use models to regress segregation over various variables, but none have built one that demonstrates conditions under which segregation increases when a choice mechanism is implemented.

Chapter 3

Modeling the Segregation Phenomenon

Later in this paper we are going to see an example of how a choice mechanism can result in a more segregated assignment than a neighborhood mechanism, even in the presence of geographical racial segregation. The observed effect is initially counter-intuitive and in fact may even seem impossible. In an effort to better visualize this phenomena, and to explain why and when it happens, I have created the following simple model.

3.1 Model Definition and Assumptions

The model assumes there are just two demographic groups, denoted group 1 and group 2, and two schools, denoted school A and school B. Without loss of generality let school A be the "better" school. This means that any student who is closer to school A will strictly prefer it to school B, and that some students who are closer to school B will still prefer school A because of its quality.

Now, let n_{A1} be the number of students in group 1 that live closer to school A

than to school B, and let n_{B1} be the number of students in group 1 that live closer to school B than to school A. Define n_{A2} and n_{B2} similarly. For example if $n_{A2} < n_{B2}$ then we would be in a situation in which the majority of group 2 students live closer to school B than to school A. As a final pair of variables, let p_1 be the probability that a student in group 1 will prefer school A over school B even if they live closer to school B and define p_2 similarly.

Lets now assume that the schools were built with a neighborhood system in mind and so have capacities $c_A = n_{A1} + n_{A2}$ and $c_B = n_{B1} + n_{B2}$. This means that no matter what mechanism we use, assuming all students are assigned, both schools will be full.

In order to help better understand the model I also created a visual example of what the environment looks like which can be seen in the figure below. Here we can see that students who live in region A are closer to school A, and that students who live in region B are closer to school B. In the example shown, 75% of students who live closer to school A are in group 1, and 75% of the students who live closer to school B are in group 2 so this specific environment represents extreme geographical racial segregation.

3.2 Mathematical Results from the Model

Lets first examine this system under a neighborhood assignment mechanism. Since we have defined the capacities to perfectly accommodate the number of students who are closer to each school we are left with a trivial assignment. School A will have n_{A1} students from group 1 and n_{A2} students from group 2. School B will have n_{B1} students from group 1 and n_{B2} students from group 2.

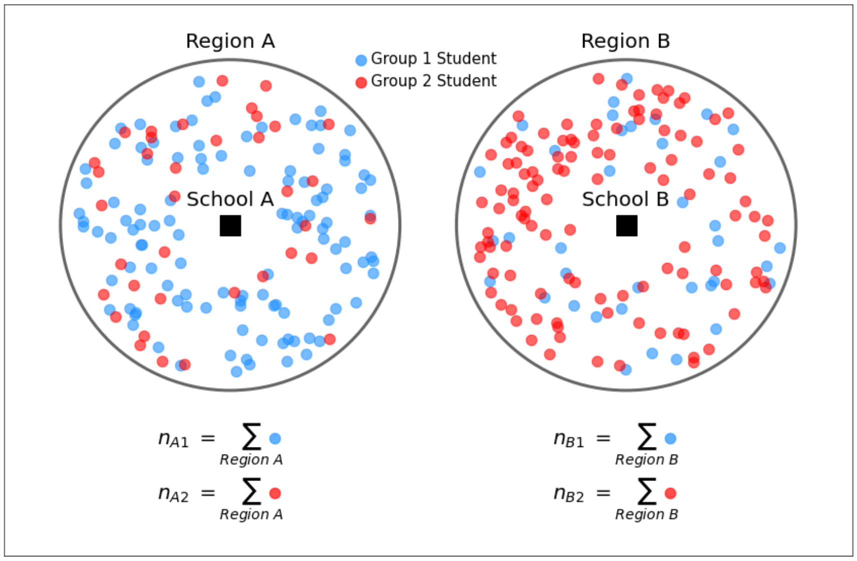

Figure 3-1: Visual Explanation of the Model as Described

Lets assume without loss of generality that students in group 2 are more densely clustered near school B than group 1 is. In our notation this means we have $n_{B2}/n_{B1} > n_{A2}/n_{A1}$. We are therefore now modeling a system in which group 1 has a systematic advantage over group 2 because a group 1 student is more likely to live closer to the better school than a group 2 student is. If the system were to become more segregated under a different mechanism, it would therefore mean we would see a decrease in the number of group 2 students who attend school A.[1]

Now lets look at this same system under a deferred acceptance system in which each family can rank the two schools, and the schools assign random priority numbers to each student. Lets look at this just from the perspective of school A. We know that all $n_{A1} + n_{A2}$ students who live closer to school A will rank it first, be we also expect to have an additional $n_{B1}p_1 + n_{B2}p_2$ applicants, representing the students

[1]Note that it is of course possible to see an increase in the number of group 2 students at school A that is so large that we become more segregated in the other direction. But we are ignoring this situation for reasons that will become clear later in this section.

who live closer to school B but prefer quality over proximity. This means we have $(n_{A1}+n_{B1}p_1+n_{A2}+n_{B2}p_2)$ applicants for only $n_{A1}+n_{A2}$ spots. The expected number of group 2 students at school A is therefore:

$$\frac{(n_{A1}+n_{A2})(n_{A2}+n_{B2}p_2)}{n_{A1}+n_{B1}p_1+n_{A2}+n_{B2}p_2}$$

If we recall, we know that the system becomes more segregated if this is lower than n_{A2}, the number of group 2 students assigned to school A under the neighborhood mechanism. This means that the choice system increases segregation whenever the following inequality holds:

$$n_{A2} > \frac{(n_{A1}+n_{A2})(n_{A2}+n_{B2}p_2)}{n_{A1}+n_{B1}p_1+n_{A2}+n_{B2}p_2}$$

If we simplify this inequality we can arrive at our key theorem:

Theorem 1 *Whenever the following inequality holds:*

$$(n_{A1}+n_{B1}p_1) > (n_{A2}+n_{B2}p_2)\left(\frac{n_{A1}}{n_{A2}}\right)$$

a choice based mechanism will in expectation increase the amount of segregation in a system when compared to a neighborhood mechanism.

This looks ugly but all the terms in Theorem 1 are somewhat intuitive. $(n_{A1}+n_{B1}p_1)$ is simply the number of students in group 1 who prefer school A, and $(n_{A2}+n_{B2}p_2)$ is the number of students in group 2 that prefer school A. The last term (n_{A1}/n_{A2}) is simply the ratio of group 1 students to group 2 students in the area surrounding school A. In a real data set we can calculate all of these values. For example to find $(n_{A1}+n_{B1}p_1)$ we would just take the number of students in group 1 who rank school A above school B.

By rearranging terms in Theorem 1 we see that it can also be rewritten in the following way:

Corollary 1.1 *Whenever the following inequality holds:*

$$\left(\frac{n_{B1}p_1}{n_{B2}p_2}\right) > \left(\frac{n_{A1}}{n_{A2}}\right)$$

a choice based mechanism will in expectation increase the amount of segregation in a system when compared to a neighborhood mechanism.

Corollary 1.1 is simpler than Theorem 1, and it gives us a much more intuitive condition for when a choice mechanism will increase segregation–in fact in hindsight we should have been able to write this down before doing any math. (n_{A1}/n_{A2}) is the demographic ratio of the students who are assigned to school A in a neighborhood system. $(n_{B1}p_1)$ is the number of group 1 students who prefer school A but are only able to apply to it in a choice mechanism. Similarly $(n_{B2}p_2)$ is the number of group 2 students who prefer school A but are only able to apply to it in a choice mechanism. This means we can think of $(n_{B1}p_1)/(n_{B2}p_2)$ as the demographic ratio of "new" students who will be able to apply to school A when we implement a choice mechanism. If that ratio is more heavily weighted towards group 1 students than the original ratio was, then we should expect the ratio of students who are assigned to school A to shift more towards group 1 under a choice mechanism, thereby increasing segregation.

The only disadvantage to corollary 1.1 when compared to Theorem 1 is that it is a harder condition to check given data because it requires us to calculate p_1 and p_2 which is more challenging to do. You can get around this by counting for example the number of group 2 students who live closer to school B but prefer school A anyway (this would give you $n_{B2}p_2$), but this is still more computationally challenging than finding the terms in Theorem 1.

A potentially more elegant way of writing corollary 1.1 is to define 3 new terms as follows. Let R_A be the ratio of group 1 to group 2 students in the region around school A: $R_A = (n_{A1}/n_{A2})$. Next, define R_B as the ratio of group 1 to group 2 students in the region around school B: $R_B = (n_{B1}/n_{B2})$. Finally let $P_{1,2}$ represent the degree to which group 1 students are more likely to prefer quality to distance when compared to group 2 students. Define $P_{1,2}$ as $P_{1,2} = p_1/p_2$. This allows us to rewrite corollary 1.1 as follows. Choice mechanisms will increase segregation compared to neighborhood mechanism in expectation whenever:

$$R_B \times P_{1,2} > R_A$$

3.3 Model Implications in Example Scenarios

In this section, I will be examining the implications of my model in three different general scenarios. The first scenario is a balanced population, meaning $n_{A1} + n_{B1} = n_{A2}+n_{B2}$. The second scenario is a majority-minority population where group 2 is the minority. To be clear, I mean minority in the sense of a systematically disadvantaged race, in this section I will use the terms minority race and group 2 interchangeably. Similarly majority race and group 1 are interchangeable terms. In the majority-minority population we have $n_{A1} + n_{B1} < n_{A2} + n_{B2}$, or in words, fewer group 1 than group 2 students. The final scenario is a majority-majority population, where we now have $n_{A1}+n_{B1} > n_{A2}+n_{B2}$, or in words, more group 1 students than group 2 students.

In each of these case I will compare the impact of a choice mechanism under varying values of $P_{1,2}$, so the extent to which majority students are more likely to prefer quality over proximity when compared to minority students. As a reminder we define $P_{1,2} = p_1/p_2$. I will also be examining the effect for varying levels of population skew. When I say a population is skewed at $x\%$, what I mean is the following; in the region where the group 1 population is more concentrated (the region around school A), $x\%$ percent of the group 1 population is in that region. In our models mathematical terms this means an $x\%$ skew is equivalent to:

$$\frac{n_{A1}}{n_{A1}+n_{B1}} = \frac{x}{100}$$

In the interest of simplicity, for all of these scenarios the skew will be the same for both groups. There is no reason this need be true in the real world, but it is a relatively innocuous assumption designed to make the results easier to interpret. In mathematical terms this means in the following scenarios we will always have:

$$\frac{n_{A1}}{n_{A1}+n_{B1}} = \frac{n_{B2}}{n_{A2}+n_{B2}}$$

In all of the following cases I will compare 3 levels of skew, 50% which means the population is not geographically segregated at all, 65% which would mean the population is relatively geographically segregated, and finally 85% which would be a very high degree of geographical segregation.

3.3.1 Balanced Population

In this scenario the general population is racially balanced, meaning exactly 50% of the students are in group 2. In mathematical terms this means we are in a world in which:

$$\frac{n_{A1}+n_{B1}}{n_{A2}+n_{B2}} = 1$$

Below in figure 3-2 we see a graph of the degree to which segregation would change if we were to switch from a neighborhood mechanism to a choice mechanism in each of the skew scenarios. The x-axis on this graph is $\log(P_{1,2})$ where as a reminder $P_{1,2}$ represents the degree to which group 1 students are more likely to prefer quality to distance when compared to group 2 students–mathematically we have $P_{1,2} = p_1/p_2$. This means as we move to the right on our graph we are seeing cases in which majority race students care much more about quality than minority race students do.

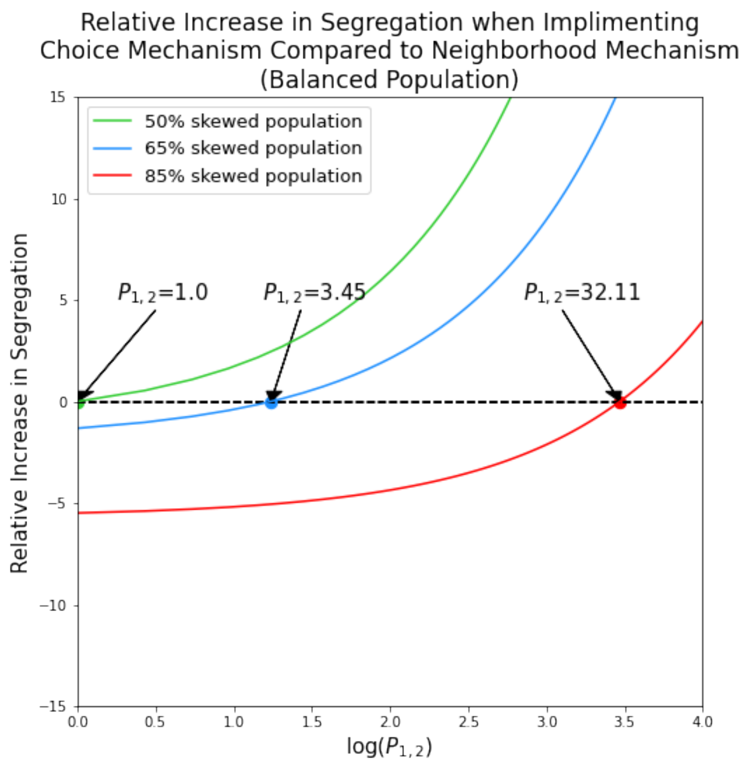

Figure 3-2: Effect of a Choice Mechanism in a Balanced Population

The y-axis is the relative change in segregation if you were to replace a neighborhood mechanism with a choice mechanism. The units aren't actually that informative because they don't directly translate to anything we care about. But generally speaking, the higher we are on the y-axis the more segregated a choice mechanism outcome becomes when compared to a neighborhood mechanism outcome.

In analyzing figure 3-2, we see that as the population becomes more and more skewed, the required $P_{1,2}$ for a choice mechanism to increase segregation becomes higher and higher. This makes a lot of sense intuitively. If we have a situation with extreme skew, meaning the population is intensely geographically segregated. It makes sense that a choice system would usually decrease segregation just because the neighborhood mechanism is going to be so segregated to begin with that it would be hard for any other mechanism to be worse.

The most reasonable skew case is probably the 65% skew. Remember that this measures the percent of group 1 students who live closer to the good school (school A). So even in cases like our example city where most clusters are more than 85% the same race, the skew wouldn't be that high because even a few good schools in regions with more minority students would substantially reduce the skew. In this 65% skew case we see that for the choice mechanism to make school segregation worse, $P_{1,2}$ would need to be at least 3.45. This is high but not inconceivable. For example, if 69% of group 1 students top choice is the better school school, but only 23% of group 2 students prefer the better school over the closer school, then this would be achieved.

You can also notice that when there is no skew, choice systems can only increase segregation. This makes tons of intuitive sense because if the demographic of each neighborhood already exactly match the demographic of the global population, then the neighborhood schools mechanism will create a perfectly non-segregated outcome, from which you can only go downhill.

3.3.2 Majority-Minority Population

The next scenario I'm going to analyze is the majority-minority scenario. The term majority-minority sounds like an oxymoron but is a relatively common phrase that is typically used to describe situations in which the majority of a population is made up from disadvantaged minorities. In this scenario we are going to say that 75% of the population is a minority student. In mathematical terms this means:

$$\frac{n_{A1} + n_{B1}}{n_{A2} + n_{B2}} = \frac{1}{3}$$

Below in figure 3-3 we see that same graph as before but now in the majority-minority case.

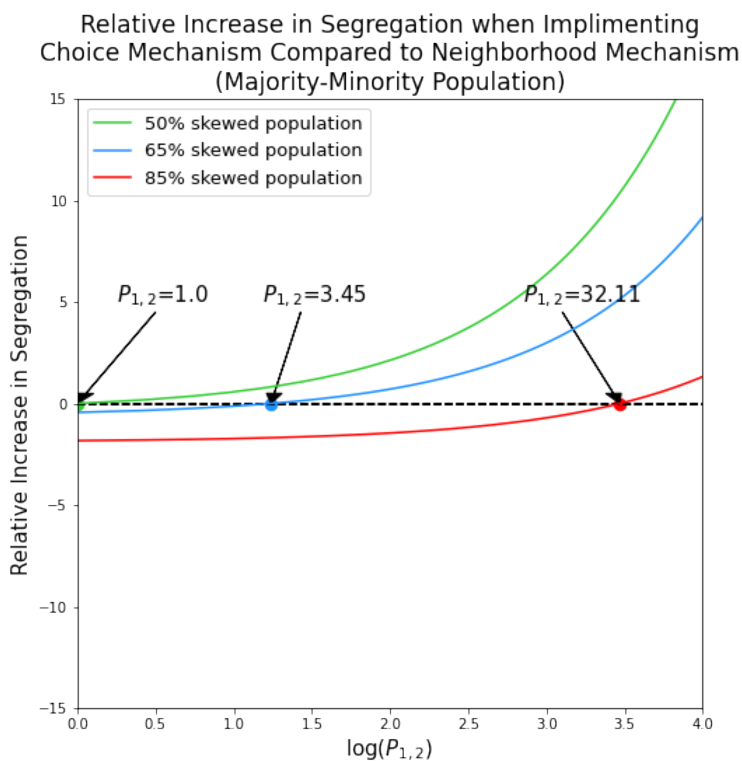

Figure 3-3: Effect of a Choice Mechanism in a Majority-Minority Population

There are two main takeaways from this graph. This first is that all of the cutoff $P_{1,2}$ values are the same. This is something that I prove mathematically later in the paper. What this means is that the underlying global demographic actually has no impact on when a choice mechanism will increase segregation which is counter-intuitive.

The other key thing to note is that all of the lines are much more gradual. What this means is that in the case where most students are in group 2, changing the system has a smaller impact on their representation in the better school, even when they care much less about school quality relative to group 1. This is to be expected because in this scenario the total size of group 1 is relatively small, meaning even if they all strictly prefer quality to proximity, they won't "steal" very many seats away from school A in a choice mechanism.

3.3.3 Majority-Majority Population

The last scenario I analyze is the majority-majority scenario. The term majority-majority is less common than majority-minority but just means the opposite. So a majority-majority population is one in which most students are not from a disadvantaged minority. In this scenario we are going to say that only 25% of the population is a minority student. In mathematical terms this means:

$$\frac{n_{A1} + n_{B1}}{n_{A2} + n_{B2}} = 3$$

Below in figure 3-4 we see that same graph as before but now in the majority-majority case.

There are again two main takeaways from this graph. This first is that again all of the cutoff $P_{1,2}$ values are the same. Again we will see a proof of this later. The other thing to note is that now all of the lines are very steep. What this means is

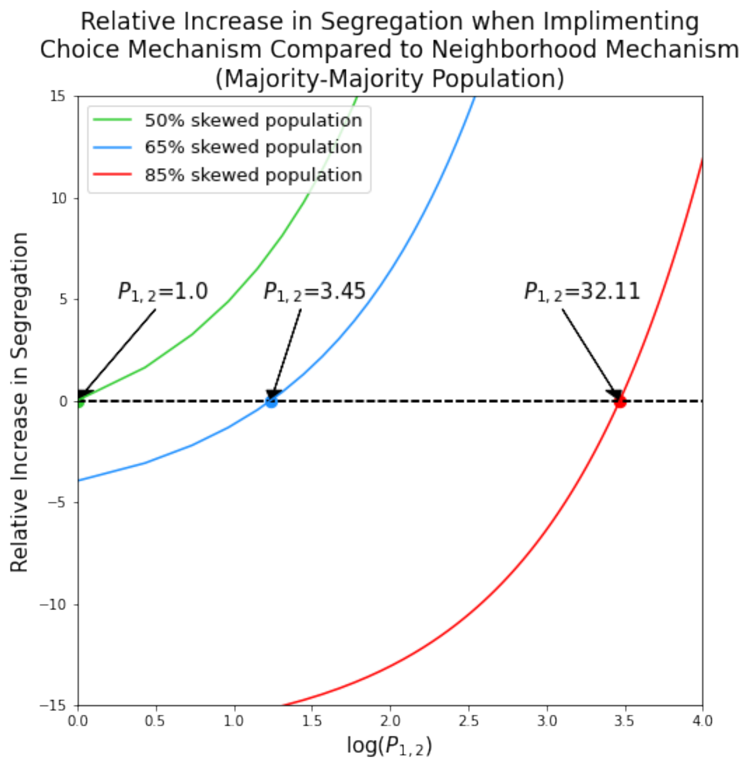

Figure 3-4: Effect of a Choice Mechanism in a Majority-Majority Population

that in the case where most students are in group 1, changing the mechanism from a neighborhood mechanism to a choice mechanism can have a huge impact on group 2 representation in the better school. This makes sense because in this scenario the total size of group 1 is massive, so if they even slightly prefer quality more than group 2 students, they have the ability "steal" away a substantial number of seats in school A that will no longer go to minority students.

3.4 Additional Corollaries from the Model

The model defined in section 3.1 is intentionally as simple as possible, but even with a model as simple as the six variable one I've defined, it is possible to derive many interesting implications. In section 3.2 we saw some of the key insights the model gives us, namely Theorem 1 and corollary 1.1. In this section I am going to examine some other interesting but arguably less significant implications of the model.

3.4.1 Insensitivity to the Number of Students

The first thing I am going to look at is how the crossover point, the point where a choice mechanism becomes more segregated than a neighborhood mechanism, is not sensitive to the total number of students of either race. Imagine we are in a world where there is some fixed n_{A1}, n_{B1}, n_{A2} and n_{B2}. We know already that as $P_{1,2}$ (the ratio of the level to which group 1 students prefer quality compared to group 2 students), increases, the choice mechanism becomes more and more likely to be more segregated than the neighborhood mechanism. Lets define $P^*_{1,2}$ as the point at which the choice and neighborhood mechanisms are equally segregated and examine how that changes as the total number of students in group 1 (defined as $n_{A1} + n_{B1}$) changes.

If we look just at Theorem 1 it initially appears as though this might depend heavily on the total, but a simple rearrangement of corollary 1.1 shows us that in fact:

$$P^*_{1,2} = \left(\frac{n_{A1}}{n_{B1}}\right)\left(\frac{n_{B2}}{n_{A2}}\right)$$

This makes it obvious now that the only thing impacting $P^*_{1,2}$ is the degree to which each group is split between each region. For example if we fix n_{A2} and n_{B2}, then the two following cases; case 1: $n_{A1} = 1000, n_{B1} = 100$ and case 2: $n_{A1} = 100, n_{B1} = 10$, result in the exact same $P^*_{1,2}$ even though the total number of group 1 students changes dramatically. This fact explains why in section 3.3 when I graph the relative segregation curves for varying population splits, the x-intercept values never change.

3.4.2 Sensitivity to Various Parameters

The model considers a situation in which one demographic, group 1, has a systematic advantage in that a group 1 student is more likely to live closer to the better school than a group 2 student is. If we recall from section 3.2, this means that we define the change in segregation (which for this section we will denote as Δ_S) from switching from a neighborhood mechanism to a choice mechanism to be:

$$\Delta_S = n_{A2} - \mathbb{E}[\text{\# of group 2 students assigned to school } A \text{ with choice mechanism }]$$

In full notation this is equivalent to:

$$\Delta_S = n_{A2} - \frac{(n_{A1} + n_{A2})(n_{A2} + n_{B2}p_2)}{n_{A1} + n_{B1}p_1 + n_{A2} + n_{B2}p_2}$$

Which simplifies to:

$$\Delta_S = \frac{n_{A2}n_{B1}p_1 - n_{A1}n_{B2}p_2}{n_{A1} + n_{B1}p_1 + n_{A2} + n_{B2}p_2} \tag{1}$$

Where if the above value is positive then the choice mechanism results in more segregation, and if the value is negative then the choice mechanism results in less segregation. In this section we will be taking the partial derivative of line (1) for each of the variables and thinking about why the derivative points in the direction it does.

Lets first look at n_{A1}, which as a reminder represents the number of group 1 students who live closer to school A than to school B. Taking the partial of Δ_S with respect to n_{A1} gives us the following corollary:

Corollary 1.2

$$\frac{\partial \Delta_S}{\partial n_{A1}} = -\frac{(n_{A2} + n_{B2}p_2)(n_{B1}p_1 + n_{B2}p_2)}{(n_{A1} + n_{B1}p_1 + n_{A2} + n_{B2}p_2)^2} < 0$$

We can observe here that this term will always be negative because all of our variables must be non-negative. p_1 and p_2 are non-negative because they're probabilities, and n_{A1}, n_{B1}, n_{A2} and n_{B2} are because they represent numbers of students. If we recall, as Δ_S gets larger, the degree to which a choice mechanism is more segregated than a neighborhood mechanism also increases. This means that as n_{A1} increases, the choice mechanism becomes more and more appealing. If we think about what this means it actually makes quite a bit of sense. This tells us that if we add more and more group 1 students to the region around school A, that a choice mechanism becomes more appealing relative to the neighborhood mechanism. In fact, if we look closely at corollary 1.1 we can see that there has to be some amount of students we can add to n_{A1} after which the choice mechanism is less segregated than the neighborhood mechanism. The reason for this intuitively has less to do with how n_{A1} impacts the choice mechanism and more to do with how it impacts the neighborhood mechanism. When n_{A1} is extremely high, the neighborhood mechanism will result in an outcome in which the vast majority of school A students are from group 1. This lowers the bar for the necessary number of group 2 students who live closer to school B but prefer school A that is required for the choice mechanism to lower segregation.

Next lets look at n_{A2}, which as a reminder represents the number of group 2 students who live closer to school A than to school B. Taking the partial of Δ_S with respect to n_{A2} gives us the following corollary:

Corollary 1.3

$$\frac{\partial \Delta_S}{\partial n_{A2}} = \frac{(n_{A1} + n_{B1}p_1)(n_{B1}p_1 + n_{B2}p_2)}{(n_{A1} + n_{B1}p_1 + n_{A2} + n_{B2}p_2)^2} > 0$$

We can observe here that this term will always be positive because all of our variables must be non-negative. This means that as n_{A2} increases, the choice mechanism becomes less appealing relative to the neighborhood mechanism. In other words, the more group 2 students who live closer to school A, the less the choice mechanism will help segregation when compared to the neighborhood mechanism. The intuitive reasoning for why this is true is similar to with n_{A1} in that it has less to do with hurting the choice mechanism and more to do with helping the neighborhood mechanism. If n_{A2} is large, meaning a sizable number of group 2 students live near school A, then segregation under the neighborhood mechanism is relatively low. When segregation is low under the neighborhood mechanism, meaning lots of group 2 students attend school A, it becomes harder for the choice mechanism to make a positive impact.

Next lets look at n_{B1}, which as a reminder represents the number of group 1 students who live closer to school B than to school A. Taking the partial of Δ_S with respect to n_{B1} gives us the following corollary:

Corollary 1.4

$$\frac{\partial \Delta_S}{\partial n_{B1}} = \frac{p_1(n_{A1} + n_{A2})(n_{A2} + n_{B2}p_2)}{(n_{A1} + n_{B1}p_1 + n_{A2} + n_{B2}p_2)^2} > 0$$

We can observe here that this term will always be positive because all of our variables must be non-negative. This means that as n_{B1} increases, the choice mechanism becomes less appealing relative to the neighborhood mechanism. In other words, the more group 1 students who live closer to school B, the less the choice mechanism will help segregation when compared to the neighborhood mechanism. The reasoning behind this has to do with who applies to school A under the choice mechanism.

When we switch from a neighborhood to a choice mechanism, school A still has the same $n_{A1} + n_{A2}$ students vying for seats in the school, but also now has all of the students who live closer to school B but prefer A anyway competing as well. Of these newcomers $n_{B1}p_1$ will be from group 1 and so as n_{B1} gets larger, the number of group 1 students who are now competing for seats in school A gets larger and larger. The more group 1 students compete for seats, the less will be assigned to group 2 students. This is why adding more group 1 students to the region around school B hurts a choice mechanism when compared to a neighborhood mechanism.

Next lets look at n_{B2}, which as a reminder represents the number of group 2 students who live closer to school B than to school A. Taking the partial of Δ_S with respect to n_{B2} gives us the following corollary:

Corollary 1.5

$$\frac{\partial \Delta_S}{\partial n_{B2}} = -\frac{p_2(n_{A1} + n_{A2})(n_{A1} + n_{B1}p_1)}{(n_{A1} + n_{B1}p_1 + n_{A2} + n_{B2}p_2)^2} < 0$$

We can observe here that this term will always be negative because all of our variables must be non-negative. This means that as n_{B2} increases, the choice mechanism becomes more appealing relative to the neighborhood mechanism. In other words, the more group 2 students who live closer to school B, the more the choice mechanism will help segregation when compared to the neighborhood mechanism. This is true for very similar logic to why n_{B1} rising hurts the choice mechanism. When we switch from a neighborhood to a choice mechanism, school A still has the same $n_{A1} + n_{A2}$ students vying for seats in the school, but also now has all of the students who live closer to school B but prefer A anyway competing as well. Of these newcomers $n_{B2}p_2$ will be from group 2 and so as n_{B2} gets larger, the number of group 2 students who are now competing for seats in school A gets larger and larger. Since the goal is to increase the number of group 2 students at school A, this is a strictly positive thing from the perspective of the choice mechanism, explaining our finding.

Our last two variables are the probabilities. We'll start with p_1. As a reminder p_1 represents the probability that a student in group 1 prefers quality over proximity, or in the context of our model, the probability that somebody in group 1 who lives closer to school B than to school A will prefer school A anyway. Taking the partial of Δ_S with respect to p_1 gives us the following corollary:

Corollary 1.6

$$\frac{\partial \Delta_S}{\partial p_1} = \frac{n_{B1}(n_{A1} + n_{A2})(n_{A2} + n_{B2}p_2)}{(n_{A1} + n_{B1}p_1 + n_{A2} + n_{B2}p_2)^2} > 0$$

We can observe here that this term will always be positive because all of our variables must be non-negative. This means that as p_1 increases, the choice mechanism becomes less appealing relative to the neighborhood mechanism. In other words, the more group 1 students care about school quality, the more likely it is that a choice mechanism would increase segregation when compared to a neighborhood mechanism. The reason this is true is very similar to why increase n_{B1} hurt the choice mechanism. As stated before, when we switch from a neighborhood to a choice mechanism, school A still has the same $n_{A1} + n_{A2}$ students vying for seats in the school, but also now has all of the students who live closer to school B but prefer A anyway competing as well. Of these newcomers $n_{B1}p_1$ will be from group 1 and so as p_1 gets larger, the number of group 1 students who are now competing for seats in school A gets larger and larger. The more group 1 students who compete for seats, the fewer seats will be assigned to group 2 students. This is why situations in which group 1 students are really quality sensitive can be more segregated under a choice mechanism than a neighborhood mechanism.

Our final variable is p_2. As a reminder, p_2 represents the probability that a student in group 2 prefers quality over proximity, or in the context of our model, the probability that somebody in group 2 who lives closer to school B than to school A will prefer school A anyway. Taking the partial of Δ_S with respect to p_2 gives us the following and final corollary:

Corollary 1.7

$$\frac{\partial \Delta_S}{\partial p_2} = -\frac{n_{B2}(n_{A1}+n_{A2})(n_{A1}+n_{B1}p_1)}{(n_{A1}+n_{B1}p_1+n_{A2}+n_{B2}p_2)^2} < 0$$

We can observe here that this term will always be negative because all of our variables must be non-negative. This means that as p_2 increases, the choice mechanism becomes more appealing relative to the neighborhood mechanism. In other words, the more group 2 students care about school quality, the more likely it is that a choice mechanism would decrease segregation when compared to a neighborhood mechanism. The rational here is very similar to why increasing n_{B2} helps the choice mechanism. when we switch from a neighborhood to a choice mechanism, school A still has the same $n_{A1}+n_{A2}$ students vying for seats in the school, but also now has all of the students who live closer to school B but prefer A anyway competing as well. Of these newcomers $n_{B2}p_2$ will be from group 2 and so as p_2 gets larger, the number of group 2 students who are now competing for seats in school A gets larger and larger. The more group 2 students who compete for seats, the more they will be assigned them. This is why situations in which group 2 students do not care about quality (i.e. p_2 is low) can be more segregated under a choice mechanism than a neighborhood mechanism.

Summary of Sensitivity Analysis:

As a general summary of what is an unfortunately verbose subsection, we can say the following about each of our variables by writing all of our derived corollaries in words as follows:

1. Having lots of group 1 students who live near school A is an indication that a choice mechanism might lower segregation when compared to a neighborhood mechanism.

2. Having lots of group 2 students who live near school A is an indication that a

choice mechanism might raise segregation when compared to a neighborhood mechanism.

3. Having lots of group 1 students who live near school B is an indication that a choice mechanism might raise segregation when compared to a neighborhood mechanism.

4. Having lots of group 2 students who live near school B is an indication that a choice mechanism might lower segregation when compared to a neighborhood mechanism.

5. Group 1 students caring a lot about school quality is an indication that a choice mechanism might raise segregation when compared to a neighborhood mechanism.

6. Group 2 students caring a lot about school quality is an indication that a choice mechanism might lower segregation when compared to a neighborhood mechanism.

Chapter 4

An Analysis of a Sample School District

The key question being investigated in this thesis is how choice mechanisms, like the one described above, can actually lead to more racial segregation in schools than neighborhood mechanisms do–particularly when there is geographic racial segregation in housing. In this chapter I use data from an actual U.S. school district as a case study of this phenomenon occurring in the wild.

4.1 The Data

The data for the project, as mentioned in the acknowledgements, was graciously given by an anonymous school district in the U.S. as part of a much larger project. Because of this there is much more data than was actually used for this thesis. In this section, I am going to lay out the data that was relevant to this work. All of the following data covers the 2014-2015 school year through the 2019-2020 school year.

The first, and most crucial data set is the assignment file. The assignment file gave the following information:

1. Student-school assignments each year
2. Student home locations as latitude, longitude tuples

The next data set is the school files. The school files cover the same year range and tell us the following information:

1. School locations as latitude, longitude tuples
2. School tiers in each year

The last large data set is the choice basket files. These files gave us the following information:

1. Student choice baskets for each year. Meaning which schools they were able to submit rankings over
2. Student rankings over their own choice basket for each year

The final piece of information was a two-column file that included just the student ID number and race of each student in the system.

4.2 The School District from Afar

In this section, I give some context for the school district whose data I use to examine these two mechanisms. As a reminder, the district requested to remain anonymous so all of the numbers I will give in this section have been generously rounded in order to obscure the identity of the city.

The school district has around 50,000 students with a little over 4,000 students each year. The student body is quite diverse and is actually majority minority, with the Black and Hispanic student populations making up almost 75% of the total student body. The remaining students are primarily White and Asian although there is a non-zero number who identify as different races, for example there is a minimal but present Native-American population.

The city itself, while diverse, is not integrated particularly well geographically speaking. This is a difficult thing to quantify but I have included a map below (see figure 4-1) that shows the percentage of students living in each geocode[1] that are either Black or Hispanic. What we see is that in the center of the city almost all students are either Black or Hispanic, but that on the edges of the city almost none are. This geographic racial segregation is a point of concern for the district, because they do not want to be left with a system in which there are some schools that are entirely White and Asian, and others that are entirely Black and Hispanic. Note that the image has been significantly distorted in an attempt to keep the city's outline unrecognizable while preserving the relative proximity of geocodes.

[1]Geocodes are small regions, usually just a few square blocks, that divide cities.

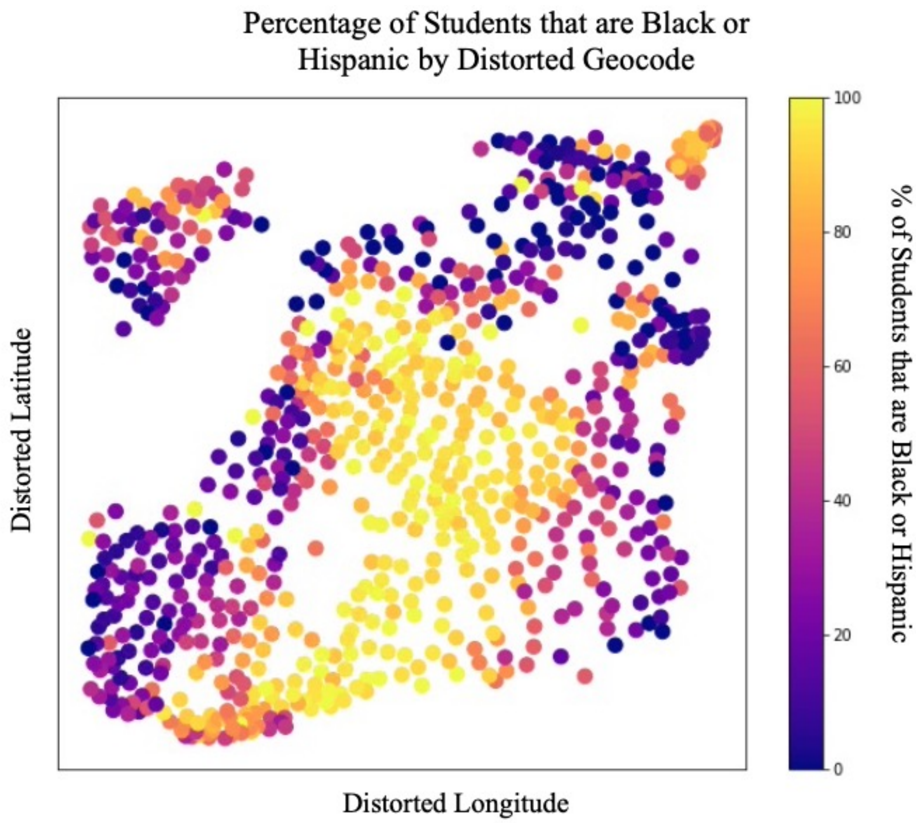

Figure 4-1: Racial Demographic Map

4.3 Measuring the Segregation of an Assignment Outcome

Measuring racial segregation in a school system is a challenging problem. We will obviously never be in a fully segregated outcome and so are instead interested in segregation as a sliding scale. The idea being that a completely unsegregated outcome would be one in which the racial demographic of every school was more or less the same as the racial demographic of the system as a whole. Some degree of segregation is unavoidable because of the geographical housing segregation (see figure 4-1), and so with the exception of a mechanism that assigns students randomly, any reasonable mechanism is going to result in some degree of segregation.

In this section I will look at two different measures of segregation. The first measure is what is known as racial isolation. A racially isolated school is defined as a school that has at least $x\%$ of its students coming from one racial subgroup. The total racial isolation of an outcome is the number of schools that are racially isolated at the selected cutoff. In this analysis, I look at two versions of the cutoff. In version 1, I consider a school to be racially isolated if at least 80% of the students are either Black or Hispanic. In version 2, I only consider a school racially isolated if at least 90% of the students are Black or Hispanic. The racial isolation measure allows us to compare the relative segregation of two outcomes by simply observing the difference in the number of racially isolated schools in each case. Obviously here we would call the mechanism that results in more racially isolated schools, the more segregated mechanism.

The second measure of segregation I will use is slightly more unorthodox. The main issue with the racial isolation measure is that it is very sensitive to the cutoff being used. Selecting a slightly different cutoff value could easily change the number of isolated schools by 5 or 6 which on a scale of about 100 schools is very large. The following metric, which I will refer to as the relative segregation metric, does not have that issue.

The relative segregation metric is a way of comparing two outcomes and deciding which of the two is more segregated. It goes as follows. Let $\bar{\pi}$ be the percent of students in the student body that are Black or Hispanic. Let $\pi_{1,i}$ be the percent of students who are Black or Hispanic at school i in outcome 1 and define $\pi_{2,i}$ similarly for outcome 2. For any school i, if $|\pi_{2,i} - \bar{\pi}| < |\pi_{1,i} - \bar{\pi}|$, or in words, if the percent of students who are Black or Hispanic at school i in outcome 2 is closer to the underlying percent than it was at school i in outcome 1, then we would say outcome 2 made school i less segregated compared to outcome 1. If outcome 2 makes at least half of the schools less segregated we would say it is the less segregated outcome, otherwise we would say outcome 1 is less segregated. In practice I report the number of schools who are made "less segregated" by each mechanism instead of just assigning a binary label.

4.4 Simulating a Neighborhood Mechanism

One interesting part of this project was creating the counterfactual assignment outcome for a neighborhood mechanism. The beauty of the neighborhood mechanism is that all you need to run it is the locations of all the students, and the locations of the schools.

The simulation goes as follows. For each student we calculate their distance to every school in the district, we then sort the schools by distance and save that order as the students "preferences" over the schools. For example, a students top preference is for the school that is closest to their home. We then take each school and calculate its distance to every student, sort all students by their distance, and then save that order as the schools "preferences" over the students. The last setup step is to estimate the capacity of each school. To do this I took the maximum observed number of enrolled students over the period we had data for and assumed that to be maximum capacity.

Once we have the preferences and capacities, we simply use the Gale-Shapley deferred acceptance algorithm to create matchings. These student-school matchings represent the outcome in which everyone is attending the closest school possible. A key property of this outcome is that if a student is not assigned to their closest school, then it must mean that every student who was assigned to the school lives closer to it than our selected student.

It should be noted that this method is a good estimate for what the outcome would have been if a neighborhood system were used but it is not entirely accurate. The main issues it ignores are that some students are assigned to special programs within schools which have different capacities, and that some students are automatically enrolled in schools for external reasons such as having family members who work in the building. It also of course ignores a large issue which is that if the neighborhood system were what was being used, families might move to guarantee their child a seat at a better school.

4.5 Segregation under the Neighborhood Mechanism vs. the Choice Mechanism

In this section, we're going to use the data provided by our sample school district to see how this segregation actually changes in a choice mechanism compared to a neighborhood mechanism. We will be using the two measures of segregation as defined in section 4.3.

4.5.1 Methodology

The methodology for this analysis goes as follows. The first step was to create a neighborhood mechanism simulation of each year of data. The year range I used here was the 2014-2015 school year up to and including the 2019-2020 school year[2]. For more information on how the neighborhood simulation was created please refer to section 4.4.

The result of the neighborhood simulation is a panel style data set where each

[2]Note that I will sometimes refer to a school year by just a single year. Whenever I do that I am referring to the calendar year when school began. For example I might refer to the 2015-2016 school year as just 2015.

row contains the following information:

1. School year
2. Student ID number
3. School assignment in the neighborhood simulation

By taking the assignment file, I was able to create the same style of data set for the actual observed outcome under the choice mechanism. I then joined each of these data sets with the race data set and created a new column which was a dummy variable for whether or not the student was either Black or Hispanic.

I then took my choice mechanism outcome file and my the neighborhood mechanism outcome file, and using them created a new single file with the following columns:

1. School year
2. School code[3]
3. Number of students actually assigned to the school in the choice mechanism
4. Percent of students assigned through the choice mechanism who are Black or Hispanic
5. Number of students assigned to school in the neighborhood mechanism simulation
6. Percent of students assigned through the neighborhood mechanism simulation who are Black or Hispanic

This data set, which I will refer to as the school level outcome data set, is what is used for the results shown in the following subsection.

[3]A school code is just a number that uniquely identifies a given school. I use it instead of the school name simply because its shorter.

4.5.2 Results

Using the school level outcome data set I collected the following 6 segregation metrics for each mechanism in each year.

1. Number of schools that had at least 80% racial isolation under the given mechanism[4]

2. Number of students assigned to schools with at least 80% racial isolation under the given mechanism

3. Number of schools that had at least 90% racial isolation under the given mechanism

4. Number of students assigned to schools with at least 90% racial isolation under the given mechanism

5. Number of schools that were less segregated under the given mechanism. To see how this is calculated refer to section 4.3

6. Number of students who are at a less segregated school under the given mechanism

We'll start this analysis by looking at the isolation results. Below you'll find a table with all of the isolation metrics (number of schools and number of students at both the 80% and 90% cutoffs), sorted by year and mechanism. There is also a line representing the average for each mechanism across all years.

Table 4.1 contains a lot of information, so I will walk through the key findings. The first thing I should point out is that the green highlighted cells represent cases where in that year, for that metric, the highlighted cells mechanism outperformed the

[4]Note that when I measure this I measure it for each school in that given year. So if a mechanism assigned 100% Black and Hispanic to a school in a year, I would consider the school to be segregated for that year, even if in years past the Black and Hispanic populations had been relatively low.

Year	Mechanism	# of Schools 80% Isolated	# of Students in 80% Isolated Schools	# of Schools 90% Isolated	# of Students in 90% Isolated Schools
2014	Choice	29	2118	17	1247
	Neighborhood	28	2059	11	874
2015	Choice	26	1668	19	1250
	Neighborhood	22	1471	12	854
2016	Choice	31	2244	8	658
	Neighborhood	25	1821	10	792
2017	Choice	29	2131	18	1401
	Neighborhood	24	1839	12	1074
2018	Choice	28	2280	11	987
	Neighborhood	26	2105	14	1262
2019	Choice	26	1880	11	801
	Neighborhood	24	1743	9	655
Average	Choice	28.2	2053.3	14.0	1057.1
	Neighborhood	24.8	1839.5	11.3	918.3

Table 4.1: Raw Results of School Racial Isolation Analysis

alternative mechanism. For example, on the bottom right of the table, the cell that says 918.3 is highlighted green because on average, in the neighborhood mechanism simulation, fewer students were assigned to schools with at least 90% racial isolation than were under the observed choice mechanism (918.3 vs. 1057.1).

The advantage of the green highlights are that they allow us to quickly see where one mechanism completely out-performs the other. The result, which was actually the observation that prompted this thesis, is that we see the neighborhood school mechanism out-performing the choice mechanism in every isolation metric on average. In fact in 4 of the 6 years, the neighborhood mechanism out performs the choice mechanism it in every metric outright.

The reason this result is so surprising is because of how extreme the geographic racial segregation had been in our sample city. Had the city been more evenly mixed you would expect the neighborhood mechanism to result in a very well balanced outcome, which would have meant it was unsurprising if the choice mechanism was worse. A key thing to note here is that even though the neighborhood mechanism is less segregated than the choice mechanism, it still results in many racially isolated

Year	Mechanism	# of Schools less Segregated w/ Mechanism	# of Students in Less Segregated Schools w/ Mechanism
2014	Choice	37	2588
	Neighborhood	39	2697
2015	Choice	36	2378
	Neighborhood	35	2295
2016	Choice	40	2833
	Neighborhood	35	2476
2017	Choice	38	3061
	Neighborhood	34	2495
2018	Choice	40	3282
	Neighborhood	36	2806
2019	Choice	35	2266
	Neighborhood	41	3127
Average	Choice	37.7	2734.3
	Neighborhood	36.7	2649.0

Table 4.2: Raw Results of School Racial Segregation Analysis

schools at both the 80% and 90% thresholds. This is an unavoidable consequence of implementing a neighborhood mechanism in a racially segregated city.

In table 4.2, we see the raw results of the other segregation metric, the relative comparison. As a reminder here we are for each school, deciding if it is less segregated under the neighborhood mechanism or the choice mechanism, as decided by which mechanism assigns a percentage of Black and Hispanic students that is closer to the global percentage in the city. Here we see the opposite result from the isolation analysis. Now it appears as though the choice mechanism slightly outperforms the neighborhood mechanism–it results in less segregation for 1 school on average.

This result is obviously smaller than the racial isolation results, but it shows two important results. First, it shows that measuring segregation is a really hard thing to do, and so even something as simple as deciding which of two mechanisms results in less segregation won't necessarily have an easy answer. Second, it shows us is that the neighborhood mechanism and choice mechanisms may be better in different

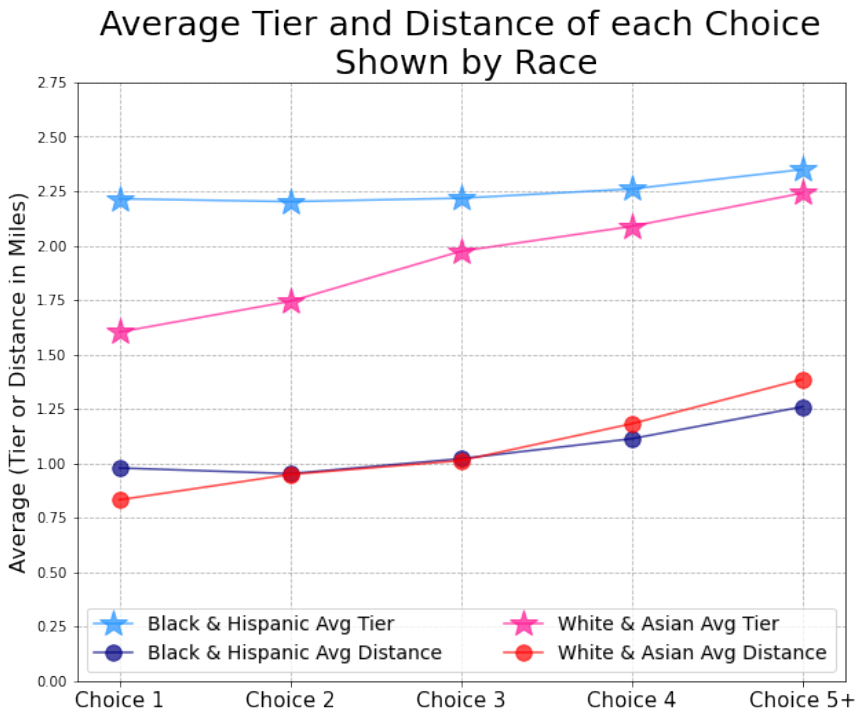

Figure 4-2: Tier and Distance of Choices by Race

ways. For example these combinations of results would suggest that the neighborhood mechanism has the potential to really cut down on the number of schools that have extreme racial isolation–but that the choice mechanism might be slightly better at lowering segregation for schools that are not already racially isolated.

4.6 Racial Differences in School Preferences

If you remember back to the model section of my thesis, you'll recall that one of the key parameters was the difference in the degree to which each subgroup preferred quality schools to nearby schools. In this section, I examine the observed differences in preferences for Black and Hispanic families when compared to White and Asian families.

In figure 4-2 we see the average tier and distance of the school for each ranking made by families of each racial subgroup. For example, the top left start point tells us that the average Tier of Black and Hispanic families first choice school is just around 2.25. The lower two lines on this graph show us the average distance away the school was. The first takeaway from this graph is that all families appear to have a very similar preference towards distance, with their first choice school typically being slightly less than a mile away and all choices being within a mile and a half.

The most apparent takeaway from this graph is that there is a huge gap in the tier lines, particularly for the first two choices. This tells us that on average, White and Asian families are much more likely to use their top two choices on better tiered schools than Black and Hispanic families are. One possible explanation for this is that White and Asian families simply live closer to the better tiered schools and so Black and Hispanic families would rank these schools higher if they lived closer.

To test this hypothesis, I for each student figured out which school was closest to their home, and then put the student into one of two bins. The first bin was for students whose closest school was a tier 1 or tier 2 school–this represented about 76% of White and Asian Students and 64% of Black and Hispanic students. The remaining students were in the second bin which represents students who's closest school was tier 3 or 4. I then remade the above graph for each bin of student.

In figure 4-3 we see students who's nearest school is a tier 1 or tier 2 school. If we look at the pink and light blue lines we see that we are left with the same result as before. White and Asian families are still choosing substantially better tiered schools for their top two choices than Black and Hispanic families are. We should also note that there is again a very marginal difference in the lower two lines, indicating that families are similar in their preferences towards distance.

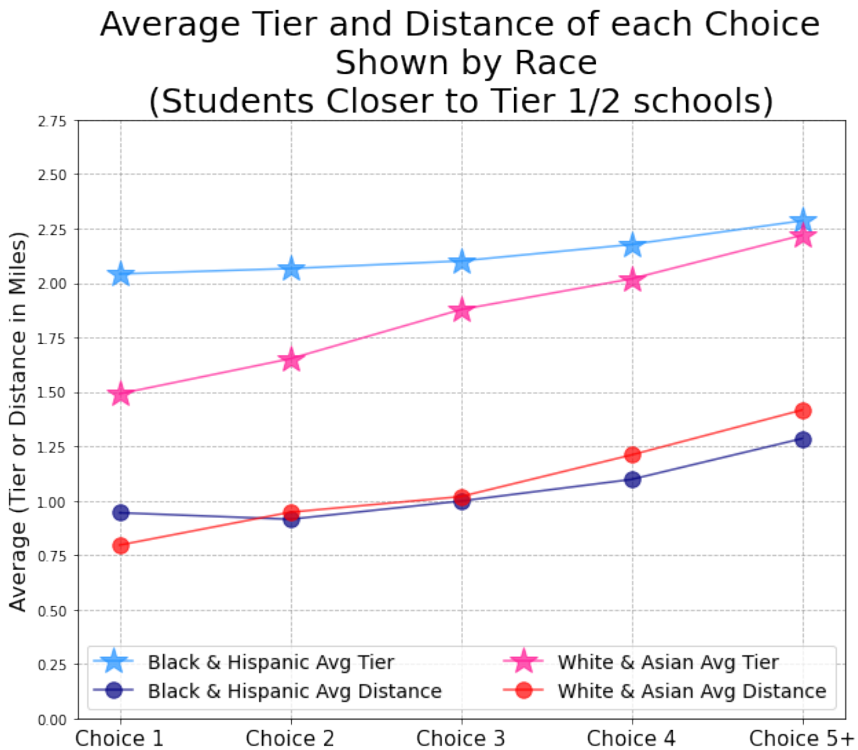

Figure 4-3: Tier and Distance of Choices by Race for Students Nearer to Tier 1 or Tier 2 Schools

In figure 4-4 we now see the inverse of our last graph. This shows us students who's nearest school is a tier 3 or tier 4 school. The takeaways here are the same as the last graph; that White and Asian families rank better tiered schools higher than Black and Hispanic families do. The new takeaway, and maybe the most significant result, is that the average first choice tier for a White or Asian student who lives closest to a bad tiered school is a better tier than the average tier of a first choice school for Black and Hispanic students who live closer to good tiered schools. This makes for a compelling argument that White and Asian families are substantially more drawn to school quality over school proximity than Black and Hispanic families are.

Putting this section together reveals a key piece of evidence for why the choice mechanism may have had a negative impact on segregation for our sample city. These graphs strongly suggest that White and Asian families place a much higher weight on

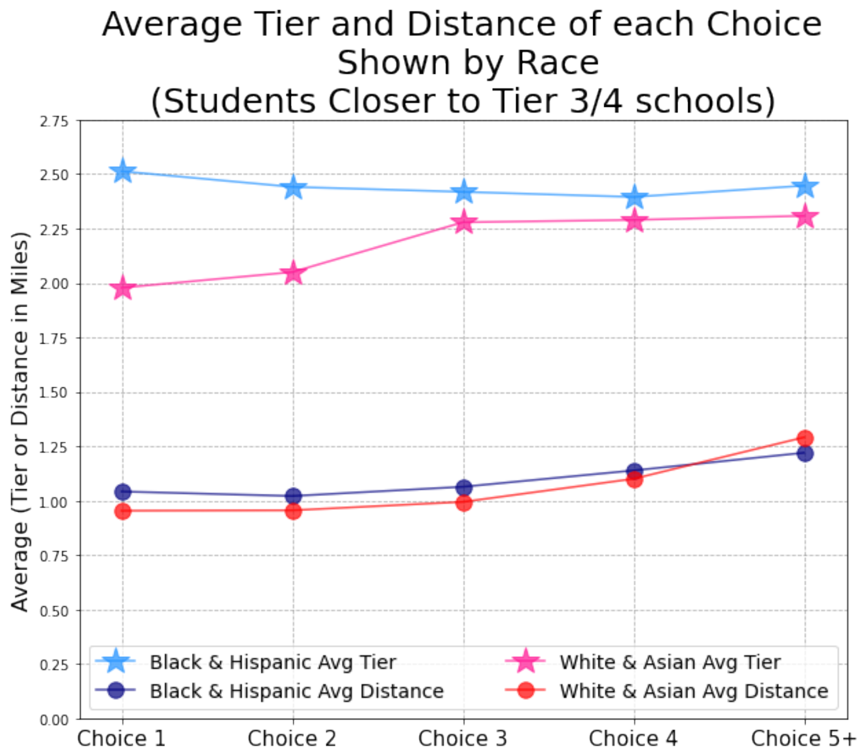

Figure 4-4: Tier and Distance of Choices by Race for Students Nearer to Tier 3 or Tier 4 Schools

the quality of a school (or at least on the listed tier) than Black and Hispanic families do. As indicated by the model outlined earlier, and by logic, this would indicate that better tiered schools are more likely to be filled by White and Asian families, leaving the remaining schools full of largely Black and Hispanic students, potentially to the point of racial isolation in some cases. This provides real world evidence of corollary 1.6, which stated in words that: "group 1 students caring a lot about school quality is an indication that a choice mechanism might raise segregation when compared to a neighborhood mechanism" were in this example Group 1 would of course be the White and Asian students.

These graphs are obviously not able to prove anything about the preference functions of different families, there are plenty of unobserved variables. But they do suggest that one of the main reasons the choice mechanism appears to cause more segregation in our sample city is simply because White and Asian families place a relatively higher weight on school quality than Black and Hispanic families do.

Chapter 5

Conclusion and Main Results

In this section I am going to give a brief summary of all of my key results. I've broken it down into the empirical results, where I look at the results of my case study city, and the theoretical results, where I look at the mathematical results that come from my model. I will also conclude with a brief overview of the project as a whole.

5.1 Empirical Results

The goal of this thesis was to examine how choice based student assignment mechanisms, which are generally seen as being more equitable, can sometimes result in more segregation than neighborhood mechanisms. A major portion of this paper was dedicated towards looking at sample city where we see this exact phenomenon occurring. The reason our sample city is particularly interesting is because it is geographically racially segregated, which is typically seen as a strong reason not to implement a neighborhood mechanism.

What we instead observe is that for most schools, the neighborhood and choice mechanisms result in very similar levels of segregation, but that at the extremes the neighborhood mechanism appears to outperform the choice mechanism. The choice mechanism resulted in on average about three more schools racially isolated at both the 80% and 90% cutoffs when compared to the neighborhood mechanism.

We also saw that one reason the choice mechanism might increase segregation in our city in particular is because of the differences in school preferences for families of different races. We found that White and Asian families tend to use their top choices on top tier schools, even if they do not live in the immediate vicinity of one, while Black and Hispanic families are much less sensitive to the tier of the school.

Even given my empirical findings for this sample city that find the neighborhood mechanism to be less segregated, I would not necessarily argue to move away from the currently implemented choice mechanism and to instead use a neighborhood mechanism. In my analysis, I am only able to simulate the neighborhood mechanism given a strong set of assumptions. Maybe the most notable is that I assume nobody will move. This is not the reality, as families with resources would inevitably shift to move away from particularly poor performing schools and shift towards particularly well performing schools–a gradient that would almost certainly worsen the geographic racial segregation. There are also benefits to choice mechanisms that I don't discuss in this paper. For example they do give families a lot more agency in selecting where their child will attend, without being forced to move. Ultimately the decision to adjust a school assignment mechanism involves much more nuanced problems that this thesis does not explore, and so it should not be considered as advice.

5.2 Theoretical Results

In section 3 I explored a simple model of the school assignment system in which there were only two schools and two types of students. This model, while simple, does shed some light on the underlying forces in the system. The main theorem of the paper finds that segregation will increase under a choice mechanism whenever the number of group 1 students who prefer the better school is greater than the number of group

2 students who prefer the better school multiplied by the ratio of group 1 students to group 2 students who are closer to the better school.

In the later portions of section 3 we define Δ_S as the amount segregation increases under a choice mechanism when compared to a neighborhood mechanism and we find that the point at which a neighborhood mechanism becomes better ($\Delta_S = 0$) depends on the ratio of each population between the two regions, but not on the total number of students in either group. We also find that Δ_S is strictly increasing, meaning choice mechanisms only get worse, as the following increase; the number of group 2 students living near the better school, the number of group 1 students living near the worse school, and the degree to which group 1 students care about school quality. On the flip side we find that Δ_S is strictly decreasing, meaning choice systems only get better, as the following increase; the number of group 1 students living near the better school, the number of group 2 students living near the worse school, and the degree to which group 2 students care about school quality.

5.3 Conclusion

In summary, this thesis identified conditions under which choice based school assignment mechanisms could actually increase segregation in schools when compared to neighborhood mechanisms, even in cities with geographic racial segregation. I demonstrate how this is possible, with the primary underlying cause appearing to be differences in school preferences between races. I find that in my real world case study, White and Asian families are much more likely to use their top choices on top tier schools than Black and Hispanic families are–even when the white and Asian families live farther away from the top tier schools. This aligns with the theoretical results, which tell us that the more the already advantaged population cares about school quality, the more likely a choice mechanism is to increase segregation when compared to a neighborhood mechanism. Thus, I conclude that under the right conditions, choice mechanisms can lead to more segregation.

Bibliography

[1] Rebecca Allen. Allocating pupils to their nearest secondary school: The consequences for social and ability stratification. *Urban Studies*, 44(4):751–770, 2007.

[2] Douglas A. Archbald. School choice, magnet schools, and the liberation model: An empirical study. *Sociology of Education*, 77(4):283–310, 2004.

[3] Jonathan Schellenberg Atila Abdulkadiroglu, Parag A. Pathak and Christopher R. Walters. Do parents value school effectiveness? *American Economic Review*, 110(5):1502–1539, 2020.

[4] Stephen J. Ball and Carol Vincent. 'i heard it on the grapevine': 'hot' knowledge and school choice. *British Journal of Sociology of Education*, 19(3):377–400, 1998.

[5] Courtney A. Bell. All choices created equal? the role of choice sets in the selection of schools. *Peabody Journal of Education*, 84(2):191–208, 2009.

[6] Holmlund H. Lindahl M Böhlmark, A. School choice and segregation: Evidence from sweden. *Working Paper, IFAU - Institute for Evaluation of Labour Market and Education Policy*, 2015.

[7] Holmlund H. Lindahl M Böhlmark, A. Parental choice, neighbourhood segregation or cream skimming? an analysis of school segregation after a generalized choice reform. *Journal of Population Economics*, 29(4):1155–1190, 2016.

[8] Moe T Chubb J. Politics, markets, and america's schools. *Washington, DC: The Brookings Institute.*, 1990.

[9] T. M. Davis. School choice and segregation: "tracking" racial equity in magnet schools. *Education and Urban Society*, 46(4):399–433, 2014.

[10] Driessena G. Sleegers P Denessen, E. Segregation by choice? a study of group-specific reasons for school choice. *Journal of Education Policy*, 20(3):347–368, 2005.

[11] Viteritti J. Choosing equality: School choice, the constitution, and civil society. *Washington, DC: Brookings Institution Press*, 1999.

[12] D. Kessel. School choice, school performance and school segregation: Institutions and design. *(Doctoral dissertation). Stockholm University*, 2018.

[13] Roope Uusitalo Martin Söderström. School choice and segregation: Evidence from an admission reform. *The Scandinavian Journal of Economics*, 112(1):55–76, 2010.

[14] S Mutgan. Free to choose?: Studies of opportunity constraints and the dynamics of school segregation. *PhD dissertation, Linköping University Electronic Press*, 2021.

[15] Salvatore Saporito. Private choices, public consequences: Magnet school choice and segregation by race and poverty. *Social Problems*, 50(2):181–203, 2014.

[16] D. Gale L. S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

[17] Lloyd Shapley and Herbert Scarf. On cores and indivisibility. *Journal of Mathematical Economics*, 1(1):23–37, 1974.