Affirmative Action and Fair School Choice Design for Minority Students

Muntasir Chaudhury

A Thesis
In the Department
of
Economics

Presented in Partial Fulfillment of the
Requirements For the Degree of
Doctor of Philosophy at Concordia University
Montréal, Québec, Canada

February 2023

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: **Mr. Muntasir Chaudhury**

Entitled: **Affirmative Action and Fair School Choice Design for Minority Students**

and submitted in partial fulfillment of the requirements for the degree of

Doctor Of Philosophy (Economics)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

______________________ Chair
Dr. Kregg Hetherington

______________________ External Examiner
Dr. Vikram Manjunath

______________________ Examiner
Dr. Dipjyoti Majumdar

______________________ Examiner
Dr. Axel Watanabe

______________________ Examiner
Dr. Huan Xie

______________________ Supervisor
Dr. Szilvia Pápai

Approved by ______________________
Dr. Christian Sigouin, Graduate Program Director
Department of Economics

__________ 2023 ______________________
Dr. Pascale Sicotte, Dean
Faculty of Arts and Science

Abstract

Affirmative Action and Fair School Choice Design for Minority Students

Muntasir Chaudhury, Ph.D.
Concordia University, 2023

In the first main chapter of this thesis we propose three basic welfare axioms for school choice mechanisms with an affirmative action policy: non-wastefulness, respecting the affirmative action policy, and minimal responsiveness, and show that none of the previously proposed mechanisms satisfy all three welfare axioms. Then we introduce a new mechanism which satisfies the three welfare axioms. This mechanism issues immediate acceptances to minority students for minority reserve seats and otherwise it is based on deferred acceptance. We analyze the fairness and incentive properties of this newly proposed affirmative action mechanism.

In the second main chapter we investigate school choice mechanisms with an affirmative action policy that have appealing responsiveness properties to changes in the strength of the affirmative action policy. We present two intuitive mechanisms which satisfy stringent responsiveness criteria and are the first ones to be studied with such properties. One is fully responsive and results in a welfare improvement for minority students when the minority reserves are increased, while the other one only ensures a welfare improvement when moving to an affirmative action policy. We study further properties of these two school choice mechanisms which suggest trade-offs.

In the third main chapter we use a stylized model to explore whether the choice of the matching mechanism in school choice has an impact on the degree of school segregation. We find that the celebrated Deferred Acceptance and the well-known Top Trading Cycle mechanisms both lead to complete segregation, while the Immediate Acceptance mechanism results in less segregation, even though it has often been replaced by the Deferred Acceptance mechanism on the recommendation of theorists. Our results suggest that in order to reduce school segregation, despite their manipulability, the use of matching mechanisms that rely more on student preferences than schools' priority rankings should be reconsidered. We also study practically relevant special cases which indicate that standardized entrance exams do not aggravate the segregation outcomes but the funding and quality gap among schools does.

Acknowledgments

The journey through the Ph.D. program is often strenuous. A successful accomplishment requires exertion and sacrifices. However, it can also become an exciting jubilant experience if the journey companions are benevolent. I was fortunate to have many benefactors. My profound gratitude to my supervisor, Dr. Szilvia Pápai, cannot be fully expressed in words. Her persistent guidance, inspiration, wisdom, and benefaction made it possible for me to complete my journey. Her help went above and beyond what anyone could expect, always being there compassionately without pause on every occasion.

I also must thank Dr. Damba Lkhagvasuren, who was always extra kind to me and helped me in my difficult times. I would also like to acknowledge the former and current Graduate Program Directors, Dr. Szilvia Pápai, Dr. Damba Lkhagvasuren, and Dr. Christian Sigouin (current), and to the former and current Chairs, Dr. Greg LeBlanc and Dr. Jorgen Hansen, who were always available to help. I also extend my gratitude to the department's kind and compassionate staff, Ms. Elise Melancon and Ms. Émilie Martel.

I must thank my mother, Shyamali Nasreen Chaudhury, for her continuous support in every possible way, providing me with everything I needed. I express my gratitude to my sister Dr. Nuzhat Choudhury for always being an overwhelming source of encouragement and emotional support to boost me up whenever I felt disheartened.

My acknowledgment will be incomplete without recognizing my loving wife, Tabassum Alam, who made the most sacrifice and throughout supported me the most with patience and endurance. I thank her for being a source of encouragement, an altruistic life partner, and a true friend. I must mention my son, Marshad Chaudhury, for giving up so many desires at a young age, understanding like grownups, and being my ultimate source of inspiration.

Contribution of Authors

Chapters 2, 3, and 4 of the thesis are joint work with my supervisor, Professor Szilvia Pápai.

Table of Content

List of Figures

List of Tables

Chapter 1

Introduction

Affirmative action policy is the subject of ongoing debate and research over many academic disciplines, including the matching theory literature. Affirmative action refers to a set of policies and practices that are aimed at increasing opportunities to reduce discrimination, which can be undertaken by employers, university admissions offices, and government agencies. These policies provide a preferential allocation of scarce positions to underrepresented groups of society, namely minorities, with the objective of actively improving their welfare status (Fryer Jr & Loury, 2005; Holzer & Neumark, 2006). In the matching theory literature the design and use of affirmative action policy in a many-to-one matching model, commonly referred to as "school choice," have been actively studied over the last decade, with many puzzles yet to solve.

The three matching mechanisms that are most discussed in the literature are the celebrated Deferred Acceptance (DA) mechanism introduced by Gale and Shapley (1962), the Top Trading Cycles (TTC) mechanisms (Shapley & Scarf, 1974; Pápai, 2000; Abdulkadiroğlu & Sönmez, 2003), and the widely used Immediate Acceptance (IA) mechanism, also known as the Boston Mechanism. Abdulkadiroğlu and Sönmez (2003) introduced a quota-based affirmative action policy which they call "controlled choice." Controlled choice allows for type-specific inflexible quotas at schools; thus, any school may have a "majority quota" restricting the number of majority students that can be assigned to that school with the aim of providing minority students preferential access to that school. Later Hafalir, Yenmez, and Yildirim (2013) introduced a flexible quota called "minority reserves," which are specifically designed for the DA mechanism and resolve the inherent wastefulness of the inflexible quotas.

Unfortunately, neither the quotas nor the minority reserves of Hafalir et al. (2013) guarantee the welfare improvement of the targeted beneficiaries when affirmative action policy is introduced or strengthened, and thus fail to achieve the main objectives of the affirmative action policy. Kojima (2012) showed that under certain situations both the DA and the TTC mechanisms with majority quotas at some schools could return a matching that is Pareto-inferior for all minority students, compared with the case of no affirmative action. In other words, the DA and the TTC are not "responsive" to quota-based affirmative action policy. Doğan (2016) defined a mechanism with affirmative action policy that is minimally responsive, which means that the implementation of a stronger affirmative action policy makes at least one minority student better off in case there is a minority student who is worse off. The DA with a minority-reserve-based affirmative action policy is also not minimally responsive (Hafalir et al., 2013; Doğan, 2016). Similarly, the IA mechanism is not minimally responsive to majority quotas (Afacan & Salman, 2016). Furthermore, both the DA and the IA mechanisms become wasteful with majority quotas; that is, in some cases a school may have unassigned school seats while there is at least one student who prefers that school to her assignment.

We study three axioms in Chapter 2 for an in-depth analysis of the effectiveness of the majority-quota and the minority-reserve-based affirmative action policies across all previously proposed affirmative action mechanisms. Along with minimal responsiveness and non-wastefulness, two important axioms, we introduce the axiom of "respecting the affirmative action policy." A mechanism fails to respect the affirmative action policy if, at any school, the number of assigned minority students is less than the minority reserve at that school (or, alternatively the remaining capacity in excess of the majority quota) while, at the same time, there exists a minority student who prefers that school to her current assignment. We show that all previously studied mechanisms with an affirmative action policy, based either on the DA or the IA, fail to satisfy all three axioms. Therefore, we introduce the Immediate and Deferred Acceptance Mechanism with Minority Reserves (IA-DA-R) which satisfies the three axioms. The IA-DA-R mechanism performs the best among the existing mechanisms based on these three fundamental criteria which aim to improve the welfare of minority students while maintaining a balance with the welfare of majority students. Further properties of IA-DA-R regarding fairness and incentives are also analyzed.

Chapter 3 builds on the previous chapter by introducing more demanding responsiveness axioms than the usually studied minimal responsiveness. The so-called "responsiveness" properties of affirmative action mechanisms are based on a comparison between stronger and weaker affirmative action policies. An affirmative action policy is deemed to be stronger than another one if at least one school has strictly more minority reserve seats (or a strictly lower majority quota) and no other school has fewer minority reserves (or a higher majority quota) when compared to the other policy. The strongest axiom requires that minority students unambiguously gain when a stronger affirmative action policy replaces a weaker one, or at least none of the minority students are hurt by an increase in affirmative action. We also introduce the concept of "elemental" responsiveness, which refers to comparing some affirmative action to the case of no affirmative action policy. The weakest of the four axioms is minimal elemental responsiveness, which only requires that at least one minority student's welfare increases weakly when no affirmative action is compared with having an affirmative action policy. In this chapter we introduce two new mechanisms, the Divided DA and the Guaranteed DA mechanisms, which are the first ones in the literature that satisfy more demanding responsiveness criteria than minimal responsiveness. We show that while the Divided DA mechanism is responsive but wasteful, the Guaranteed DA mechanism is only minimally responsive and elementally responsive but non-wasteful. We study other properties as well of these two new matching mechanisms which suggest trade-offs.

Chapter 4 investigates the impact of the school choice mechanism on school segregation. In recent years three observed trends reveal a paradoxical picture. The IA mechanism is being replaced by the DA mechanism gradually all over the world, such as in New York City in 2003 or in Boston in 2005. The adoption of the DA mechanism was expected to reduce school segregation, as many scholars argued that matching under the IA is segregated, while the DA is desegregated because of its better incentive properties (Pathak & Sönmez, 2008; Basteck & Mantovani, 2018; Calsamiglia, Martínez-Mora, & Miralles, 2021, 2017). The second observation is that recently specific effort has been made to promote minority participation to reduce school segregation. Traditionally, in the USA students were only allowed to apply to schools in their residential district, known as the neighborhood school programs. Given that residential districts remain segregated by race, as they were in the past (Logan & Parman, 2017), the district schools are also segregated. Over the last few decades many programs have been introduced, collectively known as school choice programs, allowing students to apply to schools outside their districts. The third observation is that there is no improvement in school segregation; rather, there is a trend towards more segregation in many places. In the USA the number of different policies aiming to promote racial integration did just the opposite (Roda & Wells, 2013). In recent years, reports show that introducing various school choice programs to mitigate the segregation problem leads to even more segregation by race in public schools in New York City (e.g., Mader, Hemphill, and Abbas (2018)). This failure to desegregate is not limited to the USA, as different recent studies show that the existing level of segregation is either increasing or at least shows no significant improvement despite policy changes all over the world, such as in Chile (Valenzuela, Bellei, & Ríos, 2014; Kutscher, Nath, & Urzua, 2020), England (Terrier, Pathak, & Ren, 2021), and Madrid (Gortázar, Mayor, & Montalbán, 2020), among others.

Compared to numerous empirical studies reporting on issues of school segregation, there are few theoretical studies, and no theoretical models in the matching and market design literature that could explain the observed trends discussed above. In Chapter 4 we present a stylized model that provides a theoretical explanation of school segregation observed in the last few decades. We show that the widely used DA mechanism (and also the much-studied but practically less relevant TTC mechanism) leads to segregation under two realistic assumptions; in contrast, the IA mechanism is less segregated.

In the last concluding chapter of the thesis we summarize our findings and provide a comparative picture. We argue that our mechanisms developed in Chapter 2 and Chapter 3 can provide useful policy alternatives to promote integration in schools and other institutions, and that these or similar affirmative action mechanisms and policies could play a vital role in improving the welfare of underprivileged minority communities by providing better access to better institutions.

Chapter 2

Affirmative Action Policies in School Choice: Immediate versus Deferred Acceptance

2.1 Introduction

Affirmative action policy in university admissions, school assignments, in labor markets and within organizations is a subject of continuous discourse. On the one hand, such policies involving preferential selection based on race, ethnicity or gender have generated philosophical debates. On the other hand, the ineffectiveness of preferential treatment policies in matching mechanisms has been explored over the last decade or so. This paper focuses on the latter issue from a theoretical perspective. We analyze two types of affirmative action policies in a many-to-one matching model (the theoretical model known as school choice), namely quota-based and reserve-based policies. The former has been widely used in school assignment to promote diversity and increase the participation of students from underrepresented groups. In our model we simply refer to such students as minority students, and the rest of the students are called majority students.

Gale and Shapley (1962) introduced the celebrated Deferred Acceptance (DA) mechanism which always results in a fair (or stable) matching (Gale & Shapley, 1962) which Pareto dominates all other fair matchings (Gale & Shapley, 1962; Balinski & Sönmez, 1999), and it is strategyproof (Dubins & Freedman, 1981; Roth, 1982). In the DA procedure students apply to their most preferred school first and are either temporarily accepted or rejected based on the school's priorities, and in subsequent steps rejected students apply to their next most preferred school. A temporary acceptance may turn into a rejection since a higher-priority student may apply to a school in a later step. One affirmative action policy incorporated by the DA mechanism is based on type-specific quotas at schools. Abdulkadiroğlu and Sönmez (2003) discussed such type-specific quotas using the DA mechanism, where each school has a fixed quota for students of each type, and the quota cannot be exceeded by the number of accepted students of each specific type. If there is a quota for only one type, the *majority quota* at each school, then it simply limits the maximum number of majority students that can be assigned to a school (Kojima, 2012). This is what we call the DA with Majority Quotas mechanism (DA-Q), which follows the same iterative application procedure as the standard DA, but in the DA-Q the total number of majority students is not allowed to exceed the limit set by the majority quota at each school.

Majority quotas are inefficient when there are not enough minority student applicants to fill the set-aside school seats, since these seats cannot be given to majority applicants and make the affirmative action policy inefficient and even more controversial. To remedy this problem, Hafalir et al. (2013) introduced a more flexible quota called the *minority reserve*. Schools give priority to all minority students over majority students when assigning minority reserve seats, but these seats can also be assigned to majority students in case there are not enough minority students to fill the reserved seats, which is not allowed by quota systems. In the DA with Minority Reserve (DA-R) mechanism of Hafalir et al. (2013), schools first temporarily assign minority reserve seats to minority applicants before filling all the remaining seats from the general applicant pool.[1]

The DA-R mechanism is considerably more efficient than the DA-Q mechanism, as the minority reserves successfully eliminate the wastefulness of the quota-based policy. However, another issue of effectiveness, namely, the responsiveness of the affirmative action policy in terms of welfare improvement of the minority students when such a policy is introduced or strengthened, as proposed by Kojima (2012), is not resolved by the DA-R mechanism, which remains unresponsive, similarly to the DA-Q mechanism. A mechanism is called minimally responsive to some affirmative action policy (what Kojima (2012) called 'respecting the spirit of affirmative action') if a stronger affirmative action policy - increased minority reserves or decreased majority quotas - does not lead to a Pareto-inferior outcome for minority students, that is, it benefits at least one minority student if the policy change has any impact. Kojima (2012) demonstrated that at certain preference and priority profiles lower majority quotas do not help any minority student and may even harm some, and thus the DA-Q mechanism is not minimally responsive. The same is true for the DA-R

[1] Another noteworthy contribution is Ehlers, Hafalir, Yenmez, and Yildirim (2014) that study both upper and lower bounds in quota-based affirmative action policies with multiple types of students.

mechanism, as it turns out.

Addressing this issue, Doğan (2016) introduced the Modified DA with Minority Reserves (MDA) mechanism, which is minimally responsive. The MDA mechanism achieves minimal responsiveness by treating specific minority students as majority students to benefit some minority students, resembling the efficiency improvements of the EADAM mechanism (Kesten, 2010). Another mechanisms which uses efficiency improvements was proposed by Ju, Lin, and Wang (2018), called the Efficiency Improved DA with Minority Reserves (EIDA) mechanism.

The Immediate Acceptance (IA) mechanism, originally known as the Boston mechanism, is more frequently used historically in the US and around the world than the DA to match students to schools. In the IA mechanism students are accepted permanently in each step of the iterative mechanism, as opposed to the temporary acceptances of the DA. The IA mechanism is neither strategyproof nor fair (Abdulkadiroğlu & Sönmez, 2003) which mean, respectively, that it is manipulable by students and students' priorities may be violated at schools. The IA mechanism has superior welfare properties to the DA in principle, but due to its manipulability the actual efficiency of the mechanisms is uncertain. Ergin and Sönmez (2006) carried out an equilibrium analysis exploring the manipulability of the IA mechanism.

The IA mechanism is still a popular student placement mechanism in school choice programs around the world (Kojima & Ünver, 2014), and only in the last two decades did school choice programs start opting for the DA instead. Affirmative action policies have also been implemented in conjunction with immediate acceptances; for example, the IA mechanism was in use in Boston with racial quotas prior to 1999 (Abdulkadiroğlu & Sönmez, 2003). While the effects of affirmative action policies under the DA mechanism are well studied, no attention was given to affirmative action policies in the IA mechanism until recently. Afacan and Salman (2016) studied a hybrid immediate acceptance mechanism, the IA with Minority Reserves and Majority Quotas (IA-RQ, for short) which allows for both minority reserves and majority quotas. The only other papers that consider affirmative action policies under the IA mechanism are Chen, Huang, Jiao, and Zhao (2022), and to a lesser extent Doğan and Klaus (2018). The IA mechanisms with affirmative action are interesting, as they have distinct properties, both good and bad, compared with their DA counterparts. Most notably, the IA-R mechanism (allowing for just minority reserves) is minimally responsive, unlike DA-R, as shown by Afacan and Salman (2016), but it lacks some other attributes that the IA-Q mechanism (allowing for just majority quotas) possesses.

In this paper we focus on three basic welfare axioms for affirmative action policies, namely, *non-wastefulness*, *respecting the affirmative action policy*, and *minimal responsiveness* of the affirmative action policy. Non-wastefulness is a weak efficiency property: it eliminates empty school seats that are desired by a student, regardless of whether the student is a minority or majority student. Respecting the affirmative action policy is fundamental for any mechanism with affirmative action and requires that school seats set aside for minority students should be filled with minority students, as long as there are any minority students who desire them. Finally, the minimal responsiveness axiom of Kojima (2012) requires that a stronger affirmative action policy should benefit at least one minority student, whenever some minority students are affected by the change.

We consider these three axioms basic welfare criteria for a mechanism with an affirmative action policy; nonetheless, quite surprisingly, none of the above described six affirmative action mechanisms that have been proposed in the literature satisfy all three of them.[2] Given some well-known impossibilities in the matching theory literature, one might suspect that we cannot satisfy all three axioms simultaneously. However, this is not the case. After analyzing the six aforementioned mechanisms, we propose a new mechanism with an affirmative action policy which satisfies the three welfare axioms. This mechanism, which we call Immediate and Deferred Acceptance Mechanism with Minority Reserves (IA-DA-R, for short), combines both immediate and deferred acceptances. Namely, minority reserve positions are filled with minority applicants based on immediate acceptances, while all other acceptances are temporary, as in the DA. We also study the fairness (stability) properties of the proposed IA-DA-R mechanism, and analyze its incentive properties.

2.2 Model and Definitions

There is a finite set of students S which is divided into the set of **majority students** S^M and the set of **minority students** S^m; $S^M \bigcap S^m = \emptyset$ and $S^M \bigcup S^m = S$. Assume that there are at least two students of each type: $|S^M| \geqslant 2$ and $S^m| \geqslant 2$. We denote majority students by $a \in S^M$ and minority students by $i \in S^m$. There is also a finite set of schools C such that $|C| \geqslant 3$. Each school $c \in C$ has a capacity $q_c \geqslant 1$, which is the maximum number of students that can be assigned to school c.

Each student $s \in S$ has a strict preference ordering P_s over $C \cup \{0\}$, where 0 denotes the "null school" and represents staying unmatched. If a student ranks a school below 0 then this school is considered unacceptable to the student. Let R_S denote the weak counterpart

[2]TTC-based affirmative action mechanisms have been the least studied in the literature (only Kojima (2012) studies the quota-based TTC) and we omitted them from this analysis. Nevertheless, the TTC-based mechanisms also suffer from similar problems as the other proposed mechanisms, and don't satisfy all three welfare axioms simultaneously.

of P_s, so that $c\ R_s\ c'$ if and only if either $c\ P_s\ c'$ or $c = c'$. Let a preference profile be denoted by $P \equiv (P_s)_{s\in S}$, and let $\mathcal{P}$ be the set of all preference profiles. Each school $c \in C$ has a strict priority ordering $\succ_c$ over S. Let a priority profile be denoted by $\succ \equiv (\succ_c)_{c\in C}$, and let Π be the set of all priority profiles. When S, C and $(q_c)_{c\in C}$ are fixed, a school assignment problem is given simply by a **profile** $(P, \succ) \in \mathcal{P} \times \Pi$ which consists of a preference profile and a priority profile.

A **matching** μ is a function from the set of students to $C \cup \{0\}$ such that at most q_c students are assigned to each school $c \in C$, while the "capacity" of the null school is unlimited. To simplify notation, for all $s \in S$ we denote the assignment that student s receives by μ_s, where $\mu_s \in C \cup \{0\}$. Moreover, we denote the set of students assigned to school c by μ_c, for all $c \in C$. Hence, $\mu_c \subseteq S$ and $|\mu_c| \leqslant q_c$. We will also use the notation μ_c^m to denote the set of minority students matched to c by μ. Let $\mathcal{M}$ denote the set of matchings when $S,\ C$ and $(q_c)_{c\in C}$ are fixed. A **mechanism** φ (without affirmative action) assigns a matching μ to each profile $(P, \succ) \in \mathcal{P} \times \Pi$, that is, $\varphi : \mathcal{P} \times \Pi \to \mathcal{M}$. Later on we will expand the domain of a mechanism to include the affirmative action policy, and all the properties of mechanisms defined below extend to mechanisms with an affirmative action policy naturally. We will denote the assignment of student s at $(P, \succ)$ by $\varphi_s(P, \succ)$, and the assignments of a set of students $S' \subset S$ at $(P, \succ)$ by $\varphi_{S'}(P, \succ)$. For simplicity we also use that notation P_{-s} to denote $P_{N\setminus\{s\}}$.

A basic property of a mechanism is **individual rationality,** which requires the mechanism to assign a matching to each profile such that no student prefers the null school to their assignment, that is, students are not assigned to a school that is unacceptable to them. All the mechanisms that we study in this paper satisfy individual rationality.

A matching μ Pareto-dominates another matching ν at profile $(P, \succ)$ if for all students $s \in S,\ \mu_s R_s \nu_s$, and there exists student $s' \in S$ such that $\mu_{s'} P_{s'} \nu_{s'}$. A matching is Pareto-efficient at $(P, \succ)$ if it is not Pareto-dominated at $(P, \succ)$. A mechanism is **Pareto-efficient** if it assigns to each profile a matching that is Pareto-efficient at that profile.

A mechanism φ is **strategyproof** if for all students $s \in S$, for all profiles $(P, \succ)$, and all alternative preferences P'_s for student s, $\ f_s(P, \succ)\ R_s\ f_s((P'_s, P_{-s}), \succ)$. Otherwise, if a student can report preferences P'_s such that $f_s((P'_s, P_{-s}), \succ)\ P_s\ f_s(P, \succ)$, then we will say that students s can **manipulate** φ at $(P, \succ)$. A mechanism φ is **strategyproof for a set of students** $T \subseteq S$, if for all $s \in T$, student s cannot manipulate φ at any profile $(P, \succ)$.

2.3 Affirmative Action Policies

We study two types of affirmative action policies, *majority quota* policies and *minority reserve* policies. We also introduce a common framework for these two types of affirmative action policies which we refer to as *minority allotment* policies.

2.3.1 Majority Quotas

A majority quota policy determines the maximum number of majority students that can be assigned to each school. Majority quotas were introduced by Abdulkadiroğlu and Sönmez (2003) and were later analyzed by Kojima (2012). A less than desirable feature of majority quotas is that even if the capacity of a school is not filled, the number of majority students assigned to a school cannot exceed the majority quota of the school.

Let a majority quota policy be denoted by $q^M \equiv (q_c^M)_{c \in C}$, where q_c^M is the majority quota at school c. In order to be feasible, q_c^M is required to satisfy $0 \leqslant q_c^M \leqslant q_c$ for all $c \in C$. Note that $q^M = q$ is the case of no affirmative action policy. Let $\mathcal{Q}$ denote the set of feasible majority quota policies. A **mechanism with majority quotas** is defined as $\varphi^q : \mathcal{Q} \times \mathcal{P} \times \Pi \to \mathcal{M}$, which assigns a matching μ to each majority quota policy $q^M \in \mathcal{Q}$ and each profile $(P, \succ) \in \mathcal{P} \times \Pi$.

2.3.2 Minority Reserves

A minority reserve policy reserves a number of seats for minority students at each school such that minority students are to be prioritized for the reserved seats over majority students. Minority reserve policies were proposed by Hafalir et al. (2013) in order to eliminate the wastefulness of majority quota policies. Accordingly, a crucial feature of minority reserve policies is that if there are not enough minority applicants to fill all minority reserve seats with minority students then majority students may also be assigned to reserved seats.

Let a minority reserve policy be denoted by $r \equiv (r_c)_{c \in C}$, where r_c is the number of reserved seats at school c. In order to be feasible, r_c is required to satisfy $0 \leqslant r_c \leqslant q_c$ for all $c \in C$. Note that $r = \mathbf{0}$ is the case of no affirmative action policy, where $r_c = 0$ for all $c \in C$. Let $\mathcal{R}$ denote the set of feasible minority reserve policies. A **mechanism with minority reserves** is defined as $\varphi^r : \mathcal{R} \times \mathcal{P} \times \Pi \to \mathcal{M}$, which assigns a matching μ to each minority reserve policy $r \in \mathcal{R}$ and each profile $(P, \succ) \in \mathcal{P} \times \Pi$.

2.3.3 Common Framework: Minority Allotments

The two affirmative action policies, the majority quota and minority reserve policies, while different, both aim to improve the representation and welfare of minority students by prioritizing them over majority students for a specified number of school seats. Now we unify our reference to the two policies, in order to be able to state axioms that pertain to both affirmative action policies and thus be able to evaluate them in a common framework. We capture the unified affirmative action policy with the **minority allotments**, which refer to either minority reserves or to the number of seats that remain in excess of the majority quotas, which are the seats that can only be assigned to minority students under a majority quota policy.

Let a minority allotment policy be denoted by $v \equiv (v_c)_{c \in C}$, where v_c is the number of minority allotment seats at school c. If the minority allotment policy describes a minority reserve policy then $v_c = r_c$ for each school c, that is, $v = r$, and if the the minority allotment profile describes a a majority quota policy then $v_c = q_c - q_c^M$ for each school c, that is, $v = q - q^M$. Feasibility requires in both cases that for all $c \in C$, v_c satisfies $0 \leqslant v_c \leqslant q_c$. Note that $v = \mathbf{0}$ is the case of no affirmative action policy. Let $\mathcal{V}$ denote the set of feasible minority allotment policies. A **mechanism with minority allotments** is defined as $\varphi^v : \mathcal{V} \times \mathcal{P} \times \Pi \to \mathcal{M}$, which assigns a matching μ to each minority allotment policy $v \in \mathcal{V}$ and each profile $(P, \succ) \in \mathcal{P} \times \Pi$.

2.4 Main Welfare Axioms

Now we define and discuss the three key welfare axioms that we use to assess the performance of school choice mechanisms with an affirmative action policy. We consider these minimal requirements when evaluating affirmative action policies. All three of the axioms are welfare requirements: the first one, non-wastefulness, is a general requirement, while the next two consider welfare properties of the affirmative action policy.

2.4.1 Non-Wastefulness

Non-wastefulness is an essential efficiency requirement for mechanisms. A mechanism is considered to be wasteful if there is an unassigned school seat which is preferred by at least one student to her assignment at some profile. Wastefulness is a serious drawback for any mechanism, as it implies that the mechanism is not only Pareto-dominated but it also leads to wasting valuable school seats.

Non-Wastefulness. A mechanism φ^v is non-wasteful if for all $v \in \mathcal{V}$, $(P, \succ) \in \mathcal{P} \times \Pi$, $s \in S$ and $c \in C$, if $c \; P_s \; \varphi^v_s(v, P, \succ)$ then $|\mu_c| = q_c$, where $\varphi^v\,(v, P, \succ) = \mu$.

Although non-wastefulness is independent of the affirmative action policy v, we consider it as one of our main requirements because a non-wasteful mechanisms may become wasteful when it incorporates certain affirmative action policies. Non-wastefulness is a minimal efficiency requirement and it is typically satisfied by any interesting mechanism in standard models without affirmative action or distributional constraints.[3] The reason we focus on non-wastefulness here is that quota-based mechanisms with affirmative action are wasteful, and thus the axiom of non-wastefulness is an appropriate property for evaluating mechanisms in this context, as it distinguishes quota-based affirmative action policies, which do not satisfy it, from reserve-based affirmative action policies, which do. This is the main insight of Hafalir et al. (2013) who first propose a reserve-based affirmative action policy using the DA mechanism to eliminate wasteful seats. The axiom of non-wastefulness is typically a part of standard fairness conditions (see, for example, Hafalir et al. (2013) and Doğan (2016)) and it also follows from Pareto-efficiency, which is not satisfied by most of the mechanisms, and indeed it is not satisfied even by the standard DA without affirmative action when only the students' welfare is considered.

2.4.2 Respecting the Affirmative Action Policy

We will say that a matching mechanism with a minority allotment policy v *respects the affirmative action policy* if there is no minority student who prefers a school c to his assignment at any profile such that school c has fewer than v_c minority students assigned to it.

Respecting the Affirmative Action Policy. A mechanism φ^v respects the affirmative action policy if for all $v \in \mathcal{V}$, $(P, \succ) \in \mathcal{P} \times \Pi$, $i \in S^m$ and $c \in C$, if $c \; P_i \; \mu_i$ then $|\mu^m_c| \geqslant v_c$, where $\varphi^v\,(v, P, \succ) = \mu$.

This is an intuitive axiom which simply requires that minority students should indeed be prioritized for the number of school seats specified by the minority allotment, and thus it is a basic requirement for any mechanism with an affirmative action policy that relies on minority allotments. It is not a new requirement, as some version of this is typically included in fairness (stability) axioms for affirmative action. We propose it separately as one

[3] Given any type-specific quota that sets a hard bound and is enforced independently of student preferences, such as an upper or lower limit on the types of students that are assigned to a school, there may not even exist a matching that is both non-wasteful and eliminates all priority violations (Ehlers et al., 2014).

of our main axioms because, somewhat surprisingly, not all affirmative action mechanisms studied in the literature satisfy this basic intuitive property.

The axioms of non-wastefulness and respecting the affirmative action policy are related: we might say that respecting the affirmative action policy means that no reserved seats are wasted when it comes to minority students. But the two concepts are logically independent. Consider a mechanism and a profile at which some student s would prefer a seat at school c compared to her assignment at that profile. If c has an empty seat and the minority allotment v_c of c is not filled with minority students then the mechanism is wasteful, but if s is a majority student it may respect the affirmative action policy; however, if s is a minority student then the mechanism is both wasteful and does not respect the affirmative action policy. On the other hand, if the capacity at school c is exhausted and more than $q_c - v_c$ seats are allocated to majority students, then if s is a minority student the mechanism does not respect the affirmative action policy, but this does not imply that it is wasteful.

2.4.3 Minimal Responsiveness

Minimal responsiveness to the affirmative action policy requires that at least one minority student gains if the affirmative action policy is strengthened, assuming that the policy change affects the outcome for any minority student. This axiom was first proposed for quota-based affirmative action by Kojima (2012), which he called 'respecting the spirit of quota-based affirmative action'. After Hafalir et al. (2013) introduced the concept of the minority reserve policy, Kojima's axiom was extended to reserved-based affirmative action by Doğan (2016).

In the formal definition below, $v' \geqslant v$ means that for all $c \in C$, $v'_c \geqslant v_c$ and thus v' represents a weakly stronger affirmative action policy than v.

Minimal Responsiveness. A mechanism φ^v is minimally responsive if for all $v, v' \in \mathcal{V}$ such that $v' \geqslant v$ and all $(P, \succ) \in \mathcal{P} \times \Pi$ such that $\varphi^v_{S^m}(v, P, \succ) \neq \varphi^v_{S^m}(v', P, \succ)$ there exists $i \in S^m$ such that $\varphi^v_i(v', P, \succ) \; P_i \; \varphi^v_i(v, P, \succ)$.

Minimal responsiveness stipulates that a weakly stronger minority allotment policy does not result in a Pareto-dominated outcome for minority students. This appears to be a mild and intuitive requirement for any type of affirmative action policy, yet it is not necessarily easy to satisfy.

2.5 Affirmative Action Polices and their Compliance with the Main Welfare Axioms

In this section we evaluate the performance of the different mechanisms with affirmative action policies that have been studied in the literature based on the three main axioms of non-wastefulness, respecting the affirmation action policy, and minimal responsiveness. We start with DA mechanisms, then evaluate DA mechanisms with efficiency improvements, and finally study IA mechanisms, of which we know the least.

2.5.1 DA Mechanisms with Affirmative Action

We provide first a formal definition of the DA with Majority Quotas (DA-Q) mechanism studied by Kojima (2012). The DA-Q mechanism is based on the split school model of Abdulkadiroğlu and Sönmez (2003), where the set of students is partitioned according to their types and each school has a quota for students of each type. The majority quota-based affirmative action policy for the DA mechanism is an adaptation of these mechanisms to the case where only the set of majority students has a type-specific quota, while minority students don't have a cap. The mechanism is iterative and we specify the first step and a general step t.

DA with Majority Quotas

Fix a majority quota policy q^M and a profile $(P, \succ)$.

Step 1: Every student applies to her most preferred school according to P. Each school c tentatively assigns seats to applying students up to its capacity q_c, following its priority ordering $\succ_c$. However, each school c rejects majority students when its majority quota q_c^M is reached by accepted majority students.

Step *t* $(t \geqslant 2)$**:** Every student who was rejected in step $t-1$ applies to her next most preferred acceptable school according to P. Each school c considers its tentatively assigned students from the previous step along with the new applicants and tentatively assigns seats to these students up to its capacity q_c, following its priority ordering $\succ_c$. However, each school c rejects majority students when its majority quota q_c^M is reached by accepted majority students.

The algorithm terminates when there is no more rejection by any school and all tentative matches in the final step become final matches which together constitute the matching assigned to $(P, \succ)$ when the majority quota policy is q^M. ◇

Next, we introduce the DA with Minority Reserves (DA-R) which was proposed by Hafalir et al. (2013) to enhance the efficiency of the DA-Q mechanism, as the latter is wasteful at some profiles. It should be noted that both DA-Q and DA-R simplify to the classic Deferred Acceptance algorithm (Gale & Shapley, 1962) when $q^M = q$ and $r = \mathbf{0}$, respectively, that is, without an affirmative action policy.

DA with Minority Reserves

Fix a minority reserve policy r and a profile $(P, \succ)$.

Step 1: Every student applies to her most preferred school according to P. Each school c first tentatively accepts as many as r_c minority students following its priority ordering $\succ_c$. Then each school c tentatively accepts students among the remaining applicants following its priority ordering $\succ_c$ until either its capacity q_c is filled or there are no more applicants. Any remaining applicants are rejected.

Step *t* $(t \geqslant 2)$**:** Every student who was rejected in step $t-1$ applies to her next most preferred acceptable school according to P. Each school c considers its tentatively assigned students from the previous step along with the new applicants (henceforth the *applicant set*) and first tentatively accepts as many as r_c minority students following its priority ordering $\succ_c$. Then each school c tentatively accepts students among the remaining students in the applicant set following its priority ordering $\succ_c$ until either its capacity q_c is filled or the applicant set is exhausted. Any remaining applicants are rejected.

The mechanism terminates when there is no more rejection by any school and all tentative matches in the final step become final matches which together constitute the matching assigned to $(P, \succ)$ when the minority reserve policy is r. $\diamond$

The DA-Q mechanism is wasteful, since it puts an upper limit on the majority student admissions and does not allow majority students to occupy additional seats even if some of the remaining seats are not claimed by minority students. This was the main motivation for Hafalir et al. (2013) to introduce the more flexible reserve-based DA-R mechanism which allows majority students to occupy seats that would otherwise be left empty and is therefore non-wasteful. However, the DA-Q mechanism respects the affirmative action policy, since majority students cannot be accepted by any school c in excess of q_c^M, while minority student applicants are accepted for the remaining $q_c - q_c^M$ seats. With minority reserves, if a minority student applies to a school in a step of the procedure where all the reserved seats are already filled, this minority student will be considered for a reserved seat even if some majority students have been assigned reserved seats temporarily, as no assignment is permanent before the mechanism terminates. Therefore, the DA-R mechanism also respects the affirmative action policy, and the stability condition specified by Hafalir et al. (2013) which is satisfied by the DA-R mechanism also implies this. However, neither of the two mechanisms satisfy minimal responsiveness. This was shown for the DA-Q mechanism by Kojima (2012), and pointed out for the DA-R mechanism by Hafalir et al. (2013) and Doğan (2016). We summarize these findings below.

Proposition 1. *The DA-Q mechanism respects the affirmative action policy, but it is wasteful and not minimally responsive.*

Proposition 2. *The DA-R mechanism is non-wasteful and respects the affirmative action policy, but it is not minimally responsive.*

Example 1. This example shows the intuition for why the DA-Q and DA-R mechanisms are not minimally responsive. Let $S^M = \{a_1, a_2\}$ and $S^m = \{i_1, i_2\}$ be the sets of majority and minority students. Let $C = \{c_1, c_2, c_3, c_4\}$ with capacities $q = (1, 1, 1, 1)$. Consider the profile in Table 2.1.

Table 2.1: Profile for Example 1

P_{a_1}	P_{a_2}	P_{i_1}	P_{i_2}	$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$	$\succ_{c_4}$
$\underline{c_2}$	$\underline{c_1}$	$\underline{c_3}$	c_1	a_2	a_2	a_1	a_1
$\boxed{c_3}$	$\boxed{c_2}$	$\boxed{c_1}$	c_2	i_1	a_1	i_1	a_2
c_1	0	c_2	$\underline{\boxed{c_4}}$	i_2	i_2	a_2	i_1
0		c_4	0	a_1	i_1	i_2	i_2

Let $r = (0, 0, 0, 0)$ initially, which means no affirmative action policy. At this profile the DA-R matching μ, which coincides with the DA matching in this case, is given by $\mu_{c_1} = a_2$, $\mu_{c_2} = a_1$, $\mu_{c_3} = i_1$, and $\mu_{c_4} = i_2$, as indicated by the underlined assignments in Table 2.1. Now consider the (stronger) minority reserve policy $\tilde{v} = \tilde{r} = (1, 0, 0, 0)$. The steps of the DA-R procedure are displayed in Table 2.2. The resulting DA-R matching $\tilde{\mu}$ at this profile is given by $\tilde{\mu}_{c_1} = i_1$, $\tilde{\mu}_{c_2} = a_2$, $\tilde{\mu}_{c_3} = a_1$, and $\tilde{\mu}_{c_4} = i_2$, as seen in the last step (Step 6) in Table 2.2, and as indicated by the assignments in squares in Table 2.1.

One of the minority students (i_1) is worse off when the minority reserve policy $\tilde{r}$ is implemented, and the other minority student (i_2) is indifferent. This is because in Step 1 minority student i_2 is accepted (instead of majority student a_2 when there is no affirmative action policy), and this starts a rejection chain in the DA-R procedure which leads to a Pareto-dominated outcome for minority students (and, in fact, for all students) when the affirmative action policy is implemented.

Table 2.2: Steps of the DA-Q/DA-R mechanism with $\tilde{v}$ in Example 1

	c_1	c_2	c_3	c_4
Step 1	$\cancel{a_2}$, i_2	a_1	i_1	
Step 2	i_2	$\cancel{a_1}$, a_2	i_1	
Step 3	i_2	a_2	a_1, $\cancel{i_1}$	
Step 4	i_1, $\cancel{i_2}$	a_2	a_1	
Step 5	i_1	a_2, $\cancel{i_2}$	a_1	
Step 6	i_1	a_2	a_1	i_2

Struck out students (e.g., $\cancel{a_2}$) are rejected at that step.

This example demonstrates not only that the DA-R mechanism is not minimally responsive, but also that the DA-Q mechanism is not minimally responsive either, since at the specified profile the DA-Q mechanism with $\tilde{v} = (1, 0, 0, 0)$, which corresponds to $\tilde{q}^M = (0, 1, 1, 1)$, leads to the exact same steps, and thus to the same matching, as the DA-R mechanism. $\diamond$

2.5.2 Efficiency Improved DA Mechanisms with Affirmative Action

Since the DA-R mechanism is not minimally responsive, Doğan (2016) proposes the Modified DA with Minority Reserves (MDA) mechanism which is minimally responsive. The MDA mechanism modifies the DA-R mechanism and it is somewhat similar to the EADAM (Kesten, 2010), as it is based on the concept of interferers that are similar to Kesten's interrupters. An interferer is defined to be a minority student who causes the school to reject a majority student in some step of the DA-R algorithm due to minority reserves, but in a later step this minority students is also rejected by this school. A main difference is that, instead of aiming for general efficiency improvements primarily, MDA is designed to be minimally responsive to the minority reserve policy, which is achieved by treating the relevant minority student interferers as majority students.

Modified DA with Minority Reserves:

Fix a minority reserve policy r and a profile $(P, \succ)$. An *interferer* is a minority student $i \in S^m$ at some school $c \in C$ with $r_c > 0$ if there is a majority student $a \in S^M$ with $a \succ_c i$ who is rejected by c while i is accepted due to minority reserves in some step of the DA-R mechanism at $(P, \succ)$, but then in a later step i gets rejected by school c. The MDA mechanism involves iterative rounds of the DA-R mechanism. If there are no interferers at any school in the first round, then the MDA mechanism assigns the DA-R matching to $(P, \succ)$ and the algorithm ends in one round. If there are interferers in some steps in any round of the MDA mechanism, the MDA mechanism moves to the next round. The next round runs the DA-R mechanism again, considering all interferers of the latest step of the previous round at which there are interferers, and treats each such interferer at each school c as majority students at c. These iterative rounds continue until there is no interferer at any school, which is the final round. The matching selected by the MDA mechanism at $(P, \succ)$ and minority reserve policy r is the DA-R matching of the final round of the procedure. $\diamond$

Ju et al. (2018) introduce another mechanisms which uses efficiency improvements following Kesten (2010), the Efficiency Improved DA with Minority Reserves (EIDA) mechanism. This mechanism is also an efficiency improvement mechanism over the DA-R, just like the MDA mechanism, but it is based on a straightforward EADAM-style efficiency improvement over the DA-R mechanism (while EADAM is an efficiency improvement over the DA mechanisms). Ju et al. (2018) utilize the simplified EADAM definition introduced by Tang and Yu (2014) in the description of their new mechanism.

Efficiency Improved DA with Minority Reserves

Fix a minority reserve policy r and a profile $(P, \succ)$. A school c is *under-demanded* at the profile $(P, \succ)$ if all students weakly prefer their DA-R assignment to c at $(P, \succ)$. EIDA runs iterative rounds of the DA-R mechanism such that in each round all under-demanded schools along with students assigned to these schools, as well as all unmatched students, are removed. The algorithm terminates when all students are removed. The matching selected by the EIDA mechanism at $(P, \succ)$ and minority reserve policy r is given by the assignments with which students are removed. $\diamond$

Both MDA and EIDA are non-wasteful, in fact EIDA is Pareto-efficient, similarly to EADAM. MDA is minimally responsive, as proved by Doğan (2016), and this attribute of the mechanism was the main motivation for Doğan to propose MDA. EIDA, on the other hand, has been shown by Ding, Hong, Jiao, and Luo (2019) to fail minimal responsiveness. Finally, neither of the two mechanisms satisfy the axiom of respecting the affirmative action policy which may be surprising, but we can see this from Example 8 below. Intuitively, the affirmative action policy is not respected because minority students may be removed from initially assigned reserved positions in the efficiency improvement steps.

Example 2. This example demonstrates that MDA and EIDA do not respect the affirmative action policy. Let $S^M = \{a\}$ and $S^m = \{i_1, i_2\}$ be the sets of majority and minority students. Let $C = \{c_1, c_2, c_3\}$ with capacity $q = (1, 1, 1)$ and $r = (1, 0, 0)$. Consider the profile in Table 2.3.

Table 2.3: Profile for Example 8

P_a	P_{i_1}	P_{i_2}	$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$
$\boxed{c_1}$	$\boxed{c_3}$	c_1	a	a	a
$\underline{c_3}$	$\underline{c_1}$	$\underline{\boxed{c_2}}$	i_1	i_2	i_1
c_2	0	c_3	i_2	i_1	i_2

Minority student i_2 is an interferer for school c_1 in the DA-R matching at this profile (underlined in Table 2.3), and a second round of the DA-R algorithm with the modification that i_2 is considered a majority student at school c_1 yields the matching $\mu_a = c_1$, $\mu_{i_1} = c_3$, and $\mu_{i_2} = c_2$, where μ is the MDA matching at this profile (indicated by the assignments in squares in the table). Since $c_1 \; P_{i_2} \; c_2$, $r_{c_1} = 1$, and the majority student a is matched to school c_1, MDA does not respect the affirmative action policy.

In this example EIDA yields the same matching at $(P, \succ)$ with $r = (1, 0, 0)$ as MDA, given that c_2 is the only under-demanded school after round 1, and if we remove i_2 with her assignment c_2 then i_1 and a are both assigned their respective first choices in the second round. Therefore, EIDA does not respect the affirmative action policy. $\diamond$

We summarize the properties of MDA and EIDA below.

Proposition 3. *The MDA mechanism is non-wasteful and minimally responsive, but it does not respect the affirmative action policy.*

Proposition 4. *The EIDA mechanism is non-wasteful, but it does not respect the affirmative action policy and it is not minimally responsive.*

2.5.3 IA Mechanisms with Affirmative Action

We first describe the IA mechanism with quota-based affirmative action. Ergin and Sönmez (2006) introduced the IA mechanism with a fixed quota for each type (what they call the Boston mechanisms with type-specific quotas). The majority-quota-based affirmative action policy for the IA mechanism that we define here is an adaptation of this mechanism to the case where only the set of majority students has a type-specific quota, while minority students don't have a cap. Hence, this mechanism can be viewed as the equivalent of the DA-Q mechanism (Kojima, 2012) but with immediate acceptance, and we call it the IA-Q mechanism.

IA with Majority Quotas

Fix a majority quota policy q^M and a profile $(P, \succ)$.

Step 1: Every student applies to her most-preferred school according to P. Each school c permanently assigns seats to applying students up to its capacity q_c, following its priority ordering $\succ_c$. However, each school c rejects majority students when its majority quota q_c^M is reached by the accepted majority students. Acceptances are final, and any remaining applicants are rejected.

Step *t* $(t \geqslant 2)$**:** Every student who was rejected in step $t-1$ applies to her next most preferred acceptable school according to P, the student's t^{th}-ranked school. Each school c permanently assigns seats to students applying in this step, following its priority ordering $\succ_c$. However, each school c rejects majority students when its majority quota q_c^M is reached in total, taking into account accepted majority students in all previous steps and the current step. Acceptances are final, and any remaining applicants are rejected.

The algorithm terminates when each student is either accepted by a school or has been rejected by all of her acceptable schools. The acceptances made in each step are final and together constitute the matching assigned to profile $(P, \succ)$ when the majority quota policy is q^M. $\diamond$

Afacan and Salman (2016) show, using the following example, that the IA-Q mechanism is not minimally responsive.[4]

Example 3. This example demonstrates that the IA-Q mechanism is not minimally responsive.[5] Let $S^M = \{a_1, a_2\}$ and $S^m = \{i\}$ and be the sets of majority and minority students. Let $C = \{c_1, c_2, c_3\}$ with capacities $q = (1, 1, 1)$. Consider the preference and priority profiles in Table 2.4.

Table 2.4: Profile for Example 3

P_{a_1}	P_{a_2}	P_i	$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$
$\underline{c_2}$	$\underline{\boxed{c_1}}$	c_1	a_2	a_1	a_1
$\boxed{c_3}$	0	$\underline{c_3}$	i	i	i
0		$\boxed{0}$	a_1	a_2	a_2

[4] Afacan and Salman (2016) introduce an immediate acceptance mechanism with both reserves and quotas, what we refer to as the IA-RQ mechanism, and Example 3 is about the case where $r = \mathbf{0}$, and thus it pertains to the IA-Q mechanism. We provide some discussions on the IA-RQ mechanism of Afacan and Salman (2016) in the appendix.

[5] Chen et al. (2022) show that the IA-Q mechanism becomes minimally responsive when all schools have higher priority for all minority students than all majority students.

With no affirmative action, that is, $q^M = (1,1,1)$, minority student i is matched to school c_3 by the IA-Q mechanism, the same as the IA assignment, at the specified profile (these assignments are underlined in Table 2.4). With (stronger) affirmative action policy, namely, $\tilde{q}^M = (1,0,1)$, minority student i is unassigned by IA-Q at this profile (see the assignments indicated in squares). Since i is the only minority student, this shows that the IA-Q mechanism is not minimally responsive. $\diamond$

Proposition 5. *The IA-Q mechanism respects the affirmative action policy, but it is wasteful and not minimally responsive.*

Proof.

Wasteful: It is easy to see that the IA-Q mechanism is wasteful, since if a school seat is not claimed by any minority student when the majority quota is already filled by other majority students at this school, then a majority student cannot be assigned to this seat, even if it is preferred by this student to her school assignment.

Respects the affirmative action policy: Using the IA-Q mechanism, each of the $v_c = q_c - q_c^M$ school seats at c that cannot be assigned to majority students need to be assigned to a minority student or remain empty, even if there is a majority applicant for such a seat. Thus, the IA-Q mechanism will never assign a majority student to any seat that exceeds the majority quota, while a minority applicant cannot be rejected from a school with empty seats. Therefore, the IA-Q mechanism respects the affirmative action policy.

Not minimally responsive: See Example 3. $\square$

We describe next the IA mechanism with reserve-based affirmative action, namely the IA with Minority Reserve (IA-R) mechanism, which corresponds to the IA-RQ mechanism of Afacan and Salman (2016) (see the appendix) without any majority quotas, i.e., when $q^M = q$, and this is also the same as the IA Mechanism with Affirmative-Action-Target introduced by Doğan and Klaus (2018).

IA with Minority Reserves

Fix a minority reserve policy r and a profile $(P, \succ)$.

Step 1: Every student applies to her most preferred school according to P. Each school c first permanently assigns seats to applying minority students up to its number of minority reserve seats r_c, following its priority ordering $\succ_c$. Then each school c permanently assigns its remaining seats to the remaining applicants, following its priority ordering $\succ_c$, up to its capacity q_c. Any remaining applicants are rejected.

Step *t* $(t \geqslant 2)$**:** Every student who was rejected in step $t-1$ applies to her next most preferred acceptable school according to P, the student's t^{th}-ranked school. Each school c that has fewer minority students accepted than r_c first permanently assigns seats to minority students applying in this step up to its number of minority reserve seats r_c in total,including minority students accepted in previous steps, following its priority ordering $\succ_c$. Then each school c that still has available seats permanently assigns its remaining applicants, following its priority ordering $\succ_c$, up to its capacity q_c. Any remaining applicants are rejected.

The algorithm terminates when each student is either accepted by a school or has been rejected by all of her acceptable schools. The acceptances made in each step are final and together constitute the matching assigned to profile $(P, \succ)$ when the minority reserve policy is r. ◇

Afacan and Salman (2016) show that IA-RQ mechanisms are minimally responsive to minority reserves when the majority quota profile does not change. This means in our terminology that the IA-R mechanism is minimally responsive. The next example demonstrates that the IA-R mechanism does not respect the affirmative action policy. However, it is a non-wasteful mechanism. These results are summarized in the following proposition.

Example 4. This example demonstrates that the IA-R mechanism does not respect the affirmative action policy. Let $S^M = \{a_1, a_2\}$ and $S^m = \{i\}$ be the sets of majority and minority students. Let $C = \{c_1, c_2, c_3\}$ with capacity $q_c = 1$ for all $c \in C$. Consider the preference and priority profiles in Table 3.2. Let $r = (0, 1, 0)$ be the minority reserve policy.

Table 2.5: Profile for Example 9

P_{a_1}	P_{a_2}	P_i	$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$
$\boxed{c_2}$	$\boxed{c_1}$	c_1	a_2	a_1	a_1
c_3	0	c_2	i	i	i
0		$\boxed{c_3}$	a_1	a_2	a_2

With the specified minority reserve policy r, the IA-R matching at this profile is given by $\mu_{c_1} = a_2$, $\mu_{c_2} = a_1$ and $\mu_{c_3} = i$, as indicated by the assignments in squares in Table 3.2. Given that $c_2 \; P_i \; c_3$, but the minority reserve $r_{c_2} = 1$ is not used by a minority student since a_1 is a majority student, the IA-R mechanism does not respect the affirmative action policy. ◇

Proposition 6. *The IA-R mechanism is non-wasteful and minimally responsive to minority reserves but does not respect the affirmative action policy.*

Proof.

Non-wasteful: Let ψ denote the IA-R mechanism. Fix a minority reserve policy $r \in \mathcal{R}$ and a profile $(P, \succ) \in \mathcal{P} \times \Pi$. Let $\mu = \psi(r, P, \succ)$. Suppose, by contradiction, that μ is wasteful. Then there exist $s \in S$ and $c \in C$ such that $c\ P_s\ \mu_s$ and $|\mu_c| < q_c$. Therefore, student s has applied to school c in some step t of the IA-R procedure at $(P, \succ)$ and was rejected by c in some step $t' \geqslant t$. Given that there is an empty seat at c after the procedure terminated, there must have been at least one empty seat at c in step t, since the IA-R algorithm fills as many (remaining) seats as possible up to its capacity q_c in each step, first accepting minority students for minority reserve seats and then any applicants for the remaining seats. Then, since student s was rejected by c in step t' of the procedure, it follows that school c was filled to capacity in step t' and remained filled until the end of the procedure. Therefore, $|\mu_c| = q_c$, which is a contradiction, implying that the IA-R mechanism is non-wasteful.
Does not respect the affirmative action policy: See Example 9.
Minimally responsive: This is proved by Afacan and Salman (2016). □

The IA-R mechanisms does not respect the affirmative action policy because it allows majority students to fill unoccupied minority reserve seats permanently, due to the immediate acceptances. IA-Q does not suffer from the same problem because its inflexible quota-based affirmative action policy does not allow for assigning majority students to minority allotment seats. The latter, however, leads to wastefulness. Nonetheless, this feature of the IA-Q mechanism makes majority quotas interesting again for immediate acceptance mechanisms, whereas for deferred acceptance mechanisms a reserve-based affirmative action policy is clearly preferable to a quota-based policy.

2.6 Immediate and Deferred Acceptance with Minority Reserves

The DA-R mechanism of Hafalir et al. (2013) is unambiguously preferable to the DA-Q mechanism, since it dispenses with the wastefulness of the DA-Q mechanism while it still respect the affirmative action policy. The same cannot be said about the IA-R and IA-Q mechanisms, since IA-R does not respect the affirmative action policy, and thus the two mechanisms cannot be compared unambiguously based on our three welfare axioms, even though IA-R is minimally responsive (as opposed to DA-R). In fact, none of these four mechanisms satisfy minimal responsiveness, except for IA-R, but IA-R does not respect the affirmative action policy which, one may argue, is a more fundamental requirement for an affirmative action policy than minimal responsiveness. For the same reason MDA, which is minimally responsive, is less appealing because it does not respect the affirmative action policy. It is more appealing than EIDA, however, despite EIDA being Pareto-efficient, since EIDA is neither minimally responsive, nor does it respect the affirmative action policy, which makes EIDA the only mechanism among all the proposed mechanisms that has neither desirable property of an affirmative action policy. In sum, none of the six mechanisms analyzed in Section 2.5 satisfy all three of the main welfare axioms.

In what follows, we introduce a mechanism that satisfies all three of the welfare axioms, proving in a constructive manner that they can be satisfied simultaneously. As we have seen, a major issue is that an immediate acceptance mechanisms with a minority reserve policy is not necessarily effective, since majority students who apply prior to minority students are allowed to occupy minority reserve seats permanently, as required by immediate acceptance. Due to this conflict between the principle of a minority reserve policy (providing flexibility and avoiding wastefulness) and the permanent assignments made by an immediate acceptance mechanism, there is no way to have a mechanism with immediate acceptances only that is both non-wasteful and respects the affirmative action policy. The main idea of minority reserves is to assign majority students to minority reserve seats temporarily, and only if there are not enough minority applicants to fill all reserve seats considering all steps of the procedure will such majority student assignments become final. The adaptation of a minority reserve-based affirmative action policy that ensures both non-wastefulness and respecting the affirmative action policy requires inherently that the set of accepted students can be updated when new minority students apply to a school, replacing tentatively accepted majority students who were previously allowed to temporarily occupy minority reserve seats in the absence of minority applications. This, however, is at odds with minimal responsiveness, unless minority students are not replaced by newly applying minority students, which may lead to rejection cycles that eventually do not benefit any minority students. Therefore, our proposed mechanism, the **Immediate and Deferred Acceptance Mechanism with Minority Reserves (IA-DA-R)** incorporates immediate assignments for minority students who are accepted for minority reserve seats, while all other assignments are temporary assignments, including the temporary assignment of majority students to minority reserve seats that have not been claimed by minority applicants yet.

IA-DA Mechanism with Minority Reserves

Fix a minority reserve policy r and a profile $(P, \succ)$.

Step 1: Every student applies to her most preferred school according to P. Each school c first permanently assigns seats to applying minority students up to its number of minority reserve seats r_c, following its priority ordering $\succ_c$. Then each school c tentatively accepts students among the remaining applicants following its priority ordering $\succ_c$ until either its capacity q_c is filled or the applicant set is exhausted. Any remaining applicants are rejected.

Step *t* $(t \geqslant 2)$**:** Every student who was rejected in step $t-1$ applies to her next most preferred acceptable school according to P. Each school c that has fewer minority students accepted than r_c first permanently assigns seats to minority students applying in this step up to its number of minority reserve seats r_c in total, including minority students accepted in previous steps, following its priority ordering $\succ_c$. Then each school c considers its tentatively assigned students from the previous step along with the new applicants (henceforth the *applicant set*) and tentatively accepts students among the remaining applicants following its priority ordering $\succ_c$ until either its capacity q_c is filled or the applicant set is exhausted. Any remaining applicants are rejected.

The mechanism terminates when there is no more rejection by any school and all tentative matches in the final step become final matches which together with the permanently accepted minority students constitute the matching assigned to $(P, \succ)$ when the minority reserve policy is r. ◇

In the IA-DA-R mechanism minority students are accepted permanently for reserved seats but are only accepted temporarily for unreserved seats. Majority students can only be accepted temporarily either for reserved or unreserved seats, and are only assigned temporarily to reserved seats when there are no minority applicants for the reserved seats. Therefore, this mechanism combines both immediate and deferred acceptances. Namely, reserved seats are treated as immediate acceptance seats for minority students only, while for majority students reserved seats are deferred acceptance seats. Moreover, unreserved seats are deferred acceptance seats for both minority and majority students, where a deferred acceptance seat means that accepted students can be replaced by new applicants with a higher priority, as in the DA. It is easy to see that with no affirmative action the IA-DA-R mechanism yields the same matching as the classic DA mechanism, since in this case all seats are deferred acceptance seats and minority and majority students are treated the same way. Moreover, let us remark that although this is a hybrid mechanism between the IA and DA mechanisms, it is not a member of the class of PRP rules studied by Ayoade and Pápai (2023), since minority applicants are accepted permanently for remaining minority reserve seats regardless of where they rank the school, and thus there is no preference rank partition profile for minority students that can be applied to each profile when considering reserved seats.

Example 5. This example illustrates in detail how the IA-DA-R mechanism works. Let $S^M = \{a_1, \ldots, a_5\}$ and $S^m = \{i_1, \ldots, i_4\}$ be the sets of majority and minority students. Let $C = \{c_1, \ldots, c_4\}$ with capacities $q = (3, 2, 3, 1)$. Consider the preference and priority profiles in Table 2.6.

Table 2.6: Profile for Example 5

P_{a_1}	P_{a_2}	P_{a_3}	P_{a_4}	P_{a_5}	P_{i_1}	P_{i_2}	P_{i_3}	P_{i_4}	$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$	$\succ_{c_4}$
$\underline{\boxed{c_1}}$	$\underline{c_1}$	c_1	$\underline{\boxed{c_4}}$	c_4	c_4	c_4	c_4	$\boxed{c_1}$	a_1	a_5	a_5	a_4
c_2	$\boxed{c_3}$	$\underline{\boxed{c_2}}$	c_3	$\underline{c_1}$	$\underline{c_2}$	c_2	$\underline{\boxed{c_3}}$	$\underline{c_3}$	a_5	a_3	a_2	a_1
c_3	c_2	c_3	c_2	$\boxed{c_2}$	c_1	$\boxed{c_1}$	c_2	c_2	a_2	a_1	i_3	a_2
c_4	c_4	c_4	c_1	c_3	$\boxed{c_3}$	$\underline{c_3}$	c_1	c_4	a_3	i_3	a_1	a_3
									i_1	i_1	a_3	i_1
									i_2	i_2	a_4	i_2
									a_4	a_4	i_1	i_3
									i_3	a_2	i_2	a_5
									i_4	i_4	i_4	i_4

Without an affirmative action policy, that is, with $r = (0, 0, 0, 0)$, the IA-DA-R matching μ is the same as the DA matching and it is given by $\mu_{c_1} = \{a_1, a_2, a_5\}$, $\mu_{c_2} = \{a_3, i_1\}$, $\mu_{c_3} = \{i_2, i_3, i_4\}$, and $\mu_{c_4} = \{a_4\}$ at this profile (underlined in Table 2.6). Now consider the minority reserve policy $\tilde{r} = (2, 0, 0, 0)$. The IA-DA-R matching $\tilde{\mu}$ is given by $\tilde{\mu}_{c_1} = \{a_1, i_2, i_4\}$, $\tilde{\mu}_{c_2} = \{a_3, a_5\}$, $\tilde{\mu}_{c_3} = \{a_2, i_1, i_3\}$, and $\tilde{\mu}_{c_4} = \{a_4\}$ at this profile (indicated by the squares in Table 2.6). The steps of the procedure are displayed in Table 2.7.

Table 2.7: Steps of the IA-DA-R mechanism with $\tilde{r}$ in Example 5

	$\boldsymbol{c_1}$	$\boldsymbol{c_2}$	$\boldsymbol{c_3}$	$\boldsymbol{c_4}$
Step 1	a_1, a_2, $\cancel{a_3}$, i_4			a_4, $\cancel{a_5}$, $\cancel{i_1}$, $\cancel{i_2}$, $\cancel{i_3}$
Step 2	a_1, $\cancel{a_2}$, a_5, i_4	a_3, i_1, $\cancel{i_2}$	i_3	a_4
Step 3	a_1, $\cancel{a_5}$, i_2, i_4	a_3, i_1	a_2, i_3	a_4
Step 4	a_1, i_2, i_4	a_3, a_5, $\cancel{i_1}$	a_2, i_3	a_4
Step 5	a_1, $\cancel{i_1}$, i_2, i_4	a_3, a_5	a_2, i_3	a_4
Step 6	a_1, i_2, i_4	a_3, a_5	a_2, i_1, i_3	a_4

Struck out students are rejected at that step.
Circled students are minority students permanently assigned to minority reserve seats.

Note that although i_1 has a higher priority at school c_1 than i_2 and i_4, due to the immediate acceptances made by schools for minority reserve seats i_1 is rejected by c_1, since

i_1 applies to school c_1 only in Step 5, and both minority reserve seats have already been filled permanently with minority students (i_2 and i_4) by the time i_1 applies. Observe that with the stronger affirmative action policy $\tilde{r}$, compared with no affirmative action, minority students i_2 and i_4 are better off, although minority student i_1 is worse off. $\diamond$

Theorem 1. *The IA-DA-R mechanism is non-wasteful, respects the affirmative action policy and is minimally responsive.*

Proof.

Non-Wasteful: The IA-DA-R mechanisms never rejects any new applicant in any step if there is any empty school seat remaining, and a temporarily accepted student is only rejected in any step of the IA-DA-R procedure if a new student is accepted to fill the school seat. Therefore, it is not possible to have an empty school seat in the final matching which is preferred by any student to her assignment, and thus the IA-DA-R mechanism is non-wasteful.

Respects the Affirmative Action Policy: Using the IA-DA-R mechanism the minority reserve seats at each school are assigned to available minority applicants in each step of the procedure, prioritizing them over majority applicants. Moreover, minority students are permanently assigned to seats up to the number of minority reserve seats r_c at each school, while majority students are only temporarily assigned to minority reserve seats and only if there are no competing minority applicants, and are rejected in later steps of the procedure if any new minority student applies to the school. Therefore, it is not possible for a minority student to be rejected by a school unless all of its reserved seats are filled with minority students, and therefore the IA-DA-R mechanism respects the affirmative action policy.

Minimally Responsive: Let the IA-DA-R mechanism be denoted by ψ. Now fix a profile $(P, \succ) \in \mathcal{P} \times \Pi$. Let r, r' be two minority reserve policies such that $r' \geqslant r$ and let $\mu = \psi(r, P, \succ)$ and $\mu' = \psi(r', P, \succ)$. Assume that $\mu \neq \mu'$. Let Step t be the first step in the IA-DA-R procedure at which the accepted sets of students differ at $(r, P, \succ)$ and $(r', P, \succ)$. Since Steps 1 to $t-1$ are the same and since the difference is only in the minority reserve policy, it must be the case that a new minority student applicant gets accepted by a school at $(r', P, \succ)$ for a minority reserve seat but is rejected by this same school at $(r, P, \succ)$, due to the stronger affirmative action policy r' compared with r. Given that minority students are permanently accepted for minority reserve seats by the IA-DA-R mechanism, this implies that at least one minority student is better off with matching μ' than μ: there exists $i \in S^m$ such that $\mu'_i \; P_i \; \mu_i$. Therefore, the IA-DA-R mechanism is minimally responsive. $\square$

The next statement follows immediately from Theorem 1.

Corollary 1. *There exists a mechanism with minority allotments which is non-wasteful, respects the affirmative action policy and is minimally responsive.*

We examine the fairness and incentive properties of the IA-DA-R mechanism next.

2.7 Priority Violations and Minority Fairness

In this section we focus on fairness (stability) conditions and priority violations. Given a profile $(P, \succ)$, a matching μ, students s, s' and school c, the priority of student s is said to be violated at c by student s' if $\mu_{s'} = c$, $c \; P_s \; \mu_s$ and $s \succ_c s'$. We will also say in this case that $\boldsymbol{s'}$ **violates the priority of** $\boldsymbol{s}$ in matching μ at profile $(P, \succ)$. The standard fairness (or stability) condition, satisfied by the classic DA, requires that no student violate any other student's priority at any profile (this is also known as eliminating justified envy).

A mechanism with an affirmative action policy cannot satisfy standard fairness, since the essence of affirmative action is to allow minority students to be prioritized over majority students in some instances, violating the exogenously given priorities of majority students. We define below a fairness axiom which allows for priority violations due to affirmative action and is satisfied by the IA-DA-R mechanism. Such fairness conditions in the literature often require that there is no priority violation within the sets of same types of students (here the sets of majority and minority students), but we cannot require this for minority students due to the immediate acceptances in the IA-DA-R mechanism. While the desirability of ruling out such priority violations among the same type of students is taken for granted in more general settings, we argue that it is debatable for minority students, especially when an affirmative action policy is in place. A common criticism of affirmative action policies is that they are ineffective because they only help the most privileged minority students, which means the highest-priority minority students when the priority orderings of schools are performance-based. Thus, enforcing the priorities of minority students among themselves is not desirable in such cases. The relaxation of this requirement for minority students plays a crucial role and allows us to obtain an existence result that contrasts with previous results, as we will explain below.

Minority Fairness. Given a minority allotment policy $v \in \mathcal{V}$ and a profile $(P, \succ) \in \mathcal{P} \times \Pi$, we call a matching μ **minority fair with respect to** $\boldsymbol{v}$ **and** $\boldsymbol{(P, \succ)}$ if, at profile $(P, \succ)$, it satisfies the following conditions:

1. no majority student violates another student's priority in μ,
2. at most v_c minority students violate the priority of another student at each school c, and

3. no minority student violates the priority of another minority student $i \in S^m$ at a school c if P_i ranks c first.

A mechanism φ^v is **minority fair** if for all minority allotment policies $v \in \mathcal{V}$ and all profiles $(P, \succ) \in \mathcal{P} \times \Pi$, $\varphi^v(v, P, \succ)$ is minority fair with respect to v and $(P, \succ)$.

Note that minority fairness becomes the standard fairness property (i.e., no priority violations) when there is no affirmative action, since in this case not only majority students cannot violate another student's priority by condition 1, but also minority students cannot violate another student's priority by condition 2 when $v_c = 0$ for all $c \in C$.

Affirmative action mechanisms based on deferred acceptance, namely DA-Q and DA-R, satisfy minority fairness (as is the DA), but the efficiency improvements of MDA and EIDA, which build on DA-R, destroy this property of DA-R. Clearly, the affirmative action mechanisms based on immediate acceptance, namely IA-Q and IA-R, do not satisfy minority fairness (and nor does IA), as immediate acceptances generally violate the priorities of students. However, IA-DA-R satisfies minority fairness, as we will show next. Therefore, based on Theorem 1, we can state the following result which strengthens Corollary 1.

Corollary 2. *There exists a mechanism with minority allotments which is minority fair, non-wasteful, respects the affirmative action policy and is minimally responsive.*

Proof. Observe that due to Theorem 1, it suffices to show that the IA-DA-R mechanism is minority fair. We need to show that IA-DA-R satisfies the three conditions in the definition of minority fairness. First note that majority students are always accepted temporarily only and may be rejected due to higher-priority applicants in later steps in the IA-DA-R procedure, whether for a reserved seat or an unreserved seat. Thus, no majority student violates any other students' priorities at any profile, and condition 1 holds. Moreover, minority students are treated similarly to majority students when accepted for seats after the minority reserve seats have been filled with minority students, or for seats at schools without minority reserves. This implies that no more than v_c minority students may violate the priority of another student at any school c, satisfying condition 2. Finally, we can also easily verify that a minority student's priority may not be violated by another minority student at a school if the former student ranks school c as her top choice, because such a priority violation would be ruled out by the immediate acceptances for minority reserve seats in the first step of the IA-DA-R procedure, given that these permanent acceptances follow the priority ordering of the school. This implies that the IA-DA-R mechanism satisfies condition iii). □

It is instructive to compare this possibility result to an impossibility result proved by Doğan (2016), Proposition 9, which states that there is no mechanism which satisfies weak fairness and minimal responsiveness. His minimal responsiveness property is the same as ours, while weak fairness can be expressed as the following four requirements: i) a slightly weaker property than our axiom of respecting the affirmative action policy, ii) no majority student's priority is violated at any school by any majority student, iii) no minority student's priority is violated at any school by any minority student, and iii) no student's priority is violated at any school which has no reserved seats ($r_c = 0$). Minority fairness implies ii) based on condition 1 in the definition of minority fairness, and it also implies iv), since if $v_c = r_c = 0$ then no minority student may violate another student's priority at school c, given condition 2, and thus iv) follows, taking into account condition 1. It is also clear that respecting the affirmative action policy implies i). Thus, the only requirement that is not satisfied among Doğan's requirements by the properties in Corollary 2 is iii), namely, that a minority student cannot violate another minority student's priorities. Condition 3. restricts this type of priority violations to cases where a minority student whose priority is violated at a school by another minority student (for a reserved seat, based on condition 2.) didn't rank this school first. Intuitively speaking, requirement iii) is not satisfied by IA-DA-R due to the immediate acceptance of minority students for minority reserve seats, which allows for a minority student who applies to a school in a step when all minority reserve seats are already filled to be rejected, even if the minority student has a higher priority for this school than some of the permanently accepted minority students. Therefore, the relaxation of the requirement that there should be no priority violations among minority students is responsible for our possibility result when compared with Doğan's impossibility result. While the desirability of ruling out such priority violations among minority students is debatable, we note that in the IA-DA-R mechanism minority students who rank a school higher than other minority students are more likely to be assigned to a reserved seat at that school, although the rankings of schools by minority students do not always explain priority violations between two minority students. This is because a minority student i_1 may be "stuck" with a school seat which is not a reserved seat (when all reserved seats are already permanently assigned to minority students at that school or when the school has no minority reserves) for several steps in the IA-DA-R procedure before being rejected, and thus may apply too late to another school c and lose out to a permanently accepted minority student i_2 who has both a lower priority for c than i_1 and ranks c worse than i_1.[6]

[6]This means that the PP-stability axiom of Ayoade and Pápai (2023), which relaxes standard stability by allowing for priority violations based on preference ranks, is not satisfied by IA-DA-R. This fits with the finding, as shown by Ayoade and Pápai (2023), that PRP mechanisms that treat students asymmetrically do not satisfy PP-stability.

2.8 Strategyproofness: An Impossibility Result

Now we turn to the analysis of incentive properties, namely, strategyproofness, in the context of affirmative action policies. Among the seven mechanisms only the DA-Q and DA-R are strategyproof. The strategyproofness of DA-Q follows immediately from the strategyproofness of the DA, while the same property of the DA-R mechanism was shown by Hafalir et al. (2013). The other mechanisms are all manipulable, which is not too surprising for MDA and EIDA, given that EADAM is not strategyproof, and for IA-Q and IA-R, given that IA is not strategyproof.

The IA-DA-R mechanism may be manipulated by minority students in a similar manner as the IA mechanism can be manipulated by students, by reporting a school higher for which a student has higher priority. Specifically, in the IA-DA-R mechanism a minority student may be able to obtain a school's reserved seat ahead of some other minority student by ranking the school higher than it would be ranked truthfully, and by being immediately accepted by that school before other minority students with a higher priority for that school apply to it. Since the time of the application matters in the IA mechanism, as opposed to the DA mechanism, in the IA-DA-R mechanisms the time when a minority student applies to a school with minority reserves matters and the outcome can be profitably changed by a minority students to obtain a reserved seat. This brings up the question whether the IA-DA-R mechanism can be manipulated by majority students for whom all school seats are deferred acceptance seats, as explained earlier, and thus seemingly face a similar environment to the DA mechanism which is strategyproof for all students (Dubins and Freedman (1981), Roth (1982)). Unfortunately, since by reporting different preferences majority students can alter the step of the IA-DA-R procedure in which minority students apply to schools with minority reserves, and hence can affect the distribution of minority reserves that is based on immediate acceptance, majority students can also manipulate the outcome of the IA-DA-R mechanism.

Example 6. This example demonstrates that the IA-DA-R mechanism can be manipulated by majority students. Let $S^M = \{a_1, a_2\}$ and $S^m = \{i_1, i_2\}$ be the sets of majority and minority students. Let $C = \{c_1, c_2, c_3\}$ with capacities $q = (1, 1, 1)$ and let the minority reserve policy be $r = (0, 1, 0)$. Consider the profile $(P, \succ)$ in Table 2.8.

At the specified profile the IA-DA-R matching is $\mu_{c_1} = \{a_1\}$, $\mu_{c_2} = \{i_2\}$, $\mu_{c_3} = \{i_1\}$, and $\mu_{a_2} = 0$, as indicated by the underlined assignments in Table 2.8. However, if majority student a_2 reports P'_{a_2} instead of P_{a_2}, then at the resulting profile $((P'_{a_2}, P_{-a_2}), \succ)$ the IA-DA-R matching is μ' given by $\mu'_{c_1} = \{a_1\}$, $\mu'_{c_2} = \{i_1\}$, $\mu'_{c_3} = \{a_2\}$, and $\mu'_{i_2} = 0$, as indicated by the assignments in squares in the table. In particular, majority student a_2 gets

Table 2.8: Profiles for Example 6

P_{a_1}	P_{a_2}	P_{i_1}	P_{i_2}	P'_{a_2}	$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$
c_2	$\boxed{c_3}$	c_1	c_1	c_1	a_1	a_1	a_1
$\underline{\boxed{c_1}}$	c_1	$\boxed{c_2}$	$\underline{c_2}$	c_3	a_2	i_1	i_1
0	$\underline{0}$	$\underline{c_3}$	c_3	0	i_1	i_2	a_2
		0	$\boxed{0}$		i_2	a_2	i_2

c_3 when reporting false preferences P'_{a_2}, which is preferred by a_2 to remaining unmatched when she reports P_{a_2} truthfully. Therefore, majority student a_2 can manipulate the outcome at $(P, \succ)$. Intuitively, when a_2 reports c_1 first, both minority students are rejected from c_1 in the first step, since a_2 has higher priority at c_1 than both, and thus both minority students apply to c_2 in step 2, which ensures that i_1 is accepted permanently by c_2, as opposed to when a_2 reports truthfully and i_1 applies to c_2 too late, when the minority reserve seat is already assigned to i_2 permanently. Therefore, when a_2 reports untruthfully, a_2's competition at c_3 is i_2, instead of i_1 in the next step. Since i_1 has a higher priority and i_2 has a lower priority than a_2 at school c_3, this allows a_2 to manipulate. $\diamond$

Although the IA-DA-R mechanism does not appear to have good incentive properties in terms of strategyproofness, we show in our final result next that it is not possible to reconcile all the axioms required in Corollary 2 with strategyproofness, namely minority fairness and the three main welfare axioms. In fact, we do not need to require non-wastefulness in addition to the other axioms, but the impossibility result of course still holds if we also impose non-wastefulness.

Theorem 2 (Impossibility result)**.** *There is no mechanism with minority allotments which is strategyproof, minority fair, respects the affirmative action policy and is minimally responsive.*

Proof. Suppose by contradiction that mechanism $\psi : \mathcal{V} \times \mathcal{P} \times \Pi \to \mathcal{M}$ with minority allotments is strategyproof, minority fair, respects the affirmative action policy and is minimally responsive. Let $S^M = \{a_1, a_2\}$ and $S^m = \{i_1, i_2\}$ be the sets of majority and minority students. Let $C = \{c_1, c_2, c_3\}$ with capacities $q = (1, 1, 1)$. Consider the profile in Table 2.9.

When there is no affirmative action, any minority fair mechanism yields a fair matching at each profile (i.e., no priority violations), due to conditions 1 and 2 in the definition of minority fairness. Moreover, since ψ is both strategyproof and minority fair, when minority allotment $v = \mathbf{0}$ is fixed and there is no affirmative action, the resulting restricted mechanism is both strategyproof and stable, and thus it follows from Alcalde and Barberà (1994)

Table 2.9: Profiles for the proof of Theorem 2

P_{a_1}	P_{a_2}	P_{i_1}	P_{i_2}	P'_{i_1}	$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$
$\underline{c_2}$	$\underline{c_1}$	$\underline{c_3}$	c_2	c_1	a_1	i_1	a_1
$\boxed{c_3}$	c_2	c_1	$\boxed{c_1}$	c_3	a_2	a_2	i_1
c_1	$\boxed{0}$	$\boxed{c_2}$	$\underline{0}$	c_2	i_1	a_1	a_2
					i_2	i_2	i_2

that for all profiles $(\bar{P}, \succ^-) \in \mathcal{P} \times \Pi$, $\psi(\mathbf{0}, \bar{P}, \succ^-)$ yields the DA matching at $(\bar{P}, \succ^-)$. Therefore, for the profile specified in Table 2.9, $\mu = \psi(\mathbf{0}, P, \succ)$ is the DA matching, given by $\mu_{c_1} = \{a_2\}$, $\mu_{c_2} = \{a_1\}$, $\mu_{c_3} = \{i_1\}$, and $\mu_{i_2} = 0$, as indicated by the underlined assignments in the table.

Now consider the minority allotment policy $\tilde{v} = (1, 0, 0)$, and let $\tilde{\mu} = \psi(\tilde{v}, P, \succ)$ be the matching at the specified profile under this affirmative action policy. Given that ψ is minority fair and $\tilde{v}_{c_2} = 0$, no student's priority is violated at school c_2, and thus $\tilde{\mu}_{c_2} \neq \{i_2\}$, given that P_{a_1} ranks c_2 first and $a_1 \succ_{c_2} i_2$. Thus, since $\tilde{v}_{c_1} = 1$, either i_1 or i_2 is assigned to school c_1 in $\tilde{\mu}$, otherwise $c_1 \; P_{i_2} \; \tilde{\mu}_{i_2}$ and $\tilde{\mu}^m_{c_1} < 1 = \tilde{v}_{c_1}$ would imply that ψ does not respect the affirmative action policy. Now note that if $\tilde{\mu}_{c_1} = \{i_1\}$ then ψ would not be minimally responsive, given that $\tilde{v}$ is a stronger affirmative action policy than v, $\mu_{i_1} = c_3$, $c_3 \; P_{i_1} \; c_1$, and i_2 would get at best 0 in $\tilde{\mu}$, implying that neither minority student would gain from the stronger minority reserve policy and at least i_1 would be worse off. Therefore, $\tilde{\mu}_{c_1} = \{i_2\}$.

Since ψ is minority fair and $\tilde{v}_{c_2} = \tilde{v}_{c_3} = 0$, no student's priority can be violated at schools c_2 and c_3 in $\tilde{\mu}$ at $(P, \succ)$. If a_1 is assigned to c_2 in $\tilde{\mu}$ then, since $a_2 \succ_{c_2} a_1$, a_2's priority would be violated at c_2. However, if a_2 is assigned to c_2 in $\tilde{\mu}$ then $i_1 \succ_{c_2} a_2$ would imply that i_1 is assigned to c_3, and a_1's priority would be violated at school c_3, since then a_1 would remain unmatched in $\tilde{\mu}$ and $a_1 \succ_{c_3} i_1$. Therefore, since i_2 is assigned to c_1, a_2 remains unmatched in $\tilde{\mu}$, given that $a_1 \succ_{c_3} i_1 \succ_{c_3} a_2$, and thus a_2 cannot be assigned to c_3 in $\tilde{\mu}$ without violating one of these students' priority at c_3 (note that we have not assumed that ψ is individually rational). Then minority fairness implies that c_2 cannot be assigned to a_1, given that $a_2 \succ_{c_2} a_1$, and it must be the case that c_2 is assigned to i_1 and c_3 is assigned to a_1 in $\tilde{\mu}$, in order to avoid any priority violations at schools c_2 and c_3. This yields the matching $\tilde{\mu}_{c_1} = \{i_2\}$, $\tilde{\mu}_{c_2} = \{i_1\}$, $\tilde{\mu}_{c_3} = \{a_1\}$, and $\tilde{\mu}_{a_2} = 0$, as indicated by the assignments in squares in Table 2.9.[7]

Let P'_{i_1} be as indicated in Table 2.9, and let $P' = (P'_{i_1}, P_{-i_1})$. Then, since i_1 ranks c_1 first at the new profile P', condition 3 in the definition of minority fairness implies that i_1's

[7]Note that $\tilde{\mu}$ is the IA-DA-R matching at the specified profile.

priority cannot be violated by i_2 at c_1, and thus it follows from $i_1 \succ_{c_1} i_2$ that i_2 is not assigned to c_1 in matching $\psi(\tilde{v}, P', \succ)$. Then, since ψ respects the affirmative action policy, $\psi_{i_1}(\tilde{v}, P', \succ) = c_1$. Given that $\tilde{\mu}_{i_1} = \psi_{i_1}(\tilde{v}, P, \succ) = c_2$ and $c_1 \ P_{i_1} \ c_2$, minority student i_1 can manipulate at $(P, \succ)$. This means that ψ is not strategyproof, contrary to our assumption. Therefore, there is no mechanism with minority allotments which is strategyproof, minority fair, respects the affirmative action policy and is minimally responsive at the same time. □

Doğan (2016) states a similar impossibility result (Proposition 10) with a fairness axiom called minimal fairness instead of minority fairness, while minimal responsiveness and strategyproofness are the same in the two statements. Minimal fairness includes a weak version of respecting the affirmative action policy, and thus the main difference from our axioms is that his minimal fairness axiom requires that there is no priority violation among minority students, while minority fairness only rules out such priority violations at a school for minority students who rank this school first. This is an important difference, and our result strengthens Doğan's impossibility result, although technically Theorem 2 does not directly imply his. His proof is also distinct from ours, as it demonstrates that a majority student can manipulate when all the other axioms are satisfied, while in our proof of Theorem 2 we have shown that a minority student can manipulate in a manner that is similar to the manipulation of the IA mechanism.

2.9 Conclusion

We summarize our findings, based on Propositions 1-6 and Theorem 1, on the previously studied affirmative action mechanisms and the newly proposed IA-DA-R mechanism in terms of the three main welfare criteria in Table 2.10.

Table 2.10: Comparison of mechanisms

Mechanism	**WELFARE AXIOMS**		
	Non-Wasteful	*Respects AA*	*Minimally Responsive*
DA-Q	×	✓	×
DA-R	✓	✓	×
MDA	✓	×	✓
EIDA	✓	×	×
IA-Q	×	✓	×
IA-R	✓	×	✓
IA-DA-R	✓	✓	✓

The table shows that none of the affirmative action mechanisms proposed previously satisfy all three welfare axioms.[8] If we consider minimal responsiveness the least important axiom among the three welfare criteria, the DA-R mechanism emerges as the best mechanism among the six mechanisms that have been proposed in the literature, as it is the only mechanism that satisfies the other two more fundamental criteria. If respecting the affirmative action is less important than minimal responsiveness, then MDA and IA-R stand out, since they both satisfy the other two criteria. However, as Theorem 1 demonstrates, we can design mechanisms that satisfy all three welfare axioms. The IA-DA-R mechanism is combines some of the good qualities of immediate acceptance and deferred acceptance, and results in an effective affirmative action policy, overcoming the individual weaknesses of both immediate and deferred acceptance, reflected by the fact that this mechanism satisfies all three welfare axioms. Clearly, this is not a trivial accomplishment for an affirmative action mechanism, as evidenced by Table 2.10.

When we also consider fairness properties, our minority fairness axiom is weaker than usual fairness axioms for affirmative action due to allowing some priority violations among minority students, although these are restricted at each school to the number of minority allotments. Compared to the DA-R mechanism, this is the price of achieving minimal responsiveness, and compared to the MDA and IA-R mechanisms, this is the price of respecting the affirmative action policy. The latter is an interesting trade-off between being able to allocate all minority reserves to minority applicants versus ensuring that minority students obtain reserved seats in the order of their exogenously given priorities. This reveals a choice between the overall effectiveness of the affirmative action policy versus a fair treatment among individual minority students. Moreover, one may also argue that enforcing the given priorities of minority students is not always desirable. Some critics of affirmative action policies claim that affirmative action benefits primarily the most privileged members of underrepresented groups, and this suggests that in contexts where the schools' priority rankings of students are based on achievement, fairness considerations may actually require priority violations among minority students, which would make it easy to choose the overall effectiveness of the affirmative action policy over enforcing minority student priorities. One way to de-emphasize the within-group priority ranking of minority students when assigning them to reserved seats is by relying on their relative preference ranks of schools, as done by the IA mechanism, a fairness measure that would fit well with the IA-DA-R

[8]The least-studied TTC-based affirmative action mechanisms were not included in our analysis. Kojima (2012) introduces what we may call TTC-Q, imposing quotas on majority students; this mechanism, similarly to IA-Q, is not minimally responsive. Moreover, a TTC-R mechanism, if defined in the spirit of IA-R but with TTC as the base mechanism instead of IA, does not respect the affirmative action policy, similarly to IA-R, due to its permanent acceptances, which are based on top trading cycles rather than on the priority profile in the case of TTC-based mechanisms.

mechanism as well.

Regarding incentives, the DA-Q and DA-R mechanisms are strategyproof, while all other mechanisms in our analysis, including the IA-DA-R mechanism, are manipulable. We demonstrate with Theorem 2 that it is not possible to reconcile strategyproofness with the three welfare axioms and some fairness criteria. This suggests that basic welfare considerations that include even a minimal requirement of responsiveness to the degree of affirmative action are not possible to obtain together with strong incentive properties, indicating necessary trade-offs that designers of matching mechanisms with affirmative action policies will have to take into account.

One important issue to address when comparing different types of mechanisms based on the DA, IA and EADAM (efficiency improvement type of mechanisms) is that while the DA-based DA-Q and DA-R mechanisms are strategyproof, IA-based mechanisms have been shown not only to be manipulable but also obviously manipulable (Troyan & Morrill, 2020), with EADAM-based mechanisms somewhere between the two, which are manipulable but not obviously manipulable. Thus, we cannot be sure whether IA-Q and IA-R actually satisfy the welfare axioms claimed in Propositions 5 and 6, and even the claims in Propositions 3 and 4 may not hold exactly, since all the results hold exactly only when preferences are reported truthfully. We argue, however, that the alternative, to check the properties that hold in equilibrium, primarily for IA-Q and IA-R, are also not reliable, since an equilibrium analysis does not necessarily give robust results either. Specifically, in practice participants of IA-based mechanisms may not manipulate as much as theoretically predicted (see, e.g., Featherstone and Niederle (2016)). Moreover, the assumption of truth-telling is also questionable for strategyproof mechanisms, as investigated theoretically by Li (2017) and many recent papers following it. Specifically, the DA is not obviously strategyproof and thus people who have trouble with contingent reasoning are likely to misrepresent their preferences in the DA. There is also a steadily growing experimental literature documenting the frequent misrepresentation of truthful preferences in the DA mechanism, which is often done the same way the IA mechanism is manipulated (see a brief summary of this literature in Dreyfuss, Heffetz, and Rabin (2022)). However, mechanisms with a truth-telling dominant strategy cannot be subjected to an equilibrium analysis, as it does not address the problem of untruthful reporting in the experimental lab and in real life. Thus, we cannot be sure of the accuracy of either a direct comparison or an equilibrium comparison of these mechanisms, but as a first step in a theoretical investigation with practically still very likely applicable results, we work in this paper with the hypothetical scenario where none of the mechanisms are manipulated.

Furthermore, based on the equilibrium analysis of Ergin and Sönmez (2006) which shows that any Nash-equilibrium outcome is a stable matching in the IA mechanism, keeping in mind that it is subject to the above criticisms, assuming that a focal point among the equilibria is to coordinate on the only undominated matching, the DA outcome, we can offer the following understanding of the results displayed in Table 2.10. Since IA-Q and DA-Q have the same properties, even if IA is manipulated and its outcomes become more similar to the DA outcomes, these properties will still hold in any scenario. In addition, given the properties of the DA-R mechanism, IA-R may satisfy the axiom of respecting the affirmative action more often, but it may not always be minimally responsive. In terms of distorted results compared to the direct mechanism, the IA-DA-R mechanism should fare better than either IA-Q or IA-R, since it is intuitively clear that it is less manipulable than either. As discussed in Section 2.8, only the minority students have a strong incentive to manipulate the IA-DA-R mechanism, which can be done in a similar manner to the manipulation of the IA mechanism, by placing "safe" schools with minority reserves higher in the preference ordering, while for majority students manipulation is quite difficult.

In the end, despite some uncertainty surrounding the issues of truth-telling, which applies to all the mechanisms and all the related studies in the market design literature in the absence of better behavioral foundations, we believe that our analysis offers significant insights and has practical relevance. Therefore, given our findings, the IA-DA-R mechanism should be considered as a potentially superior alternative to all previously proposed mechanisms with affirmative action not just theoretically but also for real-life use.

Appendix: IA-RQ Mechanism

Some of the relevant results on IA mechanisms with affirmative action have been established by Afacan and Salman (2016) which studies primarily the IA Mechanism with Minority Reserve and Majority Quota (what we call IA-RQ, for short), a mechanism that allows for both reserve-based and quota-based affirmative action policies. We introduce here this class of mechanisms formally.

IA with Minority Reserves and Majority Quotas

Fix a minority reserve policy r, a majority quota policy q^M and a profile $(P, \succ)$.

Step 1: Every student applies to her most preferred school according to P. Each school c permanently assigns seats to applying minority students up to its number of minority reserve seats r_c, following its priority ordering $\succ_c$. Then each school c permanently assigns its remaining seats to the remaining applicants, following its priority ordering

$\succ_c$, up to its capacity q_c. However, each school c rejects majority students when the majority quota q_c^M is reached by accepted majority students. Acceptances are final, and any remaining applicants are rejected.

Step *t* $(t \geqslant 2)$**:** Every student who was rejected in step $t-1$ applies to her next most preferred acceptable school according to P, the student's t^{th}-ranked school. Each school c that has fewer minority students accepted than r_c permanently assigns seats to minority students applying in this step up to either its number of minority reserve seats r_c in total or its capacity q_c, whichever is exhausted first, following its priority ordering $\succ_c$. Then each school that still has available seats permanently assigns its remaining applicants, following its priority ordering $\succ_c$, up to its capacity q_c. However, each school c rejects majority students when the majority quota q_c^M is reached in total by accepted majority students. Acceptances are final, and any remaining applicants are rejected.

The algorithm terminates when each student is either accepted by a school or has been rejected by all of her acceptable schools. The acceptances made in each step are final and together constitute the matching assigned to profile $(P, \succ)$ when the majority quota policy is q^M. ◇

As the name suggests, this mechanism incorporates both types of affirmative action and allows for using them at the same time. Thus, it includes both the IA-R and the IA-Q mechanisms as special cases. Given the IA-RQ mechanism with (r, q^M), the IA-R mechanism is the special case where $q^M = q$ and the IA-Q mechanism is the special case where $r = \mathbf{0}$. We call the IA-RQ mechanism restricted to using both types of affirmative action policies at the same time IA with Both Minority Reserves and Majority Quotas, denoted by IA-RQ$^\wedge$. For the restricted IA-RQ$^\wedge$ mechanism with (r, q^M) there exist two schools $c, c' \in C$ such that $r_c > 0$ and $q_{c'} - q_{c'}^M > 0$. Figure 1 depicts the restricted cases of the IA-RQ mechanism.

The definition of the IA-RQ mechanism given by Afacan and Salman (2016), which allows for both minority reserve seats and majority quotas, leaves an important question open: how is the sum of the number of minority reserve seats and the majority quota related to a school's total capacity? Since they don't specify this relationship, presumably any relationship between them is allowed. But $r_c + q_c^M < q_c$ would not make much sense, since minority students will automatically be prioritized for seats that majority students cannot be assigned to and thus, effectively, since reserve seats are set aside for minority students to be prioritized over majority students, the number of minority reserve seats must be at least

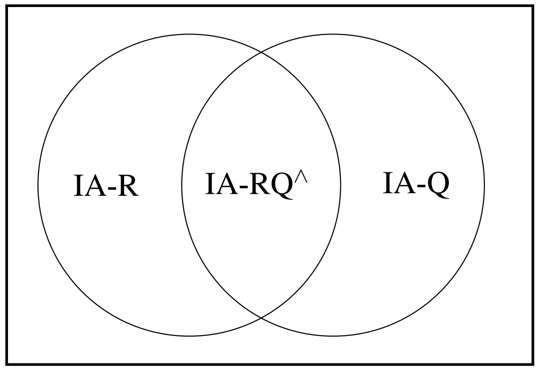

Figure 2.1: Special cases of the IA-RQ mechanism

the number of seats that cannot be occupied by majority students due to the majority quota. Indeed, not only are minority students prioritized over majority students for any school seat that is not included in the majority quota, but minority students are the only students who can occupy these seats. Thus, if one insists on having both minority reserves and majority quotas, then the number of seats kept exclusively for minority students, $q_c - q_c^M$, cannot be higher than the number of seats for which minority students are prioritized, r_c, and hence it must be the case that $q_c - q_c^M \leqslant r_c$. Furthermore, when $q_c - q_c^M = r_c$, having minority reserve seats is redundant, and the IA-RQ mechanism reduces to the IA-Q mechanism with majority quotas only. The majority-quota-based mechanism is obtained from IA-RQ mechanisms by setting the minority reserve to zero at each school according to Afacan and Salman (2016) (and subsequently according to Chen et al. (2022)) but in fact this is not necessary. This leaves the case where $q_c - q_c^M < r_c$. If $q_c^M = q_c$ then there is no restriction on the number of majority students who can be accepted by school c. This corresponds to the case where there is only minority reserve policy. When $q_c^M < q_c$ such that $q_c - q_c^M < r_c$, both majority-quota and minority-reserve-based policies play a role, and we can view these as hybrid affirmative action policies.

As we mentioned, Afacan and Salman (2016) use Example 3 to demonstrate that IA-RQ mechanisms are not minimally responsive to majority-quota-based affirmative action policies. But this argument is based on an affirmative action policy where $q_c - q_c^M > r_c$, since here $\tilde{q}_{c_2}^M = 0$ while $\tilde{r}_{c_2} = 0$ is assumed in the same affirmative action policy. However, as argued above, since $q_{c_2} = 1$, $\tilde{q}_{c_2}^M + \tilde{r}_{c_2} \geqslant 1$ should be satisfied. Thus, when $\tilde{q}_{c_2}^M = 0$, we have $\tilde{r}_{c_2} = 1$, given that $\tilde{r}_{c_2} \leqslant q_{c_2}$. Therefore, Example 3 also demonstrates that the IA-RQ mechanisms are not minimally responsive to minority-reserve-based affirmative action when the affirmative action policy is correctly interpreted.

Since the IA-RQ^ mechanism implements both minority reserve and majority quota at the same time, we need to define the responsiveness axiom in terms of a pair of affirmative actions (r, q^M). An affirmative action policy pair is deemed to be stronger if all schools in the stronger policy do not have a higher majority quota or a lower minority reserve compared to some other policy, with at least one school having a strictly lower majority quota, or strictly higher minority reserve, or both, in the stronger policy compared to the other policy. Formally, the pair of minority reserve and majority quota $(r', q^{M'})$ represents a weakly stronger affirmative action policy than (r, q^M), if $r'_c \geqslant r_c$ and $q_c^{M'} \leqslant q_c^M$ for all schools $c \in C$.

Minimal Responsiveness to Minority Reserves and Majority Quotas. A mechanism φ with a pair of affirmative action policies (r, q^M) is minimally responsive if for all profiles a weakly stronger affirmative action policy never results in a Pareto-dominated outcome for minority students.

Formally, φ is minimally responsive if, for all $(P, \succ) \in \mathcal{P} \times \Pi$ and $\big((r, q^M), (r', q^{M'})\big)$ such that $r \leqslant r'$ and $q^M \geqslant q^{M'}$, if $\varphi_{S^m}(r, q^M, P, \succ) \neq \varphi_{S^m}(r, q^{M'}, P, \succ)$ then there exists $i \in S^m$ such that $\varphi_i\,(r', q^{M'}, P, \succ)\ P_i\ \varphi_i(r, q^M, P, \succ)$.

The following example shows that the IA-RQ^ mechanism is not minimally responsive to minority reserves and majority quotas. Here both minority reserves and majority quotas are implemented, respecting that $r_c + q_c^M \leqslant q_c$ for all $c \in C$. That is, at some school $c \in C$ if $r_c > 0$ then $q_c^M = q_c$ and at some other school $c' \in C$ if $q_{c'}^M < q_c$ then $r_{c'} = 0$. In this way, we circumvent the problem that for $r_c > 0$ and $q_c^M < q_c$ when $r_c + q_c^M \geqslant q_c$ for any $c \in C$ then only the majority quota is effective and having a minority reserve is redundant. Therefore, in the following example both the minority reserves and the majority quotas are effective as affirmative action policies. This setup of not having both majority quotas and minority reserves simultaneously at any particular school $c \in C$ is aligned with the examples of Afacan and Salman (2016), where they used one policy at a time for each school.

Table 2.11: Profile for Example 7

P_{a_1}	P_{a_2}	P_{a_3}	P_{a_4}	P_i		$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$	$\succ_{c_4}$	$\succ_{c_5}$
$\underline{c_1}$	c_1	$\underline{\boxed{c_3}}$	c_3	c_3		a_1	a_4	a_3	i	a_3
$\boxed{c_5}$	c_2	c_2	$\underline{\boxed{c_2}}$	c_2		a_2	a_3	i	a_1	a_4
0	c_5	c_1	0	$\underline{c_5}$		a_3	a_2	a_4	a_2	i
	$\underline{c_4}$	0		$\boxed{c_4}$		a_4	i	a_1	a_3	a_1
	c_3			0		i	a_1	a_2	a_4	a_2
	$\boxed{0}$									

Example 7. Let $S^m = \{i\}$ and $S^M = \{a_1, a_2, a_3, a_4\}$ be the sets of minority and majority students. There are five schools $C = \{c_1, c_2, c_3, c_4, c_5\}$ with capacity $q_c = 1$ for all $c \in C$. Consider the preference and priority profiles in Table 2.11.

Consider the affirmative action policy (r, q^M) such that $r = \mathbf{0}$ and $q_c^M = 1$ for all $c \in C$. This is in fact the case of no affirmative action policy. With (r, q^M) the minority student i is assigned to school c_5 and the majority students a_1, a_2, a_3, and a_4 are respectively assigned to schools c_1, c_4, c_3, and c_2 by the IA-RQ mechanism, which coincides with the IA in this case (see the assignments underlined in Table 2.11).

Now consider a stronger affirmative action policy $(\tilde{r}, \tilde{q}^M)$ with the majority quota policy $\tilde{q}^M = (0, 1, 1, 1, 1,)$ and the minority reserve policy $\tilde{r} = (0, 0, 0, 0, 1)$. That is, only c_1 has a majority quota (for which no majority students allowed) and only c_5 has any minority reserve seat (one reserved seat). Then the minority student i is matched to c_4, the majority student a_2 is unmatched, and majority students a_1, a_3, and a_4 are respectively matched to c_5, c_3, and c_2 by the IA-RQ mechanism (these assignments are indicated by the squares in Table 2.11). $\diamond$

In light of the example above, we can show that IA-RQ^ mechanisms do not satisfy any of the three main welfare criteria.

Proposition 7. *The IA-RQ^ mechanism is wasteful, does not respect the affirmative action policy and is not minimally responsive.*

Proof. *Wastefulness:* $IA - RQ^\wedge$ mechanisms are wasteful, as we can see from Example 7: with the affirmative action policy $(\tilde{r}, \tilde{q}^M)$, the majority student a_2 is unmatched, while a seat at c_1 is empty, which is acceptable by a_2 (and, in fact, highly desired, as it is ranked first by a_2.

Not respecting the affirmative action policy: Example 7 shows $\varphi_i(P, \succ) = c_4$ for (r, q^M) and $\varphi_i(P, \succ) = c_5$ for $(\mathbf{0}, q^c)$. Thus, the $IA - RQ^\wedge$ mechanism with affirmative action policies (r, q^M) assigns majority student a_1 to c_5, while the minority reserve $\tilde{r}_{c_5} = q_{c_5} = 1$ are not filled with minority students, but the minority student i desires this seat at c_5, since $c_5 \; P_i \; c_4$. Therefore, the IA-RQ^ mechanism does not respect the affirmative action policy.

Not minimally responsive: Table 2.11 describes a profile for which the IA-RQ^ mechanism is not minimally responsive. In Example 7, $\varphi_i(P, \succ) = c_5$ for $(\mathbf{0}, q^c)$ and $\varphi_i(P, \succ) = c_4$ for $(\tilde{r}, \tilde{q}^M)$, where $c_5 \; P_i \; c_4$. Therefore, for the profile in Table 2.11 the matching $\varphi(P, \succ)$ with $(\tilde{r}, \tilde{q}^M)$ is Pareto-dominated by $\varphi(P, \succ)$ with $(\mathbf{0}, q^c)$ for minority students. Therefore, the IA-RQ^ mechanism is not minimally responsive. $\square$

Note that the IA-RQ mechanism cannot be minimally responsive since the special case of the IA-Q, and thus the IA-RQ$^{\wedge}$ mechanism, is not minimally responsive. However, Afacan and Salman (2016) claim that the IA-RQ mechanism is minimally responsive to minority-reserve-based affirmative action. But such a claim can only apply to the special case of the IA-R mechanism.

Furthermore, Chen et al. (2022) give an example to demonstrate that the IA-Q mechanism is minimally responsive. However, their model assumes the priority of all minority students over all majority students at each school, which is satisfied by the priority profile used in their example. This priority profile is equivalent to having a minority reserve at every school that is set equal to the capacity of that school. Thus, their claim does not contradict the fact that the IA-Q mechanism is not minimally responsive (Proposition 5).

In sum, even though the above-mentioned two papers make statements about the IA mechanism with affirmative action policies that appear to contradict our results, their findings are aligned with ours when properly evaluated.

Chapter 3

Responsive Affirmative Action: Axioms and Policies

3.1 Introduction

A common form of affirmative action policy in school matching is to reserve school seats for which minority students have priority. The implementation of an affirmative action policy imposes additional constraints on matching in school choice, university admissions, and other settings where diversity and helping underprivileged groups is an objective. Designing a successful algorithm to assign students to schools with set-aside seats for minority, taking into account the students' preferences, the schools' priorities prior to affirmative action, and the schools' capacity, among others, is a complex task. Among many attributes that a designer of such an assignment algorithm would like to satisfy in the context of affirmative action is for the algorithm to exhibit the expected relationship to a change in the number of reserved seats. This attribute, which understandably requires that minority student benefit from an increase in the number of reserved seats instead of being harmed by it, is surprisingly difficult to satisfy under all circumstances.

In the matching theory and market design literature Abdulkadiroğlu and Sönmez (2003) adapted the famous Student-Optimal Deferred Acceptance (DA) rule (Gale & Shapley, 1962) to accommodate so-called 'type-specific' quotas which ensure the representation of different groups of students at schools. The DA rule has several desirable properties when school seats are not controlled for types (such as minorities), namely stability or no justified envy, student-optimality, and strategyproofness. This first simple adaptation to allow for affirmative action and other priority treatment policies turned out to be problematic, as it may lead to wasting valuable school seats that are desired by non-prioritized students, making the implementation of affirmative action policies even costlier than they need be. This was shown by Hafalir et al. (2013) who propose an effective improvement over the initial approach, by allowing for the reserved seats to be occupied by majority students when minority students do not claim them.

However, in a perplexing paper Kojima (2012) demonstrates that the DA rule with an affirmative action policy, among many others, may lead to rather inferior results for the minority students. Namely, Kojima (2012) shows that when increasing the quota for minorities it is possible that no minority student gains from the stronger affirmative action policy and some may become worse off. Kojima's findings apply to most well-known matching rules, including the DA-based flexible reserve system that Hafalir et al. (2013) proposed. Most well-known matching rules with affirmative action are not minimally responsive, which means that an increase in affirmative action policy may have such an adverse impact that it may not help any minority students and even hurt some, a somewhat startling phenomenon. More specifically, minimal responsiveness requires that when the affirmative action policy is strengthened (e.g., reserves are increassed) the matching prescribed by the rule will have to improve at least one minority student's welfare, if there is any impact, no matter what the student preferences and the original school priorities may be.

In order to improve student welfare, the flexible quota system of Hafalir et al. (2013), which they call 'minority reserves,' make sure that if there are not enough minority students to fill-up the reserves then the seats can be assigned to majority students. This contrasts with the previously studied first approach to set-aside seats which consisted of rigid quotas that could lead to empty school seats that majority students could not be assigned to. The reserve type quota successfully addresses some of the inherent inefficiencies of the previously studied more rigid quota policies, but the new flexible rules of Hafalir et al. (2013) are still subject to the criticism of Kojima (2012), as their proposed rule with reserves is not minimally responsive either. Doğan (2016) addresses this issue by exploring minority reserve policies and minimal responsiveness, and presents a new algorithm that satisfies minimal responsiveness, along with some further negative results. Chaudhury and Pápai (2022) also study minimal responsiveness, and show that all the known matching rules with an affirmative action policy fail to satisfy some basic welfare criteria simultaneously, one of which is minimal responsiveness.

The current study contributes to this literature by introducing several axioms that are related to minimal responsiveness, and by studying mechanisms that satisfy more stringent criteria than minimal responsiveness for the first time. The strongest responsiveness criterion requires that when reserves are increased the minority students all gain, or at least none of them are harmed by the increase in affirmative action policy, while other axioms focus on the comparison between no affirmative action versus having some affirmative action policy. We propose two new intuitive matching rules that satisfy some of these stronger responsiveness axioms and analyze their properties, which suggest trade-offs.

3.2 School Choice with Minority Reserves

There is a finite set of students S and a finite set of schools C. Each school $c \in C$ has a a capacity $q_c \geqslant 1$, which is the maximum number of students that can be assigned to school c. A capacity list for all schools is denoted by $q = (q_c)_{c \in C}$. Each student $i \in S$ has a strict preference ordering P_i over $C \cup \{0\}$. If i is matched to 0 then student i is unmatched, which can also be interpreted as student i's outside option. Any school c that is ranked below 0 by student i is considered an unacceptable schools for i. Let R_i denote the weak preference relation for student i. Thus, given that preferences are strict, cR_ic' implies that either cP_ic' or $c = c'$. Let a preference profile for students be denoted by $P \equiv (P_i)_{i \in N}$, and let $\mathcal{P}$ be the set of all preference profiles. Each school $c \in C$ has a priority ordering $\succ_c$ over S. School priorities are assumed to be strict. We also assume that all students are acceptable to each school. Let $\succ = (\succ_c)_{c \in C}$ denote a priority profile, and let Π denote the set of all priority profiles consisting of strict priorities.

The set of students is divided into **majority students** S^M and **minority students** S^m, that is $S^M \cap S^m = \emptyset$ and $S^M \cup S^m = S$. There are **reserved seats** set aside for minority students at each school. Let the number of reserved seats at school $c \in C$ be denoted by $r_c \geqslant 0$, where $r_c \leqslant q_c$. We will denote a **reserve policy** by $r = (r_c)_{c \in C}$, which is just a list of the number of reserved seats for schools. The list $r = (0, \ldots, 0) = \mathbf{0}$ indicates that there is no affirmative action policy.

A **matching** μ is a function from the set of students to the set of schools and the outside option such that each school is assigned no more than q_c students. With a slight abuse of notation, for all $i \in S$ we denote by μ_i the assignment that student i receives, such that $\mu_i \in C \cup \{0\}$. For all $c \in C$, we denote by μ_c the set of students assigned to school c. Thus, $\mu_c \subseteq S$, and $|\mu_c| \leqslant q_c$.

A school choice problem with a reserve policy is given by $(S, C, q, r, P, \succ)$. In order to simplify the notation we assume that S, C and q and are fixed, and let r, P and $\succ$ vary only. Then a school choice problem with a reserve policy simplifies to $(r, P, \succ)$, which we will refer to simply as a **profile.** A **matching rule** φ assigns a matching to each profile, and we denote the matching for a given profile by $\varphi(r, P, \succ)$. For all $i \in S$, let $\varphi_i(r, P, \succ)$ denote agent i's assignment in $\varphi(r, P, \succ)$, and for all subsets of agents $T \subset S$, let $\varphi_T(r, P, \succ)$

denote the assignments of agents in T in matching $\varphi(r, P, \succ)$.

We introduce next some efficiency notions in order to be able to analyze the welfare properties of matching rules. A matching μ is **wasteful** at preference profile P if there exists a student $i \in N$ and a school $c \in C$ such that $c\ P_i\ \mu_i$ and $|\mu_c| < q_c$. That is, if there is an available school seat that is preferred by a student to her assignment then the matching is wasteful. A matching is **non-wasteful** if it is not wasteful. A matching rule is wasteful if it assigns a wasteful matching at P to at least one profile $(r, P, \succ)$, and a matching rule is non-wasteful otherwise.

Matching μ **Pareto-dominates** matching ν at preference profile P if for all $i \in S$, $\mu_i R_i \nu_i$, and there exists $j \in S$ such that $\mu_j P_j \nu_j$. Matching μ is **Pareto-efficient** at P if it is not Pareto-dominated by any other matching at P. Note that a Pareto-efficient matching is non-wasteful. A rule φ is Pareto-efficient if it assigns a Pareto-efficient matching to each profile $(r, P, \succ)$. Matching μ **weakly Pareto-dominates** matching ν at preference profile P if either μ Pareto-dominates ν at P or $\mu = \nu$. We will also say that matching ν is (weakly) Pareto-dominated by μ and that μ is a (weak) Pareto-improvement over ν. A rule φ Pareto-dominates another rule ψ if for all profiles $(r, P, \succ)$, $\varphi\,(r, P, \succ)$ weakly Pareto-dominates $\psi\,(r, P, \succ)$, and at least for one profile $(\hat{r}, \hat{P}, \hat{\succ})$, $\varphi\,(\hat{r}, \hat{P}, \hat{\succ}) \neq \psi\,(\hat{r}, \hat{P}, \hat{\succ})$.

Given a preference profile P, a priority profile $\succ$, a matching μ and a student-school pair (i, c), the **priority** of student $i \in S$ is **violated** (at school $c \in C$) at profile $(r, P, \succ)$ if i prefers c to μ_i and there exists a student $j \in S$ who is assigned to school c and has a lower priority at c than i, that is, if we have $c\ P_i\ \mu_i$, $j \in \mu_c$ and $i \succ_c j$. This is also commonly known as having justified envy. A **matching** is **fair** at a profile if no student's priority is violated at that profile, and a **matching rule** φ is **fair** if, for all $(r, P, \succ)$, $\varphi\,(r, P, \succ)$ is fair at $(r, P, \succ)$.

3.3 Responsiveness Axioms

Now we define the key axioms of the paper, which are related to the property of *respecting the spirit of quota-based affirmative action* introduced by Kojima (2012), also called *minimal responsiveness* by Doğan (2016). We define three more responsiveness axioms besides minimal responsiveness, all of which concern the welfare improvement of minority students when implementing a stronger affirmative action policy, that is, when increasing the number of reserved seats. Specifically, an affirmative action policy is deemed to be stronger than another one if at least one school has strictly more reserved seats for minority students and no other school has fewer reserved seats when compared to the other policy. Formally, we will use the following definition: a minority reserve policy $r = (r_c)_{c \in C}$

represents a *weakly stronger* affirmative action policy than another minority reserve policy $r' = (r'_c)_{c\in C}$ if for all schools $c \in C$, $r_c \geqslant r'_c$, which we denote by $r \geqslant r'$.

We distinguish among four axioms of responsiveness which are based on comparisons of the welfare implications for minority students when changing affirmative action policies. In two axioms we compare an affirmative action policy to no affirmative action policy (elemental and minimal elemental responsiveness), and in the other two we compare two arbitrary affirmative action policies where one is weakly stronger than the other (responsiveness and minimal responsiveness). On the other hand, a different welfare criterion is applied in the responsiveness and elemental responsiveness axioms when comparing different affirmative action policies, namely, a weak Pareto-improvement of the outcome for minority students is required, as opposed to merely not receiving a Pareto-dominated outcome for minority students in the axioms of minimal responsiveness and minimal elemental responsiveness. Therefore, in our terminology "elemental" refers to comparing to no affirmative action policy, and "minimal" refers to the weaker of the two welfare comparisons for minority students. The formal definitions are presented next.

Responsiveness. A rule φ is **responsive** if a weakly stronger affirmative action policy weakly Pareto-dominates the outcome of the initial affirmative action policy for minority students. That is, for all $i \in S^m$, $P \in \mathcal{P}$, and r, r' such that $r \geqslant r'$, $\varphi_i(r, P, \succ)\ R_i\ \varphi_i(r', P)$.

Elemental responsiveness. A rule φ is **elementally responsive** if an affirmative action policy weakly Pareto-dominates the outcome for minority students when there is no affirmative action policy. That is, for all $i \in S^m$, $P \in \mathcal{P}$ and r, $\varphi_i(r, P, \succ)\ R_i\ \varphi_i(\mathbf{0}, P)$.

Minimal responsiveness. A rule φ is **minimally responsive** if a weakly stronger affirmative action policy never results in a Pareto-dominated outcome for minority students. That is, for all $P \in \mathcal{P}$, and r, r' such that $r \geqslant r'$, if $\varphi_{S^m}(r, P, \succ) \neq \varphi_{S^m}(r', P)$ then there exists $i \in S^m$ such that $\varphi_i(r, P, \succ)\ P_i\ \varphi_i(r', P)$.

Minimal elemental responsiveness. A rule φ is **minimally elementally responsive** if an affirmative action policy never results in a Pareto-dominated outcome for minority students when compared to no affirmative action. That is, for all $P \in \mathcal{P}$ and r, if $\varphi_{S^m}(r, P, \succ) \neq \varphi_{S^m}(\mathbf{0}, P)$ then there exists $i \in S^m$ such that $\varphi_i(r, P, \succ)\ P_i\ \varphi_i(\mathbf{0}, P)$.

Note that minimal responsiveness in our terminology is the same as respecting the spirit of quota-based affirmative action in Kojima (2012) and the minimal responsiveness of Doğan (2016). Observe that responsiveness is the strongest axiom overall and implies the other three, while minimal elemental responsiveness is the weakest axiom and is implied by the other three. The logical relationships among the four responsiveness axioms are depicted in Figure 1.

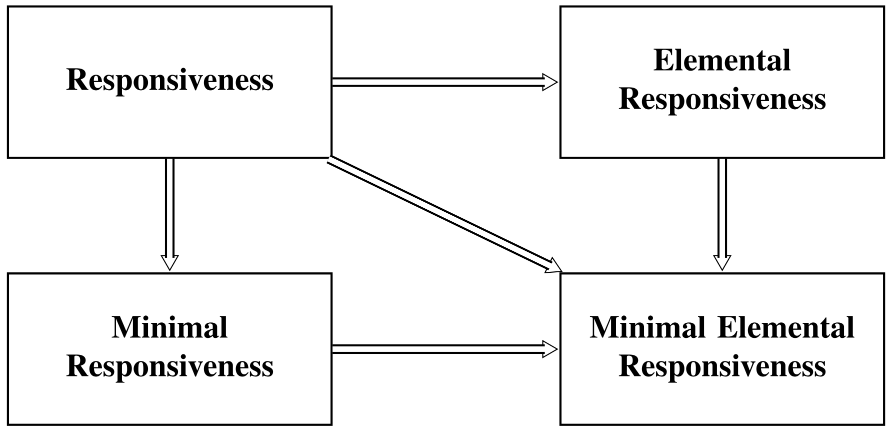

Figure 1: Logical relationships among responsiveness axioms

We provide examples of matching rules satisfying the different feasible combinations of these four axioms in the Appendix.

3.4 Minimally Responsive Rules

The Deferred Acceptance rule, introduced by Gale and Shapley (1962), is central to our analysis and is the basis for all the rules with affirmative action that we study. Given a profile $(r, P, \succ)$, we describe below the student-proposing Deferred Acceptance rule, which we will refer to from now on simply as the DA.

Deferred Acceptance (DA) rule

Fix a preference profile and a priority profile $(P, \succ)$.

Round 1: Every student $i \in S$ applies to her most preferred school. Each school $c \in C$ tentatively assigns seats to applying students up to its capacity q_c, following its priority ordering $\succ_c$. If the capacity is exhausted, all remaining applicants are rejected.

Round *k*$(k \geqslant 2)$: Every student $i \in S$ who was rejected in round $k-1$ applies to her next most preferred school. Each school $c \in C$ considers its tentatively assigned students from the previous round along with the new applicants, and tentatively assigns seats to these students up to its capacity q_c, following its priority ordering $\succ_c$. If the capacity is exhausted, all remaining students who are considered in this round are rejected.

The process terminates when no new applicant is rejected. The tentative assignments in the last round become final. Note that a student may remain unassigned if the student had applied to all acceptable schools on her preference list and was rejected by each. $\diamond$

The DA with Minority Reserves (DA-R) rule introduced by Hafalir et al. (2013) incorporates a reserve-based affirmative action policy in the DA rule. The DA-R rule is the foundation for minimally responsive rules to be discussed in this section, although DA-R itself is not minimally responsive. We introduce this rule next.

DA with Minority Reserves (DA-R) rule

Fix a profile $(r, P, \succ)$.

Round 1: Every student $i \in S$ applies to her most preferred school. Each school $c \in C$ first tentatively accepts as many as r_c minority students $i \in S^m$ with the highest priorities according to $\succ_c$. Then each school $c \in C$ tentatively accepts students $i \in S$ among the remaining applicants with the highest priorities according to schools priority $\succ_c$ until either its capacity is filled or the applicant set is exhausted. The rest of the applicants, if any remain, are rejected.

Round *k* $(k \geqslant 2)$**:** Every student $i \in S$ who was rejected in round $k-1$ applies to her next most preferred school. Each school $c \in C$ considers its tentatively assigned students from the previous round along with the new applicants (the *applicant set*) and first tentatively accepts as many as r_c minority students $i \in S^m$ with highest priorities according to $\succ_c$. Then each school $c \in C$ tentatively accepts students $i \in S$ among the remaining applicants according to schools priority $\succ_c$ until either its capacity is filled or the applicant set is exhausted. The rest of the applicants, if any remain, are rejected.

The process terminates when there are no more rejections and the tentative matching at that round is finalized. $\diamond$

The DA-R is not minimally responsive, as noted by Hafalir et al. (2013). Doğan (2016) demonstrates that this rule allows some minority students to "interfere" with the assignments without gaining from this interference, which causes the inefficiencies responsible for violating minimal responsiveness. Inspired by EADAM (Kesten, 2010), Doğan (2016) proposes to modify the DA-R rule in order to get a minimally responsive rule which he calls the **Modified DA with Minority Reserves (MDA) rule.** This rule is based on detecting so-called interfering minority students in the matching process. Doğan (2016) calls a minority student i an *interferer* at school c in round t of the DA-R rule if in round t the minority student i with a lower priority for school c than a majority student is temporarily accepted while the majority student is rejected from c, and in some later round $t' > t$ student i is also rejected from school c. MDA runs the DA-R algorithm iteratively, with the modification that it finds the last round of DA-R in which there is an interferer and it treats these minority students as majority students for the schools at which they interfere. The process is repeated until there are no more interferers.

Other minimally responsive rules are the **IA-R** (Afacan and Salman (2016), Doğan and Klaus (2018)) and the **IA-DA-R** (Chaudhury & Pápai, 2022) rules. The IA-R rule is similar to DA-R, but instead of deferred acceptances where temporarily accepted students may be rejected by a school if a higher-priority student applies to the school in a later round, the acceptances are final, as in the Immediate Acceptance rule. In the IA-DA-R rule minority students are accepted permanently for reserved seats but are only accepted temporarily for unreserved seats, while majority students can only be accepted temporarily either for reserved or unreserved seats, and are only assigned temporarily to reserved seats when there are no minority applicants for the reserved seats. It is interesting to note that without affirmative action the IA-R rule simplifies to the Immediate Acceptance rule, while the IA-DA-R rule becomes the DA rule.

We show below that the MDA and IA-R rules are not elementally responsive, and thus are not responsive. The IA-DA-R rule is also not elementally responsive, as demonstrated by Example 6 in Chaudhury and Pápai (2022), since in that example one minority student is worse off when we compare some affirmative action policy to no affirmative action policy.

Example 8 (The MDA rule is not elementally responsive)**.** Let $S^M = \{1, 2\}$ and $S^m = \{3, 4, 5, 6\}$ be the sets of majority and minority students. Let $C = \{c_1, c_2, c_3, c_4\}$ be the set of schools with capacity list $q = (2, 2, 1, 1)$. Consider the preference and priority profiles in Table 3.1, and compare reserve policies $r = (0, 0, 0, 0)$ and $r' = (1, 0, 0, 0)$.

Table 3.1: Preference and priority profiles for Example 1

P_1	P_2	P_3	P_4	P_5	P_6
$\underline{\underline{\boxed{c_1}}}$	$\underline{c_1}$	$\boxed{c_1}$	c_1	$\underline{\underline{\boxed{c_3}}}$	$\underline{c_4}$
c_3	$\boxed{c_4}$	$\underline{c_2}$	$\underline{\underline{\boxed{c_2}}}$	c_1	c_1
c_2	c_3	c_3	c_4	c_4	c_3
c_4	c_2	c_4	c_3	c_2	$\boxed{c_2}$

$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$	$\succ_{c_4}$
1	5	1	1
2	6	2	2
3	1	5	6
6	2	6	5
5	3	3	3
4	4	4	4

With reserve policy r (no affirmative action), the MDA matching, which is just the DA matching at this profile, is given by $\mu = (c_1, c_1, c_2, c_2, c_3, c_4)$[1] (underlined in the table).

[1]Henceforth, assignments in the examples will be specified by a list of the schools which indicates the assignments for students $1, 2, 3, \ldots$, respectively.

With reserve policy r' the MDA matching becomes $\mu = (c_1, c_4, c_1, c_2, c_3, c_2)$ (indicated by squares in the table). This is just the DA-R matching, since there are no interferers at this profile. While minority student 3 is better off with reserve policy r', minority student 6 is worse off with r' than with no affirmative action. Thus, the MDA rule is not elementally responsive. $\diamond$

Example 9 (The IA-R rule is not elementally responsive)**.** Let $S^M = \{1\}$ and $S^m = \{2, 3\}$ be the sets of majority and minority students. Let $C = \{c_1, c_2, c_3\}$ be the set of schools with capacity list $q = (1, 1, 1)$. Consider the preference and priority profiles in Table 3.2, and compare reserve policies $r = (0, 0, 0)$ and $r' = (1, 0, 0)$.

Table 3.2: Preference and priority profiles for Example 9

P_1	P_2	P_3	$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$
$\underline{c_1}$	$\boxed{c_1}$	c_1	1	2	1
$\boxed{c_3}$	$\underline{c_2}$	$\underline{c_3}$	2	1	3
0	0	$\boxed{c_2}$	3	3	2

With reserve policy r (no affirmative action), the IA-R matching at this profile is given by $\mu = (c_1, c_2, c_3)$ (underlined in the table). With reserve policy r', the IA-R matching becomes $\mu' = (c_3, c_1, c_2)$ (indicated by squares in the table). While minority student 2 benefits from the affirmative action policy r', minority student 3 is worse off with r' than with no affirmative action. Thus, the IA-R rule is not elementally responsive. $\diamond$

3.5 Divided DA Rule

We are now ready to introduce a new matching rule which has a responsive affirmative action policy based on reserved seats set aside for minority students. This new matching rule divides the matching procedure into two steps: in the first step the school assignment for minority students is determined based on the *minority seats*, which we will define next, and in the second step the majority students receive their school assignments. We call this new rule the **Divided Deferred Acceptance** (**Divided DA,** for short) rule.

The Divided DA rule relies on *minority seats*, denoted by $\hat{m}_c$ for school c. Fix a profile $(r, P, \succ)$ and let the DA matching at this profile be denoted by $\mu^{\text{DA}}(P, \succ)$. Note that the DA matching does not depend on r, since it is the standard DA outcome when all school seats are available and all students participate in the matching, regardless of reserved seats and the minority status of students. Let $m_c^{DA}(P, \succ)$ denote the number of minority students assigned to school c in $\mu^{DA}(P, \succ)$.

Minority seats. The number of minority seats at each school $c \in C$ is

$$\hat{m}_c(r, P, \succ) = \max\big(r_c,\, m_c^{DA}(P, \succ)\big).$$

For each school the number of minority seats at a given profile $(r, P, \succ)$ is either the number of reserved seats r_c at school c or, if the number of seats given to minority students in $\mu^{DA}(P)$ at school c exceeds the number of reserved seats at the school, then this larger number is the number of minority seats. Note that the number of minority seats cannot be less than the number of reserved seats at any school, and that this number depends on the preference and priority profiles $(P, \succ)$.

Divided DA rule ($\varphi^{\mathbf{DDA}}$)

Fix a profile $(r, P, \succ)$.

Step 1 - Minority matching: Run the DA rule with minority students S^m only, while restricting the number of seats at each school $c \in C$ to the number of minority seats $\hat{m}_c(r, P, \succ)$. For each school c, let μ_c^m denote the set of minority students matched to c in this step.

Step 2 - Majority matching: Run the DA rule with majority students S^M for the remaining seats (the ones not assigned in Step 1) at each school. Formally, the number of seats available to majority students is $\bar{q}_c = q_c - |\mu_c^m|$ for all $c \in C$. For each school c, let μ_c^M denote the set of majority students matched to school c in this step.

In both steps the matching is final, and the set of students matched to school $c \in C$ is $\mu_c = \mu_c^m \cup \mu_c^M$ $\diamond$

Therefore, for all profiles $(r, P, \succ)$, and for all $i \in S^m$, $\varphi_i^{\text{DDA}}(r, P, \succ) = c$ if $i \in \mu_c^m(r, P, \succ)$ and for all $i \in S^M$, $\varphi_i^{\text{DDA}}(r, P, \succ) = c$ if $i \in \mu_c^M(r, P, \succ)$, where $\mu^m(r, P, \succ)$ denotes the minority matching obtained in Step 1 of the Divided DA rule at $(r, P, \succ)$ and $\mu^M(r, P, \succ)$ denotes the majority matching obtained in Step 2 of the Divided DA rule at $(r, P, \succ)$.

Example 10 (Illustration of the Divided DA rule)**.** This example demonstrates how the Divided DA rule works, using the setup of Example 8. Now consider the profile $(r', P, \succ)$ with reserve policy $r' = (1, 0, 0, 0)$. The DA matching is $\mu^{\text{DA}} = (c_1, c_1, c_2, c_2, c_3, c_4)$ for this profile. Thus, the DA list for minority students is $m^{DA}(P, \succ) = (0, 2, 1, 1)$ and the minority seat list is $\hat{m}_c(r, P, \succ) = (1, 2, 1, 1)$.

The first step of the Divided DA rule is to run the DA with the set of minority students $\{3, 4, 5, 6\}$ only, using $\hat{m}_c(r, P, \succ)$ as the capacity list. This yields the minority matching $\mu_3^m = \{c_1\}$, $\mu_4^m = \{c_2\}$, $\mu_5^m = \{c_3\}$, and $\mu_6^m = \{c_4\}$. The second step of the Divided DA rule is to run the DA rule on the remaining capacity of the schools, $\bar{q}_c = (1, 1, 0, 0)$, with the set of majority students $\{1, 2\}$ only. This step yields the majority matching $\mu_1^M = \{c_1\}$, $\mu_2^M = \{c_2\}$. The final matching of the Divided DA rule is $\mu_c = \mu_c^m \cup \mu_c^M$ for all $c \in C$, which gives the matching $(c_1, c_2, c_1, c_2, c_3, c_4)$. ◇

Note that although the number of minority seats is minimally the number of reserved seats and it may exceed it, minority seats should not be viewed as extending the set of reserved seats that are set aside for minority students. We interpret minority seats as the number of seats that should be offered to minority students, since it would not be fair to offer only reserved seats to minority students if there are too few reserved seats, or if some minority students could get higher-ranked assignments in the DA without minority reserves. This is illustrated by Example 10, where there are four minority students but only one reserved seat in total, and thus offering only the reserved seat to minority students would prevent three minority students from being assigned to a school, even though there are enough school seats for all students. Furthermore, even if there were enough reserved seats for minority students in total, the distribution of reserved seats matters. For instance, if some school c has no reserved seats but the standard DA rule assigns minority students to this school, then the Divided DA rule should not be prevented from assigning minority students to school c just because there are too few reserved seats at c. This would be the case for schools c_2, c_3, and c_4 in Example 10. On the other hand, the restriction of not exceeding the minority seats (as opposed to the reserves) at each school ensures that the Divided DA rule does not give full priority to minority students, thereby preserving the rights of majority students. This can be seen in Example 8 as well, since minority student 4 would prefer to get into school c_1 compared to her assignment, but majority student 1 is matched to c_1.

A simple observation on the Divided DA without affirmative action is presented in the following example.

Example 11 (The Divided DA outcome without affirmative action is not necessarily the DA outcome)**.** Let the set of schools be $C = \{c_1, c_2, c_3\}$ with capacity $q_c = 1$ for each $c \in C$. Let the set of students be $S = \{1, 2, 3\}$ and consider the preference and priority profiles in Table 3.3. Then $\varphi^{\text{DA}}(P) = (c_1, c_2, c_3)$ (underlined in the table). If we let $S^M = \{3\}$ and

$S^m = \{1, 2\}$ then $\varphi^{\text{DDA}}(\mathbf{0}, P) = (c_2, c_1, c_3)$ (indicated by squares in the table). Note that $\varphi^{\text{DDA}}(\mathbf{0}, P)$ remains the same if $S^M = \{1, 2\}$ and $S^m = \{3\}$, since in both cases only the priority of 3 is violated at school c_1 in the Divided DA matching. This is, however, ruled out by the DA. $\diamond$

Table 3.3: Preference and priority profiles for Example 11

P_1	P_2	P_3	$\succ_{c_1}$	$\succ_{c_2}$
$\boxed{c_2}$	$\boxed{c_1}$	c_1	1	2
$\underline{c_1}$	$\underline{c_2}$	$\boxed{\underline{c_3}}$	3	$\vdots$
			2	

In accordance with the example, we can show in general that for any preference profile each student is at least as well off under the Divided DA rule without affirmative action as under the DA rule, or in other words the Divided DA rule without affirmative action weakly Pareto-dominates the DA rule. The example demonstrates that the DA rule does not satisfy the standard consistency axiom of Thomson (2011), since students 1 and 2 together get the same assignments in both $\varphi^{DA}(P)$ and $\varphi^{DDA}(\mathbf{0}, P)$. However, a weaker axiom called *weak consistency*, which requires that for any subset of students who are assigned the same seats collectively, each student weakly prefers her assignment in this reduced problem to their initial assignment, is satisfied by the DA rule. This property, which is closely related to the observation below, was shown by Chen (2017).

Observation 1. *For all preference profiles P and priority profiles $\succ$, $\varphi^{DDA}(\mathbf{0}, P, \succ)$ weakly Pareto-dominates $\varphi^{DA}(P, \succ)$.*

Proof. Fix P and $\succ$. Given the definition of minority seats, for each school $c \in C$, $\hat{m}_c(\mathbf{0}, P, \succ) = m_c^{DA}(P, \succ)$.

Step a: Since the DA is a fair rule (Gale and Shapley, 1962), there are no priority violations for $\varphi^{DA}(P, \succ)$, and hence $\varphi^{DA}_{S^m}(P, \succ)$ is also a fair matching when we restrict attention to S^m. Then, since the DA rule is optimal with respect to fairness (Gale and Shapley, 1962), the minority matching step (Step 1) of the Divided DA rule implies that for all $i \in S^m$, $\varphi_i^{DDA}(\mathbf{0}, P, \succ)\ R_i\ \varphi_i^{DA}(P, \succ)$. Note that the minority students are assigned collectively the same number of school seats at each school in $\varphi^{DDA}(\mathbf{0}, P, \succ)$ as in $\varphi^{DA}(P, \succ)$.

Step b: If we remove the minority students and the seats assigned to them in these matchings, this leaves the majority students the same set of school seats available under these two scenarios. Since there are no priority violations at $\varphi^{DA}(P, \succ)$, $\varphi^{DA}_{S^M}(P, \succ)$ is a fair matching when we restrict attention to S^M. Thus, by the optimality of the DA rule with respect to fairness, the majority matching step (Step 2) of the Divided DA rule implies that for all $i \in S^M$, $\varphi_i^{DDA}(\mathbf{0}, P, \succ)\ R_i\ \varphi^{DA}(P, \succ)$.

Alternatively, given that for all $c \in C$, $\hat{m}_c(\mathbf{0}, P, \succ) = m_c^{DA}(P, \succ)$, and that the minority students are assigned collectively the same number of school seats at each school in $\varphi^{DDA}(\mathbf{0}, P, \succ)$ as in $\varphi^{DA}(P, \succ)$, Step a also follows from the weak consistency property of the DA rule and, given this, so does Step b. $\square$

We show next that the Divided DA rule satisfies responsiveness, the strongest of the four responsiveness axioms introduced in Section 3.

Theorem 3. *The Divided DA rule is responsive.*

Before proving the theorem, we define *resource monotonicity*, an axiom which will be used in the proof. First we expand the argument of φ to include the capacity list q, allowing for a variable capacity list in the model, in order to define resource monotonicity which requires a variable resource environment.

A rule φ is **resource monotonic** if for any two capacity lists q and q' such that $q \geqslant q'$, for all $i \in S$ and for all profiles $(r, P, \succ)$, $\varphi_i(q, r, P)\ R_i\ \varphi_i(q', r, P)$.

Resource monotonicity is a solidarity axiom which requires that all agents weakly benefit from increasing the resources. In our context this means that if more school seats become available (i.e., we have at least as many seats at each school as before) then each student gets either the same or a better school assignment than before. It is well known that the DA rule is resource monotonic (Chambers and Yenmez (2017); see also Ehlers and Klaus (2016)).

Proof of Theorem 3: Fix P and $\succ$, and let r, r' be two reserve policies such that $r \geqslant r'$. Then, given that the number of minority seats is $\hat{m}_c(r, P, \succ) = \max\ (r_c,\ m_c^{DA}(P, \succ))$ at school $c \in C$ when the reserve policy is r, and $\hat{m}_c(r', P, \succ) = \max\ (r'_c,\ m_c^{DA}(P, \succ))$ when the reserve policy is r', we have $\hat{m}_c(r, P, \succ) \geqslant \hat{m}_c(r', P, \succ)$. Then the resource monotonicity of the DA rule implies that in Step 1 of the Divided DA algorithm φ^{DDA} the minority students are weakly better off at r compared to r', that is, for all $i \in S^m$, $\varphi_i^{\text{DDA}}(r, P, \succ)\ R_i\ \varphi_i^{\text{DDA}}(r', P, \succ)$. This means that the Divided DA rule is responsive. $\square$

Example 8 shows that the MDA rule is not elementally responsive. Using the same preference and priority profiles we expand this example below to illustrate that the Divided DA rule is responsive at the given preference and priority profiles, based on comparisons among different reserve policies.

Example 12 (Illustration that the Divided DA rule is responsive)**.** Consider the setup in Example 8. In addition to $r^0 = (0, 0, 0, 0)$ and $r^1 = (1, 0, 0, 0)$, consider also the reserve policy $r^2 = (2, 0, 0, 0)$.

Let MDA^t refer to the MDA matching when the reserve policy is r^t for $t = 0, 1, 2$. Similarly, let DDA^t refer to the Divided DA matchings when the reserve policy is r^t for $t = 0, 1, 2$. The preference ranking of the corresponding matching for each agent is displayed in Table 3.4.

Table 3.4: Preference rankings of matching outcomes in Example 12

i	MDA^0	MDA^1	MDA^2	DDA^0	DDA^1	DDA^2
1	1	1	1	1	1	3
2	1	2	3	1	4	4
3	2	1	1	2	1	1
4	2	2	2	2	2	1
5	1	1	4	1	1	1
6	1	4	1	1	1	1

As we can see from Table 3.4, the Divided DA rule is responsive at the given priority and preference profiles based on the rank comparisons of the minority students (student 3 to 6). By contrast, the rankings show that MDA is not elementally responsive, and hence not responsive. ◇

However, the Divided DA rule turns out to be wasteful. We illustrate this with the example below.

Example 13 (The Divided DA rule is wasteful)**.** Let $S^M = \{1, 4, 5\}$ and $S^m = \{2, 3\}$ be the sets of majority and minority students. Let the set of schools be $C = \{c_1, c_2, c_3, c_4, c_5\}$ with capacity list $q = (1, 1, 1, 1, 1)$. Consider the preference and priority profiles in Table 3.5 and let the reserve policy be $r = (1, 0, 0, 0, 0)$.

The DA matching at this profile is $\varphi_i^{\text{DA}}(P, \succ) = (c_3, c_2, c_4, c_1, c_5)$. Given the DA matching and the reserve policy r, the minority seat list is $\hat{m}_c(r, P, \succ) = (1, 1, 0, 1, 0)$. Therefore, the Divided DA matching is $\varphi^{\text{DDA}}(r, P) = (c_2, c_1, c_4, c_5, 0)$ (underlined in the table). Here minority student 3 is matched to school c_4. However, this student prefers school c_3 to c_4, and school c_3 has an empty seat. Therefore, this example proves that the Divided DA rule is wasteful. ◇

Note that, as the example demonstrates, wastefulness can only occur if a minority student desires an unassigned seat. Majority students cannot experience such wastefulness,

Table 3.5: Preference and priority profiles for Example 13

P_1	P_2	P_3	P_4	P_5	$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$	$\succ_{c_4}$	$\succ_{c_5}$
$\underline{c_2}$	$\underline{c_1}$	c_3	c_1	c_1	4	2	1	3	4
c_3	c_2	$\underline{c_4}$	c_4	c_2	2	1	5	4	5
c_1	c_3	c_1	$\underline{c_5}$	c_4	5	3	3	1	3
c_4	c_4	c_2	c_3	c_5	3	5	2	2	2
c_5	c_5	c_5	c_2	$\underline{0}$	1	4	4	5	1

since all seats that are not assigned to minority students are available to them in Step 2 of the Divided DA algorithm, and the DA itself is non-wasteful. This contrasts with the wastefulness of majority quotas, which lead to majority students envying empty school seats that are set aside exclusively for minority students.

3.6 Guaranteed DA Rule

Since the Divided DA rule is wasteful, we introduce next a non-wasteful rule which is both elementally and minimally responsive. However, as we will show, this rule is not responsive.

Guaranteed DA rule ($\varphi^{\mathbf{GDA}}$)

Fix a profile $(r, P, \succ)$.

Step 1 - Determining the minority guarantees: Find the DDA matching $\varphi^{\text{DDA}}(r, P, \succ)$. The assignments made by $\varphi^{\text{DDA}}(r, P, \succ)$ will be the guaranteed minimum assignments to minority students in the matching obtained in Step 2.

Step 2 - DA matching with minority guarantees: First, modify the priorities based on the DDA matching $\varphi^{\text{DDA}}(r, P, \succ)$. For each school $c \in C$, let the set of minority students in $\varphi_c^{\text{DDA}}(r, P, \succ)$ be ranked at the top of the priority ordering of school c. That is, each minority student assigned to c in $\varphi^{\text{DDA}}(r, P, \succ)$ is among the top q_c-ranked students in the priority ordering of school c (the exact new priority rank of a minority student within the top q_c ranks is irrelevant). Let all other priorities remain the same. Denote this priority profile by $\hat{\succ}$. Then $\varphi^{\text{GDA}}(r, P, \succ) = \varphi^{\text{DA}}(P, \hat{\succ})$, that is, the GDA matching at $(r, P, \succ)$ is the DA matching at the fixed preference profile P and the modified priority profile $\hat{\succ}$, where $\hat{\succ}$ is constructed at each $(r, P, \succ)$ as described above. $\diamond$

Example 14 (Illustration of the Guaranteed DA rule)**.** Consider the setup of Example 13: $S^M = \{1, 4, 5\}$, $S^m = \{2, 3\}$, and $C = \{c_1, c_2, c_3, c_4, c_5\}$ with capacity list $q = (1, 1, 1, 1, 1)$. When the reserve policy is $r = (1, 0, 0, 0, 0)$, $\varphi_i^{\text{DDA}}(r, P, \succ) = (c_2, c_1, c_4, c_5, 0)$, as shown in Example 13 (underlined in Table 3.5). The modified priority profile $\succ^{\hat{}}$ is presented in Table 3.6.

Table 3.6: Preference profile and the modified priority profile for Example 14

.

P_1	P_2	P_3	P_4	P_5		$\succ^{\hat{}}_{c_1}$	$\succ^{\hat{}}_{c_2}$	$\succ^{\hat{}}_{c_3}$	$\succ^{\hat{}}_{c_4}$	$\succ^{\hat{}}_{c_5}$
$\boxed{\underline{c_2}}$	$\boxed{\underline{c_1}}$	$\boxed{c_3}$	c_1	c_1		**2**	2	1	**3**	4
c_3	c_2	$\underline{c_4}$	$\boxed{c_4}$	c_2		4	1	5	4	5
c_1	c_3	c_1	$\underline{c_5}$	c_4		5	3	3	1	3
c_4	c_4	c_2	c_3	$\boxed{c_5}$		3	5	2	2	2
c_5	c_5	c_5	c_2	$\underline{0}$		1	4	4	5	1

In order to find the Guaranteed DA matching, we run the DA with the modified priority profile $\succ^{\hat{}}$. This yields $\varphi_i^{\text{GDA}}(r, P, \succ) = (c_2, c_1, c_3, c_4)$ (indicated by squares in the table). Note that the Guaranteed DA matching is non-wasteful, correcting the wastefulness of the Divided DA matching at the same profile. $\diamond$

Since the DA rule is non-wasteful, the Guaranteed DA rule is also non-wasteful by construction.

Proposition 8. *The Guaranteed DA rule is non-wasteful.*

Next, we present a series of observations on the Guaranteed DA rule.

Observation 2. *For all profiles, the Guaranteed DA assignment weakly Pareto-dominates the Divided DA assignment for minority students: that is for all $i \in S^m$ and for all profiles $(r, P, \succ)$, $\varphi_i^{GDA}(r, P, \succ)\ R_i\ \varphi_i^{DDA}(r, P, \succ)$.*

Proof. Given that in the DA algorithm a student whose priority rank is among the top q_c students in $\succ_c$ is never rejected by c, the Divided DA assignment is indeed "guaranteed" for minority students in the sense that a minority student cannot get a less preferred assignment in the Guaranteed DA matching than her Divided DA matching at the same profile. Moreover, if a minority student $i \in S^m$ is unassigned in $\varphi^{\text{DDA}}(r, P, \succ)$, then the individual rationality of the Guaranteed DA rule implies that if $\varphi_i^{\text{GDA}}(r, P, \succ) \neq 0$ then $\varphi_i^{\text{GDA}}(r, P, \succ)$ is an acceptable school to s, that is $\varphi_i^{\text{GDA}}(r, P, \succ)\ P_i\ 0$. $\square$

Observation 3. *For all preference profiles P and priority profiles $\succ$, $\varphi^{GDA}(\mathbf{0}, P, \succ)$ weakly Pareto-dominates $\varphi^{DA}(P, \succ)$ for minority students: for all $i \in S^m$ and for all $(P, \succ)$, $\varphi_i^{GDA}(\mathbf{0}, P, \succ)\ R_i\ \varphi_i^{DA}(P, \succ)$.*

Proof. Observation 1 and Observation 2 together imply the statement. $\square$

Observation 4. *Without an affirmative action policy, the Divided DA and Guaranteed DA matchings are the same for minority students: for all preference profiles p and priority profiles $\succ$, $\varphi_{S^m}^{DDA}(\mathbf{0}, P, \succ) = \varphi_{S^m}^{GDA}(\mathbf{0}, P, \succ)$.*

Proof. Let $\hat{\succ}$ denote the modified priority profile at $(r, P, \succ)$, as specified in the definition of the Guaranteed DA rule, so that $\varphi^{\mathrm{GDA}}(r, P, \succ) = \varphi^{\mathrm{DA}}(P, \hat{\succ})$. Since the DA rule satisfies weak consistency, for all $i \in S^m$, $\varphi_i^{\mathrm{DDA}}(\mathbf{0}, P, \hat{\succ})\ R_i\ \varphi_i^{\mathrm{DA}}(P, \hat{\succ})$. Then Observation 2 implies the result. $\square$

Observation 5. *Even if $\varphi^{DDA}(r, P, \succ)$ is non-wasteful at profile $(r, P, \succ)$, it is possible that $\varphi^{GDA}(r, P, \succ)$ Pareto-dominates $\varphi^{DDA}(r, P, \succ)$ for the minority students.*

Proof. See Example 15. $\square$

Example 15 (The Guaranteed DA matching may Pareto-dominate a non-wasteful Divided DA matching)**.** Let $S^M = \{1, 4\}$ and $S^m = \{2, 3\}$ be the sets of majority and minority students. Let the set of schools be $C = \{c_1, c_2, c_3, c_4\}$ with capacity list $q = (1, 1, 1, 1)$. Consider the preference profile and priority profile $\succ$ in Table 3.7, and let the minority policy be $r = (0, 0, 1, 0)$.

Table 3.7: Preference and priority profiles for Example 15

P_1	P_2	P_3	P_4	$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$	$\succ_{c_4}$	$\hat{\succ}_{c_1}$	$\hat{\succ}_{c_2}$	$\hat{\succ}_{c_3}$	$\hat{\succ}_{c_4}$
c_3	$\boxed{c_4}$	$\boxed{\underline{c_3}}$	$\boxed{c_1}$	3	2	4	4	**2**	2	**3**	4
$\boxed{\underline{c_2}}$	$\underline{c_1}$	c_1	$\underline{c_4}$	2	1	1	2	3	1	4	2
0	c_2	c_2	c_3	1	3	2	1	1	3	1	1
	c_3	0	c_2	4	4	3	3	4	4	2	3

The DA matching at this profile is $\varphi_i^{\mathrm{DA}}(P) = (c_3, c_2, c_1, c_4)$. Given the DA matching and reserve policy r, the minority seat list is $\hat{m}_c(r, P, \succ) = (1, 1, 1, 0)$. Therefore, the Divided DA matching is $\varphi_i^{\mathrm{DDA}}(r, P) = (c_2, c_1, c_3, c_4)$ (underlined in the table). The modified priority profile $\hat{\succ}$ is presented in Table 3.7.

In order to find the Guaranteed DA matching, we run the DA with the modified priority profile $\hat{\succ}$. This yields $\varphi_i^{\text{GDA}}(r, P, \succ) = (c_2, c_4, c_3, c_1)$ (indicated by squares in the table). Thus, the Guaranteed DA matching Pareto-dominates the Divided DA matching for the minority students at this profile, even though the Divided DA matching is non-wasteful. $\diamond$

Theorem 4. *The Guaranteed DA rule is both elementally responsive and minimally responsive. However, it is not responsive.*

Proof. See Example 16 for the proof that the Guaranteed DA rule is not responsive. We prove the other two properties below.

Elemental Responsiveness: Fix a profile $(r, P, \succ)$. We need to show that for all $i \in S^m$, $\varphi_i^{\text{GDA}}(r, P, \succ) \; R_i \; \varphi_i^{\text{GDA}}(\mathbf{0}, P, \succ)$.

1. *For all $i \in S^m$, $\varphi_i^{GDA}(r, P, \succ) \; R_i \; \varphi_i^{DDA}(r, P, \succ)$.*

   This statement follows from Observation 2.

2. *For all $i \in S^m$, $\varphi_i^{DDA}(r, P, \succ) \; R_i \; \varphi_i^{DDA}(\mathbf{0}, P, \succ)$.*

   This statement follows from the (elemental) responsiveness of the Divided DA rule (Theorem 3).

3. *For all $i \in S^m$, $\varphi_i^{DDA}(\mathbf{0}, P, \succ) = \varphi_i^{GDA}(\mathbf{0}, P, \succ)$.*

   This statement follows from Observation 4.

Points 1-3 together imply that the Guaranteed DA rule is elementally responsive.

Minimal Responsiveness: Fix a preference profile P and a priority profile $\succ$, and let r, r' be two minority reserve policies such that $r \geqslant r'$. Then, since the Divided DA rule is responsive by Theorem 3, for all $i \in S^m$, $\varphi_i^{\text{DDA}}(r, P, \succ) \; R_i \; \varphi_i^{\text{DDA}}(r', P)$. This implies that the guaranteed seat under r for each minority student in Step 2 of the Guaranteed DA algorithm is weakly preferred by each minority student to the guaranteed seat under r'. Given the DA algorithm, this means that if a minority student is assigned to her guaranteed school in the Guaranteed DA rule, then this student cannot be worse off under r than under r'.

Let $\mu = \varphi^{\text{GDA}}(r, P, \succ)$ and $\hat{\mu} = \varphi^{\text{GDA}}(r', P, \succ)$. Assume that $\mu_{S^m} \neq \hat{\mu}_{S^m}$. Let round k be the first round in the DA algorithm that is run in the second step of the GDA algorithm at which the accepted sets of students differ at $(r, P, \succ)$ and $(r', P, \succ)$. Since rounds 1 to $k-1$ are the same, and since only minority students' priorities have increased relative to other students' priorities at their respective guaranteed schools when only the minority reserves are different, it follows that a new minority applicant $i \in S^m$ gets accepted by a school at $(r, P, \succ)$ who is rejected by this same school at $(r', P, \succ)$, and this school is the guaranteed school for i at $(r, P, \succ)$. This school is μ_i^m such that $\mu_i^m \; P_i \; \hat{\mu}_i^m$, where μ_i^m and $\hat{\mu}_i^m$ are the guaranteed schools for i at $(r, P, \succ)$ and $(r', P, \succ)$ respectively. Given that i is guaranteed to get a seat at school μ_i^m, i cannot be rejected by this school and thus $\varphi_i^{\text{GDA}}(r, P, \succ) = \mu_i^m$. Since i is rejected by school μ_i^m at $(r', P, \succ)$, this implies that $\varphi_i^{\text{GDA}}(r, P, \succ) \; P_i \; \varphi_i^{\text{GDA}}(r', P, \succ)$. Given that i is a minority student, this means that the Guaranteed DA rule is minimally responsive. $\square$

Example 16 (The Guaranteed DA rule is not responsive)**.** Let $S^M = \{1, 4, 5\}$ and $S^m = \{2, 3\}$ be the sets of majority and minority students. Let $C = \{c_1, \ldots, c_6\}$ be the set of schools with capacity list $q = (1, \ldots, 1)$. Consider the preference and priority profiles in Table 3.8. We will compare the weaker reserve policy $r = (1, 0, 0, 0, 0, 0)$ with the stronger reserve policy $\tilde{r} = (1, 0, 0, 0, 0, 1)$.

Table 3.8: Preference and priority profiles for Example 16

P_1	P_2	P_3	P_4	P_5
$\boxed{\underline{c_2}}$	$\boxed{c_6}$	$\underline{c_3}$	$\boxed{c_1}$	$\underline{c_6}$
c_3	$\underline{c_1}$	$\boxed{c_4}$	$\underline{c_4}$	$\boxed{c_3}$
c_1	c_2	c_1	c_5	c_1
c_4	c_5	c_6	c_3	c_2
c_5	c_4	c_2	c_2	c_4
c_6	c_3	c_5	c_6	c_5

$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$	$\succ_{c_4}$	$\succ_{c_5}$	$\succ_{c_6}$
4	2	5	3	4	5
2	1	1	5	2	2
3	3	3	4	1	1
1	4	2	2	3	3
5	5	4	1	5	4

The DA matching at this profile is $\mu^{\text{DA}}(P, \succ) = (c_3, c_2, c_4, c_1, c_6)$. Thus, the DA list for minority students is $m^{DA}(P, \succ) = (0, 1, 0, 1, 0, 0)$ and the minority seat lists for the two reserve policies are $\hat{m}_c(r, P, \succ) = (1, 1, 0, 1, 0, 0)$ and $\hat{m}_c(\tilde{r}, P, \succ) = (1, 1, 0, 1, 0, 1)$.

The respective Divided DA matchings are $\varphi^{\text{DDA}}(r, P, \succ) = (c_2, c_1, c_4, c_5, c_6)$ and $\varphi^{\text{DDA}}(\tilde{r}, P, \succ) = (c_2, c_6, c_4, c_1, c_3)$. The corresponding updated priority profiles, $\succ^{\hat{}}$ for r and $\succ^{\tilde{}}$ for $\tilde{r}$, are shown in Table 3.9.

Based on the updated priority profiles, the Guaranteed DA matchings that correspond to the two reserve policies are $\varphi^{\text{GDA}}(r, P, \succ) = (c_2, c_1, c_3, c_4, c_6)$ (underlined in Table 3.6) and $\varphi^{\text{GDA}}(\tilde{r}, P, \succ) = (c_2, c_6, c_4, c_1, c_3)$ (indicated by squares in Table 3.6). Note that $\varphi^{\text{GDA}}(r, P, \succ) \; P_3 \; \varphi^{\text{GDA}}(\tilde{r}, P, \succ)$, since 3 prefers c_3 to c_4. Therefore, minority student 3 is worse off under the stronger minority reserve policy $\tilde{r}$ compared with the weaker policy $\hat{r}$. Hence, this example proves that the Guaranteed DA rule is not responsive. $\diamond$

Table 3.9: Updated priority profiles for Example 16

$\hat{\succ}_{c_1}$	$\hat{\succ}_{c_2}$	$\hat{\succ}_{c_3}$	$\hat{\succ}_{c_4}$	$\hat{\succ}_{c_5}$	$\hat{\succ}_{c_6}$		$\tilde{\succ}_{c_1}$	$\tilde{\succ}_{c_2}$	$\tilde{\succ}_{c_3}$	$\tilde{\succ}_{c_4}$	$\tilde{\succ}_{c_5}$	$\tilde{\succ}_{c_6}$
2	2	5	**3**	4	5		4	2	5	**3**	4	**2**
4	1	1	5	2	2		2	1	1	5	2	5
3	3	3	4	1	1		3	3	3	4	1	1
1	4	2	2	3	3		1	4	2	2	3	3
5	5	4	1	5	4		5	5	4	1	5	4

As we can see from the above example, the Guaranteed DA rule does not preserve the responsiveness of the Divided DA rule because it eliminates the wastefulness of the Divided DA rule, and by doing so some minority student who was able to get a higher-ranked assignment than their guaranteed assignment at some profile may not be able to do so at the same profile when the affirmative action policy is stronger.

3.7 Incentive Properties

A rule is strategyproof if reporting the true preferences is a weakly dominant strategy for each student. Formally, φ is **strategyproof** if, for all $i \in S$ and profiles $(r, P, \succ)$, and for all alternative preferences P_i' for student i, $\varphi_i(r, P, \succ)\ R_i\ \varphi_i(r, (P_i', P_{-i}), \succ)$. Otherwise, if there exists student $i \in S$, $(r, P, \succ)$ and P_i' such that $\varphi_i(r, (P_i', P_{-i}), \succ)\ P_i\ \varphi_i(r, P, \succ)$, then we will say that the rule can be **manipulated** by student i (at profile $(r, P, \succ)$).

The Divided DA and Guaranteed DA rules are not strategyproof. The following examples show that both majority and minority students can manipulate them at some profiles. Example 17 demonstrates that the Divided DA rule can be manipulated my majority students, and Example 18 shows that it can also be manipulated by minority students. Since the Guaranteed DA matching is the same as the Divided DA matching in the two examples, they also demonstrate that the Guaranteed DA rule can be manipulated by both majority and minority students. As we will see in the examples, the intuition behind these negative results is that both rules rely on the DA matching, and subsequent steps are built on the DA assignments which are calculated based on the reported preferences. While the DA outcome itself cannot be manipulated since the DA rule is strategyproof, it is possible to obtain a lower-ranked outcome in the DA, which then may alter subsequent steps in the Divided DA and Guaranteed DA algorithms in a way that the manipulating student benefits in the end.

Example 17 (The Divided DA and Guaranteed DA rules can be manipulated by majority students)**.** Let $S^M = \{2, 3\}$ and $S^m = \{1, 4, 5\}$ be the sets of majority and minority students. Let the set of schools be $C = \{c_1, c_2, c_3\}$ with capacities $q = (2, 2, 1)$. Consider the preference and priority profiles presented in Table 3.10. P'_3 is an untruthful preference ordering for majority student 3, while P_3 represents her true preferences. Let the minority reserve list for this problem be $r = (1, 0, 0)$.

Both the Divided DA and the Guaranteed DA rules yield the same matching at this profile when majority student 3 reports truthfully: $\mu = (c_3, c_1, c_2, c_1, c_2)$ (see the underlined assignments in Table 3.10). When 3 reports P'_3, while all other students report their true preferences, the matching for both rules is $\mu' = (c_1, c_3, c_1, c_2, c_2)$ (see the assignments in squares in Table 3.10). Since student 3 prefers c_1 to c_2, which means that $\mu'_3 \ P_3 \ \mu_3$, the Divided DA and Guaranteed DA rules are both manipulable by majority students. $\diamond$

Table 3.10: Preference and priority profiles for Example 17

P_1	P_2	P_3	$\boldsymbol{P'_3}$	P_4	P_5
$\underline{c_3}$	$\boxed{c_3}$	$\boxed{c_1}$	$\boldsymbol{c_1}$	$\underline{c_1}$	c_1
$\boxed{c_1}$	$\underline{c_1}$	$\underline{c_2}$	$\boldsymbol{c_3}$	c_3	$\underline{\boxed{c_2}}$
c_2	c_2	c_3	$\boldsymbol{c_2}$	$\boxed{c_2}$	c_3

$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$
1	1	3
2	2	4
3	4	2
4	5	1
5	3	5

Example 18 (The Divided DA and Guaranteed DA rules can be manipulated by minority students)**.** The setup of the problem is the same as in Example 17, but we consider here different preference and priority profiles, which are presented in Table 3.11. In this example minority student 4 can manipulate the outcome by reporting untruthful preferences P'_4, while her true preferences are given by P_4. All other students report their true preferences.

Both the Divided DA and the Guaranteed DA rules yield the same matching at this profile when minority student 4 reports truthfully: $\mu = (c_1, c_3, c_1, c_2, c_2)$ (see the underlined assignments in Table 3.11). When 4 reports P'_4, while all other students report their true preferences, the matching for both rules is $\mu' = (c_3, c_1, c_2, c_1, c_2)$ (see the assignments in squares in Table 3.11). Since student 4 prefers c_1 to c_2, which means that $\mu'_4 \ P_4 \ \mu_4$, the Divided DA and Guaranteed DA rules are both manipulable by minority students. $\diamond$

3.8 Conclusion

In this paper we investigate the possibility of constructing matching rules with an affirmative action policy that have appealing responsiveness properties to changes in the

Table 3.11: Preference and priority profiles for Example 18

P_1	P_2	P_3	P_4	$\boldsymbol{P'_4}$	P_5
$\boxed{c_3}$	$\underline{c_3}$	$\underline{c_1}$	$\boxed{c_1}$	$\boldsymbol{c_1}$	c_1
$\underline{c_1}$	$\boxed{c_1}$	$\boxed{c_2}$	$\underline{c_2}$	$\boldsymbol{c_3}$	$\underline{\underline{\boxed{c_2}}}$
c_2	c_2	c_3	c_3	$\boldsymbol{c_2}$	c_3

$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$
1	1	4
2	2	2
3	4	3
4	5	1
5	3	5

strength of the affirmative action policy and study specific matching rules with demanding responsiveness properties for the first time. Going beyond the minimal responsiveness condition, which only requires one minority student to benefit form a stronger affirmative action while all others may be worse off, we present a matching rule, the Divided DA rule, which leads to a weak Pareto-improvement for minority students when the minority reserves are increased. It is surprisingly difficult to find a reasonable rule with this property, and it turns out that the Divided DA rule may waste school seats that are desired by students. Therefore, we also construct another matching rule which has good responsiveness properties, namely, it is both minimally responsive and elementally responsive, where the latter means that introducing an affirmative action policy does not harm any of the minority students. This rule, the Guaranteed DA rule, is non-wasteful, but it does not satisfy the more stringent property of responsiveness that the Divided DA rule satisfies.

Both of these matching rules are intuitive and have other appealing properties as well. For example, both respect the affirmative action policy in the sense that if a minority student prefers a school to her assignment, then the minority reserves of that school have been exhausted. In other words, all reserve positions that are desired by minority students are filled by minority students. Other good attributes include fairness properties, such as not allowing priority violations within either group of students. On the other hand, both matching rules that we study are manipulable by both majority and minority students. However, it is well-known in this strand of the literature that strategyproofness is rarely satisfied by rules that meet other desirable criteria, with the exception of the DA-R rule which is, however, not even minimally responsive. Trade-offs in the form of possibility and impossibility results involving responsiveness, welfare, fairness, and incentive properties are worth investigating, similarly to Doğan (2016) and Chaudhury and Pápai (2022), but with more demanding responsiveness requirements, such as the ones studied here. This exploration is left for future research.

Appendix

We provide examples of matching rules satisfying the different feasible combinations of the four responsiveness axioms discussed in Section 3.3. Note that none of these rules are strategyproof.

- **Minimal elemental responsiveness only:** No known matching rule. We can construct one as follows. Take any DA-SPR rule (Pápai & Sayedahmed, 2022) and modify it to default to the DA outcome at every profile where all minority students are assigned either the same or a worse school than the school obtained in the DA outcome at this profile. At every other profile, use the DA-SPR outcome. This rule satisfies minimal elemental responsiveness by construction, but it is not elementally responsive, since there are profiles at which a DA-SPR rule makes some minority students worse off and some better off compared to their DA outcomes. Moreover, these rules are not minimally responsive, since DA-SPR rules are not minimally responsive in general.

- **Minimal responsiveness but not elemental responsiveness:** MDA (Example 8), IA-R (Example 9), and IA-DA-R (see Chaudhury and Pápai, 2022).

- **Elemental responsiveness but not minimal responsiveness:**

  No known matching rule. We can construct one as follows. Take any DA-SPR rule (Pápai & Sayedahmed, 2022) and modify it to default to the DA outcome at every profile where at least one minority student is assigned a worse school than the school obtained in the DA outcome at this profile. At every other profile, use the DA-SPR outcome. This rule is elementally responsive by construction but not Minimally Responsive, since DA-SPR rules are not minimally responsive in general.

- **Elemental responsiveness and minimal responsiveness but not responsiveness:** Guaranteed DA rule (Theorem 4)

- **Responsiveness:** Divided DA rule (Theorem 3)

Alternative proof for Claim 1:

Claim 1. *For all preference profiles $P \in \mathcal{P}$, $\varphi^{DDA}(\mathbf{0}, P)$ weakly Pareto-dominates $\varphi^{DA}(P)$.*

Proof. Fix a preference profile $P \in \mathcal{P}$ and consider alternative preferences $\hat{P}_{S^M}$ for majority students that rank their DA assignment first, while keeping the minority students' preferences the same. Then by the IR monotonicity property of the DA (Kojima and Manea,

2010) the minority students can only gain, that is, for all $i \in S^m$, $\varphi^{DA}(\hat{P}_{S^M}, P_{S^m}) R_i \varphi^{DA}(P)$, where for all $j \in S^M$, $\hat{P}_j$ ranks $\varphi_j^{DA}(P)$ first. Since all majority students get the same assignments at the two profiles by IR monotonicity (i.e., $\varphi_{S^M}^{DA}(\hat{P}_{S^M}, P_{S^m}) = \varphi_{S^M}^{DA}(P)$), one can easily verify that $\varphi_{S^m}^{DA}(\hat{P}_{S^M}, P_{S^m}) = \varphi_{S^m}^{DDA}(\mathbf{0}, P)$, since the minority students are assigned collectively the exact same school seats under these two scenarios, and both of them make assignments to minority students following the DA algorithm without any "interference" from majority students. Then it follows that for all $i \in S^m$, $\varphi_i^{DDA}(\mathbf{0}, P) R_i \varphi^{DA}(P)$, and the majority students have collectively the exact same set of school seats available under these two scenarios. Then, by the optimality property of the DA (Gale and Shapley, 1962), the majority matching in Step 2 of the Divided DA rule at $(\mathbf{0}, P)$ leads to the most preferred fair assignment of the same school seats that were assigned collectively to majority students by the DA at P, and this implies that for all $i \in S^M$, $\varphi_i^{DDA}(\mathbf{0}, P) R_i \varphi^{DA}(P)$.

□

Chapter 4

Do School Choice Mechanisms Affect School Segregation?

4.1 Introduction

Segregation is a major issue in public school assignments and high school placements in many cities all over the world, dividing the population along race, ethnicity, religion or income, and sometimes based on other characteristics such as social status or education level. Despite some specific efforts to eliminate or at least reduce segregation, the problem of schools being segregated is not just past history. Traditionally, neighbourhood school programs were used for matching students to public schools, and this is still in use in many places. Since neighbourhood school programs limit students to apply only to schools within their district, existing residential segregation implies school segregation when the students are assigned to neighborhood schools only. For instance, residential segregation by race at the household level has increased dramatically in the twentieth-century in all areas of the US, whether urban or rural, in the north or in the south (Logan & Parman, 2017).

The more recent era of US school choice programs with open enrollment laws allow students to attend public schools other than their neighbourhood school, and the options have multiplied due to charter schools, magnet schools, virtual schools, and homeschooling, among others. School choice programs eliminate the limitations of neighborhood schools and promote racial and ethnic integration, yet its impact on desegragation turned out to be the opposite in some places (Roda & Wells, 2013). Reports show that introducing various school choice programs that are expected to mitigate the segregation problem leads to even more segregation by race in public schools in New York City (see, e.g., Mader et al. (2018)). The trend of increasing segregation is not limited to the US. For example, the magnitude of school segregation measured in Chile in 2008 was already very high and it has increased over the following decade (Valenzuela et al., 2014).

The causes of school segregation were explored across many disciplines. These causes include but are not limited to sorting theories explaining the cross-section distribution of students based on socioeconomic and demographic characteristics, economic conditions and the type of residential area, parental education, housing patterns, school zoning policy, curricular tracking, and the strategic manipulation of school matching mechanisms (Logan & Parman, 2017; Bischoff, 2008; Yang Hansen & Gustafsson, 2016).

4.1.1 Overview

Our paper studies the design of school choice mechanisms and its impact on school segregation. The most popular school choice mechanisms are the celebrated Deferred Acceptance (DA) mechanism (Gale & Shapley, 1962), the Top Trading Cycles (TTC) mechanism (Shapley & Scarf, 1974; Pápai, 2000; Abdulkadiroğlu & Sönmez, 2003), and the Immediate Acceptance (IA) mechanism (Abdulkadiroğlu & Sönmez, 2003), also known as the Boston Mechanism. In addition, we study the class of Parallel mechanisms (Chen & Kesten, 2017), which includes the Deferred Acceptance and Intermediate Acceptance mechanisms as its two extreme members, since these are further examples of mechanisms that take into account directly the students' preference rankings over schools, in addition to the Immediate Acceptance mechanism. School choice mechanisms have unique advantages and drawbacks that have been widely studied. While this literature mostly explores efficiency, fairness, and incentive properties of the mechanisms, we focus on the issue of segregation that results from the choice of a matching mechanism. This topic has not been studied, for the most part, from a market design theory perspective, while there are some theoretical studies looking at manipulation and segregation. There are also many empirical and some experimental papers analyzing and reporting on school segregation and school choice mechanisms, in addition to countless studies that have appeared in education and sociology on school segregation more generally.

We can observe three trends in recent years regarding school segregation: an increase in school segregation, an increase in efforts to reduce school segregation by introducing different programs, and the IA mechanism being replaced by the DA mechanism gradually all over the world, for example in New York City in 2003, in Boston in 2005, and in England in 2008, to name only a few. The IA mechanism was also replaced by different Parallel mechanisms in most provinces of China by 2012, which are mechanisms that are closer to the DA compared to the IA mechanism. These three facts together reveal a paradoxical picture.

The design of the matching mechanism plays a crucial role in the many-to-one school assignment matching, when students are assigned to schools in a centralized system. Matching mechanisms use the preferences of students and the priorities of schools to select a student-school matching. Given a specific profile of preferences and priorities, different mechanisms return different outcomes in general. Therefore, a natural question to explore is whether the mechanism itself plays a role in school segregation when the preferences and priorities reflect realistic assumptions for segregated environments. This paper explores the role of matching mechanisms in school segregation using a theoretical framework aligned with empirical reports and aims to provide a conceptual explanation for variations in school segregation from this angle. Although some school choice programs and other centralized school assignment mechanisms incorporate an affirmative action policy, we investigate mechanisms without any affirmative action, in order to understand their impact when such a corrective policy is not in place. This is also a practically relevant scenario, since many school assignment systems do not employ affirmative action, and in fact many places, including several US states, have banned race-based or other types of affirmative action policies.

We set up a simple stylized model to analyze the impact of prominent school choice mechanisms on school segregation, namely the DA, the TTC, and IA mechanisms, and obtain unambiguous results using a natural measure of school segregation. The simplicity of our model and assumptions allow us to draw clear conclusions about the impact of different school choice mechanisms on school segregation. Our findings offer an explanation to the paradox that school segregation is on the rise despite efforts to mitigate it, while the IA mechanism is being increasingly replaced by the DA mechanism. In particular, we show that the DA and the TTC mechanisms are completely segregated, whereas the IA mechanism is more desegregated by comparison. We then revisit these results in the practically important special cases of homogeneous priorities and homogeneous preferences.

4.1.2 On the Manipulability of Mechanisms

Comparing strategyproof mechanisms, the DA and TTC, to the manipulable IA mechanism appears problematic, since manipulation may modify the results we have discovered, and an equilibrium analysis might seem more appropriate when working with the IA mechanism. Nonetheless, we compare the mechanisms while treating the IA mechanism as a direct mechanism for several reasons and, since it raises some important questions, we address these here up front.

Although the IA mechanism has been shown to be obviously manipulable theoretically (Troyan & Morrill, 2020), nonetheless the manipulation of the IA mechanism depends crucially on the information available to participants. With incomplete information, given completely symmetric students and schools, the IA mechanism induces truth-telling in equilibrium and assigns more students to their first choices than other mechanisms. Truth-telling may become a Bayesian Nash equilibrium for the IA mechanism with incomplete information, equally large schools, and independently-uniformly distributed preferences (Featherstone & Niederle, 2016).

With complete information, which would require knowledge of the preferences of other participants, Ergin and Sönmez (2006) prove that the IA Nash-equilibrium outcomes are the set of stable matchings. Given that the DA selects the stable optimal matching at each profile, the only undominated Nash-equlibrium is the DA matching, based on this analysis. Therefore, we could think of the IA outcome as one that is likely to approximate the DA matching when correctly manipulated by students, since the unique optimal DA matching, which Pareto-dominates all other fair matchings, is a focal point. This would imply that the IA outcome may be less segregated with respect to the truthful preferences compared with the reported potentially untruthful preferences, while the more realistic incomplete information environment may make manipulation less likely but also less predictable. However, regardless of how much and what kind of manipulation takes place, the IA mechanism will still remain less segregated on the whole than the strategyproof DA and TTC mechanisms, given that the latter are completely segregated, as we will show. This may reduce the gap between the relative segregation levels of the DA and IA, and TTC and IA, but we cannot be sure by how much, which makes staying with the direct IA mechanism a good option, indicating a likely gap between DA and IA as a benchmark.

Finally, there is lots of evidence in the literature that equilibrium analyses may be quite unreliable, which makes it questionable whether the segregation gap is reduced by manipulation. Evidence from experimental studies suggests that the strategic manipulation of mechanisms is often not straightforward and optimal manipulation may be hard to achieve even in very simple environments (Featherstone & Niederle, 2016). As attested by the rapidly growing strand of the literature on obvious strategyproofness, initiated by Li (2017), mechanisms that are not transparent may be subject to misrepresentation regardless of having dominant strategy equilibria. Complementing this literature there is also increasing experimental evidence demonstrating that many people don't report honestly even when using the strategyproof DA or TTC mechanisms (see, for example, Dreyfuss et al. (2022) for a brief review of the experiemental literature on manipulating the DA). This calls into question the validity of an equilibrium analysis. Thus, in the absence of clear-cut answers, we choose to analyze the IA mechanism as a direct mechanism, while keeping in mind that the results may be altered to some extent by the manipulation of not only the IA mechanism, but in fact all the mechanisms.

4.1.3 Review of the Related Literature

The IA mechanism is Pareto-efficient, while the DA outcome may be Pareto-dominated when considering only the student side of the matching, but it is neither fair[1] nor strategyproof (Abdulkadiroğlu & Sönmez, 2003). In contrast, the DA matching is student optimal, which means that it is most preferred by every student among all fair matchings (Gale & Shapley, 1962). The DA is also strategyproof (Dubins & Freedman, 1981; Roth, 1982). Due to these superior properties of the DA mechanism, Ergin and Sönmez (2006) recommended adopting the DA mechanism to replace the IA mechanism, arguing that such a policy switch would lead to unambiguous efficiency gains when there is complete information. Recent developments in this literature have added some new insights by considering more recent real-life scenarios, and pointed out that the so-called efficiency gain may not be unambiguous, arguing that the earlier literature should have explored more options before ruling against the IA mechanism (Abdulkadiroğlu, Che, & Yasuda, 2011; Featherstone & Niederle, 2011; Leo & Van der Linden, 2018; Ponzini, 2022).

Several papers that look at school segregation and matching mechanisms point out that the manipulation of the IA mechanism makes matching under the IA segregated, while the DA is desegregated because it is strategyproof (Pathak & Sönmez, 2008; Basteck & Mantovani, 2018; Calsamiglia et al., 2021, 2017; Terrier et al., 2021). The IA mechanism is not strategyproof, which means that a student can get admitted to a better school by misrepresenting her true preferences. Such incentives for strategic play create an advantageous situation for 'sophisticated' students, resulting in matchings under the IA mechanism in which poor-quality schools are segregated with 'strategically naive' students (Pathak & Sönmez, 2008). An experimental study showed that the worst schools are overrepresented by participants with lower cognitive ability under the IA mechanism, whereas under the DA mechanism schools are harmonized by cognitive ability (Basteck & Mantovani, 2018). Another similar paper reports that optimal strategies in IA might lead to ex-ante identical schools being 'cardinally segregated'[2] by socioeconomic type (Calsamiglia et al., 2021), and the IA mechanism sorts students between two identical schools to make one

[1]Fairness is also known in the literature as stability, and in this context it means no priority violations of the student placements.

[2]Cardinal segregation arises as students with identical preferences get different schools based on heterogeneous strategies associated with heterogeneous levels of risk aversion.

school elite. Meanwhile, the DA mechanism is resilient to such segregation and elitism (Calsamiglia et al., 2017). Our findings contradict much of the above, indicating that the IA mechanism would have a better impact on segregation than DA. An empirical study in favor of the IA mechanism explores data from England and reports on a case where local school authorities switched to the DA in 2008. It shows that the elimination of IA encourages high-SES (socioeconomic status) parents to report their true preferences, which increases competition for top schools, crowding out low-SES students (Terrier et al., 2021).

Another paper concerning the heterogeneity of manipulation ability concluded that the IA mechanism might not harm but rather benefit participants who may not strategize well, by providing guarantees with a higher probability than the DA to good schools for those lacking priorities at those schools (Abdulkadiroğlu et al., 2011). The paper showed, using a Bayesian model with von-Neumann Morgenstern utilities, that the naive students with the same utility value are (at least weakly) better off under the IA mechanism than the DA when true preferences are reported such that students know their cardinal preferences but only have a probability distribution for the preferences of others. This paper is in line with our findings on segregation, although the methodology, the modeling and the focus of this paper are very different from ours.

In general, the small literature on matching mechanisms and school segregation does not offer direct causality. The connection between mechanism and segregation is via manipulability properties when a heterogeneous ability to manipulate is assumed. However, it is not entirely clear how one can relate socially disadvantaged groups to their manipulation ability. Another issue with these results is that if manipulation is the reason for school segregation then it cannot explain why schools are still segregated with an upward trend in segregation under the DA with an open enrollment program. In contrast to our paper, these papers don't focus on the role preferences and priorities play and on the mechanism itself contributing to segregation without any strategic play.

Empirical and theoretical papers on school segregation without involving strategic gameplay present mixed results. Jeong (2018) showed that under the IA mechanism schools are more segregated by income both with and without open-enrollment programs compared to the DA which is resilient to segregation. Between 2012-2013, Madrid (Spain) introduced inter-district school choice under the IA mechanism, which reduced segregation by parental education but vastly increased segregation by immigrant status. However, the effects fade out when controlling for residential stratification (Gortázar et al., 2020). Different empirical evidence emerges from Chile, where the newly established school matching system adopted the DA mechanism. According to the findings of Kutscher et al. (2020), on average the DA mechanism has no statistically significant impact on school segregation with districts that are segregated by income, but it increases within-school integration from low socioeconomic backgrounds in areas with high levels of existing residential segregation.

4.2 Preliminaries

We first introduce our stylized model. We believe that it captures enough salient features of school segregation that we are interested in to derive meaningful results, while this model is simple enough at the same time to allow for clear-cut conclusions. There are two groups of students in our model, *majority* students and *minority* students, representing groups of students with socioeconomically privileged and underprivileged backgrounds respectively. There are also two types of schools, what we refer to simply as *high-quality schools*, typically located in privileged neighborhoods, and *low-quality schools*, typically located in underprivileged neighborhoods. Schools have rankings over students, which we call *priorities*, and students rank schools according to their *preferences*.

4.2.1 Assumptions

Our simplifying assumptions pertain to the priorities of high-quality schools and the preferences of majority students, and we justify them based on a variety of literature from economics, education, and sociology.

Assumption 1. *Majority students are ranked higher than minority students by all high-quality schools.*

While there are always some exceptions in real life, this assumption is supported in general by a lot of evidence in different literatures and due to various reasons. Without going into detail, it is evident from the relevant education and sociology literature that schools usually prefer students from higher socioeconomic backgrounds (K. B. Smith & Meier, 1995). This may be based on both test scores and grades earned in prior education. Standardized test scores are used in many centralized high school placement systems and in university admissions. For example, in China there is a National College Entrance Examination for students planning to attend college, and students are ranked according to their test scores (Zhu, 2014). In the US and Canada, universities consider SAT and GRE scores one of the primary criteria to rank applicants. It is well known that standardized test scores favor privileged students. A report by the Brookings Institution illustrates SAT score discrepancy by race. The data show that the average scores of black students (454) are significantly lower than those of white students (547) (E. Smith & Reeves, 2020). Another report shows that wealthier Americans from more educated families tend to do far better on the SAT, with an average combined score of 1,714 for students from families earning more than $200,000 per year, while students from families earning under $20,000 per year have an average combined score of 1,326 (Goldfarb, 2014). To make things worse, this discrepancy has been increasing over the years. Using information from a dozen large US national studies conducted between 1960 and 2010, Reardon (2013) concluded that the rich-poor gap in test scores is about 40% larger now than it was 30 years ago.

Schools can apply their own criteria regarding their preferences over students. For example, some schools may be so-called exam schools that administer their own tests or use other merit-based criteria in their assessment of their applicants. In 2010, Chicago Public Schools adopted a priority structure based on socioeconomic characteristics to place students into four tiers, but even there within each tier the tie-breaker is based on a composite score that combines the results of a specialized entrance exam, prior standardized test scores, and grades in prior coursework (Ellison & Pathak, 2021). The relevant literature is vast, and in general much evidence supports our assumption that majority students have higher priority at high-quality schools than minority students, regardless of whether the schools' priority is based on test scores or other performance-based criteria.

Even if such bias does not apply against underprivileged minority students, when school priorities are based on local laws and school funding systems, our assumption reflects reality quite well. It is often required by law for US schools, for example, to have higher priority for students from the district where the school is located. The Boston Public School (BPS) choice is an example of such a priority system. As reported by Abdulkadiroğlu, Pathak, Roth, and Sönmez (2005), the BPS assignment process in 2004 prioritized students living in the area above all other students. They had mainly four indifference classes for students who (1) have siblings at that school and live in the walk zone, (2) have siblings at the school, (3) live in the walk zone, and (4) all other students. Even today, in the era of school choice programs, the BPS offers a customized list of schools to every family based on their home address, which includes schools within a one-mile radius of their home for which they are prioritized (see BPS Welcome Services).

Thus, high-quality schools, located in privileged neighborhoods, have higher priority for the majority students who live in the walk zone of the school. While this doesn't support our assumption entirely on ranking all majority students above all minority students, it does point in the same direction as test scores and other achievement-based criteria, and when there is only one "rich" neighborhood and one "poor" neighborhood in a town, this observation matches our assumption.

Assumption 2. *High-quality schools are ranked higher than low-quality schools by all majority students.*

This assumption probably does not need much justification. Traditionally, in the US public education system and elsewhere, the amount of wealth in a school district affects the quality of its schools, and thus higher-quality schools are typically located in richer districts. Since schools are funded primarily by local property taxes, the wealthiest districts spend as much as three times more per pupil than in the most economically disadvantaged districts (Kozol, 1991; Slavin, 1999). Many states have gradually changed such heavy reliance on local property tax over the last few decades, but such policy change is hardly enough; as recently as in 2021, 45% of the funding still comes from the local community, while the federal funding is only 8% of total school funding (Irwin et al., 2021), leading to a significant disparity. The wealthiest US public schools spend at least 10 times more than the poorest schools (Darling-Hammond, 2007). Naturally, this leads to superior school quality for schools located in rich neighborhoods, what we call high-quality schools, and thus it is not surprising that students rank high-quality schools higher than low-quality schools in their preference orderings. This would be particularly true for majority students, for whom some or all of these high-quality schools are located in their own neighborhoods, or in neighborhoods they feel comfortable with, as opposed to low-quality schools which may be located further, in less desirable or sometimes downright dangerous neighborhoods.

When schools differ in terms of quality substantially, regardless of their socioeconomic status, students (supported by their parents) want high-quality education to secure a better future. Since high-quality education requires more resources, schools located in richer districts are better in quality. Therefore, it is likely that many minority students also prefer high-quality schools to low-quality schools, depending on their background; however, we need not assume that minority students rank high-quality schools ahead of low-quality schools in order to obtain our results.

4.2.2 Model and Definitions

There is a finite set of students, denoted by S. The set of students is divided into the set of majority students, S^M, and the set of minority students, S^m. Thus, $S^M \bigcap S^m = \emptyset$ and $S^M \bigcup S^m = S$. Let $a \in S^M$ and $i \in S^m$ denote majority students and a minority students respectively.

There is a finite set of n schools, denoted by C. Each school $c \in C$ has a capacity $q_c \geqslant 1$, which is the maximum number of students that can be assigned to school c. Let

$q = (q_c)_{c \in C}$ denote the capacity list. The set of schools is divided into the set of high-quality schools C^H and the set of low-quality schools C^L. Thus, $C^M \bigcap C^L = \emptyset$ and $C^H \bigcup C^L = C$.

In the following we will refer to the majority or minority status of a student as the student's type, and the quality of the school as the school's type. Let s^M denote the number of majority students, and let s^m denote the number of minority students. Then $s^M + s^m = |S|$. For simplicity, we assume that $\sum_{c \in C^H} q_c = s^M$. Then $\sum_{c \in C^L} q_c = s^m$, and the total number of school seats equals the number of students: $\sum_{c \in C} q_c = |S|$.

Each student $s \in S$ has a strict preference ordering P_s over C. We will denote the weak counterpart of P_s by R_s, so $c \ R_S \ c'$ holds if and only if either $c \ P_S \ c'$ or $c = c'$. We assume that each school is acceptable to each student. Let a preference profile be denoted by $P = (P_s)_{s \in N}$, and let $\mathcal{P}$ be the set of all preference profiles. We assume that majority students prefer any high-quality school to any low-quality school, as explained in Section 2.1. Formally, for all $a \in S^M$, $c \in C^H$ and $c' \in C^L$, $c \ P_a \ c'$. This is our Assumption 2. Note that majority students' (and minority students') preferences may differ over the different high-quality and low-quality schools, and thus the preferences are not necessarily homogeneous across students.

Each school $c \in C$ has a strict priority ordering $\succ_c$ over S. Let a priority profile be denoted by $\succ = (\succ_c)_{c \in C}$, and let Π denote the set of all priority profiles consisting of a strict priority ordering of each school. We assume that all high-quality schools have higher priority for any majority student than for any minority student, as discussed in Section 2.1. Formally, for all $c \in C^H$, $a \in S^M$ and $i \in S^m$, $a \succ_c i$. This is our Assumption 1. The priority ranks of schools over students (including high-quality schools) may be heterogeneous for schools.

A **matching** μ is a function from the set of students to the set of schools such that each school is assigned at most q_c students. With a slight abuse of notation, for all $s \in S$ we denote the assignment that student s receives by μ_s, Also, for all $c \in C$, we denote the set of students assigned to school c by μ_c, where $\mu_c \subseteq S$. For given S, C and $(q_c)_{c \in C}$, let $\mathcal{M}$ denote the set of matchings. Let $(P, \succ) \in \mathcal{P} \times \Pi$ denote a **profile**, consisting of of a preference and a priority profile. A **matching mechanism** is $\varphi : \mathcal{P} \times \Pi \rightarrow \mathcal{M}$, which assigns a matching μ to each profile $(P, \succ)$.

A matching μ is **non-wasteful** at $(P, \succ)$ if for all $s \in S$ and $c \in C$, $c \ P_s \ \mu_s$ implies that the capacity of school c is exhausted, that is, $|\mu_c| = q_c$. Given our assumptions that each school is acceptable to all students and $\sum_{c \in C} q_c = |S|$, non-wastefulness implies that each school c is assigned q_c students and thus each student is assigned to a school. That is, for all $c \in C$, $|\mu_c| = q_c$ and, for all $s \in S$, $\mu_s \in C$.

Student s is said to have **justified envy** in matching μ at profile $(P, \succ)$ if there exist a student $s' \in S$ and a school $c \in C$ such that $c \; P_s \; \mu_s$, $s \succ_c s'$ and $\mu_{s'} = c$. We also say in this case that s has justified envy for school c in μ, and that matching μ has justified envy at $(P, \succ)$. A matching μ is **fair** at $(P, \succ)$ if μ has no justified envy and is non-wasteful at $(P, \succ)$.[3] A matching mechanism φ is fair if it assigns a fair matching to each profile $(P, \succ)$.

A matching mechanism φ is **strategyproof** if for all profiles $(P, \succ)$, students $s \in S$, and alternative preferences P'_s for student s, $\varphi_s(P, \succ) \; R_s \; \varphi_s((P'_s, P_{-s}), \succ)$. Otherwise, if $\varphi_s((P'_s, P_{-s}), \succ) \; P_s \; \varphi_s(P, \succ)$, we will say that student s can **manipulate** φ at $(P, \succ)$ by reporting P'_s and φ is **manipulable.**

4.2.3 Segregation Measure and Comparisons

We will measure segregation simply by the number of minority students that are assigned to high-quality schools. The fewer minority students who can get into a high-quality school, the more segregated the matching is. This very simple and intuitive measure lets us quickly assess the extent of segregation, and it allows for complete comparability of segregation in school assignments since it is possible to compare the relative level of segregation for any two different matchings. Naturally, if the distribution of minority students across different schools is relevant, not just the types of the schools in the aggregate, then a more sophisticated measure of segregation that measures distribution would be called for, of which there are plenty in the sociology and empirical economics literature. However, for our purposes this very simple measure of segregation will suffice, since it is unlikely that minority students would be concentrated in high-quality schools, and thus this rough measure is also a good proxy for a more complicated distribution criterion both in practice and in our stylized model.

Formally, we define the level of segregation for matching $\mu \in \mathcal{M}$ by $d(\mu)$, which is the number of minority students who are assigned to a high-quality school in μ. Then matching μ is **less segregated** than matching μ' if $d(\mu) > d(\mu')$. We will also say that matching μ is **weakly less segregated** than matching μ' if $d(\mu) \geqslant d(\mu')$.

Now we can define the comparison of matching mechanisms in terms of the level of segregation. A **matching mechanism $\boldsymbol{\varphi}$** is **less segregated** than another matching mechanism φ' if, for all profiles $(P, \succ) \in \mathcal{P} \times \Pi$, $\varphi(P, \succ)$ is weakly less segregated than $\varphi'(P, \succ)$, and there exists at least one profile $(\tilde{P}, \tilde{\succ}) \in \mathcal{P} \times \Pi$ such that $\varphi(\tilde{P}, \tilde{\succ})$ is less segregated than $\varphi'(\tilde{P}, \tilde{\succ})$. That is, φ is less segregated than φ' if $d\big(\varphi(P, \succ)\big) \geqslant \big(\varphi'(P, \succ)\big)$ for all

[3] The property of *fairness* often includes individual rationality. A matching μ is **individually rational** at $(P, \succ)$ if, for all $s \in S$, $\mu_s \; R_s \; 0$, where 0 represents staying unmatched. In our setup individual rationality is automatically satisfied, since all schools are acceptable to all students.

$(P, \succ) \in \mathcal{P} \times \Pi$ with at least one strict inequality. If a matching mechanism φ is less segregated than another matching mechanism φ', then φ' is more segregated than φ.

A **matching** μ is **completely segregated** if no minority students are assigned to high-quality schools in μ, that is, $d(\mu) = 0$. A **matching mechanism** φ is **completely segregated** if no minority students are assigned to high-quality schools at any profile, that is, $d\big(\varphi(P, \succ)\big) = 0$ for each profile $(P, \succ) \in \mathcal{P} \times \Pi$.

4.3 Matching Mechanisms

In this section we provide descriptions of the main mechanisms that we study. Each mechanism is run in iterative steps. We describe the first step and a general step t for each procedure.

4.3.1 Deferred Acceptance Mechanism

The first mechanism is the well-known and practically much relevant Deferred Acceptance mechanism which was introduced by Gale and Shapley (1962).

Deferred Acceptance (DA) Mechanism (φ^{DA}):

Fix a profile $(P, \succ)$.

Step 1: Each student $s \in S$ applies to her most preferred school according to P_s. Each school $c \in C$ tentatively assigns seats to applying students up to its capacity q_c, following its priority ordering $\succ_c$. Once the capacity is exhausted, all remaining applicants are rejected.

Step *t* $(t \geqslant 2)$**:** Each student $s \in S$ who was rejected in step $t-1$ applies to her next most preferred school. Each school $c \in C$ considers its tentatively assigned students from the previous step along with the new applicants, and tentatively assigns seats to these students up to its capacity q_c, following its priority ordering $\succ_c$. Once the capacity is exhausted, all the remaining students who are considered in this step are rejected.

The mechanism terminates when there are no more new applications to any school. All tentative matches in the final step become permanent. $\diamond$

The DA mechanism is fair (Gale & Shapley, 1962), strategyproof (Dubins & Freedman, 1981; Roth, 1982), and Pareto-dominates all other fair matchings (Gale & Shapley, 1962; Balinski & Sönmez, 1999). It is not Pareto-efficient (when considering the students' welfare).

4.3.2 Serial Dictatorship Mechanism

A Serial Dictatorship, introduced by Satterthwaite and Sonnenschein (1981), is a sequential choice mechanism that has a priority sequence σ of all students $s \in S$, where σ is a permutation of S, such that $\sigma = (\sigma_1, \ldots, \sigma_{|S|})$, where $\sigma_t \in S$ is the t^{th} ranked student according to σ.

Serial Dictatorship (SD) Mechanism ($\varphi^{\mathbf{SD}}$):

Fix a profile $(P, \succ)$. In each step of the procedure a school is *available* if its current capacity is at least 1. At the beginning of the procedure each schools starts with its initial capacity q_c and thus each school is available.

Step 1: Assign student σ_1 to her most preferred school according to P_{σ_1}.

Step *t* $(t \geqslant 2)$**:** Reduce by 1 the capacity of the school that student σ_{t-1} is assigned to in step $t-1$. If the updated capacity is zero then the school is no longer an available school in this step.

Assign student $\sigma_t \in S$ to her most preferred available school according to P_{σ_t}.

The procedure ends when either all students are assigned to a school or there is no available school left. $\diamond$

The SD mechanism is Pareto-efficient and strategyproof (Satterthwaite & Sonnenschein, 1981). However, it is not fair, unless the priority profile is homogeneous and each school ranks students according to σ.

4.3.3 Top Trading Cycles Mechanism

Shapley and Scarf (1974) was the first to introduce the Top Trading Cycles mechanism, known as Gale's TTC, with one-to-one endowments in a one-to-one matching model. Pápai (2000) introduced the Fixed Endowment Hierarchical Exchange Rules, a wide class of TTC matching mechanisms that generalize both Serial Dictatorships and Gale's TTC. These rules were extended from one-to-one to many-to-one matching by Abdulkadiroğlu and Sönmez (2003). We introduce this mechanism here for the many-to-one school choice model as the Top Trading Cycles (TTC) mechanism.

The TTC mechanism starts by "endowing" each student with the schools that rank the student at the top of their priority ordering. Any student can be endowed with multiple schools or no schools, depending on the priority profile. When a school with more than one seat is assigned to a student, the school is inherited in a later step, after this student receives her assignment, by the next highest-ranked student who is still unassigned, in accordance with the priority ranking of the school. Moreover, if a student who has traded in a step leaves behind a school due to multiple school endowments, then these schools are also inherited.

Top Trading Cycles (TTC) Mechanism (φ^{TTC}):

Fix a profile $(P, \succ)$. In each step of the procedure a school is *available* if its current capacity is at least 1. At the beginning of the procedure each schools starts with its initial capacity q_c and thus each school is available.

Step 1: Endow each school $c \in C$ to the student $s \in S$ who has the top priority for c according to $\succ_c$. Let each student s point to the student endowed with her most preferred school according to P_s. This determines a directed graph, in which the directed edges are given by the pointing students. In this graph there is necessarily at least one cycle, which we call a top trading cycle. Note that a top trading cycle may be a student pointing to herself, if her most preferred school is endowed to her. Assign each student who is in a top trading cycle to her most preferred school and remove each from the market.

Step *t* $(t \geqslant 2)$**:** Reduce the capacity of each schools that was in a top trading cycle in step $t-1$ by 1. If the updated capacity is zero then the school is no longer an available school in this step.

Endow each school $c \in C$ that is available after step $t-1$ to the student $s \in S$ who has the top priority for c according to $\succ_c$ among all remaining students. Let each remaining student s point to the student endowed with her most preferred school among the available schools according to P_s. Assign each student who is in a top trading cycle to her most preferred school among the available schools and remove each from the market.

The procedure ends when either all students are removed or there is no available school left. $\diamond$

The TTC mechanism is Pareto-efficient and strategyproof (Pápai, 2000; Abdulkadiroğlu & Sönmez, 2003). However, it is not a fair mechanism, as there may be justified envy at some profiles due to trading the priorities in top trading cycles.

4.3.4 Immediate Acceptance Mechanism

The Immediate Acceptance mechanism, which is also known as the Boston mechanism (Abdulkadiroğlu & Sönmez, 2003), is one of the most popular matching mechanisms used in practice for school assignments, centralized university admissions and other matching problems.

Immediate Acceptance (IA) Mechanism ($\varphi^{\mathbf{IA}}$):

Fix a profile $(P, \succ)$.

Step 1: Each student $s \in S$ applies to her most preferred school according to P_s. Each school $c \in C$ permanently assigns seats to applying students up to its capacity q_c, following its priority ordering $\succ_c$. Remaining applicants are rejected when all school seats are assigned and thus the capacity is exhausted.

Step *t* $(t \geqslant 2)$**:** Each student $s \in S$ who was rejected in step $t - 1$ applies to her next most preferred school. Each school $c \in C$ permanently assigns its remaining seats to applying students up to its capacity q_c, following its priority ordering $\succ_c$. Remaining applicants are rejected if all school seats are assigned and thus the capacity is exhausted.

The procedure ends when either all students are assigned to a school or there is no available school left. $\diamond$

The IA mechanism is neither fair nor strategyproof. However, if the students do not manipulate the mechanism and submit their true preferences then it is Pareto-efficient (Abdulkadiroğlu & Sönmez, 2003).

4.3.5 Parallel Mechanisms

Parallel mechanisms are called Application-Rejection mechanisms by Chen and Kesten (2017) who analyze these mechanisms. Variations of Parallel mechanisms are used for centralized university admissions in almost all provinces of China and also for high school assignments. This is a family of mechanisms that spans the DA and IA mechanisms and are parameterized by k, where $k = 1, \ldots, n$.

In a Parallel mechanism students' preference ranks are partitioned such that each ranked school belongs to a preference class, and a school in a higher preference class is ranked higher than a school in a lower preference class. For each Parallel mechanism the preference partition is homogeneous across students (i.e., each student has the same partition of her preference ranks), and each student has an equal number of schools in each preference class within the preference rank partition. The distinguishing feature of these mechanisms is the size of each member of its partition. By varying the number k of schools in each preference class, the family of Parallel mechanisms comprises a range of rules, with the IA and DA mechanisms as its two extreme members. We define the k-Parallel Mechanism for $k = 1, \ldots, n$ below based on the definition provided by Ayoade and Pápai (2023).

***k*-Parallel Mechanisms:** $\left(\boldsymbol{\varphi}^{\mathbf{PA}^k}(\boldsymbol{k} = \mathbf{1}, \ldots, \boldsymbol{n})\right)$**:**

Fix a profile $(P, \succ)$.

Step 1: Each student $s \in S$ applies to her most preferred school according to P_s. Each school tentatively assigns its seats to its applicants according to the following selection rule. The selection is based on the preference classes of the applicants primarily, and on the given strict priorities of the schools secondarily. Specifically, each applicant that ranks the school in a higher preference class is chosen before each applicant that ranks the school in a lower preference class. If there is a preference class that contains a school c for multiple applicants who cannot all be selected due to the school's capacity, then the selection is done among these applicants according to the priority ordering $\succ_c$, and any remaining applicants are rejected.

Step *t* $(t \geqslant 2)$**:** Each student $s \in S$ who was rejected in step $t-1$ applies to her next most preferred school according to P_s. Each school considers the students who are tentatively assigned to the school, if any, together with its new applicants, and tentatively assigns its seats according to the same selection rule as described for Step 1, based on the preference classes of the applicants primarily, and on the given strict priorities of the schools secondarily. Any remaining applicants are rejected.

The mechanism terminates when there are no more new applications to any school. All tentative matches in the final step become permanent. $\diamond$

Observe that if $k = 1$ then each preference class contains only one school and hence the 1-Parallel Mechanism is equivalent to the IA mechanism in which all assignments in each step are final. On the other hand, when $k = n$, where n is the number of schools, then there is only one preference class containing all schools for each student, and the selection of students is based only on the strict priorities of schools. Thus, the n-Parallel Mechanism is equivalent to the DA mechanism.

4.3.6 Serial Dictatorship Mechanism for Schools

Lastly, we introduce the Serial Dictatorship mechanism for schools, as opposed to students, which we introduced in subsection 4.3.2 and is the standard form of a serial dictatorship in this context. In a Serial Dictatorship for Schools, schools get to choose their favorite remaining students according to a fixed ordering of schools. Let ρ be a permutation of C, such that $\rho = (\rho_1, \ldots, \rho_n)$, where $\rho_t \in C$ is the t^{th}-ranked school according to ρ.

Serial Dictatorship (SD^C) Mechanism for Schools (φ^{SD^C}):

Fix a profile $(P, \succ)$.

Step 1: Assign to school ρ_1 the q_{ρ_1} highest-priority students according to $\succ_{\rho_1}$. Remove the students assigned to the school.

Step *t* $(t \geqslant 2)$**:** Assign to school ρ_t the q_{ρ_t} highest-priority students according to $\succ_{\rho_1}$ among the remaining students. Remove the students assigned to the school.

The procedure ends when either all students are assigned to a school or there are no more schools left. $\diamond$

4.4 Impact of the DA and TTC Mechanisms on Segregation

We show first that the DA and the TTC mechanisms lead to a completely segregated matching regardless of the profile.

Theorem 5. *The DA mechanism is completely segregated.*

Proof. Fix a profile $(P, \succ) \in \mathcal{P} \times \Pi$. Note first that since the DA mechanism is non-wasteful, each student is assigned to a school and hence each school seat is assigned to a student in matching $\varphi^{\text{DA}}(P, \succ)$, since each school is acceptable to all students and $\sum_{c \in C} q_c = |S|$. Suppose, by contradiction, that there exists a minority student $i \in S^m$ such that $\varphi_i^{\text{DA}}(P, \succ) \in C^H$. Then, since $\sum_{c \in C^H} q_c = s^M$, there exists $a \in S^M$ such that $\varphi_a^{\text{DA}}(P, \succ) \notin C^H$. Therefore, given that each student is assigned to a school in $\varphi^{\text{DA}}(P, \succ)$, it follows that $\varphi_a^{\text{DA}}(P, \succ) \in C^L$. Let $c = \varphi_i^{\text{DA}}(P, \succ)$. Then $c \in C^H$, and given $\varphi_a^{DA}(P, \succ) \in C^L$, Assumption 2 implies that $c \ P_a \ \varphi_a^{\text{DA}}(P, \succ)$. Moreover, since $c \in C^H$, $i \in S^m$ and $a \in S^M$, Assumption 1 implies that $a \succ_c i$. This means that a has justified envy for school c, which contradicts the fact that the DA rule is fair (Gale &

Shapley, 1962). Therefore, for all $i \in S^m$, $\varphi_i^{\text{DA}}(P, \succ) \in C^L$. This implies that the DA mechanism is completely segregated. $\square$

Theorem 6. *The TTC mechanism is completely segregated.*

Proof. Fix a profile $(P, \succ) \in \mathcal{P} \times \Pi$. Suppose, by contradiction, that there exists $i \in S^m$ who is endowed with a high-quality school $c \in C^H$ in some step $t \geqslant 1$ of the TTC procedure applied to profile $(P, \succ)$. Then each majority student receives her assignment prior to step t, since for each majority student $a \in S^M$, $a \succ_c i$, by Assumption 1. Thus, given the TTC procedure, none of the majority students prefer c to their assignment, otherwise c would not be still available after all majority students have received their assignments. Therefore, majority students cannot be assigned to a low-quality school at $(P, \succ)$, since any such majority student would prefer school c in that case, given Assumption 2. This implies that each majority student is assigned to a high-quality school at $(P, \succ)$. However, this is a contradiction, since each majority student is assigned to a school prior to step t, and given that $\sum_{c \in C^H} q_c = s^M$, all high-quality school seats are assigned prior to step t and student i therefore cannot be endowed with a high-quality school $c \in C^H$ in step t of the procedure. Hence, any minority student i can only be endowed with a low-quality school in the TTC procedure applied to profile $(P, \succ)$ in any step $t \geqslant 1$ of the procedure. This means that all high-quality schools are endowed to majority students. Then, by Assumption 2, which says that all majority students prefer each high-quality school to each low-quality school, no majority student trades a high-quality school for a low-quality school in any step of the procedure, and hence for all $i \in S^m$, $\varphi_i^{\text{TTC}}(P, \succ) \in C^L$. This implies that the TTC mechanism is completely segregated. $\square$

We provide an example next which illustrates that the DA and TTC mechanisms are completely segregated, as shown by the two theorems.

Example 19. Let $S^M = \{a_1, \ldots, a_7\}$ and $S^m = \{i_1, i_2, i_3\}$ be the sets of majority students and minority students. Let $C^H = \{c_1, c_2, c_3\}$ and $C^L = \{c_4\}$ respectively be the set of high-quality and low-quality schools with capacities $q = (3, 2, 2, 3)$. Consider the preference and priority profiles in Table 4.1.

The DA matching at this profile is $\mu_{c_1}^{\text{DA}} = \{a_2, a_5, a_7\}$, $\mu_{c_2}^{\text{DA}} = \{a_1, a_3\}$, $\mu_{c_3}^{\text{DA}} = \{a_4, a_6\}$, and $\mu_{c_4}^{\text{DA}} = \{i_1, i_2, i_3\}$ (underlined in the table). The TTC matching at this profile is $\mu_{c_1}^{\text{TTC}} = \{a_2, a_4, a_5\}$, $\mu_{c_2}^{\text{TTC}} = \{a_3, a_7\}$, $\mu_{c_3}^{\text{TTC}} = \{a_1, a_6\}$, and $\mu_{c_4}^{\text{TTC}} = \{i_1, i_2, i_3\}$ (indicated by squares in the table). For this profile, both the DA and the TTC mechanisms return

Table 4.1: Profile for Example 19

P_{a_1}	P_{a_2}	P_{a_3}	P_{a_4}	P_{a_5}	P_{a_6}	P_{a_7}	P_{i_1}	P_{i_2}	P_{i_3}
c_1	$\underline{\boxed{c_1}}$	c_1	$\boxed{c_1}$	$\underline{\boxed{c_1}}$	$\underline{\boxed{c_3}}$	$\boxed{c_2}$	c_1	c_2	c_3
$\underline{c_2}$	c_3	$\underline{\boxed{c_2}}$	$\underline{c_3}$	c_2	c_2	$\underline{c_1}$	c_2	c_3	c_1
$\boxed{c_3}$	c_2	c_3	c_2	c_3	c_1	c_3	c_3	c_1	c_2
c_4	c_4	c_4	c_4	c_4	c_4	c_4	$\underline{\boxed{c_4}}$	$\underline{\boxed{c_4}}$	$\underline{\boxed{c_4}}$

$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$	$\succ_{c_4}$
a_7	a_6	a_7	a_5
a_2	a_2	a_4	a_3
a_5	a_3	a_6	a_1
a_4	a_4	a_2	a_2
a_3	a_5	a_3	a_6
a_6	a_1	a_1	i_1
a_1	a_7	a_5	a_7
i_3	i_1	i_2	a_4
i_2	i_2	i_1	i_3
i_1	i_3	i_3	i_2

completely segregated matchings, as all minority students are matched to the low-quality school c_4. $\diamond$

The inherent reason for Theorems 5 and 6 is that both the DA and the TTC rely on the priorities substantially, although in different ways. This is a good property of a mechanism in other contexts, but not when segregation is an issue. At profiles where all majority students prefer high-quality schools to low-quality schools, and all majority students have higher priority at high-quality schools than minority students, the DA and TTC mechanisms yield a matching which is completely segregated, since the competition for high-quality schools is always resolved in favor of majority students over minority students. It is clear why the DA results in complete segregation, as the DA does not allow for justified envy and thus strictly enforces the priorities, and since majority students prefer high-quality schools to low quality schools their high priority at high-quality schools allows each of them to be matched to a high-quality school. On the other hand, we can attribute the fact that the TTC mechanism is completely segregated to a feature of the TTC that one may think of as "elitist," which allows highly-ranked students in various school priorities to trade among themselves while making lower-priority students wait for their turn. More precisely, since high-quality schools are endowed to majority students first, and since all majority students prefer high-quality schools to low-quality schools, majority students end up trading high-quality schools among themselves which further increases their welfare, while minority students, endowed at best with low-quality schools, can never be part of a trading cycle with majority students and thus do not get a chance to obtain a seat at a high-quality school.

4.5 Impact of the IA and Parallel Mechanisms on Segregation

Now we turn to the IA mechanism and also analyze Parallel mechanisms. The next example demonstrates that the IA mechanism is not completely segregated at every preference profile.

Example 20. Consider the setup and profile specified in Example 19. The IA matching at this profile is $\mu^{\text{IA}}_{c_1} = \{a_2, a_4, a_5\}$, $\mu^{\text{IA}}_{c_2} = \{a_7, i_2\}$, $\mu^{\text{IA}}_{c_3} = \{a_6, i_3\}$, and $\mu^{\text{IA}}_{c_4} = \{a_1, a_3, i_1\}$ (indicated in circles in Table 4.1). For this profile, the IA mechanism yields a matching that is not completely segregated, as two minority students, i_2 and i_3, are matched to high-quality schools, namely c_2 and c_3 respectively. $\diamond$

While Example 20 is not quite as extreme, the IA mechanism may be completely desegregated in the sense that each minority student is assigned to a high-quality school. Our next example demonstrates that the above result extends to k-Parallel mechanisms, except for $k = n$. In other words, none of the Parallel mechanisms are completely segregated, with the exception of the DA. One might conjecture that a Parallel mechanism with smaller preference bands is less segregated than another one with broader preference bands, since the IA mechanism uses the smallest preference band ($k = 1$) and the DA mechanism uses the highest ($k = n$). As we will see in Example 21 below, the IA matching is completely segregated at the given profile, while the matching produced by the 2-Parallel mechanism (also known as the Shanghai mechanism) is less segregated. Therefore, a smaller k for k-Parallel mechanisms does not necessarily imply less segregation.

Example 21. Let $S^M = \{a_1, \ldots, a_7\}$ and $S^m = \{i_1, i_2\}$ be the sets of majority students and minority students. Let $C^H = \{c_1, \ldots, c_5\}$ and $C^L = \{c_6\}$ respectively be the set of high-quality and low-quality schools with capacities $q = (2, 2, 1, 1, 1, 2)$. Consider the preference and priority profiles in Table 2. The priority ordering is the same for each school $c \in C$, as specified by $\tilde{\succ}_c$.

The IA matching is $\mu^{\text{IA}}_{c_1} = \{a_1, a_5\}$, $\mu^{\text{IA}}_{c_2} = \{a_2, a_3\}$, $\mu^{\text{IA}}_{c_3} = \{a_6\}$, $\mu^{\text{IA}}_{c_4} = \{a_7\}$, $\mu^{\text{IA}}_{c_5} = \{a_4\}$, and $\mu^{\text{IA}}_{c_6} = \{i_1, i_2\}$ (underlined in the table). Thus, the IA mechanism returns a matching that is completely segregated: $d(\varphi^{\text{IA}}(P, \tilde{\succ})) = d(\varphi^{\text{PA}^1}(P, \tilde{\succ})) = 0$. The 2-Parallel matching is $\mu^{\text{PA}^2}_{c_1} = \{a_1, a_4\}$, $\mu^{\text{PA}^2}_{c_2} = \{a_2, a_3\}$, $\mu^{\text{PA}^2}_{c_3} = \{a_5\}$, $\mu^{\text{PA}^2}_{c_4} = \{a_6\}$, $\mu^{\text{PA}^2}_{c_5} = \{i_2\}$, and $\mu^{\text{PA}^2}_{c_6} = \{a_7, i_1\}$ (indicated by squares in the table). Hence, the 2-Parallel mechanism yields a matching that is less segregated: $d(\varphi^{\text{PA}^2}(P, \tilde{\succ})) = 1$. $\diamond$

Table 4.2: Profile for Example 21

P_{a_1}	P_{a_2}	P_{a_3}	P_{a_4}	P_{a_5}	P_{a_6}	P_{a_7}	P_{i_1}	P_{i_2}	$\tilde{\succ}_c$
$\underline{\underline{\boxed{c_1}}}$	$\underline{\underline{\boxed{c_2}}}$	$\underline{\underline{\boxed{c_2}}}$	c_2	$\underline{c_1}$	c_1	c_1	c_2	c_1	a_1
c_2	c_3	c_3	$\boxed{c_1}$	c_2	c_2	c_2	c_1	c_2	a_2
c_5	c_1	c_4	$\underline{c_5}$	$\boxed{c_3}$	$\underline{c_3}$	c_3	c_3	$\boxed{c_5}$	a_3
c_4	c_5	c_1	c_3	c_4	$\boxed{c_4}$	$\underline{c_4}$	c_4	c_3	a_4
c_3	c_4	c_5	c_4	c_5	c_5	c_5	c_5	c_4	a_5
c_6	c_6	c_6	c_6	c_6	c_6	$\boxed{c_6}$	$\underline{\underline{\boxed{c_6}}}$	$\underline{c_6}$	a_6
									a_7
									i_1
									i_2

The matching returned by the Shanghai mechanism ($k = 2$) is not completely segregated in Example 21. Since we could easily construct similar examples for $k > 2$, based on this example and on Example 20, we can state the following observation.

Proposition 9. *The k-Parallel mechanisms are not completely segregated for any k such that $1 \leqslant k < n$. In particular, the IA mechanism $(k = 1)$ is not completely segregated.*

As shown by Theorems 5 and 6, the DA and the TTC mechanisms are completely segregated at every profile. Therefore, given Proposition 9, the IA mechanism, and the Parallel mechanisms more generally, are less segregated than the DA and TTC mechanisms. We summarize this comparison result below which, in light of the previous two examples and the proposition based on them, together with the two theorems in Section 4.4, does not need a proof.

Corollary 3. *The k-Parallel mechanisms for $1 \leqslant k < n$ are less segregated than the DA and the TTC mechanisms. In particular, the IA mechanism $(k = 1)$ is less segregated than the DA and the TTC mechanisms.*

Although the IA mechanism is still widely used in real-world matching problems, this is a somewhat surprising result in a broader context, as the IA mechanism is generally viewed by both the theoretical literature and by informed market designers in practice as an inferior mechanism compared to the DA and TTC mechanisms. When we focus on the segregation of different types of students, and specifically on the school placement of the underrepresented and underprivileged, however, the IA mechanism can be seen as less elitist than either the DA or TTC, which are both completely segregated due to their extensive reliance on school priorities. By contrast, the IA mechanism makes assignments primarily based on the preferences and uses the priorities only for tie-breaking. Thus, it assigns a minority student to a high-quality school as long as the minority student ranks this high-quality school higher in her preferences than all of the majority students who don't have their assignments yet in some step of the procedure. More generally, the Parallel mechanisms, especially the ones with small preference bands such as the Shanghai mechanism, rely less on the priorities. When a school selects among its applicants in a Parallel mechanism, students who ranked the school in a higher preference class gain priority over students who ranked the school in a lower preference class, and the priorities are only used in student selection for students who rank the school in the same preference class, which explains why minority students get a chance of being assigned to a high-quality school by these mechanisms. This sheds light on the contrasting impact on segregation between the DA and TTC on one hand, and the IA and other Parallel mechanisms with small preference bands on the other hand.

4.6 Homogeneous Priorities and Segregation

We explore next the effect of homogeneous priorities on segregation. A priority profile is homogeneous when all schools have the same ranking of all students. Homogeneous priorities would arise, for example, when there is a centralized test which is the sole criterion for admission decisions, and a high overall correlation is likely in school systems where priorities are based on student performance, even if evaluated independently by different schools.

Formally, $\tilde{\succ} \in \Pi$ is a homogeneous priority profile if for all $c, c' \in C$, $\tilde{\succ}_c = \tilde{\succ}_{c'}$. Let $\tilde{\Pi}$ denote the set of homogeneous priority profiles. Note that due to Assumption 1, when priorities are homogeneous all majority students are ranked above all minority students not just by high-quality schools, but also by low-quality schools.

It follows from Theorems 5 and 6 that the DA and TTC mechanisms are completely segregated, as homogeneous priority profiles are a subset of all priority profiles, and we have shown that regardless of the priority and preference profiles, the DA and the TTC matchings are always completely segregated. Moreover, with homogeneous priorities the DA and the TTC mechanisms both correspond to the Serial Dictatorship mechanism. Kesten (2006) showed that when the priority profile is acyclic (which is satisfied by a homogeneous priority profile), the DA and TTC mechanisms coincide.

It is also easy to see intuitively that the DA mechanism is the same as the SD mechanism for all $\tilde{\succ} \in \tilde{\Pi}$. As the priority of all schools is the same in $\tilde{\succ}$, each school's priorities are given by the same permutation of S, which we can denote by σ. Then when students apply to schools in the DA procedure, each conflict is resolved according to σ, where a conflict means that there is a larger applicant pool for a school than the school's capacity. This implies that each student is assigned her favorite school subject to available school seats, where the availability of school seats is determined following the student ordering σ. This means that the DA procedure simplifies to the SD mechanism based on σ. Furthermore, it was shown by Pápai (2000) in the one-to-one model that the TTC mechanism is the same as the SD mechanism for all $\tilde{\succ} \in \tilde{\Pi}$. The intuition which also applies to the many-to-one case is that with homogeneous priorities each student will inherit all the schools that have available capacity, one at a time, according to the common priority ordering $\tilde{\succ}_c = \sigma$. Thus, in each step the next student in σ can choose a school with available capacity according to her preference by pointing to herself, and hence the TTC procedure also simplifies to the SD mechanism based on σ.

Therefore, for all homogeneous priority profiles $\tilde{\succ} \in \tilde{\Pi}$, for all preference profiles $P \in \mathcal{P}$, $\varphi^{\text{DA}}(P, \tilde{\succ}) = \varphi^{\text{TTC}}(P, \tilde{\succ}) = \varphi^{\text{SD}}(P, \tilde{\succ})$, and it follows that the DA and TTC are not only completely segregated when restricted to homogeneous priorities, but they also coincide with each other and the SD mechanism, which is also completely segregated. On the other hand, the following example demonstrates that the IA mechanism is not completely segregated even under the assumption of homogeneous priorities.

Example 22. Consider the setup in Example 19 and let the preference profile be the same as in Table 4.1. Now consider the homogeneous priority profile in Table 4.2 (Example 21). The IA matching is $\mu^{\text{IA}}_{c_1} = \{a_1, a_2, a_3\}$, $\mu^{\text{IA}}_{c_2} = \{a_7, i_2\}$, $\mu^{\text{IA}}_{c_3} = \{a_6, i_3\}$, and $\mu^{\text{IA}}_{c_4} = \{a_4, a_5, i_1\}$. The IA matching for this example with homogeneous priorities is not completely segregated, as two minority students, i_2 and i_3, are matched to high-quality schools: $d(\varphi^{\text{IA}}(P, \tilde{\succ})) = 2$.

Note also that in this example the DA and the TTC matchings are the same, namely $\mu^{\text{DA}}_{c_1} = \mu^{\text{TTC}}_{c_1} = \{a_1, a_2, a_3\}$, $\mu^{\text{DA}}_{c_2} = \mu^{\text{TTC}}_{c_2} = \{a_5, a_7\}$, $\mu^{\text{DA}}_{c_3} = \mu^{\text{TTC}}_{c_3} = \{a_4, a_6\}$, and $\mu^{\text{DA}}_{c_4} = \mu^{\text{TTC}}_{c_4} = \{i_1, i_2, i_3\}$. Observe that the DA and the TTC matchings are different than in Example 19 and correspond to the SD matching, but still completely segregated: $d(\varphi^{\text{DA}}(P, \tilde{\succ})) = d(\varphi^{\text{TTC}}(P, \tilde{\succ})) = d(\varphi^{\text{SD}}(P, \tilde{\succ})) = 0$. $\diamond$

Due to the immediate acceptance property of the IA mechanism it does not coincide with the SD mechanism, since a student with an arbitrary priority rank will be matched to any school if she applies to the school earlier than some other student. This implies that the IA mechanism does not yield a completely segregated matching for all $(P, \tilde{\succ})$, where $\tilde{\succ} \in \tilde{\Pi}$. Therefore, the IA mechanism is still less segregated than the DA and TTC mechanisms, as illustrated by Example 22. In addition to Example 22, Example 20 shows that the 2-Parallel mechanism may also return a desegregated matching when the priorities are homogeneous. These and similar examples for larger preference bands in Parallel mechanisms, in conjunction with Theorems 5 and 6, imply the following results.

Theorem 7. *For homogeneous priorities the k-Parallel mechanisms for $1 \leqslant k < n$ are less segregated than the DA and the TTC mechanisms. In particular, the IA mechanism $(k = 1)$ is less segregated than the DA and the TTC mechanisms.*

Therefore, we can conclude that having a homogeneous priority profile does not alter the general segregation outcomes of matching mechanisms compared to the broader heterogeneous case. This is good news, since this tells us that the use of centralized or standardized entrance exams, among others, do not have a negative impact on segregation through the use of specific matching mechanisms, although their direct impact on segregation is very likely a different matter.

4.7 Homogeneous Preferences and Segregation

We explore the effect of homogeneous preferences next. The assumption of homogeneous preferences is used as a simplification for highly correlated student preferences over schools. If school quality differs widely and there is consensus about the performance of schools, given that it is an important determinant of the submitted preference orderings of schools, the preference profile is likely to be highly correlated. Different ways of measuring school quality is an issue of ongoing debate, but arguably standardized testscores are frequently viewed as the primary way to evaluate a school's success, and schools are often ranked based on such scores. For example, many US states use standardized tests to evaluate school quality. An empirical analysis from England concluded that testscore measures of school quality dominate parental satisfaction with the school (Gibbons & Silva, 2011). A similar conclusion was reached by Burgess, Greaves, Vignoles, and Wilson (2015), who found that most families have strong preferences for the schools' academic performance which override other variables such as the location and other characteristics. There is much evidence that most of the variation in preferences for school quality across socioeconomic groups arises from differences in the quality of 'accessible schools' rather than differences in parents' preferences. These findings imply that correlated preferences are rather likely in school choice and high school placement if there are substantial discrepancies in the quality of schools. Universities may also differ widely in the quality of the education they offer in most countries, and if the ranking of schools and universities is public knowledge then assuming homogeneous preferences is quite realistic due to the very high degree of correlation in the preferences.

Formally, $\tilde{P} \in \mathcal{P}$ is a homogeneous preference profile if for all $s, s' \in S$, $\tilde{P}_s = \tilde{P}_{s'}$. Let $\tilde{\mathcal{P}}$ denote the set of all homogeneous preference profiles. Note that for all $\tilde{P} \in \tilde{\mathcal{P}}$, all high-quality schools are ranked above all low-quality schools, as specified by Assumption 2, which is true for all students due to the assumption of homogeneous preferences.

Given that the set of homogeneous preferences are a subset of the set of all preferences ($\tilde{\mathcal{P}} \subset \mathcal{P}$), Theorems 5 and 6 imply that the DA and the TTC mechanisms are completely segregated for all $(\tilde{P}, \succ) \in \tilde{\mathcal{P}} \times \Pi$. Now we will show that all k-Parallel mechanisms, including the IA mechanism, are also completely segregated for all $(\tilde{P}, \succ) \in \tilde{\mathcal{P}} \times \Pi$. We prove this by showing that all Parallel mechanisms and the TTC mechanism coincide with the Serial Dictatorship mechanism for schools (not for students) when the preference profile is homogeneous.

Theorem 8. *Given homogeneous preferences, both the k-Parallel mechanisms (including the DA and IA mechanisms) and the TTC mechanism are completely segregated, and each mechanism gives the same matching at every profile.*

Proof. Fix a profile $(\tilde{P}, \succ) \in \tilde{\mathcal{P}} \times \Pi$. We will show that the k-Parallel mechanism with any k such that $1 \leqslant k \leqslant n$ yields the same matching at this profile as the SD^C mechanism with $\rho = \tilde{P}_s$, where $\tilde{P}_s$ is the common preference ordering for all students. Fix $k \in \{1, \ldots, n\}$. In the k-Parallel mechanism in each step all students who are not yet assigned to a school in a previous step apply to the same school c, the most preferred school according to their common preference ordering $\tilde{P}_s$ that has not yet rejected them. Since each student ranks c in the same preference class, school c selects the highest-priority q_c students among the applicants. These are final acceptances, as all remaining students who have not been assigned yet apply to c in this step, and thus it is not possible for a higher-priority student to apply to c in a later step of the procedure. This procedure leads to the SD^C matching, which assigns q_c students who were not yet assigned to a school in a previous step to each school c, following the priority ordering $\succ_c$, according to the school permutation $\tilde{P}_c$. Note that this argument holds for an arbitrary $k \in \{1, \ldots, n\}$. Since the n-Parallel mechanism is the DA, this proves that each Parallel mechanism, including the IA, leads to a completely segregated matching at $(\tilde{P}, \succ)$ due to Theorem 5.

Considering the TTC mechanism, in each step of the procedure all students who were not assigned to a school in a previous step point to the agent who is endowed with school c that is the most preferred school according to their common preference ordering $\tilde{P}_s$ with available seats left. School c is endowed to the highest-ranked student s according to $\succ_c$ who is still unassigned, and hence student s forms a top trading cycle by herself and is assigned to school c. This process continues until the capacity q_c is exhausted, and then all unassigned students point to the student who is endowed with the next most preferred school according to their common preference ranking $\tilde{P}_s$ with available seats left after school c. Therefore, the TTC procedure also leads to the SD^C matching.

To summarize, for all values $1 \leqslant k \leqslant n$, and for all profiles with homogeneous preferences $(\tilde{P}, \succ) \in \tilde{\mathcal{P}} \times \Pi$, $\varphi^{\text{SD}^C}(\tilde{P}, \succ) = \varphi^{\text{PA}^k}(\tilde{P}, \succ) = \varphi^{\text{TTC}}(\tilde{P}, \succ)$, and thus each mechanism yields the same completely segregated matching. $\square$

The following example demonstrates that each mechanism in Theorem 8 is completely segregated when the preferences are homogeneous.

Example 23. Consider the setup of Example19 and the priority profile in Table 4.1. Let $\tilde{P}$ be a homogeneous preference profile with $c_2\ \tilde{P}_s\ c_1\ \tilde{P}_s\ c_3\ \tilde{P}_s\ c_4$, for all $s \in S$. Then all k-Parallel mechanisms, including the DA and the IA, as well as the TTC mechanism yield the same matching as the SD^C mechanism. Specifically, $\mu_{c_1}^{\text{SD}^C} = \{a_4, a_5, a_7\}$, $\mu_{c_2}^{\text{SD}^C} = \{a_2, a_6\}$, $\mu_{c_3}^{\text{SD}^C} = \{a_1, a_3\}$, and $\mu_{c_4}^{\text{SD}^C} = \{i_1, i_2, i_3\}$. This matching is completely segregated. $\diamond$

Our findings provide some complex policy implications that can be summarized as follows. Since highly correlated preferences arise typically when schools differ substantially in quality but not in other attributes, segregation may be mitigated by decreasing the funding and quality differences among schools in different neighborhoods, as school quality is public knowledge and it is well-known to contribute to the similarity of preference rankings of schools by parents. Another way to mitigate segregation is by increasing either the specialization of schools or any other attributes that appeal differently to different individual tastes. Both of these would induce more heterogeneity in the preferences over schools, which is desirable if we are to use mechanisms that rely more on preferences than priorities and thus could lead to more desegregated school placements under more heterogeneous school rankings.

4.8 Summary and Policy Implications

We have shown that the DA and TTC mechanisms are completely segregated for all profiles under the assumptions that high-quality schools rank majority students above minority students and the majority students prefer high-quality schools. Moreover, the IA mechanism is less segregated on the whole and may be completely desegregated at some profiles. We obtain the same results if we consider the set of homogeneous priority profiles only. When we focus on the restricted set of homogeneous preference profiles, however, the IA mechanism is also completely segregated, just like the DA and TTC mechanisms. These findings are summarized in Table 4.3.

The choice of the school matching mechanism depends on the policy objectives. Over the last few decades many school districts have implemented a wide range of school choice

Table 4.3: Segregation impact of different mechanisms

Mechanisms	**Heterogeneous Preferences & Priorities**	**Homogeneous Priorities**	**Homogeneous Preferences**
DA	Segregated	Segregated	Segregated
TTC	Segregated	Segregated	Segregated
IA	Not Segregated	Not Segregated	Segregated

programs to diversify schools and allow for more flexible schooling options. School desegregation appears to be among the policy objectives for opening school choice programs. Despite these efforts, schools remain segregated by race, ethnicity and economic status, among others, in may parts of the world. Roda and Wells (2013) commented that new school enrollment policies in the US are 'color blind,' referring to earlier work by Wells (1993) that pointed out that newer school choice policies are not designed to address issues of racial segregation specifically. Mickelson, Bottia, and Southworth (2008) reported that 'choice schools' in many US cities are often more segregated, and one of the primary reasons is that choice programs formally and informally allow schools to select students. In the context of our formal analysis this indicates the prominence of the schools' priorities over the student's preferences in the school matching process, which generally leads to more segregation according to our findings.

Specifically, in the DA and the TTC mechanisms the priorities of schools play a more consequential role than the reported preferences, while the IA mechanism primarily relies on the preferences of the students. As we have shown, given that good schools' priorities favor socioeconomically advantageous groups, the DA and the TTC mechanisms are prone to school segregation. We argue that, given such an essentially unavoidable priority structure in the absence of any corrective affirmative action policy, the IA mechanism, and more generally Parallel mechanisms with small preference bands, may be a better alternative if desegregation is a real policy objective, not just empty rhetoric. This may explain recent US evidence that schools are becoming more segregated as more and more districts are replacing the IA mechanism with the DA mechanism.

The case of a homogeneous priority profile is meant to capture school placement systems in which the schools' priorities are based on test scores and other performance-based criteria, and this clearly also favors socioeconomically advantageous groups, as discussed in subsection 4.2.1. In such situations the IA mechanism remains preferable to the DA or TTC, as the segregation outcomes are similar to the general case of heterogeneous school rankings. This result appears counterintuitive because the criteria based on achievement benefit students with a higher socioeconomic status. However, our results pertain to the choice of the school choice mechanism once the school priorities are set, and our findings do not imply that performance-based criteria in the selection make no difference compared to other criteria, but rather that if admissions are based on homogeneous priority rankings the preference-based IA and Parallel mechanisms still perform well, as these mechanisms allow for selecting lower-priority students over higher-priority students even under homogeneous priorities.

A more damaging scenario for school segregation is when the preferences are homogeneous or highly correlated. Under such circumstances the DA, TTC, IA and all other Parallel mechanisms are segregated since then each mechanism uses the priority profile for selection, making endowments or for tie-breaking (whichever is applicable) in each step, making all mechanisms select students based on the priority profile and hence are outcome-equivalent. As a result, given that most students prefer good-quality schools, while good-quality schools prefer students from advantageous backgrounds, all mechanisms result in concentrating disadvantaged students in poor-quality schools.

Therefore, one way to mitigate segregation is to reduce the correlation in the preference profile by equalizing the quality of schools, or at least by helping to reduce the funding gap. Preferences would also be more diverse if schools had distinguishing attributes that appeal to different interests, such as specializations. While our simple theoretical results on the impact of school choice mechanisms lead to these implications, these solutions are clearly beyond the scope of our paper and involve complex political and social dynamics. Another politically sensitive but obvious solution would be to use affirmative action policies that are designed specifically to place disadvantaged students in good-quality schools. Lastly, but very importantly, if the underlying causes for school segregation (and for segregation in general) were reduced or eliminated then the school choice mechanism would obviously not have the same importance for the segregation outcome. This would have a much more fundamental impact on school segregation than the choice of the mechanism could ever have, but it is also true that in the absence of such radical societal changes one important variable that may be relatively easy to adjust to potentially great effect is the design of the school choice mechanism.

Chapter 5

Conclusion

The effectiveness of affirmative action policies in many-to-one matching has been discussed in the literature for at least a decade. Many proposed solutions turned out not to be adequate to ensure the welfare improvement of minority groups. The effectiveness of matching mechanisms with affirmative action is evaluated based on many different criteria, including non-wastefulness and minimal responsiveness, incorporating various fairness and incentive properties. We simplified the criteria to three very basic requirements. In addition to non-wastefulness and minimal responsiveness, we introduced a third fundamental criterion, "respecting the affirmative action policy," to evaluate the effectiveness of affirmative action policies. While not a new criterion, imposing it in this simple form highlights this as a basic requirement that any affirmative action policy should satisfy. We believe that mechanisms that satisfy all three axioms provide a more effective affirmative action policy than others that do not, although trade-offs are always present and should be considered. Unfortunately, none of the existing proposed mechanisms satisfy all three axioms. Therefore, we designed the IA-DA-R mechanism that satisfies all three axioms. We introduced the minority fairness condition that only allows minimum priority violations required for affirmative action that focuses on preferential access for underprivileged groups. We also showed that it is impossible for any mechanisms to be strategyproof, minority fair, respect the affirmative action policy, and be minimally responsive. The IA-DA-R mechanism is minority fair but it is not strategyproof for the members of either the minority or the majority group.

Next, we introduced new axioms of responsiveness, in addition to minimal responsiveness which was formally proposed by Kojima (2012) and studied later on by Doğan (2016). Minimal responsiveness is weaker than 'responsiveness,' the strongest of four related axioms. Responsiveness guarantees that no minority student will be worse off with affirmative action when some stronger policy is compared with some weaker policy. We introduced two new mechanisms, namely the Divided DA and the Guaranteed DA, and showed that the Divided DA mechanism is responsive but wasteful, whereas the Guaranteed DA mechanism is non-wasteful but both minimally responsive and elementally responsive. However, it is well-known that strategyproofness cannot be satisfied by most mechanisms with an affirmative action policy that meet other desirable criteria. Trade-offs in the form of possibility and impossibility results involving welfare, fairness, responsiveness and incentive properties are worth investigating, and we hope that this initial study, which is the first one to investigate mechanisms with affirmative action policies that meet stringent criteria of responsiveness, will be followed by further studies establishing more general results.

Finally, we studied the issue of segregation in institutions such as schools, and argued that mechanisms themselves could play a vital role. Previous studies on segregation emphasized mostly the incentive properties, socioeconomic demographics, and laws to explain segregation. Specific policies to promote racial integration worldwide hardly had any notable success. Under two realistic assumptions regarding preferences and priorities, we showed that the two most recommended mechanisms, namely the DA and the TTC, are mechanisms that lead to segregation. In contrast, contrary to popular belief, the IA mechanism induces less segregation. Our stylized model explains the recent paradoxical picture of increasing segregation despite specific efforts to reduce school segregation, and despite the adaptation of strategyproof matching mechanisms observed in many school districts worldwide.

One objective of affirmative action policies is to decrease segregation among schools. However, it is easy to see that affirmative action policies used by existing centralized mechanisms such as the DA and the IA mechanisms have difficulty reaching successful outcomes. These mechanisms with majority quotas and minority reserves cannot guarantee better access for minorities to better institutions as they are either wasteful, or not responsive even in a minimal sense to changes in the policy, or worse yet, don't allow access to all the set-aside positions for minority group members. Thus, a first step towards successfully employing affirmative action policies to alleviate school segregation is to explore the design and properties of mechanisms with affirmative action polices, in order to be more prepared to tackle the very complex issues of school segregation with corrective affirmative action policies, assuming that such policies are politically viable. Apart from the insufficiency of existing mechanisms with an affirmative action policy, another reason for studying school segregation without an affirmative action policy in place, as we have done, is that this may not be politically viable. As we know, affirmative action is highly controversial, and in fact it has been banned in several US states as discriminatory practice and has been restricted in other places. Subject to these practical limitations, we believe that our segregation model and result can not only shed some light on the failures of desegregation efforts but also that, combined with the the mechanisms and attributes of affirmative action policies that we introduced in the earlier chapters, our studies may be able to provide some insight and inspiration for further studies of policy alternatives for policymakers aiming to reduce segregation.

One limitation of our theoretical findings is that they are limited to only two types of agents, majority and minority students. An obvious future research direction is to extend the mechanisms studied in this thesis and its findings to multiple types of underprivileged groups in a more general model. For example, Ehlers et al. (2014) study a model that allows for multiple types of groups, but their focus is not on responsiveness of affirmative action or segregation. Their model could be extended to study these topics as well.

In sum, we believe that our findings are interesting not only theoretically but also that our proposed mechanisms may have some practical implications and relevance. We hope that our studies will contribute to designing effective affirmative action and racial policies and help societies achieve the desired socioeconomic and racial integration.

References

Abdulkadiroğlu, A., Che, Y.-K., Yasuda, Y. (2011). Resolving conflicting preferences in school choice: The "Boston mechanism" reconsidered. *American Economic Review*, *101*(1), 399–410.

Abdulkadiroğlu, A., Pathak, P. A., Roth, A. E., Sönmez, T. (2005). The Boston public school match. *American Economic Review*, *95*(2), 368–371.

Abdulkadiroğlu, A., Sönmez, T. (2003). School choice: A mechanism design approach. *American Economic Review*, *93*(3), 729–747.

Afacan, M. O., Salman, U. (2016). Affirmative actions: The Boston mechanism case. *Economics Letters*, *141*, 95–97.

Alcalde, J., Barberà, S. (1994). Top dominance and the possibility of strategy-proof stable solutions to matching problems. *Economic Theory*, *4*(3), 417–435.

Ayoade, N., Pápai, S. (2023). School choice with preference rank classes. *Games and Economic Behavior*, *137*, 317–341.

Balinski, M., Sönmez, T. (1999). A tale of two mechanisms: student placement. *Journal of Economic Theory*, *84*(1), 73–94.

Basteck, C., Mantovani, M. (2018). Cognitive ability and games of school choice. *Games and Economic Behavior*, *109*, 156–183.

Bischoff, K. (2008). School district fragmentation and racial residential segregation: How do boundaries matter? *Urban Affairs Review*, *44*(2), 182–217.

Burgess, S., Greaves, E., Vignoles, A., Wilson, D. (2015). What parents want: School preferences and school choice. *The Economic Journal*, *125*(587), 1262–1289.

Calsamiglia, C., Martínez-Mora, F., Miralles, A. (2017). Sorting in public school districts under the Boston mechanism. *Working paper 17/10, Division of Economics, University of Leicester*.

Calsamiglia, C., Martínez-Mora, F., Miralles, A. (2021). School choice design, risk aversion and cardinal segregation. *The Economic Journal*, *131*(635), 1081–1104.

Chambers, C. P., Yenmez, M. B. (2017). Choice and matching. *American Economic Journal: Microeconomics*, *9*(3), 126–47.

Chaudhury, M., Pápai, S. (2022). Affirmative action policies in school choice: Immediate versus deferred acceptance. *Working paper*.

Chen, Y. (2017). New axioms for deferred acceptance. *Social Choice and Welfare*, *48*, 393–408.

Chen, Y., Huang, Y., Jiao, Z., Zhao, Y. (2022). A comparison study on responsiveness of three mechanisms to affirmative action in school choice. *Operations Research Letters*, *50*(5), 488–494.

Chen, Y., Kesten, O. (2017). Chinese college admissions and school choice reforms: A theoretical analysis. *Journal of Political Economy*, *125*(1), 99–139.

Darling-Hammond, L. (2007). Race, inequality and educational accountability: The irony of 'no child left behind'. *Race Ethnicity and Education*, *10*(3), 245–260.

Ding, F., Hong, S., Jiao, Z., Luo, X. (2019). Corrigendum to "Affirmative action in school choice: A new solution". *Mathematical Social Sciences*, *97*, 61–64.

Doğan, B. (2016). Responsive affirmative action in school choice. *Journal of Economic Theory*, *165*, 69–105.

Doğan, B., Klaus, B. (2018). Object allocation via immediate-acceptance: Characterizations and an affirmative action application. *Journal of Mathematical Economics*, *79*, 140–156.

Dreyfuss, B., Heffetz, O., Rabin, M. (2022). Expectations-based loss aversion may help explain seemingly dominated choices in strategy-proof mechanisms. *American Economic Journal: Microeconomics*, *14*(4), 515–55.

Dubins, L. E., Freedman, D. A. (1981). Machiavelli and the Gale-Shapley algorithm. *The American Mathematical Monthly*, *88*(7), 485–494.

Ehlers, L., Hafalir, I. E., Yenmez, M. B., Yildirim, M. A. (2014). School choice with controlled choice constraints: Hard bounds versus soft bounds. *Journal of Economic Theory*, *153*, 648–683.

Ehlers, L., Klaus, B. (2016). Object allocation via deferred-acceptance: Strategy-proofness and comparative statics. *Games and Economic Behavior*, *97*, 128–146.

Ellison, G., Pathak, P. A. (2021). The efficiency of race-neutral alternatives to race-based affirmative action: Evidence from Chicago's exam schools. *American Economic Review*, *111*(3), 943–75.

Ergin, H., Sönmez, T. (2006). Games of school choice under the Boston mechanism. *Journal of Public Economics*, *90*(1-2), 215–237.

Featherstone, C. R., Niederle, M. (2011). School choice mechanisms under incomplete information: An experimental investigation. *Unpublished Manuscript, Harvard Business School*.

Featherstone, C. R., Niederle, M. (2016). Boston versus Deferred Acceptance in an interim setting: An experimental investigation. *Games and Economic Behavior*, *100*, 353–375.

Fryer Jr, R. G., Loury, G. C. (2005). Affirmative action and its mythology. *Journal of Economic Perspectives*, *19*(3), 147–162.

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

Gibbons, S., Silva, O. (2011). School quality, child wellbeing and parents' satisfaction. *Economics of Education Review*, *30*(2), 312–331.

Goldfarb, Z. A. (2014, March 5). These four charts show how the SAT favors rich, educated families. *The Washington Post*. `https://www.washingtonpost.com/news/wonk/wp/2014/03/05/these-four-charts-show-how-the-sat-favors-the-rich-educated-families/`.

Gortázar, L., Mayor, D., Montalbán, J. (2020). School choice priorities and school segregation: Evidence from Madrid. *Working Paper 1/2020, SOFI, Stockholm University*.

Hafalir, I. E., Yenmez, M. B., Yildirim, M. A. (2013). Effective affirmative action in school choice. *Theoretical Economics*, *8*(2), 325–363.

Holzer, H. J., Neumark, D. (2006). Affirmative action: What do we know? *Journal of Policy Analysis and Management*, *25*(2), 463–490.

Irwin, V., Zhang, J., Wang, X., Hein, S., Wang, K., Roberts, A., … others (2021). Report on the condition of education 2021. NCES 2021-144. *National Center for Education Statistics*.

Jeong, B. (2018). School choice in context: Can open enrollment cure segregation? *Available at SSRN 4031432*.

Ju, Y., Lin, D., Wang, D. (2018). Affirmative action in school choice: A new solution. *Mathematical Social Sciences*, *92*, 1–9.

Kesten, O. (2006). On two competing mechanisms for priority-based allocation problems. *Journal of Economic Theory*, *127*(1), 155–171.

Kesten, O. (2010). School choice with consent. *The Quarterly Journal of Economics*, *125*(3), 1297–1348.

Kojima, F. (2012). School choice: Impossibilities for affirmative action. *Games and Economic Behavior*, *75*(2), 685–693.

Kojima, F., Ünver, M. U. (2014). The Boston school-choice mechanism: An axiomatic approach. *Economic Theory*, *55*(3), 515–544.

Kozol, J. (1991). *Savage inequalities: Children in america's schools. New York city.* Crown Publishers.

Kutscher, M., Nath, S., Urzua, S. (2020). *Centralized admission systems and school segregation: Evidence from a national reform.* IZA Discussion Paper no. 13305.

Leo, G., Van der Linden, M. (2018). Competing for priorities in school choice. *Available at SSRN 3252745*.

Li, S. (2017). Obviously strategy-proof mechanisms. *American Economic Review*, *107*(11), 3257–3287.

Logan, T. D., Parman, J. M. (2017). The national rise in residential segregation. *The Journal of Economic History*, *77*(1), 127–170.

Mader, N., Hemphill, C., Abbas, Q. (2018). The paradox of choice: How school choice divides New York city elementary schools. *Center for New York City Affairs*.

Mickelson, R. A. M., Bottia, M. A. S., Southworth, M. A. (2008). *School choice and segregation by race, class, and achievement.* `http://nepc.colorado.edu/files/CHOICE-08-Mickelson-FINAL-EG043008.pdf`. (Accessed: 2022-12-7)

Pápai, S. (2000). Strategyproof assignment by hierarchical exchange. *Econometrica*, *68*(6), 1403–1433.

Pápai, S., Sayedahmed, D. (2022). Targeted priority reserve policies. *Working Paper*.

Pathak, P. A., Sönmez, T. (2008). Leveling the playing field: Sincere and sophisticated players in the Boston mechanism. *American Economic Review*, *98*(4), 1636–52.

Ponzini, R. (2022). Public school funding in the United States and its systemic inequities. *Economia Aziendale Online*, *13*(1), 143–147.

Reardon, S. F. (2013, April 27). No rich child left behind. *The New York Times*. `https://archive.nytimes.com/opinionator.blogs.nytimes.com/2013/04/27/no-rich-child-left-behind/`.

Roda, A., Wells, A. S. (2013). School choice policies and racial segregation: Where white parents' good intentions, anxiety, and privilege collide. *American Journal of Education*, *119*(2), 261–293.

Roth, A. E. (1982). The economics of matching: Stability and incentives. *Mathematics of Operations Research*, *7*(4), 617–628.

Satterthwaite, M. A., Sonnenschein, H. (1981). Strategy-proof allocation mechanisms at differentiable points. *The Review of Economic Studies*, *48*(4), 587–597.

Shapley, L., Scarf, H. (1974). On cores and indivisibility. *Journal of Mathematical Economics*, *1*(1), 23–37.

Slavin, R. E. (1999). How can funding equity ensure enhanced achievement? *Journal of Education Finance*, *24*(4), 519–528.

Smith, E., Reeves, R. (2020, Dec). *SAT math scores mirror and maintain racial inequity.* `https://www.brookings.edu/blog/up-front/2020/12/01/sat-math-scores-mirror-and-maintain-racial-inequity/`. (Accessed: 2022-9-24)

Smith, K. B., Meier, K. J. (1995). Public choice in education: Markets and the demand for quality education. *Political Research Quarterly*, *48*(3), 461–478.

Tang, Q., Yu, J. (2014). A new perspective on Kesten's school choice with consent idea. *Journal of Economic Theory*, *154*, 543–561.

Terrier, C., Pathak, P. A., Ren, K. (2021). From immediate acceptance to deferred acceptance: Effects on school admissions and achievement in England. Discussion paper no. 1815. *Centre for Economic Performance*.

Thomson, W. (2011). Consistency and its converse: An introduction. *Review of Economic Design*, *15*(4), 257–291.

Troyan, P., Morrill, T. (2020). Obvious manipulations. *Journal of Economic Theory*, *185*, 104970.

Valenzuela, J. P., Bellei, C., Ríos, D. d. l. (2014). Socioeconomic school segregation in a market-oriented educational system. The case of Chile. *Journal of Education Policy*, *29*(2), 217–241.

Wells, A. S. (1993). *Time to choose: America at the crossroads of school choice policy. New frontiers of education.* Putnam Publishing Group, East Rutherford, NJ 07073.

Yang Hansen, K., Gustafsson, J.-E. (2016). Causes of educational segregation in Sweden–school choice or residential segregation. *Educational Research and Evaluation*, *22*(1-2), 23–44.

Zhu, M. (2014). College admissions in China: A mechanism design perspective. *China Economic Review*, *30*, 618–631.