Affirmative Action Policies in School Choice: Immediate versus Deferred Acceptance[*]

Muntasir Chaudhury[†] Szilvia Pápai[‡]

May 17, 2024

Abstract

We study three basic welfare axioms for school choice mechanisms with a reserve or quota-based affirmative action policy, namely non-wastefulness, respecting the affirmative action policy, and minimal responsiveness, and show that none of the previously proposed mechanisms satisfy all of them. Then we introduce a new mechanism which satisfies these three axioms. This mechanism issues immediate acceptances to minority students for minority reserve seats and otherwise it employs deferred acceptance. We analyze the fairness and incentive properties of this newly proposed affirmative action mechanism and provide possibility and impossibility results which highlight the trade-offs.

Keywords: school choice, affirmative action, minority reserves, non-wastefulness, minimal responsiveness, deferred acceptance, immediate acceptance, priority violations, strategyproofness

JEL classification: C78, D47, D63, D78

[*] We would like to thank Mustafa Afacan, Haris Aziz, Battal Doğan, Lars Ehlers, Sean Horan, Michihiro Kandori, Fuhito Kojima, Pinaki Mandal, Vikram Manjunath, Ariel Procaccia, Alex Teytelboym, and William Thomson for their constructive feedback that have led us to further explorations. We would also like to acknowledge numerous seminar and conference audiences that offered both positive and critical comments.

[†] East West University, Bangladesh; e-mail: muntasir@ewubd.edu

[‡] Concordia University and CIREQ; e-mail: szilvia.papai@concordia.ca

1 Introduction

Affirmative action policy at educational institutions, 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 other attributes have generated philosophical and political debates. On the other hand, the effectiveness of preferential treatment administered by centralized matching mechanisms has been explored over the last decades. This study focuses on the latter from an axiomatic perspective. We analyze quota-based and reserve-based affirmative action policies in a many-to-one matching model. Quotas and reserves have been used widely in university admissions and school assignment to promote diversity and help disadvantaged groups.

1.1 Background

Gale and Shapley (1962) introduced the celebrated Deferred Acceptance (DA) mechanism which is strategyproof (Dubins and Freedman, 1981; Roth, 1982), thus students cannot gain from misrepresenting their true preferences. In the DA algorithm students are only accepted tentatively in each step of the iterative algorithm (hence the name "deferred" acceptance), which ensures that students who apply in a later step can still be accepted if they have a higher priority at that school, and thus the DA mechanism always yields a matching which respects the students' priorities at each school. One affirmative action policy that can easily be incorporated by the DA mechanism is based on type-specific quotas at schools. Abdulkadiroğlu and Sönmez (2003) discussed such type-specific quotas for the DA mechanism with a fixed quota for students of each type which cannot be exceeded by the matching. To analyze affirmative action policies, in our simple model the students from under-represented racial, ethnic, religious or economically disadvantaged groups are the *minority students,* while the rest of the students are the *majority students*. Thus, in the context of affirmative action, there is a quota at each school for majority students only. This leads to what we call the DA with Majority Quotas mechanism (DA-Q) which, apart from complying with the majority quotas, follows the same iterative procedure as the standard DA.

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 student applicants. This wastefulness, which only affects majority students, makes the affirmative action policy more controversial. To remedy this problem, Hafalir et al. (2013) introduced a flexible quota system called minority reserves. Schools give priority to 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-based systems. In the DA with Minority Reserve (DA-R) mechanism of Hafalir et al. (2013), schools first tentatively assign minority reserve seats to minority applicants before filling all the remaining seats (including unused minority reserve seats) from the general applicant pool. The minority reserve policy of the DA-R mechanism successfully eliminates the wastefulness of the quota-based policy of DA-Q.

Another issue of effectiveness, which was first raised by Kojima (2012), is the responsiveness of the welfare of minority students to a strengthened affirmative action policy. A mechanism is called *minimally responsive* to an affirmative action policy (what Kojima (2012) called "respecting the spirit of affirmative action") if a stronger affirmative action policy - decreased majority quotas or increased minority reserves - does not lead to a Pareto-dominated outcome for minority students compared to the outcome before. Kojima (2012) demonstrated that under any quota-based mechanism that satisfies a stability condition consistent with majority quotas, it is always possible to find certain preference and priority profiles at which lower majority quotas do not help any minority student and may even harm some. Not only the DA-Q mechanism is not minimally responsive, but the same is true for the DA-R mechanism (Hafalir et al., 2013). Addressing this issue, Doğan (2016) introduced the Modified DA with Minority Reserves (MDA) mechanism, which achieves minimal responsiveness by treating some minority students at certain schools as majority students in iterations of the DA-R mechanism that resemble the efficiency improvements of EADAM (Kesten, 2010) in some respects. Another mechanism which uses efficiency improvements, called the Efficiency Improved DA with Minority Reserves (EIDA) mechanism, was proposed by Ju et al. (2018), but it is not minimally responsive (Ding et al., 2019).

The Immediate Acceptance (IA) mechanism, originally known as the Boston mechanism (Abdulkadiroğlu and Sönmez, 2003; Kojima and Ünver, 2014), has been used frequently in the US and around the world to match students to schools. It is still a popular student placement mechanism, although in the past two decades many school choice programs started opting for the DA instead due to its better normative and incentive properties. In the IA mechanism students are accepted permanently (or "immediately") in each step of the iterative procedure, which is a crucial difference from the tentative acceptances of the DA. The IA mechanism is neither strategyproof nor fair, since it is manipulable and the schools' priorities over students are not respected by the mechanism. However, as opposed to the DA, it is Pareto-efficient for the students. Affirmative action policies can also be implemented in conjunction with immediate acceptances. Afacan and Salman (2016) analyzed a hybrid immediate acceptance mechanism which allows for both majority quotas and minority reserves. Other papers that consider affirmative action policies under the IA mechanism are Chen et al. (2022), and to a lesser extent Doğan and Klaus (2018). The IA mechanisms with a quota or reserve-based affirmative action policy have distinct properties compared to their DA counterparts. Most notably, the IA-R mechanism (the IA-based mechanism with minority reserves) is minimally responsive (Afacan and Salman, 2016), unlike DA-R. This is an interesting finding which justifies looking at the discredited IA mechanism again in the context of affirmative action policies. However, the IA-R mechanism lacks some further good attributes that other mechanisms, including the IA-Q mechanism (the IA-based mechanism with majority quotas), possess.

In addition to the already cited closest papers to ours, there are many other recent studies on affirmative action or similar reserve policies. These papers tend to study either choice rules with reserves or affirmative action policies and related applications in specific contexts, or sometimes both. There are too many to cite all of them, but see for example Westkamp (2013), Echenique and Yenmez (2015), Kominers and Sönmez (2016), Dur et al. (2018), Aziz et al. (2020), Dur et al. (2020), Abdulkadiroğlu and Grigoryan (2021), Aygün and Bó (2021), Aziz and Sun (2021), Pathak et al. (2022), Sönmez and Yenmez (2022) and Pathak et al. (2023).

1.2 Overview

We initially 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 mild efficiency property: it eliminates the possibility that a school seat remains empty when it is preferred by any student to her school assignment, regardless of whether the student is a minority or majority student. Respecting the affirmative action policy is fundamental for any mechanism with affirmative action, since it requires that school seats set aside for minority students be filled with minority students, as long as there is any minority student who desires them. As already discussed, minimal responsiveness ensures that a stronger affirmative action policy benefits at least one minority student whenever any minority student is affected by the change.

We consider these three axioms basic welfare requirements for a mechanism with a quota or reserve-based affirmative action policy. Surprisingly, none of the above described six affirmative action mechanisms which have been proposed in the literature satisfy all three of them. Given some well-known impossibilities in the axiomatic matching theory literature, one might suspect that we cannot satisfy all three axioms simultaneously. However, this is not the case, as we will show. 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. Specifically, minority reserve positions are filled with minority applicants based on immediate acceptances, while all other acceptances are tentative, as in the DA. We study the fairness (stability) and incentive properties of the proposed IA-DA-R mechanism, and we explore more generally the class of mechanisms that satisfy the three welfare axioms together with a simple fairness axiom that we introduce. Our analysis identifies some interesting compatibilities and trade-offs, as demonstrated by the possibility and impossibility results that we present.

2 Affirmative Action with Quotas and Reserves

2.1 Model

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; $S^M \cap S^m = \emptyset$ and $S^M \cup S^m = S$. For notational clarity, we denote majority students by $a \in S^M$ and minority students by $i \in S^m$. There is a finite set of schools C and each school $c \in C$ has a capacity $q_c \geq 1$. Let $q = (q_c)_{c \in C}$. To avoid trivialities, we assume that $|S^M| \geq 3$, $|S^m| \geq 3$ and $|C| \geq 4$.

Each student $s \in S$ has a strict preference ordering P_s over $C \cup \{0\}$. School 0 denotes the "null school" and represents staying unassigned. If a student ranks a school below 0, this school is considered unacceptable to the student. We assume that $q_0 = |S|$. Let R_s denote the weak counterpart of P_s, that is, $c \; R_s \; c'$ if and only if either $c \; P_s \; c'$ or $c = c'$. We will sometimes specify a preference ordering in a list form, e.g., $P_s : (c_2, c_1, 0)$ means that c_2 is ranked first, c_1 is ranked second, and there are no other acceptable schools for s. Let a preference profile be denoted by P, where $P = (P_s)_{s \in S}$, 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 $\succ$, where $\succ = (\succ_c)_{c \in C}$, and let Π be the set of all priority profiles. When S, C and q are fixed, a problem is given simply by $(P, \succ) \in \mathcal{P} \times \Pi$, consisting of a preference profile and a priority profile. In the following, we refer to a pair $(P, \succ)$ as a *profile.*

Given fixed S, C and q, a **matching** μ is a function from the set of students S 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 allows for any student to remain unassigned. For all $s \in S$ we denote the assignment that student s receives in matching μ by μ_s, where $\mu_s \in C \cup \{0\}$. Moreover, abusing notation, for all $c \in C$ we denote the set of students assigned to school c in matching μ by μ_c. Hence, $\mu_c \subseteq S$ and $|\mu_c| \leq q_c$. We will also use the notation μ_c^m to denote the set of minority students matched to c in μ. Let $\mathcal{M}$ denote the set of matchings.

2.2 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.

Majority Quotas

A majority quota policy determines the maximum number of majority students that can be assigned to each school. Quotas were studied by Abdulkadiroğlu and Sönmez (2003) and later analyzed by Kojima (2012), among others. Let a majority quota policy be denoted by $q^M = (q_c^M)_{c \in C}$, where q_c^M is the majority quota at school c which satisfies $0 \leq q_c^M \leq q_c$ for all $c \in C$. Note that $q^M = q$ corresponds to no affirmative action policy.

Minority Reserves

A minority reserve policy specifies a number of seats at each school for which the minority students are prioritized over majority students. An essential 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. Minority reserve policies were first proposed by Hafalir et al. (2013) in order to eliminate the wastefulness of majority quota policies in the DA. Let a minority reserve policy be denoted by $r = (r_c)_{c \in C}$, where r_c is the number of reserved seats at school c which satisfies $0 \leq r_c \leq q_c$ for all $c \in C$. Note that $r = (0, \ldots, 0)$ corresponds to no affirmative action policy.

Common Framework: Minority Allotments

The two affirmative action policies, majority quotas and minority reserves, 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 types of affirmative action policies and hence allow us to evaluate them in a common framework. We call this unified notion of the affirmative action policies **minority allotments**, which refer to the seats that are set aside for minority students due to majority quotas or minority reserves.

Let a minority allotment policy be denoted by $v = (v_c)_{c \in C}$, where v_c is the number of minority allotment seats at school c. For each school $c \in C$, if the minority allotment policy is a majority quota policy then $v_c = q_c - q_c^M$, and if it is a minority reserve policy then $v_c = r_c$. Feasibility requires in both cases that, for all $c \in C$, v_c satisfies $0 \leq v_c \leq q_c$. Let $\mathcal{V}$ denote the set of feasible minority allotment policies.

Given fixed S, C and q, a **mechanism (with minority allotments)** assigns a matching μ to each minority allotment policy v and profile $(P, \succ)$, defined as $\varphi : \mathcal{V} \times \mathcal{P} \times \Pi \rightarrow \mathcal{M}$. We refer to a minority allotment policy v together with a profile $(P, \succ)$ as an *aa-profile* (where aa stands for affirmative action) and thus a mechanism φ assigns a matching to each aa-profile. Furthermore, we refer to $(S, C, q, v, P, \succ)$ as a **market**, where S, C and q are fixed, but the aa-profile may vary. We denote the assignment of student s at aa-profile $(v, P, \succ)$ by $\varphi_s(v, P, \succ)$, and the assignments of a set of students $\hat{S} \subset S$ by $\varphi_{\hat{S}}(v, P, \succ)$. To ease the notation, in the following when we refer to all aa-profiles $(v, P, \succ) \in \mathcal{V} \times \mathcal{P} \times \Pi$, we will write simply all $(v, P, \succ)$.

2.3 Properties of Mechanisms

A basic property of a mechanism φ is **individual rationality,** which requires that a student is never assigned to a school that is unacceptable to her: for all $(v, P, \succ)$ and $s \in S$, $\varphi_s(v, P, \succ) \; R_s \; 0$. All the mechanisms that we study in this paper satisfy individual rationality.

A matching ν is **Pareto-dominated** by matching μ at $(v, P, \succ)$ 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 $(v, P, \succ)$ if it is not Pareto-dominated at $(v, P, \succ)$. A mechanism is **Pareto-efficient** if it assigns to each aa-profile a Pareto-efficient matching at that aa-profile.

The priority of student s is said to be violated at school c in μ at $(v, P, \succ)$ if there exists student s' such that $c \; P_s \; \mu_s$, $\mu_{s'} = c$, and $s \succ_c s'$. We will say in this case that **s' violates the priority of s** at school c in μ at the given aa-profile. A matching in which no student violates another student's priority at any school at $(v, P, \succ)$ is a fair matching at $(v, P \succ)$. A mechanism is **fair** if it assigns to each aa-profile a fair matching at this aa-profile.

A mechanism φ is **strategyproof** if for all students $s \in S$, for all aa-profiles $(v, P, \succ)$, and all alternative preference orderings P'_s for student s, $\varphi_s(v, P, \succ) \; R_s \; \varphi_s(v, (P'_s, P_{-s}), \succ)$, where P_{-s} stands for $P_{S\setminus\{s\}}$. Otherwise, if a student has a preference ordering P'_s such that $\varphi_s(v, (P'_s, P_{-s}), \succ) \; P_s \; \varphi_s(v, P, \succ)$, then we will say that students s can **manipulate** φ at $(v, P, \succ)$ and P'_s is a **manipulation strategy** for s when the true preference ordering of s is P_s. 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 aa-profile.

3 Main Welfare Axioms

Now we define and discuss the three key axioms that we use to assess the performance of school choice mechanisms with minority allotments. All three of the axioms are welfare criteria: non-wastefulness is a general requirement, while the other two consider welfare properties of the affirmative action policy. We consider each of these axioms a minimal requirement when evaluating mechanisms with affirmative action policies that set aside seats for minority students.

3.1 Non-Wastefulness

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 aa-profile. Wastefulness is a serious drawback for a mechanism, as it means that valuable resources may be wasted, and thus *non-wastefulness* is a basic efficiency requirement.

Non-Wastefulness. A mechanism φ is non-wasteful if for all $(v, P, \succ)$, $s \in S$ and $c \in C \cup \{0\}$, if $c \; P_s \; \varphi_s(v, P, \succ)$ then $|\mu_c| = q_c$, where $\varphi(v, P, \succ) = \mu$.

Non-wastefulness holds for most matching mechanisms of interest without affirmative action, but the axiom becomes more difficult to satisfy when an affirmative action policy or similar distributional constraints are imposed, even if non-wasteful matchings exist that comply with the distributional objectives. In our setting it is not necessary to impose a hard upper bound on majority student assignments, since the objective is to prioritize minority students over majority students for some seats, not to limit the number of assigned majority students, and thus non-wasteful matchings that satisfy the affirmative action objectives exist. Although quota mechanisms are simple and hence popular, the quotas mean hard upper limits which imply wastefulness.[1] Non-wastefulness is often required as a part of fairness (stability) axioms for mechanisms with affirmative action (see, for example, Hafalir et al. (2013) and Doğan (2016)).

A quick note on the connection between non-wastefulness and individual rationality is in order. Since $q_0 = |S|$, the definition of non-wastefulness implies individual rationality. We do not explicitly require individual rationality in our analysis, but all the mechanisms studied in this paper satisfy it, including the wasteful ones.

3.2 Respecting the Affirmative Action Policy

We say that a matching mechanism with a minority allotment policy v *respects the affirmative action policy* if the matching assigned to all aa-profiles is such that there is no minority student who prefers a school c to her assignment in this matching, while at the same time school c has fewer than v_c minority students assigned to it.

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

This axiom simply asks that minority students be prioritized for the school seats set aside for them by the minority allotment policy, and thus it is an essential stipulation 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 that allow for affirmative action. We propose it separately as one of our main axioms because not all affirmative action mechanisms proposed in the literature satisfy it.

3.3 Minimal Responsiveness

Minimal responsiveness (to an affirmative action policy) requires that at least one minority student gains if the affirmative action policy is strengthened, assuming that the

[1] See Ehlers et al. (2014) for an analysis of both hard and soft bounds.

policy change affects the outcome for any minority student. This axiom was first proposed for quota-based affirmative action policies by Kojima (2012). After Hafalir et al. (2013) introduced the concept of a minority reserve policy, Kojima's axiom was extended to reserved-based affirmative action policies by Doğan (2016). Complementing the findings on general priority structures, Doğan (2016) and Chen et al. (2022) provide further results on the minimal responsiveness of affirmative action mechanisms when priority profiles are restricted, while Jiao and Tian (2019) study a stronger responsiveness axiom.

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

Minimal Responsiveness. A mechanism φ is minimally responsive if, for all $v, v' \in \mathcal{V}$ such that $v' \geq v$, and for all profiles $(P, \succ)$ such that $\varphi_{S^m}(v, P, \succ) \neq \varphi_{S^m}(v', P, \succ)$, there exists $i \in S^m$ such that $\varphi_i(v', P, \succ)\ P_i\ \varphi_i(v, P, \succ)$.

Minimal responsiveness stipulates that a weakly stronger minority allotment policy does not result in a Pareto-dominated outcome for minority students when compared to the original outcome. This appears to be a basic intuitive requirement for any type of affirmative action policy, yet it is not easy to satisfy. This axiom is in the spirit of *resource monotonicity* (Luce and Raiffa, 1957),[2] a solidarity property which requires that an increase in resources only benefit (not harm) the agents who receive the resources, while minimal responsiveness considers, naturally, the impact of an increase in minority allotments on minority students only. However, minimal responsiveness is not only less stringent in its welfare implication than resource monotonicity (since it does not require Pareto-improvement for the minority students, it only calls for avoiding a Pareto-inferior outcome for them) but it is also even more compelling, given that the explicit objective of affirmative action is to increase the representation of minority students and hence benefit them.

4 Affirmative Action Mechanisms and their Compliance with the Main Welfare Axioms

We first evaluate the performance of the different mechanisms with minority allotment policies that have been studied in the literature based on the three main axioms of non-

[2] For papers that study resource monotonicity in the matching context see, for example, Ehlers and Klaus (2003) and Kojima and Ünver (2014).

wastefulness, respecting the affirmation action policy, and minimal responsiveness. The definition of each mechanism is provided in Appendix A.

4.1 DA Mechanisms with Minority Allotments

The DA with Majority Quotas (DA-Q) mechanism was first 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 other DA-based mechanism, the DA with Minority Reserves (DA-R) was proposed by Hafalir et al. (2013). It should be noted that both DA-Q and DA-R simplify to the standard DA algorithm when $q^M = q$ and $r = (0, \dots, 0)$, respectively, that is, without affirmative action.

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. 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 tentatively, as the assignments are not permanent before the mechanism terminates. Therefore, the DA-R mechanism also respects the affirmative action policy, and the stability notion 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 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.*

The following example illustrates why the DA-Q and DA-R mechanisms are not minimally responsive, namely, due to the possibility of rejection cycles in the iterative algorithms.

Example 1. DA-Q and DA-R are not minimally responsive.[3]
Let $S^M = \{a_1, a_2\}$ and $S^m = \{i_1, i_2\}$. Let $C = \{c_1, c_2, c_3, c_4\}$ with capacities $q = (1, 1, 1, 1)$.

Table 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

Consider profile $(P, \succ)$ in Table 1. If there is no affirmative action ($v = (0, 0, 0, 0)$), the DA-Q and DA-R matchings both coincide with the DA matching which is given by (c_2, c_1, c_3, c_4)[4] at $(v, P, \succ)$ (underlined in Table 1). Now consider the (stronger) minority allotment policy $\tilde{v} = (1, 0, 0, 0)$. The steps of the DA-R procedure are displayed in Table 2. The resulting DA-R matching at $(\tilde{v}, P, \succ)$ is (c_3, c_2, c_1, c_4), as seen from the final step (step 6) in Table 2 (and also indicated by the squares in Table 1).

One of the minority students (i_1) is worse off when the minority allotment policy is $\tilde{v}$, 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), and this starts a rejection cycle 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 $\tilde{v}$ is used, compared to no affirmative action.

This example demonstrates not only that the DA-R mechanism is not minimally responsive, but also that the DA-Q mechanism suffers from the same problem, since at

[3] The specific markets used in the examples and some of the proofs can easily be generalized to markets of arbitrary size by embedding them in a larger market, and thus we omit these straightforward constructions.

[4] Throughout the paper, the assignments of students are listed in the order in which the students appear in the preference profile specified in a table.

Table 2: Steps of the DA-Q/DA-R mechanism with $\tilde{v}$ in Example 1

	$\boldsymbol{c_1}$	$\boldsymbol{c_2}$	$\boldsymbol{c_3}$	$\boldsymbol{c_4}$
step 1	~~a_2~~ i_2	a_1	i_1	
step 2	i_2	~~a_1~~ a_2	i_1	
step 3	i_2	a_2	a_1 ~~i_1~~	
step 4	i_1 ~~i_2~~	a_2	a_1	
step 5	i_1	a_2 ~~i_2~~	a_1	
step 6	i_1	a_2	a_1	i_2

Struck out students are rejected in the corresponding step.

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. $\diamond$

4.2 Efficiency Improved DA Mechanisms with Minority Allotments

Doğan (2016) proposed the Modified DA with Minority Reserves (MDA) mechanism which is minimally responsive. The MDA mechanism iteratively modifies the DA-R mechanism based on the concept of interferers which resembles Kesten's concept of interrupters for EADAM (Kesten, 2010). Instead of general efficiency improvements, MDA is designed to be minimally responsive to the minority reserve policy, which is achieved by treating relevant minority student interferers identified in specific steps of the DA-R algorithm as majority students at some schools in the improvement rounds of the algorithm. Ju et al. (2018) introduced another mechanism, the Efficiency Improved DA with Minority Reserves (EIDA) mechanism, which performs exact EADAM-style efficiency improvement rounds over the DA-R mechanism. Ju et al. (2018) utilize the simplified EADAM definition introduced by Tang and Yu (2014) in the description of their proposed mechanism. Both MDA and EIDA are non-wasteful. MDA is minimally responsive, as proved by Doğan (2016); this attribute of the mechanism was the main motivation for proposing MDA. EIDA, on the other hand, has been shown by Ding et al. (2019) to fail minimal responsiveness. Surprisingly, neither of the two mechanisms satisfy the axiom of respecting the affirmative action policy, which we demonstrate next by Example 2. For both mechanisms the intuitive reason for this is that in the improvement steps minority students are prevented from initiating a rejection cycle when applying to a school that initially assigns them to a reserved position but rejects them later on. This results in efficiency improvement but at the same time disregards the main intent of the affirmative action policy.

Example 2. MDA and EIDA do not respect the affirmative action policy. Let $S^M = \{a\}$ and $S^m = \{i_1, i_2\}$. Let $C = \{c_1, c_2, c_3\}$ with capacities $q = (1, 1, 1)$, and let $r = (1, 0, 0)$.

Table 3: Profile for Example 2

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
$\underline{c_3}$	$\underline{c_1}$	$\underline{\underline{\boxed{c_2}}}$	i_1	i_2	i_1
c_2	0	c_3	i_2	i_1	i_2

Consider profile $(P, \succ)$ in Table 3. When the minority reserve policy is r, minority student i_2 is an interferer for school c_1 in the DA-R matching at this profile (underlined in Table 3), and the second round of the DA-R algorithm, with the modification that i_2 is considered a majority student at school c_1, yields the MDA matching (c_1, c_3, c_2) (indicated by the squares in the table). Since the minority reserve is $r_{c_1} = 1$ at school c_1 and $c_1 \; P_{i_2} \; c_2$, while majority student a is assigned to c_1, MDA does not respect the affirmative action policy.

In this example EIDA yields the same matching at $(r, P, \succ)$ 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. ◇

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.*

4.3 IA Mechanisms with Minority Allotments

Ergin and Sönmez (2006) introduced the Boston mechanism with type-specific quotas, which is an IA mechanism with a fixed quota for each type of students. The IA-Q mechanism is an adaptation of this mechanism that only has a type-specific quota for majority students, similar to the DA-Q mechanism but uses immediate acceptances instead of deferred acceptances. Afacan and Salman (2016) analyzed an immediate acceptance mechanism with both quotas and reserves, and showed that the IA-Q mechanism is not minimally responsive. We demonstrate this finding with the next example.

Example 3. IA-Q is not minimally responsive.
Let $S^M = \{a_1, a_2\}$ and $S^m = \{i\}$. Let $C = \{c_1, c_2, c_3\}$ with capacities $q = (1, 1, 1)$.

Table 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

Consider profile $(P, \succ)$ in Table 4. With no affirmative action ($q^M = (1, 1, 1)$), minority student i is matched to school c_3 at the specified profile by the IA-Q mechanism (see the underlined matching in Table 4). With the stronger affirmative action policy $\tilde{q}^M = (1, 0, 1)$, minority student i is unassigned by the IA-Q mechanism at the same profile (indicated by the squares in Table 4). Since i is the only minority student, this shows that the IA-Q mechanism is not minimally responsive. ◇

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 no additional majority student can be assigned to this seat.

Respects the affirmative action policy: The IA-Q mechanism does not assign majority students to $v_c = q_c - q_c^M$ seats at any school c, since it never assigns majority students in excess of the majority quota. At the same time, it does not reject a minority applicant from a school with empty seats. Therefore, the IA-Q mechanism respects the affirmative action policy.

Not minimally responsive: See Example 3. □

We explore next the IA with Minority Reserve (IA-R) mechanism, which corresponds to the special case of the hybrid mechanism studied by Afacan and Salman (2016) where there are no majority quotas. This is also the same as the IA Mechanism with Affirmative-Action-Target introduced by Doğan and Klaus (2018).

Example 4. IA-R does not respect the affirmative action policy.
Let $S^M = \{a_1, a_2\}$ and $S^m = \{i\}$. Let $C = \{c_1, c_2, c_3\}$ with capacities $q = (1, 1, 1)$.

Table 5: Profile for Example 4

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_2	a_1
c_3	0	c_2	i	a_1	a_2
0		$\boxed{c_3}$	a_1	i	i

Consider Profile $(P, \succ)$ in Table 5. The IA-R matching at this profile with minority reserves $r = (0, 1, 0)$ is (c_2, c_1, c_3) (as indicated by the squares in Table 5). Given that the minority reserve is $r_{c_2} = 1$ at school c_2 and $c_2 \; P_i \; c_3$, while majority student a_1 is assigned to the only seat at c_2, it follows that the IA-R mechanism does not respect the affirmative action policy. ◇

The most remarkable feature of the IA-R mechanism is that it is minimally responsive, as shown by Afacan and Salman (2016), which contrasts interestingly with the fact that the DA-R mechanism is not minimally responsive. However, the IA-R mechanism does not respect the affirmative action policy, as shown by Example 4, because it allows majority students to fill unoccupied minority reserve seats permanently, due to the immediate acceptances. Interestingly, the IA-Q mechanism 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, but at the same time this feature leads to wastefulness. This differs from the DA mechanisms for which a reserve-based affirmative action policy is a clear improvement over a quota-based policy.

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

Proof.

Non-wasteful: The IA-R algorithm fills as many remaining seats as possible up to its capacity q_c in each step, by first accepting minority students for remaining minority reserve seats and then by accepting any remaining applicants for the remaining seats, and all acceptances are permanent.

Does not respect the affirmative action policy: See Example 4.

Minimally responsive: See Afacan and Salman (2016). □

4.4 Summary of the Previous Affirmative Action Mechanisms

As we have just shown, none of the previous mechanisms satisfy all three of the main welfare axioms.[5] Table 6 summarizes the findings of Propositions 1-6.

Table 6: Comparison of mechanisms

Mechanism	**Welfare Axioms**		
	Non-Wasteful	*Respects AA*	*Minimally Responsive*
DA-Q		✓	
DA-R	✓	✓	
MDA	✓		✓
EIDA	✓		
IA-Q		✓	
IA-R	✓		✓

[5] TTC-based affirmative action mechanisms have also been studied in the literature. We omitted these from our detailed analysis since they are less relevant for the paper, given the focus on immediate and deferred acceptance, but nevertheless the two TTC-based affirmative action mechanisms do not meet all three of our welfare axioms either. The TTC-Q mechanism (Kojima, 2012), which imposes quotas, has similar properties to DA-Q and IA-Q: it respects the affirmative action policy, but it is wasteful and not minimally responsive (for the latter, see Kojima (2012) or Example 3 in which IA-Q and TTC-Q yield the same matchings). The TTC-R mechanism with minority reserves, defined by Hafalir et al. (2013), while non-wasteful and respects the affirmative action policy, does not satisfy minimal responsiveness, just like DA-R (the latter can also be demonstrated by means of a simple example: see Chen et al. (2022)).

If we consider minimal responsiveness the least important axiom among the three welfare criteria, DA-R emerges as the best one among the six mechanisms. If, on the other hand, respecting the affirmative action is less important than minimal responsiveness, then MDA and IA-R stand out. The DA-R mechanism is unambiguously preferable to the DA-Q mechanism, as it dispenses with the wastefulness of DA-Q and it still respects 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, while at the same time it is minimally responsive in contrast to DA-R. Among the four DA-based or IA-based mechanisms only IA-R satisfies minimal responsiveness, but the IA-R mechanism does not respect the affirmative action policy which is in some sense an even more fundamental requirement than minimal responsiveness. The same goes for MDA which, despite being minimally responsive, is arguably less appealing than DA-R because it does not respect the affirmative action policy.

4.5 On Manipulability

An important issue to address when comparing different types of mechanisms based on the DA, IA and EADAM is their different vulnerability to manipulation. While the DA-based DA-Q and DA-R mechanisms are strategyproof, IA-based mechanisms can be shown not only to be manipulable but also obviously manipulable, with EADAM somewhere between them, as it is manipulable but not obviously manipulable (Troyan and Morrill, 2020). Thus, in strategic settings we cannot be certain to what extent the axioms are satisfied, since the exact theoretical results hold only for the reported preferences which are not necessarily the true preferences. Does this mean that only the properties of DA-R and DA-Q hold exactly, while the theoretical properties of IA-Q and IA-R are much less reliable due to strategic distortions? This is far from clear, since a steadily growing experimental literature documents the frequent misrepresentation of the true preferences even in strategyproof mechanisms, and specifically in the DA, which has been demonstrated in many different environments under a variety of treatments (Chen and Sönmez (2006), Pais et al. (2011), Klijn et al. (2013), Echenique et al. (2016) and Chen and Kesten (2019), among others[6]), and there is also experimental evidence that the same holds when minority reserves are in place (Klijn et al., 2016). Moreover, surprisingly, there are some experimental treatments

[6] See Hakimov and Kübler (2021) for a comprehensive survey of experimental results in the school choice and college admissions models.

for which the truth-telling rates of DA and IA are comparable, typically depending on the informational setting (Pais and Pintér (2008), Featherstone and Niederle (2016), Ding and Schotter (2019)). The experimental findings are complemented by some field evidence (e.g., Hassidim et al. (2018), Chen and Pereyra (2019) and Fack et al. (2019)), which may not be highly reliable, however, since the true preferences are difficult to observe or estimate. On the other hand, experimental treatments have their own problems, as experimental subjects may find too little incentive to perform well, compared to making real life choices, and experiments may provide too little context or leave too short a time to learn about the mechanisms.

Misrepresentation in the DA is often done the same way the IA mechanism is manipulated, by placing "safe" schools higher in the preference ordering, where a school is considered "safe" for a student if the school ranks the student high in its priority ordering. This may be connected to risk or loss aversion, but there are also other explanations that attribute the use of often dominated choices to the complexity of understanding optimal strategies or to biased beliefs, among others. The theoretical approaches that attempt to offer various reasons for the lack of truth-telling (see e.g., Li (2017), Ashlagi and Gonczarowski (2018), Velez and Brown (2019) and Dreyfuss et al. (2022)) are too wide-ranging to summarize here. Suffice it to say, not only our comparison of mechanisms with a varying degree of manipulability is difficult, but even the received wisdom of assuming truthful reporting under strategyproof mechanisms cannot be maintained any more.

One usual way to deal with manipulable mechanisms is to check whether the results hold in equilibrium, but the above arguments should make it clear that such an equilibrium analysis would not address the strategic issues adequately, as no straightforward analysis can address the problem of untruthful reporting, given that even seemingly robust dominant strategies are not followed by participants. There is also evidence in the matching context that non-truth-telling equilibria are unlikely to be reached (Featherstone and Niederle, 2016). In sum, we cannot be certain of the accuracy of either a direct comparison (based on the assumption that students are sincere) or an equilibrium comparison of these mechanisms. Due to the lack of an appropriate way to address strategic considerations, we view our results as a first step in a theoretical investigation, keeping in mind that the results apply exactly only in non-strategic settings.

5 A New Mechanism: IA-DA-R

We now introduce a mechanism that satisfies all three of the welfare axioms, proving in a constructive manner that they can be satisfied simultaneously. In order to motivate the proposed new mechanism, first we make several observations based on the previous analysis of the already known mechanisms.

i. Deferred (tentative) acceptances allow for the non-wasteful minority reserves to respect the affirmative action policy, but when applied to reserved seats they violate minimal responsiveness, since tentative acceptances may lead to rejection cycles. Thus, there is no way to have a minimally responsive mechanism with deferred acceptances only.

ii. Immediate (permanent) acceptances allow for the minority reserves to be minimally responsive, but when applied to majority students and reserved seats they do not respect the affirmative action policy, since majority students may be permanently accepted for reserved seats ahead of minority students who apply later. Immediate (permanent) acceptances with majority quotas, on the other hand, ensure that the affirmative action policy is respected, but they lead to wastefulness. Thus, there is no way to have a non-wasteful mechanism which respects the affirmative action policy with immediate acceptances only.

iii. A non-wasteful mechanism that respects the affirmative action policy requires that the set of accepted students be updated when new minority students apply to a school, replacing tentatively accepted majority students who were previously allowed to occupy minority reserve seats in the absence of minority applicants. This is the essence of minority reserves, but this is at odds with minimal responsiveness, *unless minority students assigned to reserved seats are not replaced by newly applying minority students,* since such replacements can lead to rejection cycles that may, in the end, result in not benefiting any minority students and potentially harming some.

In light of the above considerations, our proposed mechanism, the **Immediate and Deferred Acceptance Mechanism with Minority Reserves (IA-DA-R)** incorporates immediate acceptances of minority students for minority reserve seats, while all other acceptances are deferred (i.e., tentative) acceptances.

We will need the following notation in order to formally define the IA-DA-R mechanism. For all $t \geq 1$, let $A_c(t)$ denote the set of students applying to school c in step t, and let

$T_c(t)$ denote the set of tentatively accepted students at school c in step t. Let $m_c(t)$ denote the number of permanently accepted minority students at school c in total in steps 1 to t. In the following we will denote the IA-DA-R mechanism by ψ, to distinguish it from a general mechanism denoted by φ.

IA-DA Mechanism with Minority Reserves (IA-DA-R)

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

Step 1: Every student applies to her most preferred (acceptable) school according to P.

> **Substep 1.a: Applying minority students are permanently assigned to reserved seats.**
>
> If $r_c > 0$, school c permanently assigns seats to minority students in $A_c(1)$ according to its priority ordering $\succ_c$, up to its number of minority reserve seats r_c.
>
> **Substep 1.b: Remaining applicants are tentatively assigned to unreserved seats and remaining reserved seats.**
>
> Each school c tentatively accepts students among the remaining applicants in $A_c(1)$ according to its priority ordering $\succ_c$, up to its capacity q_c in total. Any remaining unassigned students in $A_c(1)$ are rejected.

Step *t* ($t \geq 2$): Every student who was rejected in step $t-1$ applies to her next most preferred acceptable school according to P.

> **Substep *t*.a: Applying minority students are permanently assigned to remaining reserved seats.**
>
> If $r_c - m_c(t-1) > 0$, school c permanently assigns seats to minority students in $A_c(t)$ according to its priority ordering $\succ_c$, up to its remaining number of minority reserve seats $r_c - m_c(t-1)$.
>
> **Substep *t*.b: Remaining applicants and tentatively assigned students in the previous step are tentatively assigned to unreserved seats and remaining reserved seats.**
>
> Each school c tentatively accepts students remaining unassigned in $A_c(t) \cup T_c(t-1)$, that is, among the remaining applicants in step t and the tentatively assigned applicants in step $t-1$, according to its priority ordering $\succ_c$, up to its capacity q_c in total. Any remaining unassigned students in $A_c(t) \cup T_c(t-1)$ are rejected.

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

Note that for all $t \geq 1$, substep t.b may include the tentative assignment of majority student applicants to remaining reserved seats if less than r_c minority students have applied to school c up to step t, and thus $r_c > m_c(t)$. Alternatively, it is also possible that remaining minority student applicants are tentatively assigned to unreserved seats if all reserved seats are already filled, i.e., if $r_c = m_c(t)$. In the IA-DA-R mechanism minority students are accepted permanently for reserved seats, but are only accepted tentatively for unreserved seats and can be replaced at unreserved seats by new applicants with a higher priority later, similarly to majority students. Majority students can only be accepted tentatively for both reserved and unreserved seats, and are only assigned tentatively to reserved seats when there are no minority applicants for the reserved seats.

It is easy to see that with no affirmative action the IA-DA-R mechanism is equivalent to the DA. At the other extreme, if all the seats are minority reserve seats then minority students are prioritized for all the seats over majority students, while majority students are assigned tentatively to remaining empty seats only. In principle, in the unlikely scenario that there are enough minority students and acceptable schools for minority students to take up all the seats when all seats are reserved, the IA-DA-R mechanism becomes equivalent to the IA mechanism.[7]

Example 5. Illustration of the IA-DA-R mechanism.
Let $S^M = \{a_1, \ldots, a_5\}$ and $S^m = \{i_1, \ldots, i_4\}$. Let $C = \{c_1, \ldots, c_4\}$ with capacities $q = (3, 2, 3, 1)$.

Consider the profile in Table 7. Without affirmative action, the IA-DA-R matching μ is the same as the DA matching and it is given by $(c_1, c_1, c_2, c_4, c_1, c_2, c_3, c_3, c_3)$ (underlined in Table 7). With minority reserve policy $\tilde{r} = (2, 0, 0, 0)$, the IA-DA-R matching is $(c_1, c_3, c_2, c_4, c_2, c_3, c_1, c_3, c_1)$ (indicated by the squares in Table 7). The steps of the IA-DA-R algorithm are displayed in Table 8. $\diamond$

[7] Note that the IA-DA-R mechanism is not a member of the class of PRP mechanisms studied by Ayoade and Pápai (2023), since minority applicants are accepted permanently for remaining minority reserve seats in the step when they apply, regardless of where they rank the school, and hence there are no preference rank partitions for minority students that can be applied to each profile when considering reserved seats.

Table 7: 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}	$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$	$\succ_{c_4}$
$\underline{\underline{\boxed{c_1}}}$	$\underline{c_1}$	c_1	$\underline{\underline{\boxed{c_4}}}$	c_4	c_4	c_4	c_4	$\boxed{c_1}$	a_1	a_5	a_5	a_4
c_2	$\boxed{c_3}$	$\underline{\underline{\boxed{c_2}}}$	c_3	$\underline{c_1}$	$\underline{c_2}$	$\boxed{c_1}$	$\underline{\underline{\boxed{c_3}}}$	c_2	a_5	a_3	a_2	a_1
c_3	c_2	c_3	c_2	$\boxed{c_2}$	c_1	c_2	c_2	c_4	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	$\underline{c_3}$	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

Table 8: IA-DA-R steps at $(\tilde{r}, P, \succ)$ in Example 5

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

Struck out students are rejected in the corresponding step.

Circled students are minority students who are permanently assigned to minority reserve seats.

Theorem 1 (IA-DA-R satisfies the three main welfare axioms). *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 a new applicant in any step if there is any empty school seat remaining, and a tentatively 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: The minority reserve seats at each school are assigned to available minority applicants in each step of the IA-DA-R algorithm, prioritizing minority applicants over majority applicants for the minority reserve seats. Moreover, minority students are permanently assigned to seats up to the number of minority reserve seats r_c at each school c, while majority students are only tentatively assigned to minority reserve seats and only if there are not enough minority applicants to fill the minority reserve seats, and they are rejected in later steps of the procedure if new minority students apply to the school. Therefore, it is not possible for a minority student to be rejected by a school unless all reserved seats are already filled with minority students at this school, and hence the IA-DA-R mechanism respects the affirmative action policy.

Minimally Responsive: Let r, r' be two minority reserve policies such that $r \leq r'$ and fix a profile $(P, \succ)$. Let $\mu = \psi(r, P, \succ)$ and $\mu' = \psi(r', P, \succ)$, where ψ denotes the IA-DA-R mechanism. Assume that $\mu \neq \mu'$. Let step t be the first step in the IA-DA-R algorithm 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 (when $t \geq 2$) and the difference is only in the minority reserve policy, it must be the case that a 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 matching μ: there exists $i \in S^m$ such that $\mu'_i \; P_i \; \mu_i$. Therefore, the IA-DA-R mechanism is minimally responsive. □

Given Theorem 1, the axioms of non-wastefulness, respecting the affirmative action policy, and minimal responsiveness are compatible. We establish a stronger existence result next based on the fairness properties of the IA-DA-R mechanism.

6 Minority Fairness: An Existence Result

A mechanism with an affirmative action policy cannot satisfy standard fairness, since the main objective of affirmative action is to allow minority students to be prioritized over majority students in some instances, violating the priorities of majority students. We define below a fairness axiom which allows for priority violations due to affirmative action. It permits priority violations only by minority students, and priority violations at each school are limited to the minority allotment seats, balancing the requirements of affirmative action with the rights of majority students.

Minority Fairness. A matching μ is **minority fair at $(\boldsymbol{v}, \boldsymbol{P}, \succ)$** if it satisfies the following two 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 in μ at each school c.

A mechanism φ is **minority fair** if, for all $(v, P, \succ)$, $\varphi(v, P, \succ)$ is minority fair at $(v, P, \succ)$.

Minority fairness is a simple fairness axiom for quota or reserve-based affirmative action mechanisms. It 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 any other student's priority at any school, by condition 2, given $v_c = 0$ for all $c \in C$.

DA-Q, DA-R and MDA satisfy minority fairness (as does DA), while the efficiency improvements of EIDA destroy this property of DA-R. Immediate acceptances generally violate the priorities of students and thus IA-Q and IA-R allow minority and majority students alike to violate other students' priorities. Consequently, IA-Q and IA-R are not minority fair (and neither is IA). However, IA-DA-R satisfies minority fairness, as we will show next. Therefore, given Theorem 1, we can state the following possibility result.

Theorem 2 (**Possibility result**).
There exists a mechanism that satisfies the following properties:

- *non-wastefulness*
- *respecting the affirmative action policy*
- *minimal responsiveness*
- *minority fairness*

Proof. Given Theorem 1, it suffices to show that the IA-DA-R mechanism is minority fair, that is, it satisfies the two conditions in the definition of minority fairness. First note that majority students are always accepted tentatively only and without violating any previous or current applicant's priority at that school, and are 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 student's priorities at any aa-profile and condition 1 holds. Majority students are also rejected due to new minority student applicants if the reserved seats are not yet filled. Moreover, minority students are treated similarly to majority students when applying for seats after the minority reserve seats have been filled with minority students. This implies that no more than v_c minority students violate the priority of another student at any school c, satisfying condition 2. □

An implication of condition 1 in the definition of minority fairness is that there is no priority violation within the group of majority students. Fairness axioms in the literature often require that there is no priority violation within the sets of same types of students, which would also imply that there are no priority violations among minority students either. However, minority fairness does not imply this, and the relaxation of this requirement for minority students plays a crucial role in our analysis. It allows us to obtain an existence result which contrasts with an important impossibility result of Doğan (2016). Since the axioms are grouped and named differently in his paper from ours, in order to present a clear comparison we state a variation of Proposition 1 in Doğan (2016) using our axioms, and give a direct proof to demonstrate this result and to provide further intuition (see also Proposition 9 in Doğan (2016) which further strengthens this impossibility result).

No Within-Minority-Group Priority Violations. A matching μ satisfies no within-minority-group priority violations at $(v, P, \succ)$ if no minority student violates another minority student's priority in μ at $(v, P, \succ)$. A mechanism φ satisfies no within-minority-group priority violations if for all $(v, P, \succ)$, $\varphi(v, P, \succ)$ satisfies no within-minority-group priority violations at $(v, P, \succ)$.

Proposition 7 (Impossibility result with the addition of no within-minority–group priority violations).
There is no mechanism that satisfies the following properties:

- *non-wastefulness*
- *respecting the affirmative action policy*
- *minimal responsiveness*
- *minority fairness*
- *no within-minority-group priority violations*

Proof. Suppose for a contradiction that mechanism φ with minority allotments is non-wasteful, respects the affirmative action policy, is minimally responsive, minority fair and satisfies no within-minority-group priority violations. Let $S^M = \{a\}$ and $S^m = \{i_1, i_2\}$. Let $C = \{c_1, c_2\}$ with capacities $q = (1, 1)$.

Consider profile $(P, \succ)$ in Table 9. When there is no affirmative action, a minority fair mechanism yields a fair matching at each profile. Let $v = (0, 0)$. Then (a, c_2) at $(v, P, \succ)$,[8] otherwise a's priority would be violated at c_2, contradicting minority fairness. Then non-wastefulness implies that $\varphi(v, P, \succ) = (c_2, 0, c_1)$ (underlined in Table 9).

[8] We denote the assignment of student s to school c by (s, c), and write $(s, 0)$ when s remains unassigned.

Table 9: Profile for the proof of Proposition 7

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

Now consider the minority allotment policy $\tilde{v} = (0, 1)$. If (a, c_2) then φ does not respect the affirmative action policy, and thus either (i_1, c_2) or (i_2, c_2) holds. If (i_2, c_2) then non-wastefulness (specifically individual rationality) implies that $(i_1, 0)$ holds and thus, compared to the matching obtained at $(v, P, \succ)$, i_1 is indifferent and i_2 is worse off. This is ruled out by minimal responsiveness, and therefore (i_1, c_2) holds. Since $\tilde{v}_{c_1} = 0$, minority fairness implies that there are no priority violations at school c_1, and hence (a, c_1). This means that $(i_2, 0)$ and $\varphi(\tilde{v}, P, \succ) = (c_1, c_2, 0)$ (as indicated by the squares in Table 9), implying that the axiom of no within-minority-group priority violations is not satisfied, since $i_2 \succ_{c_2} i_1$. This is a contradiction, which proves the impossibility result. □

Comparing Theorem 2 to Proposition 7, the only additional requirement in Proposition 7 is the axiom of no within-minority-group priority violations. This indicates that if we allow minority students to violate another minority student's priorities (up to the number of reserved seats at each school, as limited by minority fairness), we turn the impossibility result into a possibility result, and hence this 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. This is an interesting insight and it fits well with what we know about the IA-DA-R mechanism, as it is a salient feature of IA-DA-R that minority students may violate another minority student's priority due to the immediate acceptance of minority students for minority reserve seats. The possibility of priority violations between minority students is also demonstrated by Example 5. Although i_1 has a higher priority at school c_1 than i_2 and i_4, due to the immediate acceptances for minority reserve seats i_1 is rejected by school c_1, since i_1 applies to c_1 only in step 4, and the two minority reserve seats have already been filled permanently with minority students i_2 and i_4 in previous steps.

It is possible to satisfy the axioms simultaneously in Theorem 2 at the cost of relaxing the fairness requirement within the minority student group. One may argue that under some circumstances Doğan's impossibility result can be avoided at possibly little cost. Although the desirability of ruling out priority violations among the same type of agents is usually taken for granted, this is not always obvious in the case of minority students, especially when an affirmative action policy is in place. 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, which may be an attractive feature for allocating minority reserve seats.[9] Some critics of affirmative action policies claim that affirmative action benefits primarily the most privileged members of disadvantaged groups, which may very well mean the highest-priority minority students when the priority orderings of schools are based on achievement. Therefore, fairness considerations may actually favor not respecting the priorities among minority students in certain contexts.

7 Impossibility Theorems

We now turn to the analysis of strategyproofness for affirmative action mechanisms that have set-aside seats for minorities. Among the seven analyzed 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 established by Hafalir et al. (2013). The other mechanisms are all manipulable, which is not too surprising for either MDA and EIDA, given that EADAM is not strategyproof, or for IA-DA-R, IA-Q and IA-R, given that IA is not strategyproof. In particular, the IA-DA-R mechanism can be manipulated by minority students, since minority students may be able to obtain a school's reserved seat by ranking this school higher than in their true preferences, thereby gaining immediate (permanent) acceptance at this school before other minority students with a higher priority apply to it. This is the same type of manipulation that all students may be able to carry out in the IA mechanism and have been shown to be *obvious manipulation* strategies by Troyan and Morrill (2020). However, it is not possible to reconcile strategyproofness for minority students with the three welfare axioms and minority fairness, as we will establish next. This implies that the IA-DA-R mechanism

[9] However, 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 being assigned to an unreserved school seat for multiple steps in the IA-DA-R procedure before being rejected by this school, 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 lower than i_1 does. 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.

is necessarily manipulable by minority students, since it satisfies all the other axioms.

Theorem 3 (**Impossibility of strategyproofness for minority students**).
There is no mechanism which satisfies the following properties:

- *non-wastefulness*
- *respecting the affirmative action policy*
- *minimal responsiveness*
- *minority fairness*
- *strategyproofness for minority students*

Proof. Suppose for a contradiction that mechanism φ is non-wasteful, respects the affirmative action policy, is minimally responsive, minority fair and strategyproof for minority students. Let $S^M = \{a_1\}$ and $S^m = \{i_1, i_2, i_3\}$. Let $C = \{c_1, c_2, c_3\}$ with capacities $q = (1, 1, 1)$.

Table 10: Profiles for the proof of Theorem 3

P_{a_1}	P_{i_1}	P_{i_2}	P_{i_3}	P_{a_1}	P_{i_1}	P_{i_2}	P'_{i_3}	$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$
$\underline{c_2}$	$\underline{c_1}$	c_3	$\underline{c_3}$	$\underline{c_2}$	c_1	c_3	$\underline{c_1}$	a_1	a_1	i_1
$\boxed{c_1}$	c_2	$\boxed{c_2}$	c_1	$\boxed{c_1}$	$\boxed{c_2}$	c_2	$\boxed{c_3}$	i_3	i_1	i_3
0	$\boxed{c_3}$	$\underline{0}$	$\boxed{0}$	0	$\underline{c_3}$	$\boxed{\underline{0}}$	0	i_1	i_2	i_2
	0				0			i_2	i_3	a_1

Consider profile $(P, \succ)$ in Table 10. When there is no affirmative action a minority fair mechanism yields a fair matching at each profile. Thus, when $v = (0, 0, 0)$ there is no priority violation and φ yields (a_1, c_2). Then, given that i_2 is ranked last according to $\succ_{c_3}$ among the remaining students, individual rationality and minority fairness imply $(i_2, 0)$. If (i_1, c_3) and (i_3, c_1) then both i_1 and i_3 can manipulate at $(v, P, \succ)$ by reporting $\hat{P}_{i_1} : (c_1, 0)$ and $\hat{P}_{i_3} : (c_3, 0)$, respectively, since in either case the only non-wasteful fair matching that remains is $(c_2, c_1, 0, c_3)$. Since φ is strategyproof for minority students, this is a contradiction, and non-wastefulness implies that $\varphi(v, P, \succ) = (c_2, c_1, 0, c_3)$ (underlined in Table 10 in preference profile P).

Now assume that the minority allotment policy is $\tilde{v} = (0, 1, 0)$. Respecting the affirmative action policy combined with non-wastefulness (specifically, individual rationality) requires (i_1, c_2) or (i_2, c_2), since otherwise (i_2, c_3) would hold, and this would violate the priority of i_3 at c_3. If (i_1, c_2) at $(\tilde{v}, P, \succ)$ then minimal responsiveness implies that (i_2, c_3), which has already been ruled out. Hence, (i_2, c_2). Then minority fairness implies that (a_1, c_1), and thus (i_1, c_3). Therefore, $\varphi(\tilde{v}, P, \succ) = (c_1, c_3, c_2, 0)$ (indicated by the squares in Table 10 in preference profile P).

Next, consider the preference profile where i_3 reports $P'_{i_3} : (c_1, c_3, 0)$ and all other students report the same as at P, as shown by Table 10. Let $P' = (P'_{i_3}, P_{-i_3})$. When there is no affirmative action, the matching at P' is fair, and fairness and non-wastefulness imply that (a_1, c_2), and consequently (i_3, c_1). Then the non-wastefulness of φ implies that c_3 is assigned and, by minority fairness, $\varphi(v, P', \succ) = (c_2, c_3, 0, c_1)$ (underlined in Table 10 in preference profile P').

Consider again the minority allotment policy $\tilde{v} = (0, 1, 0)$. Respecting the affirmative action policy implies that either (i_1, c_2) or (i_2, c_2) must hold, otherwise (i_1, c_1) would be required and i_3's priority would be violated at c_1. Then, by minority fairness and non-wastefulness, $\varphi(\tilde{v}, P', \succ)$ is either $(c_1, c_2, 0, c_3)$ or $(c_1, c_3, c_2, 0)$. Suppose for a contradiction that $\varphi(\tilde{v}, P', \succ) = (c_1, c_3, c_2, 0)$. Then i_1 could report $\check{P}_{i_1} : (c_1, c_2, 0)$ at P', and minority fairness and non-wastefulness would imply that $\varphi(v, (\check{P}_{i_1}, P'_{-i_1}), \succ) = (c_2, 0, c_3, c_1)$. Then, if $\varphi_{i_1}(\tilde{v}, (\check{P}_{i_1}, P'_{-i_1}), \succ) = 0$, either respecting the affirmative action policy is violated (when (i_2, c_2) does not hold) or minimal responsiveness is violated (when (i_2, c_2) holds), which is a contradiction, and thus $\varphi_{i_1}(\tilde{v}, (\check{P}_{i_1}, P_{-i_1}), \succ) \neq 0$. Then i_1 can manipulate at $(\tilde{v}, P', \succ)$ by reporting $\check{P}_{i_1}$. This is a contradiction, since i_1 is a minority student and cannot manipulate. Therefore, $\varphi(\tilde{v}, P', \succ) = (c_1, c_2, 0, c_3)$ (indicated by the squares in Table 10 in preference profile P'). However, this means that minority student i_3 can manipulate at $(\tilde{v}, P, \succ)$ by reporting P'_{i_3}. This is a contradiction. □

An interesting question is whether the IA-DA-R mechanism can be manipulated by majority students, as they are only accepted tentatively for both reserved and unreserved seats, and thus they seem to face a similar environment to that of the strategyproof DA mechanism. Surprisingly, majority students are also able to manipulate the IA-DA-R mechanism. However, majority students can only obtain a better school assignment by manipulating indirectly, through affecting the immediate acceptances of minority students for minority reserve seats in some step of the IA-DA-R algorithm, which in turn can lead to a better assignment for the misreporting student in a later step. This kind of manipulation may also be carried out by minority students, but majority students can only manipulate this way. In the next example we show the manipulability of the IA-DA-R mechanism in this subtle manner.

Example 6 (**IA-DA-R can be manipulated by majority students**).
Let $S^M = \{a_1, a_2\}$ and $S^m = \{i_1, i_2\}$. Let $C = \{c_1, c_2, c_3\}$ with capacities $q = (1, 1, 1)$ and let the minority reserve policy be $r = (0, 1, 0)$.

Table 11: Profiles for Example 6

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

Given the profile $(P, \succ)$ specified in Table 11, the IA-DA-R matching is $\psi(r, P, \succ) = (c_1, 0, c_3, c_2)$. If majority student a_2 reports P'_{a_2} instead of P_{a_2} then the IA-DA-R matching is $\psi(r, (P'_{a_2}, P_{-a_2}), \succ) = (c_1, c_3, c_2, 0)$ (both matchings are indicated in Table 11). Thus, majority student a_2 gets c_3 with the false report P'_{a_2}, which is preferred by a_2 to remaining unassigned. Therefore, majority student a_2 can manipulate ψ at $(r, P, \succ)$ by reporting P'_{a_2}. 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 of them, and thus both minority students apply to c_2 in step 2, which ensures that i_1 is accepted permanently by c_2. By contrast, if a_2 reports truthfully, i_1 applies to c_2 too late, when the minority reserve seat is already assigned to i_2 permanently. Thus, a_2's competition at c_3 is i_1 in the next step, as opposed to i_2 when a_2 reports untruthfully. Since i_1 has a higher priority than a_2 at school c_3, while i_2 has a lower priority, this misrepresentation allows a_2 to successfully manipulate. Note that the same manipulation would occur if a_2 was a minority student in this example. $\diamond$

For comparison, note that MDA is also manipulable by both minority and majority students. However, MDA is robust to manipulation in certain limited information environments, similarly to EADAM (Doğan, 2016), which does not hold for IA-DA-R. We analyze the incentive properties of IA-DA-R in Section 8.

In light of Example 6, it makes sense that a similar impossibility result to Theorem 3 can be obtained for majority students, as we state below.

Theorem 4 (**Impossibility of strategyproofness for majority students**).
There is no mechanism that satisfies the following properties:

- *non-wastefulness*

- *respecting the affirmative action policy*
- *minimal responsiveness*
- *minority fairness*
- *strategyproofness for majority students*

Proof. Suppose for a contradiction that mechanism φ is non-wasteful, respects the affirmative action policy, is minimally responsive, minority fair and strategyproof for majority students. Let $S^M = \{a_1, a_2, a_3\}$ and $S^m = \{i_1, i_2\}$. Let $C = \{c_1, \ldots, c_4\}$ with capacities $q = (1, 1, 1, 1)$.

Table 12: Profiles for the proof of Theorem 4

P_{a_1}	P_{a_2}	P_{a_3}	P_{i_1}	P_{i_2}
$\underline{c_2}$	$\underline{c_3}$	$\underline{c_4}$	$\underline{c_1}$	c_4
$\boxed{c_1}$	$\boxed{c_4}$	c_3	c_2	$\boxed{c_2}$
0	0	c_1	$\boxed{c_3}$	$\underline{0}$
		$\boxed{0}$	0	

P_{a_1}	P_{a_2}	P'_{a_3}	P_{i_1}	P_{i_2}
$\underline{c_2}$	$\boxed{c_3}$	c_3	c_1	c_4
$\boxed{c_1}$	$\underline{c_4}$	$\underline{c_1}$	$\boxed{c_2}$	c_2
0	0	$\boxed{c_4}$	$\underline{c_3}$	$\boxed{\underline{0}}$
		0	0	

$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$	$\succ_{c_4}$
a_1	a_1	i_1	a_2
a_3	i_1	a_2	i_1
i_1	i_2	a_3	a_3
a_2	a_3	a_1	i_2
i_2	a_2	i_2	a_1

Consider profile $(P, \succ)$ in Table 12. When there is no affirmative action a minority fair mechanism yields a fair matching at each profile. Thus, when $v = (0, 0, 0, 0)$ there is no priority violation and φ yields (a_1, c_2) at $(v, P, \succ)$. Note that i_2 cannot be assigned c_4, given that a_3 has a higher priority for c_4 and ranks it first. Thus, $(i_2, 0)$. Now suppose that (a_2, c_3) does not hold. Then (i_1, c_3), otherwise a_2's priority would be violated at c_3. Then (a_2, c_4), otherwise a_2's priority would be violated at c_4. Consequently, non-wastefulness implies (a_3, c_1). Now consider the scenario that a_2 reports preferences $\hat{P}_{a_2} : (c_3, 0)$ instead of P_{a_2}. At $(\hat{P}_{a_2}, P_{-a_2})$ we still have (a_1, c_2) as before, and individual rationality and minority fairness imply (a_3, c_4). Then, by non-wastefulness, (i_1, c_1), and hence non-wastefulness implies (a_2, c_3). This means that if $\varphi_{a_2}(v, P, \succ) \neq c_3$ then majority student i_2 can manipulate at $(v, P, \succ)$ by reporting $\hat{P}_{a_2}$, which would be a contradiction. Therefore, (a_2, c_3) holds. By individual rationality and minority fairness (a_3, c_4) follows, and thus non-wastefulness yields (i_1, c_1). In sum, $\varphi(v, P, \succ) = (c_2, c_3, c_4, c_1, 0)$ (underlined in Table 12 in profile P).

Now assume that the minority allotment policy is $\tilde{v} = (0, 1, 0, 0)$. If (i_1, c_2) at $(\tilde{v}, P, \succ)$ then minimal responsiveness would imply (i_2, c_4). However, this would contradict minority fairness, since a_3 has a higher priority for c_4 and ranks it first. Therefore, respecting the affirmative action policy requires (i_2, c_2). Then non-wastefulness and minority fairness imply (a_1, c_1), then (i_1, c_3) and (a_2, c_4). Therefore, $(a_3, 0)$. This means that $\varphi(\tilde{v}, P, \succ) = (c_1, c_4, 0, c_3, c_2)$ (indicated by the squares in Table 12 in preference profile P).

Next, consider the preference profile where a_3 reports $P'_{a_3} : (c_3, c_1, c_4, 0)$ and all other students report the same as at P, as shown by Table 12. Let $P' = (P'_{a_3}, P_{-a_3})$. Non-wastefulness and minority fairness imply (a_1, c_2), as before. By minority fairness, a_3 cannot be assigned c_3, given that a_2 has a higher priority for c_3 and ranks it first. Then, given that a_1 is already assigned, non-wastefulness and minority fairness imply (a_3, c_1), and thus (i_1, c_3) and (a_2, c_4). Hence, $(i_2, 0)$. In sum, $\varphi(v, P', \succ) = (c_2, c_4, c_1, c_3, 0)$ (underlined in Table 12 in preference profile P').

Consider again the minority allotment policy $\tilde{v} = (0, 1, 0, 0)$. We show first that either (i_1, c_2) or (i_2, c_2) must hold at $(\tilde{v}, P', \succ)$. Suppose otherwise. Then, since φ respects the affirmative action policy, (i_1, c_1) and (i_2, c_4), and non-wastefulness implies (a_1, c_2). Thus, non-wastefulness and minority fairness lead to (a_2, c_3). This implies $(a_3, 0)$, which contradicts minority fairness, since $a_3 \succ_{c_1} i_1$ and $a_3 \succ_{c_4} i_2$. Therefore, there are two cases to consider:

Case 1: (i_1, c_2)
Non-wastefulness and minority fairness imply that (a_1, c_1) and (a_2, c_3), and thus (a_3, c_4). However, $(i_2, 0)$.

Case 2: (i_2, c_2)
Non-wastefulness and minority fairness imply that (a_1, c_1) and thus (i_1, c_3). Hence, (a_2, c_4) and $(a_3, 0)$.

If Case 2 holds then a_2 could report $\hat{P}_{a_2} : (c_3, 0)$, and since a_3 is a majority student and cannot manipulate, at this new aa-profile $(\tilde{v}, (\hat{P}_{a_2}, P'_{-a_2}), \succ)$ we would have $(a_2, 0)$. However, in this case non-wastefulness and minority fairness would imply (i_1, c_3) and thus, given $\tilde{v} = (0, 1, 0, 0)$, respecting the affirmative action policy would require (i_2, c_2). Then, by non-wastefulness and minority fairness, (a_1, c_1) and (a_3, c_4). On the other hand, note that with no affirmative action policy strategyproofness requires $(a_2, 0)$ at $(v, (\hat{P}_{a_2}, P'_{-a_2}), \succ)$, and then non-wastefulness and minority fairness imply (i_1, c_3). We also have (a_1, c_2), and then non-wastefulness implies (a_3, c_1) and hence (i_2, c_4). Note that at preference profile $(\hat{P}_{a_2}, P'_{-a_2})$ we have (i_1, c_3) and (i_2, c_4) with $v = (0, 0, 0, 0)$ and (i_1, c_3) and (i_2, c_2) with $\tilde{v} = (0, 1, 0, 0)$, contradicting minimal responsiveness. This is a contradiction for Case 2, and therefore Case 1 must hold. However, if Case 1 holds then majority student a_3 can manipulate at $(\tilde{v}, P, \succ)$ by reporting P'_{a_3}, which is a contradiction. □

These impossibility results underscore that manipulability by both minority and majority students is inevitable under very mild welfare and fairness criteria, indicating necessary trade-offs that market designers face. An immediate consequence of both Theorem 3 and Theorem 4 is that the normative axioms in these theorems cannot be satisfied together with strategyproofness. In contrast to Theorem 2 versus Proposition 7, allowing for within-minority-group priority violations does not reverse the impossibility for strategyproofness.[10]

The independence of the axioms for both impossibility theorems is established in Appendix B. The examples of mechanisms satisfying all but one of the axioms suggest that the main trade-off is between incentives and minimal responsiveness, as represented by the comparison between DA-R and IA-DA-R. The strategyproof DA-R mechanism satisfies all the axioms in the two theorems except for minimal responsiveness, and we can infer from the theorems that even a minimal requirement of responsiveness to the degree of the affirmative action is not possible to obtain together with strategyproofness for either type of students, in combination with the other basic properties. On the other hand, IA-DA-R satisfies minimal responsiveness in addition to the other normative properties in the theorems, and these two impossibility theorems explain why the IA-DA-R mechanism is manipulable by both minority and majority students. The theorems pinpoint minimal responsiveness as the main "culprit" for the strategic vulnerability of IA-DA-R, since the other normative properties may be considered more fundamental, which can also be seen from the fact that when we drop one of these axioms the others can be satisfied by a mechanism that is unappealing as an affirmative action mechanism.

8 Incentive Properties of Minority Fair Mechanisms

We focus on less demanding incentive properties next to establish some positive results for the IA-DA-R and other minority fair mechanisms. Although strategyproofness is not possible to attain together with the other desirable properties, it may be difficult for majority students to manipulate the IA-DA-R mechanism, as Example 6 suggests. The next theorem generalizes this intuition: it shows that minority fair mechanisms which are non-wasteful and respect the affirmative action policy are not obviously manipulable by majority students according to the formal definition of Troyan and Morrill (2020). Obvious manipulation is defined in the "local" sense by fixing the minority allotment policy v and

[10] See Proposition 10 of Doğan (2016) for a related impossibility result with strategyproofness, which imposes no within-minority-group priority violations.

the priority profile $\succ$, which leads to a strong version of non-obvious manipulation. For notational ease, we suppress v and $\succ$ in the notations for the rest of this section.

Non-obvious manipulation. Given fixed $(v, \succ)$, under mechanism φ for any true preference ordering P_s of student s, let $(\min|P_s)_{P_{-s}}\,[\varphi_s(P'_s, P_{-s})]$ denote the worst assignment that s can obtain under any reported preferences P_{-s} of all the other students when student s reports P'_s. Similarly, under mechanism φ let $(\max|P_s)_{P_{-s}}\,[\varphi_s(P'_s, P_{-s})]$ denote the best assignment that s can obtain under any reported preferences P_{-s} of all the other students when student s reports P'_s.

Given $(v, \succ)$, a manipulation strategy P'_s for student s is **obvious** under mechanism φ if one of the following holds:

(i) P'_s makes s strictly better off than truthful reporting in the *worst* case:

$$(\min|P_s)_{P_{-s}}\,[\varphi_s(P'_s, P_{-s})]\ P_s\ (\min|P_s)_{P_{-s}}\,[\varphi_s(P_s, P_{-s})]$$

(ii) P'_s makes s strictly better off than truthful reporting in the *best* case:

$$(\max|P_s)_{P_{-s}}\,[\varphi_s(P'_s, P_{-s})]\ P_s\ (\max|P_s)_{P_{-s}}\,[\varphi_s(P_s, P_{-s})]$$

If there is an obvious manipulation strategy for s under φ for a given $(v, \succ)$ then φ is **obviously manipulable** by s. Otherwise, if there is no obvious manipulation strategy for s under φ for any $(v, \succ)$ then φ is **not obviously manipulable** by s.

Theorem 5 (**Non-obvious manipulation**).
Let mechanism φ satisfy the following properties:

- *non-wastefulness*
- *respecting the affirmative action policy*
- *minority fairness*

Then φ is not obviously manipulable by majority students.

Proof. Let φ satisfy non-wastefulness, respecting the affirmative action policy and minority fairness.

(i) *Worst assignment:* First we show that under φ the worst assignment from truth-telling is always weakly better than the worst assignment from any other strategy for any majority student. Formally, given $(v, \succ)$ and given the true preference ordering P_a for a majority student $a \in S^M$, we will show that for all strategies P'_a for a, $(\min|P_a)_{P_{-a}}[\varphi_a(P_a, P_{-a})]\ R_a$ $(\min|P_a)_{P_{-a}}\,[\varphi_a(P'_a, P_{-a})]$. Let the true preference ordering be P_a for a majority student $a \in S^M$, and fix a preference profile $\bar{P}_{-a}$ for all the other students such that $(\min|P_a)_{P_{-a}}\,[\varphi_a(P_a, P_{-a})] =$

$\varphi_a(P_a, \bar{P}_{-a})$. For notational convenience, let $\mu = \varphi(P_a, \bar{P}_{-a})$. For each $c \in C$ such that $c \; P_a \; \mu_a$, let μ_c^l denote the set of students assigned to c in μ who have a lower priority for c than a, that is, $\mu_c^l = \{s \in S : \mu_s = c \text{ and } a \succ_c s\}$. Thus, for each $c \in C$ such that $c \; P_a \; \mu_a$ and for each $s \in S$ such that $\mu_s = c$, either $s \succ_c a$ or $s \in \mu_c^l$. Since φ satisfies minority fairness, $\mu_c^l \subseteq S^m$ and $|\mu_c^l| \leq r_c$. Now consider the preference profile $\hat{P}_{-a}$ for all the students other than a such that for all $s \in S \setminus \{a\}$, $\hat{P}_s : (\mu_s, 0)$ if $\mu_s \in C$ and $\hat{P}_s : (0)$ otherwise. Since for each $c \in C$ such that $c \; P_a \; \mu_a$, $|\mu_c^l| \leq r_c$, given that φ respects the affirmative action policy, for each $s \in \mu_c^l$, $\varphi_s(P_s, \hat{P}_{-s}) = c$. Furthermore, for each $c \in C$ such that $c \; P_a \; \mu_a$ and for each $s \in S^M$ such that $\varphi_s(P_s, \bar{P}_{-s}) = c$, we have $s \succ_c a$, and thus minority fairness (condition 1) and non-wastefulness imply that $\varphi_s(P_s, \hat{P}_{-s}) = c$. This means that for each $c \in C$ such that $c \; P_a \; \mu_a$, and for each $s \in S$ such that $\varphi_s(P_a, \bar{P}_{-a}) = c$, $\varphi_s(P_a, \hat{P}_{-a}) = c$. Since φ is non-wasteful, this implies that for all $c \in C$ such that $c \; P_a \; \mu_a$, $\varphi_a(P_a, \hat{P}_{-a}) \neq c$ and thus, by non-wastefulness, $\varphi_a(P_a, \hat{P}_{-a}) = \mu_a$.

Let P'_a be different from the true preference ordering P_a. For notational convenience, let $\mu' = \varphi(P'_a, \hat{P}_{-a})$ and $c' = \mu'_a$. Suppose for a contradiction that $c' \; P_a \; \mu_a$. Then the non-wastefulness of φ implies that there exists $\tilde{s} \in S$ such that $\mu_{\tilde{s}} = c'$ but $\mu'_{\tilde{s}} \neq c'$. Note that $\hat{P}_{\tilde{s}} : (c', 0)$ and thus individual rationality implies that $\mu'_{\tilde{s}} = 0$. Given that a is a majority student, it follows from minority fairness (condition 1) that $a \succ_{c'} \tilde{s}$, and thus $\tilde{s} \in \mu_{c'}^l$, implying that $\tilde{s}$ is a minority student. However, since $|\mu_{c'}^l| \leq r_{c'}$ and the only minority students who find c' acceptable at preference profile $(P'_a, \hat{P}_{-a})$ are the students in $\mu_{c'}^l$, this means that φ does not respect the affirmative action policy, which is a contradiction. Therefore, $\mu_a \; R_a \; c'$. This implies that, for all preference orderings P'_a for a which differ from the true preference ordering P_a, $(\min|P_a)_{P_{-a}} \; [\varphi_a(P_a, P_{-a})] \; R_a \; (\min|P_a)_{P_{-a}} \; [\varphi_a(P'_a, P_{-a})]$. Given that the same argument holds for each majority student $a \in S^M$ and for each true preference ordering P_a of student a, the proof is completed.

(ii) *Best assignment:* We also need to show that for mechanism φ the best assignment from truth-telling is always weakly better than the best assignment from a manipulation strategy for any majority student. In our setting this holds trivially, since if there is no competition for a particular student's first-ranked school (e.g., all other students report this school unacceptable), then the student will be assigned to her first-ranked school due to the non-wastefulness of φ when reporting truthfully. □

An implication of Theorem 5 is that a stable mechanism is not obviously manipulable, since if there is no affirmative action then the premises of the theorem simplify to stability (a conjunction of no priority violation, non-wastefulness and individual rationality) and minority students are treated the same way as majority students, so the conclusion holds for all students. This makes the non-obvious manipulation result of stable mechanisms (Theorem 3) of Troyan and Morrill (2020) a special case of Theorem 5, as the implication holds for the school choice model, and this result immediately extends to a model with strategic agents on both sides.

Non-obvious manipulation is closely connected to maximin strategies in our setting. A maximin strategy is a risk-averse strategy that maximizes the worst-case outcome when comparing different preference reports.

Maximin strategies. A preference ordering $\tilde{P}_s$ for student s with true preference ordering P_s is a **maximin strategy under mechanism $\boldsymbol{\varphi}$** if, for all $(v, \succ)$ and all preference orderings $\bar{P}_s$ for s, $(\min|P_s)_{P_{-s}} [\varphi_s(\tilde{P}_s, P_{-s})] \; R_s \; (\min|P_s)_{P_{-s}} [\varphi_s(\bar{P}_s, P_{-s})]$.

Theorem 6 (**Maximin strategy**).
Let mechanism φ satisfy the following properties:

- *non-wastefulness*
- *respecting the affirmative action policy*
- *minority fairness*

Then truth-telling is a maximin strategy for majority students under φ.

Proof. Note that we did not have to assume that the alternative strategy P'_a is a manipulation strategy in the first part of the proof of Theorem 5 pertaining to the worst assignment. Since the proof of case (i) holds for an arbitrary report P'_a that differs from the true preference ordering, it also implies this result. □

Theorems 5 and 6 demonstrate that two of the welfare axioms and minority fairness are not only compatible with but in fact imply good incentive properties for majority students. Each of the three axioms in the statements of the theorems are needed, as we show in Appendix C. While minimal responsiveness is not required to reach the conclusions of the theorems, it is compatible with the other axioms and thus also with the implied incentive properties for majority students. Although weaker than strategyproofness (which leads to an impossibility result when minimal responsiveness is also required, see Theorems 3 and 4), having truth-telling as a maximin strategy shows that the mechanisms satisfying the required axioms are incentive compatible for majority students if they are averse to uncertainty.

Since the IA-DA-R mechanism satisfies the axioms required by Theorems 5 and 6, we can state the following corollary.

Corollary 1 (**Incentive properties of IA-DA-R**).

1. *The IA-DA-R mechanism is not obviously manipulable by majority students.*
2. *Truth-telling is a maximin strategy for majority students under the IA-DA-R mechanism.*

Corollary 1 shows that the IA-DA-R mechanism restricts obvious manipulations to minority students. Given that majority students are typically a great majority of the students and that minority students face similar incentives to majority students with respect to obtaining unreserved seats, the IA-DA-R mechanism provides much better incentives than IA-Q or IA-R, for which both majority and minority students have obvious manipulation strategies.

As for the incentives of minority students, it is likely that many minority (i.e., disadvantaged) families do not have the resources or the time to learn the information needed to successfully manipulate, unlike families with a higher socioeconomic status, and thus they may simply submit their true preference ordering even if the highly manipulable IA mechanism were used. The unfair disparity between the assignments of sincere and sophisticated players in the IA mechanism, demonstrated by Pathak and Sönmez (2008), was put forth as an argument in favor of the DA mechanism over IA in order to "level the playing field"(see also Basteck and Mantovani (2018, 2023) for related experimental results). The crucial difference from this for the IA-DA-R mechanism is that unlike in the IA mechanism, where majority students may be able to take advantage of the sincerity of minority student reports, when using the IA-DA-R mechanism majority students face very different incentives from the incentives in IA, as shown by Corollary 1. The relative truth-telling rates of IA-DA-R compared to DA-R in real life would be difficult to predict, especially given that misrepresentation is frequent even in the strategyproof DA mechanism in experimental settings. As a theoretical exercise only, we consider the (unrealistic) scenario in which all students are perfectly sophisticated and informed, and show that the DA-R matching is one of the multiple Nash-equilibrium outcomes of the preference revelation game induced by the IA-DA-R mechanism; this analysis is relegated to Appendix D.

9 Conclusion

This paper calls for a systematic analysis of mechanisms with a quota or reserve-based affirmative action policy. We show that even though all previously studied mechanisms with such an affirmative action policy violate at least one of three essential welfare axioms (Propositions 1-6, Table 6), it is possible to design mechanisms that meet all three criteria. The novel IA-DA-R mechanism which we propose overcomes the individual weaknesses of both immediate and deferred acceptances and ensures that the three welfare axioms are satisfied (Theorem 1), and these features of our mechanism are also quite intuitive. The IA-DA-R mechanism has desirable fairness properties as well (Theorem 2). The minority fairness axiom that we introduce is weaker than usual fairness axioms for affirmative action, as it allows for priority violations among minority students when they obtain reserved seats. This may be viewed as the price of achieving minimal responsiveness compared to the DA-R mechanism, and as the price of respecting the affirmative action policy compared to the MDA mechanism, while in some settings not enforcing the priorities among minority students may even be desirable (see Section 6 for details).

The main trade-off is between incentive properties and minimal responsiveness, as demonstrated by Theorems 3 and 4. This trade-off is most clearly seen when contrasting the DA-R mechanism to IA-DA-R. DA-R only fails minimal responsiveness in these impossibility results, while IA-DA-R fails the strategyproofness requirements respectively, but satisfies all the other properties in the two theorems. The other mechanisms violate at least two of the axioms, with the most appealing among them, MDA, violating not only the strategyproofness properties but also the very fundamental axiom of respecting the affirmative action policy. Based on our findings, DA-R stands out when enforcing the priorities within the minority student group is desired and superior incentive properties are important. On the other hand, if not enforcing the priorities among minority students is acceptable (or even preferable), given its good incentive properties for majority students (Theorems 5 and 6; Corollary 1) the IA-DA-R mechanism is a strong competitor to DA-R, as it also ensures minimal responsiveness. Failing to satisfy minimal responsiveness means that strengthening, or even just introducing, the affirmative action policy may harm students without helping any minority student, which is a shortcoming of the DA-R mechanism. Thus, notwithstanding the issues surrounding strategic reporting, which apply to all the mechanisms, including the strategyproof DA-R mechanism, our results suggest that the IA-DA-R mechanism is a noteworthy alternative to previously proposed mechanisms with a quota or reserve-based affirmative action policy. More generally, this study offers new theoretical insights that lead to a deeper understanding of compatibilities and inevitable trade-offs among welfare, fairness, and incentive properties, providing essential information for designers of affirmative action mechanisms.

Appendix

A Definitions of Mechanisms

DA with Majority Quotas (DA-Q)

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 $\succ_c$, subject to the restriction that each school c rejects majority students when its majority quota q_c^M is reached by accepted majority students.

Step *t* ($t \geq 2$)**:** 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 $\succ_c$, subject to the restriction that 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 $(q - q^M, P, \succ)$.

DA with Minority Reserves (DA-R)

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* ($t \geq 2$)**:** 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 (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 $(r, P, \succ)$.

Modified DA with Minority Reserves (MDA)

Fix a minority reserve policy r and a profile $(P, \succ)$. An *interferer* at some school c with $r_c > 0$ is a minority student $i \in S^m$ if there is a majority student $a \in S^M$ with $a \succ_c i$ who is rejected by c when i is accepted in some step of the DA-R mechanism at $(r, P, \succ)$ due to the minority reserves, but then in a later step i gets rejected by school c. Then the minority student i is considered an interferer in the step where i is rejected by c. The MDA mechanism consists of iterative rounds of the DA-R mechanism with specific modifications as follows. If there are no interferers at any school when running the DA-R algorithm, then the MDA mechanism assigns the DA-R matching to $(r, P, \succ)$ and the algorithm ends in one round. If there are interferers in some steps of the DA-R, the MDA mechanism moves to the next round and runs the DA-R mechanism again, but treats all interferers in the latest step of the first round at which there are interferers at school c as majority students at c. Further DA-R rounds are iterated similarly until no interferer remains at any school. The matching selected by the MDA mechanism at $(r, P, \succ)$ is the DA-R matching of the final round of the procedure.

Efficiency Improved DA with Minority Reserves (EIDA)

Fix a minority reserve policy r and a profile $(P, \succ)$. A school c is *under-demanded* at $(v, P, \succ)$ if all students weakly prefer their DA-R assignment to c at $(r, P, \succ)$. 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 are removed (including the removal of unassigned students in the first round). Since there is always an under-demanded school, at least one student is removed in each round. The algorithm terminates when all students are removed. The matching selected by the EIDA mechanism at $(r, P, \succ)$ is given by the assignments with which students are removed.

IA with Majority Quotas (IA-Q)

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 $\succ_c$, subject to the restriction that 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* $(t \geq 2)$**:** 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 up to its capacity q_c, following its priority ordering $\succ_c$, subject to the restriction that each school c rejects majority students when its majority quota q_c^M is reached, 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 $(q - q^M, P, \succ)$.

IA with Minority Reserves (IA-R)

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* ($t \geq 2$)**:** 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 $(r, P, \succ)$.

B Independence of the Axioms in Theorems 3 and 4

We show that the axioms in Theorems 3 and 4 are independent, that is, if we drop one axiom at a time, then there exists a mechanism which satisfies all the other axioms.

We refer to the mechanism which only allocates minority allotment seats to minority students using the DA algorithm as the **Minority-Restricted DA** mechanism. This mechanism is minimally responsive, since the DA satisfies resource-monotonicity.

Non-wastefulness: Minority-Restricted DA

Respecting the affirmative action policy: DA

Minimal responsiveness: DA-R

Minority fairness: Serial Dictatorship with a permutation that ranks all minority students ahead of all majority students.

Strategyproofness for minority/majority students: IA-DA-R

C Necessity of the Axioms in Theorems 5 and 6

We show that each of the three axioms in Theorems 5 and 6 is needed to reach the conclusion: if we drop one axiom at a time, then there exists a mechanism which satisfies the other two axioms and is obviously manipulable by a majority student a due to a manipulation strategy P'_a that satisfies (i) in the definition of an obvious manipulation strategy (the worst assignment case). This proves the necessity of each axiom in Theorem 5. To see the necessity of each axiom in Theorem 6, note that the existence of such an obvious manipulation strategy for majority student a based on the worst assignment implies that truth-telling is not a maximin strategy for student a.

> **Non-wastefulness:** If a majority student a reports one acceptable school only: DA-R matching; otherwise: Minority-Restricted DA matching.
>
> Let $\succ_c$ rank a first, and let a's true preference ordering be $P_a = (c, c', 0)$, where $v_c < q_c$. Then $P'_a = (c, 0)$ is an obvious manipulation strategy for a based on the worst assignment.
>
> **Respecting the affirmative action policy:** If a majority student a reports one acceptable school only: DA matching; otherwise: DA-R matching.
>
> Let $\succ_c$ rank only majority students in the first q_c positions, with a in position q_c, and let a's true preference ordering be $P_a = (c, c', 0)$, where $0 < v_c < q_c$. Then $P'_a = (c, 0)$ is an obvious manipulation strategy for a based on the worst assignment.
>
> **Minority fairness:**
> *Step 1:* Use the DA mechanism to allocate the minority allotment seats to minority students, as in the Minority-Restricted DA mechanism.
> *Step 2:* Use the IA mechanism to allocate all the remaining seats (including any unused minority allotment seats in Step 1) to the remaining unassigned students (including any unassigned minority students in Step 1 and all the majority students).
>
> Then majority students have an obvious manipulation strategy based on the worst assignment at some $(v, \succ)$.

D Nash-Equilibrium Outcomes of IA-DA-R

We study the Nash-equilibrium outcomes of the IA-DA-R revelation game as a theoretical exercise, since reaching Nash-equilibria requires complete information and coordinating untruthful reporting among players, which makes the results questionable when it comes to predicting actual behavior.

A mechanism φ and an aa-profile $(v, P, \succ)$ induce a strategic game in which the players are the students and their strategies are the reported preference orderings. Let $\langle \varphi, v, P, \succ \rangle$ denote the strategic game induced by φ and $(v, P, \succ)$. Strategy profile $\tilde{P}$ is a **Nash-equilibrium** of strategic game $\langle \varphi, v, P, \succ \rangle$ if, for all $s \in S$ and all P'_s, $\varphi_s(v, \tilde{P}, \succ)$ R_s $\varphi_s(v, (P'_s, \tilde{P}_{-s}), \succ)$. This means that no student s can profitably deviate at a Nash-equilibrium profile by reporting different preferences, when all the other students' preferences are unchanged.

Strong Minority Fairness. A matching μ is **strongly minority fair at $(\boldsymbol{v}, \boldsymbol{P}, \succ)$** if it satisfies both minority fairness and no within-minority-group priority violations. A mechanism φ is **strongly minority fair** if for all $(v, P, \succ)$, $\varphi\,(v, P, \succ)$ is strongly minority fair at $(v, P, \succ)$.

Theorem 7 (Nash-equilibrium matchings of IA-DA-R).
For each market $(S, C, q, r, P, \succ)$, all strongly minority fair matchings are Nash-equilibrium matchings of the strategic game induced by the IA-DA-R mechanism ψ and $(r, P, \succ)$. In particular, the DA-R matching at $(r, P, \succ)$, which Pareto-dominates all other strongly minority fair matchings at $(r, P, \succ)$, is a Nash-equilibrium matching of the strategic game $\langle \psi, r, P, \succ \rangle$.

Proof. Following Hafalir et al. (2013) who use a similar construction, we work with the *augmented priority-modified market* (*apm-market*, in short) corresponding to each market $(S, C, q, v, P, \succ)$. In the apm-market each school c is split into a "reserved school" c^r and a "regular school" c^g, with capacities r_c and $q_c - r_c$ respectively. Reserved schools rank minority students ahead of majority students in their priority ordering, while they keep all other priority orderings the same within the set of majority students as well as within the set of minority students. Regular schools have the same priorities as the original school. The reserved and regular schools for the same original school are consecutive in the preference ordering of each student, replacing school c in the original preference ordering by c^r and c^g in this order. A matching in the original market naturally corresponds to a matching in the corresponding apm-market and vice versa.

One can easily verify that a matching is strongly minority fair in market $(S, C, q, r, P, \succ)$ if and only if it is a fair matching in the apm-market corresponding to this market. Thus, given a market $(S, C, q, r, P, \succ)$, we need to show that all fair matchings in the apm-market corresponding to this market are Nash-equilibrium matchings of the strategic game

$\langle \psi, r, P, \succ \rangle$, where ψ denotes the IA-DA-R mechanism. Our proof follows a similar reasoning to that of Ergin and Sönmez (2006), which shows that in the strategic game induced by the IA mechanism all stable matchings under the true preferences are Nash-equilibrium outcomes, but since IA-DA-R has an affirmative action policy we focus on the augmented apm-markets instead of the original markets.

Fix a fair matching μ in the apm-market corresponding to $(S, C, q, r, P, \succ)$. Let $\tilde{P}$ be a preference profile where each student s ranks μ_s first. Since ψ is the IA-DA-R mechanism, the resulting matching is μ at $(r, \tilde{P}, \succ)$. Suppose for a contradiction that there exists a student s who can unilaterally deviate at this aa-profile to obtain a school c such that $c \ P_s \ \mu_s$. Given that μ is fair in the apm-market, regardless of whether s is assigned in the apm-market to a reserved or regular school, s cannot be assigned to c when s reports different preferences, since μ is non-wasteful and all students who are assigned to c^r and c^g have a higher priority for their respective school than s, and all these students already rank c first. Thus, under the IA-DA-R mechanism student s with true preferences P_s cannot profitably deviate from reporting $\tilde{P}_s$ at $(r, \tilde{P}, \succ)$, and thus the first part of the claim on strongly minority fair matchings holds.

For the second part of the statement, note that the DA-R matching is not only a strongly minority fair matching, but since it corresponds to the student-optimal fair matching in the apm-market (Hafalir et al., 2013), it is the unique strongly minority fair matching that Pareto-dominates all other strongly minority fair matchings at each aa-profile. □

Theorem 7 demonstrates that if we view IA-DA-R as a strategic game played by sophisticated players who can coordinate to play some Nash-equilibrium by reporting untruthfully, then the DA-R matching, along with all other strongly minority fair matchings, is an equilibrium matching. However, similarly to the DA which may also have other, even unstable, Nash-equilibrium outcomes (Haeringer and Klijn, 2009), IA-DA-R may have Nash-equilibrium outcomes which don't satisfy strong minority fairness. This is not surprising, since the IA-DA-R mechanism is more similar to the DA than to the IA, and if there is no affirmative action the IA-DA-R mechanism is equivalent to the DA and minority fairness simplifies to fairness (note that the standard stability axiom corresponds to the conjunction of fairness and non-wastefulness in our model). Moreover, just like for the DA which may have Nash-equilibria that Pareto-dominate the student-optimal stable matching at some profiles (Dur and Morrill, 2020), the DA-R matching is not necessarily an undominated Nash-equilibrium in the strategic game induced by IA-DA-R. Whether this strategic game would make it a focal point for sophisticated players with complete information to reach the DA-R outcome is not clear, but nonetheless having the DA-R as a prominent Nash-equilibrium matching is a positive theoretical feature of the IA-DA-R mechanism (albeit not a strong one due to the multiplicity of equilibria), as the DA-R arguably has the best properties among all the previous mechanisms on the basis of the three main welfare criteria. However, we want to emphasize that this is merely a theoretical attribute, not a realistic scenario, and untruthful Nash-equilibria are unlikely to be realized in most circumstances. We expect the IA-DA-R outcome to differ from the DA-R and other Nash-equilibrium outcomes in realistic settings, where minority students do not have the information, the resources or the interest to identify their best responses.

References

Abdulkadiroğlu, A. and Grigoryan, A. (2021). Priority-based assignment with reserves and quotas. *Working paper.*

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

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

Ashlagi, I. and Gonczarowski, Y. A. (2018). Stable matching mechanisms are not obviously strategy-proof. *Journal of Economic Theory*, 177:405–425.

Aygün, O. and Bó, I. (2021). College admission with multidimensional privileges: The Brazilian affirmative action case. *American Economic Journal: Microeconomics*, 13(3):1–28.

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

Aziz, H., Gaspers, S., and Sun, Z. (2020). Mechanism design for school choice with soft diversity constraints. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 153–159.

Aziz, H. and Sun, Z. (2021). Multi-rank smart reserves. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 105–124.

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

Basteck, C. and Mantovani, M. (2023). Aiding applicants: Leveling the playing field within the immediate acceptance mechanism. *Review of Economic Design*, 27(1):187–220.

Chen, L. and Pereyra, J. S. (2019). Self-selection in school choice. *Games and Economic Behavior*, 117:59–81.

Chen, Y., Huang, Y., Jiao, Z., and 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. and Kesten, O. (2019). Chinese college admissions and school choice reforms: An experimental study. *Games and Economic Behavior*, 115:83–100.

Chen, Y. and Sönmez, T. (2006). School choice: An experimental study. *Journal of Economic Theory*, 127(1):202–231.

Ding, F., Hong, S., Jiao, Z., and Luo, X. (2019). Corrigendum to "Affirmative action in school choice: A new solution" [*Mathematical Social Sciences,* 92 (2018) 1–9]. *Mathematical Social Sciences*, 97:61–64.

Ding, T. and Schotter, A. (2019). Learning and mechanism design: An experimental test of school matching mechanisms with intergenerational advice. *The Economic Journal*, 129(623):2779–2804.

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

Doğan, B. and 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., and 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. and Freedman, D. A. (1981). Machiavelli and the Gale-Shapley algorithm. *American Mathematical Monthly*, 88(7):485–494.

Dur, U., Kominers, S. D., Pathak, P. A., and Sönmez, T. (2018). Reserve design: Unintended consequences and the demise of Boston's walk zones. *Journal of Political Economy*, 126:2457–2479.

Dur, U., Pathak, P. A., and Sönmez, T. (2020). Explicit vs. statistical targeting in affirmative action: Theory and evidence from chicago's exam schools. *Journal of Economic Theory*, 187:104996.

Dur, U. M. and Morrill, T. (2020). What you don't know can help you in school assignment. *Games and Economic Behavior*, 120:246–256.

Echenique, F., Wilson, A. J., and Yariv, L. (2016). Clearinghouses for two-sided matching: An experimental study. *Quantitative Economics*, 7(2):449–482.

Echenique, F. and Yenmez, M. B. (2015). How to control controlled school choice. *American Economic Review*, 105(8):2679–2694.

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

Ehlers, L. and Klaus, B. (2003). Coalitional strategy-proof and resource-monotonic solutions for multiple assignment problems. *Social Choice and Welfare*, 21(2):265–280.

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

Fack, G., Grenet, J., and He, Y. (2019). Beyond truth-telling: Preference estimation with centralized school choice and college admissions. *American Economic Review*, 109(4):1486–1529.

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

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

Haeringer, G. and Klijn, F. (2009). Constrained school choice. *Journal of Economic Theory*, 144(5):1921–1947.

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

Hakimov, R. and Kübler, D. (2021). Experiments on centralized school choice and college admissions: A survey. *Experimental Economics*, 24(2):434–488.

Hassidim, A., Romm, A., and Shorrer, R. I. (2018). "Strategic" behavior in a strategy-proof environment. *Working paper.*

Jiao, Z. and Tian, G. (2019). Responsive affirmative action in school choice: A comparison study. *Economics Letters*, 181:140–145.

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

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

Klijn, F., Pais, J., and Vorsatz, M. (2013). Preference intensities and risk aversion in school choice: A laboratory experiment. *Experimental Economics*, 16:1–22.

Klijn, F., Pais, J., and Vorsatz, M. (2016). Affirmative action through minority reserves: An experimental study on school choice. *Economics Letters*, 139:72–75.

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

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

Kominers, S. D. and Sönmez, T. (2016). Matching with slot-specific priorities: Theory. *Theoretical Economics*, 11(2):683–710.

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

Luce, R. D. and Raiffa, H. (1957). *Games and decisions: Introduction and critical survey.* New York: Wiley.

Pais, J. and Pintér, Á. (2008). School choice and information: An experimental study on matching mechanisms. *Games and Economic Behavior*, 64(1):303–328.

Pais, J., Pintér, Á., and Veszteg, R. F. (2011). College admissions and the role of information: An experimental study. *International Economic Review*, 52(3):713–737.

Pathak, P. A., Rees-Jones, A., and Sönmez, T. (2022). Immigration lottery design: Engineered and coincidental consequences of H-1B reforms. *Review of Economics and Statistics*, forthcoming.

Pathak, P. A. and Sönmez, T. (2008). Leveling the playing field: Sincere and sophisticated players in the Boston mechanism. *American Economic Review*, 98(4):1636–1652.

Pathak, P. A., Sönmez, T., Ünver, M. U., and Yenmez, M. B. (2023). Leaving no ethical value behind: Triage protocol design for pandemic rationing. *Management Science*, forthcoming.

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

Sönmez, T. and Yenmez, M. B. (2022). Affirmative action in India via vertical, horizontal, and overlapping reservations. *Econometrica*, 90(3):1143–1176.

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

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

Velez, R. A. and Brown, A. L. (2019). Empirical strategy-proofness. *Working paper*.

Westkamp, A. (2013). An analysis of the german university admissions system. *Economic Theory*, 53:561–589.