Fair Implementation of Diversity in School Choice*

JOB MARKET PAPER

Inacio Bo†

This version: January, 2014‡

Abstract

Many school districts have objectives regarding how students of different race, ethnicity or religious backgrounds should be distributed across schools. A growing literature in mechanism design introduce school choice mechanisms that attempt to satisfy those requirements. We show that these mechanisms may fail to a great extent to satisfy those objectives, and introduce a new one, which satisfies two properties. First, it produces assignments that satisfy a fairness criterion which incorporates the diversity objectives as an element of fairness. Second, it approximates optimally the diversity objectives while still satisfying the fairness criterion. We do so by embedding “preference” for those objectives into the schools’ choice functions in a way that satisfies the substitutability condition and then using the school-proposing deferred acceptance procedure. This leads to the equivalence of stability with the desired definition of fairness and the maximization of those diversity objectives among the set of fair assignments. We complement the paper with a comparative analysis that shows analytically that the mechanism that we provide has a general ability to satisfy those objectives, while in many familiar classes of scenarios the alternative ones yield segregated assignments.

JEL classification: C78, D63, D78, D82

Keywords: Mechanism Design, Matching, School Choice, Affirmative Action, Diversity.

*Preliminary. I am grateful to Utku Ünver and Tayfun Sönmez for their advice and comments in this work. I also thank Samson Alva, Orhan Aygün, Onur Kesten, Vikram Manjunath, Bertan Turhan and the seminar participants at Boston College and the Royal Economic Society Postgraduate Meeting for their helpful comments. All errors are mine, despite their efforts.

†Address: Department of Economics, Boston College, 140 Commonwealth Ave, Chestnut Hill MA 02467; website: http://www2.bc.edu/inacio-bo/; e-mail: boin@bc.edu.

‡First version: September 2013
1 Introduction

Racial and ethnic diversity in school cohorts is believed to be a key condition to obtain more cohesive communities and a less segregated society.\(^1\) Over the last five decades a multitude of policies has been implemented, with varying degree of success, to reduce historical and emerging racial, religious and ethnic segregation at the school level. In the United States, numerous school districts implemented desegregation efforts through methods ranging from the ability of choosing which school to apply to forced reallocation of students.\(^2\) More recent such efforts include Brazil’s racial and income-based public university reserves (Aygün and Bo, 2013) and the attempt to increase religious diversity in british schools (Coldron et al., 2008).

Most policies used to achieve that objective consist in either establishing maximum quotas for the so-called majority students (that is, a maximal number of majority students that are allowed at a certain school) or giving higher priority to minority students in either all or part of the seats available.

Failure to design adequate mechanisms may have severe consequences. In their effort to increase the proportion of white students in schools that were attended predominantly by black students, the school district of Kansas City reserved a significant number of seats exclusively for white students in order to satisfy a court-ordered ratio of 60/40 black/white students. The result is detailed in Cioitti (2001):

An overzealous commitment to their desegregation plan sometimes led proponents of the plan to take positions seemingly at odds with their ultimate goal of helping inner-city blacks. At one point the Landmark Legal Foundation had to go to court to stop the district from enforcing a quota that allowed desks to sit empty in new magnet schools (waiting for whites who never came) while some overcrowded all-black schools had to house their students in trailers. If a white suburban student wanted to go to a magnet school, admission was automatic because that brought the district closer to the 60/40 black/white ratio ordered by the judge. If a black student wanted to go to the same school, however, that student often ended up on a waiting list. As a result, some black parents registered their children as white in order to get them into certain schools. Finally, the dis-

\(^{1}\)For an analysis of the effects of ethnic and racial segregation on community cohesion and the role of school segregation in the UK, for example, see Ministerial Group on Public Order and Community Cohesion (2001)
\(^{2}\)For a comprehensive survey of methods used in those efforts until the 1980s, see Welch and Light (1987).
trict had discovered that it was easier to meet the court’s 60/40 integration ratio by letting black students drop out than by convincing white students to move in. As a result, nothing was done in the early days of the desegregation plan about the district’s appalling high school dropout rate, which averaged about 56 percent in the early 1990s (when desegregation pressures were most intense) and went as high as 71 percent at some schools (for black males it was higher still.)

(...) Twenty-five percent of the KCMSD’s 37,000 students were white. Thus, to meet the court-mandated ratio of 40 percent white to 60 percent black, the district needed to attract 10,000 additional white students.

Since the seminal work on the subject by Abdulkadiroğlu and Sönmez (2003), a growing number of papers have used mechanism design principles to obtain school assignments that achieve some balance between diversity objectives, fairness, efficiency and other properties. One class of such mechanisms, which we denote affirmative action mechanisms, expands the set of schools that certain types of students have access to by giving them higher priority and/or reserving some seats in the schools to be filled by those students, making the seats available to everyone otherwise.\(^3\) Examples of affirmative action mechanisms include bonuses on exam scores for students from public schools in university admissions (Matos et al., 2012) and giving higher priority for racial minorities on a number of seats in schools.\(^4\) Another class of mechanisms take diversity as an objective instead, and accommodates other properties, such as fairness or constrained efficiency. We denote that class of mechanisms diversity implementation mechanisms. Mechanisms with majority quotas (which enforce a maximum number of “majority type” students in each school) or others that enforce certain ratios among types of students are examples of diversity implementation mechanisms.

For problems such as university admissions, where, in many cases, the determinant of admission is a student’s performance in tests and high-school grades, affirmative action mechanisms could increase diversity of cohorts by improving access of minority students to more competitive universities. In the case of school choice, however, that is not necessarily the case. Typically the criteria for admission rely on aspects such as residence location, presence of siblings in the school, special needs, etc (see, for example, Coldron et al. (2008).) That is, minority students are not necessarily disadvantaged with respect to others in their access to

---

\(^3\)The term affirmative action is normally used in a broader sense across the literature. Mechanisms such as those in Abdulkadiroğlu and Sönmez (2003) and Abdulkadiroğlu (2005) are denoted in those papers as implementing affirmative action while in our terminology those are diversity implementation mechanisms.

\(^4\)Other examples of affirmative action mechanisms can be found in Hafalir et al. (2013) and Ehlers et al. (2011).
desired schools, and thus the use of such mechanisms may not help in obtaining more diverse groups of students. In fact, in section 4 we show that in many scenarios those can lead to a completely segregated distribution of students.

We introduce a new diversity implementation mechanism that differs from others available in the literature in two main aspects: the incorporation of diversity objectives as an element of fairness and a more pragmatic interpretation of those objectives, where a given distribution of types in a school is used as a target instead of a strict objective.

Consider first the definition of fairness. One that is used in the literature for school choice with diversity concerns is same-type fairness\(^5\): a school assignment is fair if no student \(s\) of type \(t\) assigned to a school \(c\) would prefer to be assigned to a school \(c'\) where there is another student \(s'\) of type \(t\), while \(s\) has a higher priority than \(s'\) at \(c'\). Using a common terminology in the literature, no student justifiably envies another student of the same type. Although that definition allows, for example, minority students with low priority to be assigned to a school instead of some majority students with high priority, the definition fails to capture the diversity objectives as an element of fairness, giving instead property rights to a set of seats to students of a certain type, regardless of those objectives. In fact, if the diversity objectives are, for example, to have an equal number of minority and majority students in each school, one school with only minority students and another with only majority students satisfies same-type fairness even if there are students of different types that would prefer to be assigned to each other’s school.

The definition of fairness that we use (and that the mechanism we propose satisfies) is instead that of fairness with diversity.\(^6\) An assignment is fair with diversity if it is individually rational, non-wasteful, and no student \(s\) of type \(t\) assigned to a school \(c\) would prefer to be assigned to a school \(c'\), where one of the following is true:

- There is a student \(s'\) assigned to \(c'\), who has a lower priority than \(s\), and replacing \(s'\) with \(s\) would not affect the satisfaction of the diversity objectives.

- It is possible to replace a student \(s'\) assigned to \(c'\) by \(s\), and as a result \(c'\) would strictly improve how much a certain diversity objective is satisfied in that school, without negatively affecting another diversity objective in \(c'\).

Therefore, an assignment that is fair with diversity incorporates the diversity objectives as an element of fairness. In the example above, the assignment in which one school with only

---

\(^5\)See Ehlers et al. (2011) and Troyan and Fragiadakis (2013).

\(^6\)A similar definition of fairness is presented in the context of an affirmative action mechanism in Ehlers et al. (2011).
minority students and another with only majority students could only be fair with diversity if every student prefers its assignment to being assigned to the other school. Otherwise, that wouldn’t be fair with diversity.

Another way by which the mechanism presented here differs from the other diversity implementation mechanisms in the literature is by the fact that it doesn’t use the diversity objectives as binary objective (that is either satisfied or not). It instead quantifies the satisfaction of those objectives at a given assignment, which allows for the maximization of those subject to the fairness definition and the actual distribution of types in the population, without having to rely on assumptions over that distribution. The quote above on the problems encountered during the desegregation process in Kansas City gives an example of why being able to adapt the diversity objectives to the actual population distribution is fundamental in practical applications.

From a theoretical perspective, the key novelty in this paper is to use the school-proposing deferred acceptance procedure while at the same time “designing preferences” for the schools, in the form of a choice function that satisfies the substitutability condition. While marriage markets and college admissions problems are two-sided matching problems in which the welfare and incentives of both sides are under consideration, in a school choice problem the seats in the schools are simply objects to be allocated to students. Therefore the school’s choice function, instead of representing some sort of preference that schools have over students sets, can be designed in a way such that the property of stability and the relationship between the school’s choice function and the set of stable allocations presented in Roth (1984)\(^7\) induce the desired properties on the allocation.

While the use of the school-proposing version of the deferred acceptance procedure allows the use of the result in Roth (1984) to select among stable assignments with some degree of arbitrariness, the student-proposing deferred acceptance procedure is strategy-proof for the students and has positive well-known welfare properties. We argue that the choice of which procedure to use depends on the desired allocation along the trade-off presented by these options.

Many school districts give higher priority on a school to students that have their residence in an area that surrounds it (in districts in which school choice mechanisms aren’t implemented students are actually restricted to schools in their neighborhoods). It is also the case that there is, especially in urban areas, a high degree of residential segregation on their occupation\(^8\), forming racial, income-based and religious enclaves. The combination of these

---

\(^7\)More specifically, Lemma 4 presented in section 3 of this paper.

\(^8\)See Steinmetz et al. (2002) for an analysis of recent trends on residential segregation in the U.S.
two elements usually leads, also, to a high degree of segregation in school cohorts (Clotfelter, 1998).

We complement our paper with an analytic evaluation of how the outcomes of our mechanism compares to three other mechanisms for assigning students to schools: Neighborhood-based Student Assignment (any mechanism that restricts students to schools in their neighborhood), Top Trading Cycles (a Pareto efficient mechanism) and an affirmative action mechanism that generates assignments that are fair with diversity. We show that in every scenario analyzed the mechanism proposed in this paper is able to minimize the segregation of school assignments, while the other three alternatives often lead to segregated assignments.

1.1 Main Results

Our first results shows that, for a school choice problem with diversity objectives, the traditional concept of fairness (using priority-based no-envy) and even that of fairness with diversity are incompatible with the strict enforcement of those diversity objectives (Propositions 1 and 3). Despite those negative results, we propose a mechanism (Diversity Deferred Acceptance, or DDA) that generates an assignment that satisfies, in a well-defined maximal way, those diversity objectives without being wasteful or compromising fairness (Theorem 1). The mechanism we proposed may be vulnerable to strategic manipulation by students (Example 3). However, there is no mechanism that implements diversity with fairness and is strategy-proof (Theorem 3). Moreover, we show that when the number of students is large or students have “low information” on others’ preferences and priorities, gains from manipulating their preferences disappear (Theorems 4 and 5).

Given that the objective of using such mechanisms is to reduce the segregation of students across schools without jeopardizing fairness, we compare the assignments generated by DDA and 3 other mechanisms. These are representative of the available alternatives currently used or proposed in the literature: Neighborhood School Assignment (NSA), Top Trading Cycles (TTC) and Deferred Acceptance Minority Reserves (DAMR) under some symmetric population distributions. We show that the DDA mechanism is able to minimize segregation regardless of schools’ priorities or students’ preferences (Theorem 2). We show that, on the other hand, the DAMR mechanism yields segregated assignments for some familiar preference profiles (Propositions 5, 6 and 7). When students’ geographic distribution is segregated (that is, if their residence location is correlated with their types) and schools give higher priority for students who live in that school’s neighborhood, we show that the use of NSA and TTC results in totally segregated school cohorts (Proposition 9).
1.2 Relation with the literature

The concept of stable matchings was first introduced by Gale and Shapley (1962). Inspired by the problem of college admissions, the authors also present two procedures that produce stable matchings: the Student-Proposing Deferred Acceptance and the College-Proposing Deferred Acceptance (SPDA and CPDA). They showed that the outcome of the SPDA is student-optimal in the sense that the matching it generates is preferred by every student to any other stable matching. Similarly, the outcome of the CPDA is college-optimal in the sense that its outcome is preferred by every college to any other stable matching.

When one of those mechanisms is used, a game is induced on the participants, where the stated preferences are the strategies and their matches the outcomes. While Dubins and Freedman (1981) show that when using the SPDA no student or group of students can be made better-off by misrepresenting their preferences, Gale and Sotomayor (1985) show that this is not normally the case if using CPDA. Furthermore, Roth (1985) show that there is no stable mechanism that is immune to manipulation by colleges.9

The incentive and welfare properties of both mechanisms come into play in the context of college admissions in Balinski and Sönmez (1999). Arguing that colleges should not be considered strategic agents but their seats simple objects to be consumed by the students, they eliminate the need for strategic or welfare considerations on the part of colleges. Moreover, they show that when colleges preferences are based on natural priorities (e.g. exam scores), stability is equivalent to an intuitive notion of fairness. As a result, the SPDA is suggested as the ideal mechanism for the student placement problem.

The subsequent literature on college admissions and school choice, as well as their applications, focuses on the use of the SPDA procedure (see Abdulkadiroğlu and Sönmez (2003), Abdulkadiroğlu et al. (2005), Abdulkadiroğlu et al. (2006) and Abdulkadiroğlu et al. (2009)). When concerns about diversity on the distribution of students across schools were introduced in the mechanisms, that choice persisted. Abdulkadiroğlu and Sönmez (2003) uses the SPDA associated with a maximum quota for certain types. After Kojima (2012) showed that maximal quotas may hurt every minority student, focus shifted to minority reserves, used for example in Ehlers et al. (2011), Echenique and Yenmez (2012), Erdil and Kumano (2012), Kominers and Sönmez (2012) and Hafalir et al. (2013). These papers attempt to satisfy the diversity constraints by embedding them into the schools’ choice function.

9Whereas Dubins and Freedman (1981) assumes that the colleges’ choices over the students can be represented by ranking them and choosing the most preferred ones up to a capacity constraint, similar results are shown for more general choice functions in, among others, Hatfield and Kojima (2010) and Abdulkadiroğlu (2005).
When students have strict preferences, however, diversity objectives and students’ welfare (when measured in terms of these preferences) may go in opposite directions. As a result, unless students’ preferences and those objectives “agree” with each other, the use of the SPDA procedure may fail to satisfy them (section 4 analyzes scenarios where this may happen).

By combining the use of the CPDA procedure with a choice function that satisfies substitutability, we are able to obtain assignments that implement (or approximate) those diversity objectives in a wider range of scenarios, while still satisfying a fairness criterion. To the best of our knowledge this is the first paper to suggest the use of the CPDA procedure in this way.

While not using the CPDA procedure, two other papers make a similar attempt to satisfy distributional concerns in school assignments without having to rely on students’ preferences to do so. Ehlers et al. (2011) presents, to the best of our knowledge, the first mechanism that applies the diversity constraints as a distributional objective instead of an advantage for some students. Since they look for mechanisms that are able to perfectly satisfy those objectives, however, their results depend on a consistency between them and the distribution of types in the population. Moreover, the outcome of their mechanism puts the responsibility for the implementation of the diversity objectives partially on the students themselves: a student may not be able to be assigned to a school she prefers, which has an empty seat available, if by doing so the diversity objectives in her assigned school would be violated. Our mechanism presents improvements in both issues.

Troyan and Fragiadakis (2013) also presents a mechanism which has the objective of implementing diversity objectives, focusing on the property, absent in Ehlers et al. (2011) and in the present paper, of strategy-proofness. Strategy-proofness, however, comes at a cost: their proposed mechanisms satisfies same-type fairness but is *wasteful*, that is, schools may end up with empty seats that are desired by some students. Our mechanism, while not strategy-proof, has fairness requirements that are stronger than same-type fairness and has good incentives properties in large markets and low information scenarios.

The paper proceeds as follows. Section 2 introduces the model and the basic definitions of fairness and implementation of diversity. Section 3 presents the DDA mechanism and its general properties. Section 4 compares outcomes of the DDA mechanism to those generated other school choice mechanisms. Section 5 discusses the incentives in the game induced by the DDA mechanism. Section 6 concludes. Proofs omitted from the main text are found in the Appendix.

---

10This statement refers to the “hard-bounds” mechanism in Ehlers et al. (2011).

11More specifically, every assignment that is fair in our model is fair on theirs, but not vice-versa.
2 Model

A school choice with diversity problem consists of a tuple \( \langle S, C, T, \tau, q, d >, s_c >, C > \rangle \):

1. A finite set of students \( S = \{ s_1, \ldots, s_{|S|} \} \)
2. A finite set of schools \( C = \{ c_1, \ldots, c_{|C|} \} \)
3. A finite set of types \( T = \{ t_1, \ldots, t_k \} \)
4. A function \( \tau : S \to T \) where \( \tau (s) \) is the type of student \( s \). We denote by \( S^t (I) \) the set of students in \( I \subseteq S \) of type \( t \), that is, \( S^t (I) = \{ s \in I : \tau (s) = t \} \).
5. A capacity vector \( q = (q_{c_1}, \ldots, q_{c_m}) \) where \( q_c \) is the capacity of school \( c \in C \).
6. For each school \( c \), a vector \( q^T_c = (q^t_{c_1}, \ldots, q^t_{c_k}) \) of diversity objectives, where \( q^t_c \) is the minimum desired number of students with type \( t \) at school \( c \), where \( \sum_{t \in T} q^t_c \leq q_c \). Let \( \underline{q} = (q^T_{c_1}, \ldots) \).
7. Students’ preference profile \( >S = (>_1, \ldots, >_{|S|}) \), where \( >_s \) is a strict ranking over \( C \cup \{ s \} \), where \( s \) represents remaining unmatched to any school. If \( s >_c s \), school \( c \) is deemed unacceptable to student \( s \).
8. Schools’ priority profile \( >C = (>_c, \ldots, >_{c|C|}) \), which is a collection of complete and strict rankings over the students in \( S \cup \emptyset \). If \( \emptyset >_c s \), student \( s \) is deemed unacceptable to school \( c \).

An assignment \( \mu \) is a function from \( C \cup S \) to subsets of \( C \cup S \cup \{ \emptyset \} \) such that:

- \( \mu (s) \in C \cup \{ s \} \) and \( |\mu (s)| = 1 \) for every student \( s \)
- \( |\mu (c)| \leq q_c \) and \( \mu (c) \subseteq S \) for every school \( c \)
- \( \mu (s) = c \) if and only if \( s \in \mu (c) \)

For a student \( s \), \( \mu (s) \) is the school to which \( s \) is assigned under \( \mu \), and for a school \( c \), \( \mu (c) \) is the set of students that are assigned to school \( c \) under \( \mu \). For a given school choice with diversity problem, we will denote by \( M \) the set of all assignments. A set of students \( I \subseteq S \) enables diversity at school \( c \) if for all \( t \in T \), \( |S^t (I)| \geq q^t_c \). An assignment \( \mu \) fully implements diversity if for every school \( c \), \( \mu (c) \) enables diversity at \( c \). If there is

\[ 12 \text{We will abuse notation and consider } \mu (s) \text{ as an element of } C, \text{ instead of a set with an element of } C. \]
an assignment $\mu^* \in \mathcal{M}$ where $\mu^*$ fully implements diversity, we say that diversity objectives are feasible. We say that a student $s$ justifiably claims an empty seat at school $c$ under the assignment $\mu$ if $|\mu(c)| < q_c$ and $c >_s \mu(s)$. An assignment $\mu$ is non-wasteful if no student justifiably claims an empty seat at some school. An assignment is individually rational if for every student $s$, $\mu_p s q$ and $c = \mu_p c q$. An assignment $\mu$ is non-wasteful if no student justifiably claims an empty seat at some school. An assignment is individually rational if for every student $s$, $\mu_p s q$ and for every school $c$ and every student $s' \in \mu(c)$, $s' > c \emptyset$. A traditionally desirable condition for an assignment to satisfy is that of having no student that justifiably envies another. We define that formally, using Balinsky and Sönmez’s (1999) notion of fairness:

**Definition 1.** A student $s$ justifiably envies student $s'$ under the assignment $\mu$, where $c = \mu(s')$, if and only if $c >_s \mu(s)$ and $s >_c s'$. An assignment $\mu$ satisfies no justified envy if no student justifiably envies another under $\mu$. An assignment $\mu$ is fair if it is non-wasteful and satisfies no justified envy.

A mechanism that chooses only assignments that satisfy no justified envy is one that uses the priority ordering of a school as both an implementation of property rights and as a public and verifiable instrument by which the public can assess the fairness of the outcome. Although an assignment that is fair always exists (Gale and Shapley, 1962; Balinski and Sönmez, 1999; Abdulkadiroğlu and Sönmez, 2003), an assignment that is fair and fully implements diversity may not exist:

**Proposition 1.** There may be no fair assignment that fully implements diversity, even if diversity objectives are feasible.

**Proof.** Consider the following school choice with diversity problem:

$$S = \{s_1, s_2\}$$
$$T = \{t_1\}$$
$$C = \{c_1, c_2\}$$
$$S_{t_1}(S) = \{s_1\}$$

$>_{s_1} = c_2 c_1$ $>_{c_1} = s_2 s_1$

$>_{s_2} = c_2 c_1$ $>_{c_2} = s_2 s_1$

Capacities are $q_{c_1} = q_{c_2} = 1$, diversity objectives are $q_{c_1}^T = (q_{c_1}^{d_1}) = (0)$ and $q_{c_2}^T = (q_{c_2}^{d_1}) = (1)$. Consider the following assignments:

$$\mu = \begin{pmatrix} c_1 & c_2 \\ s_1 & s_2 \end{pmatrix}$$
$$\mu' = \begin{pmatrix} c_1 & c_2 \\ s_2 & s_1 \end{pmatrix}$$

Assuming that both the priority ranking and the assignment are public information, a student can verify whether a school that she ranked higher than her assigned school incorrectly accepted one with lower priority.
Diversity objectives are feasible, since the assignment $\mu'$ fully implements diversity. The unique fair assignment is $\mu$, which doesn’t fully implement diversity.

Notice that in the example used in Proposition 1, the reason why the assignment that fully implements diversity isn’t fair is that the fairness criterion ignores diversity. That is, if the definition of fairness incorporated a higher priority for student with type $t$ whenever the diversity objective associated with $t$ isn’t yet satisfied, $\mu'$ would be a fair allocation. In order to accommodate these concerns, we first define a modification of the concept of justified envy as follows:

**Definition 2.** A student $s$ **justifiably demands a seat in school** $c$ if $c >_s \mu(s)$ and either:

1. $|\tau(s) (\mu(c))| < q^t_c$.
2. There is a student $s' \in \mu(c)$ such that $\tau(s') = \tau(s)$ and $s >_c s'$.
3. There is $t' \in T$ and $s' \in S^{t'} (\mu(c))$ such that $|S^{t'} (\mu(c))| > q''_c$ and $s >_c s'$.

An assignment is **fair with diversity** if it is individually rational, non-wasteful and no student justifiably demands a seat in any school.\(^{14}\)

Put more informally, a student $s$ justifiably demands a seat in a school $c$, which is preferred by $s$ to her assigned school, under 3 circumstances:

- She has a type associated with a diversity objective that isn’t currently being satisfied at $c$.
- She has a higher priority than another student of her type who is assigned to $c$.
- She has a higher priority at $c$ than some student whose acceptance wasn’t determined by a diversity objective.

In order to obtain an assignment that is fair with diversity, therefore, a mechanism must first focus on students that have a type that satisfies some diversity objective in a school up to the point in which that objective is satisfied. Whenever the number of students of a type is higher than the diversity objectives for that type, their normal priorities become the criterion

\(^{14}\)The reader may verify that in the example used in proposition 1 the unique fair with diversity assignment fully implements diversity.
over which those students are selected. Every seat in a school that isn’t assigned because of a diversity objective has priorities as the sole criterion of selection.

The set of fair with diversity assignments is a superset of the set of fair assignments that fully implement diversity:

**Proposition 2.** If $\mu$ is fair and fully implements diversity, then $\mu$ is also fair with diversity.

*Proof.* Since $\mu$ is non-wasteful, we only need to show that if $\mu$ is fair and fully implements diversity then no student justifiably demands a seat in any school. Let $s$ and $c$ be such that $\mu (s) \neq c, c \succ_s \mu (s)$. We show why none of the three conditions in definition 2 is satisfied:

- **Condition 1:** Let $t = \tau (s)$. Since $\mu$ fully implements diversity, $|S^t(\mu(c))| \geq q^t_c$.
- **Conditions 2 and 3:** Since $\mu$ is fair, there is no $s' \in \mu (c)$ such that $s \succ_c s'$.

The following result, however, shows that a fair with diversity assignment that fully implements diversity may not exist even when diversity objectives are feasible:

**Proposition 3.** There may not be an assignment that is fair with diversity and fully implements diversity, even if diversity objectives are feasible.

*Proof.* Consider the following school choice with diversity problem:

$$
S = \{s_1, s_2\} \\
T = \{t_1\} \\
C = \{c_1, c_2\} \\
S^t_1 (S) = \{s_1\} \\
>_{s_1}: c_1 c_2 \\
>_{s_2}: c_2 c_1
$$

Capacities are $q_{c_1} = q_{c_2} = 1$, diversity objectives are $q^{t_1}_{c_1} = (q^{t_1}_{c_2}) = (0)$ and $q^{t_1}_{c_2} = (q^{t_1}_{c_1}) = (1)$. Consider the following assignments:

$$
\mu = \begin{pmatrix}
    c_1 & c_2 \\
    s_1 & s_2
\end{pmatrix} \\
\mu' = \begin{pmatrix}
    c_1 & c_2 \\
    s_2 & s_1
\end{pmatrix}
$$

Diversity objectives are feasible, since the assignment $\mu'$ fully implements diversity. The unique fair with diversity assignment is $\mu$, which doesn’t fully implements diversity.  

\[\square\]
Even though it is not always possible to obtain an assignment that fully implements diversity, we would like to have the alternative of choosing one that is as “close” to that objective as possible. In order to achieve that, we first propose the following partial order:

**Definition 3.** Let $\trianglerighteq$ be the partial order over the set of assignments $\mathcal{M}$ such that $\mu' \trianglerighteq \mu$ if:

1. For all $c \in C$ and $t \in T$ such that $|S^t(\mu(c))| \leq q^t_c$, $|S^t(\mu'(c))| \geq |S^t(\mu(c))|$ and

2. There are $c' \in C$ and $t' \in T$ such that $|S^{t'}(\mu(c'))| < q^{t'}_{c'}$ and $|S^{t'}(\mu'(c'))| > |S^{t'}(\mu(c'))|$.

Denote $\mu' \trianglerighteq \mu$ if $\mu' >\trianglerighteq \mu$ is false.

In other words, $\mu' >\trianglerighteq \mu$ if $\mu'$ has less seats “reserved” for students of certain types occupied by students that are not of those types when compared to $\mu$. Notice that if $\mu' >\trianglerighteq \mu$ then it cannot be the case that both $\mu'$ and $\mu$ enable diversity in every school. As an example, suppose that there is only one school $c$ with 100 seats and the diversity objective says that at least 50 of them should be occupied by minority students. If $\mu$ assigns 40 minority students to $c$ and $\mu'$ assigns 45, then $\mu' >\trianglerighteq \mu$. However, If $\mu''$ assigns 51 minority students to $c$ and $\mu'''$ assigns 55, both $\mu'' >\trianglerighteq \mu'''$ and $\mu''' >\trianglerighteq \mu''$. Increasing the number of minority students after the diversity objective is satisfied doesn’t make an assignment greater with respect to $\trianglerighteq$. That is why we believe that this is the intuitive and correct method to compare assignments with respect to how they satisfy diversity objectives.

We now define formally what it means to implement diversity in this framework.

**Definition 4.** An assignment $\mu$ **implements diversity** if $\mu$ is fair with diversity and there is no assignment $\mu'$ such that $\mu'$ is fair with diversity and $\mu' >\trianglerighteq \mu$. A mechanism implements diversity if for every school choice with diversity problem the assignment it generates implements diversity.

An assignment $\mu$ implements diversity, thus, if either $\mu$ fully implements diversity or $\mu$ doesn’t fully implements diversity but there is no assignment that is fair with diversity and “further satisfies” some diversity objective is some school without jeopardizing another in the same or some other school.

### 3 The mechanism

We start by defining the choice function $\mathcal{C}_c : 2^S \rightarrow 2^S$ at school $c$ for the school choice with diversity problem $\langle S, C, T, \tau, q, q, >_S, >_C \rangle$. Fix any $S' \subseteq S$ and let $I \subseteq S'$ be the set of students **acceptable to** $c$ among $S'$. $\mathcal{C}_c(S')$ is defined by the following procedure.
1. **Step 0**: If $|I| \leq q_c$, $C_c(S') = I$.

2. **Step 1**: If $|S^t(I)| < q^t_c$, accept all students in $S^t(I)$. Otherwise accept the top $q^t_c$ students in $S^t(I)$ with respect to $>_c$. Denote by $\Psi_{t_1}(I)$ the set of students accepted in this step.

3. **Step 1 < \ell \leq k** (*the step associated with* $t_\ell$): If $|S^{t_\ell}(I)| < q^{t_\ell}_c$, accept all students in $S^{t_\ell}(I)$. Otherwise accept the top $q^{t_\ell}_c$ students in $S^{t_\ell}(I)$ with respect to $>_c$. Denote by $\Psi_{t_\ell}(I)$ the set of students accepted until step $\ell$.

4. **Final step**: If $|\Psi_{t_k}(I)| < q_c$, accept the top $q_c - |\Psi_{t_k}(I)|$ students in $I \setminus \Psi_{t_k}(I)$ with respect to $>_c$.

The choice function above is, following the definitions in Echenique and Yenmez (2012), a choice rule *generated by reserves*. It is also a generalization for multiple types of the choice function used in Hafalir et al. (2013). We will proceed now to show some important properties of $C_c$.

**Definition 5.** A choice function $C$ satisfies the *substitutability condition* if for all $z, z' \in X$ and $Y \subseteq X$:

$$z \notin C(Y \cup \{z\}) \implies z \notin C(Y \cup \{z, z'\})$$

**Lemma 1.** The function $C_c$ satisfies the substitutability condition.

**Proposition 4.** Let $I \subseteq S$ and $c \in C$ be such that every student in $I$ is acceptable to school $c$ and for every $t \in T$, $S^t(I) \geq q^t_c$. Then $C_c(I)$ enables diversity at $c$.

**Proof.** Suppose not. Then there is at least one type $t \in T$ such that $|S^t(C_c(I))| < q^t_c$. Suppose first that $|I| < q_c$. Then the choice procedure would accept all students in $I$ in step 0, which contradicts with $|S^t(I)| \geq q^t_c$, so it must be that $|I| \geq q_c$ and that the choice procedure will run until the final step. Notice, however, that in the step associated with type $t$, since $|S^t(I)| \geq q^t_c$, the top $q^t_c$ students with respect to $>_c$ are accepted, which is a contradiction with $|S^t(C_c(I))| < q^t_c$. $\square$
The proposition above shows that whenever the set of students available is such that there are subsets of them which enable diversity at \( c \), \( \mathcal{C}_c \) selects one of them. As we will show in section 4, however, this does not guarantee that the outcome of a deferred acceptance procedure fully implements diversity.

**Lemma 2.** Let \( |\mathcal{C}_c (I)| < q_c \) and \( I' \subseteq I \) be the set of acceptable students for school \( c \) in \( I \). Then the following are true:

1. \( \mathcal{C}_c (I) = I' \).
2. \( \mathcal{C}_c (I \cup \{ s \}) = \mathcal{C}_c (I) \cup \{ s \} \) for any \( s \in S \) if \( s \) is acceptable to \( c \).
3. \( |\mathcal{C}_c (I \cup J)| > |\mathcal{C}_c (I)| \) for any \( J \subseteq S \) such that for every \( s \in J \), \( s \) is acceptable to \( c \), \( I \cap J \neq \emptyset \) and \( I \neq J \).

The results presented in the lemma above come easily from the definition of the procedure for \( \mathcal{C}_c \), so we omit the proof.

An assignment \( \mu \) is **blocked** by a student \( s \) if \( s \succ_s \mu (s) \), and by a school \( c \) if \( \mu (c) \neq \mathcal{C}_c (\mu (c)) \). Similarly, \( \mu \) is **blocked by a student-school pair** \( (s, c) \) if \( \mu (s) \neq c \), \( c \succ_s \mu (s) \) and \( s \in \mathcal{C}_c (\mu (c) \cup \{ s \}) \). An assignment \( \mu \) is **pairwise stable** if it is not blocked by any individual agent or any student-school pair. The following comes easily from Lemma 2:

**Corollary 1.** If \( \mu \) is pairwise stable then \( \mu \) is non-wasteful.

The result below establishes an identity between the set of assignments that are fair with diversity and which are pairwise stable. This allows us to use well-known results and properties of stable matchings in our analysis.

**Lemma 3.** An assignment \( \mu \) is fair with diversity if and only if \( \mu \) is pairwise stable.

The Diversity Deferred Acceptance (DDA) mechanism, that we propose, consists in applying a modified version of the school-proposing deferred acceptance procedure described in Roth (1984) using \( \mathcal{C}_c \) as the schools’ choice function:

1. **Step 1**
   
   (a) Each school \( c \) proposes to the students in \( \mathcal{C}_c (S) \).
   
   (b) Each student \( s \) that received a proposal from one or more schools accepts her most preferred acceptable one according to \( \succ_s \) and rejects the rest of the schools.
2. Step k

(a) Each school $c$ proposes to the students in $C_c(S \setminus r(c))$, where $r(c)$ is the set of students that rejected $c$ at some step before $k$.

(b) Each student $s$ that received a proposal from one or more schools accepts her most preferred acceptable one according to $>_{s}$ and rejects the rest of the schools.

The procedure terminates at any step $t$ in which no rejections are issued, and the resulting assignment $\mu$ is such that for every school $c$, $\mu(c) = C_c(S \setminus r(c))$ as defined above. Students that are not in the choice set of any school are left unmatched.

The following somewhat surprising result is based on a more general theorem in Roth (1984):

**Lemma 4.** (Roth, 1984) Suppose that students’ preferences are strict and that the schools choice function satisfies the substitutability condition. Then the assignment $\mu^C$ which is the outcome of the school-proposing deferred acceptance procedure is pairwise stable and school-optimal in the sense that for each school $c$ and every pairwise stable assignment $\mu$, $\mu^C(c) = C_c(\mu^C(c) \cup \mu(c))$.

**Lemma 5.** Let $\mu$ and $\mu'$ be fair with diversity assignments. If for every $c \in C$, $\mu(c) = C_c(\mu(c) \cup \mu'(c))$ then $\mu' \not\geq 2 \mu$.

*Proof.* Suppose not. Then $\mu' >^2 \mu$ and therefore there is $c \in C$ and $t \in T$ such that $|S^t(\mu(c))| < q^t_c$ and $|S^t(\mu'(c))| > |S^t(\mu(c))|$. By Lemma 3, both $\mu(c)$ and $\mu'(c)$ contain only acceptable students for $c$, $\mu(c) = C_c(\mu(c))$ and $\mu'(c) = C_c(\mu'(c))$.

If $|\mu(c)| < q_c$, then by Lemma 2 and the fact that $\mu(c) \neq \mu'(c)$, $|C_c(\mu(c) \cup \mu'(c))| > |C_c(\mu(c))|$. But then $|C_c(\mu(c) \cup \mu'(c))| > |\mu(c)|$ which implies $\mu(c) \neq C_c(\mu(c) \cup \mu'(c))$, a contradiction. Thus, $|\mu(c)| = q_c$ and the procedure for $C_c(\mu(c) \cup \mu'(c))$ finishes after the final step.

Since $|S^t(\mu'(c))| > |S^t(\mu(c))|$, there is at least one student $s' \in \mu'(c)$ such that $\tau(s') = t$ and $s' \notin \mu(c)$ . Since $|S^t(\mu(c))| < q^t_c$, student $s'$ is accepted in the step associated with $t$ in $C_c(\mu(c) \cup \mu'(c))$. As a consequence, $s' \in C_c(\mu(c) \cup \mu'(c))$ and thus $\mu(c) \neq C_c(\mu(c) \cup \mu'(c))$, a contradiction. 

Putting it all together we get our main result:

**Theorem 1.** The DDA mechanism implements diversity.
Proof. Let $\mu^C$ be the outcome of the DDA mechanism. By Lemmas 1 and 4, $\mu^C$ is pairwise stable and thus, by Lemma 3, $\mu^C$ is fair with diversity. By Lemmas 4 and 5, for any fair with diversity assignment $\mu', \mu' \succ^2 \mu^C$. As a consequence, $\mu^C$ implements diversity. 

4 Comparative Analysis

The purpose of the DDA mechanism is to attain a school assignment that not only is fair but approximates, as much as possible, the assignment to the diversity objectives. In this section we show how it compares to other school choice mechanisms available in the literature in its ability to satisfy those objectives. The first one is “Neighborhood School Assignment” (NSA). It consists in assigning students to schools in their own neighborhood.\footnote{Since in this section we will assume that the number of seats in schools and of students in a neighborhood are equal it will turn out not to be necessary to know the specifics of how students are assigned in their neighborhoods’ schools.} The second is the Top Trading Cycles (TTC), introduced and specified for the domain of school assignment problems in Abdulkadiroğlu and Sönmez (2003). It is the only Pareto Efficient mechanism analyzed here. Neither NSA or TTC embody any specific attempt to respond to diversity concerns and are added to the analysis as representing the maintenance of the status-quo.

The last mechanisms that we compare incorporates explicit attempts to increase the diversity of cohorts in school assignments. The Deferred Acceptance with Minority Reserves (DAMR), proposed in Hafalir et al. (2013)\footnote{We use a simple extension of the DAMR mechanism to accommodate for more than one type of student (Hafalir et al. (2013) consider only two types: minority and majority). This extension can also be found in Echenique and Yenmez (2012) and is a special case of the Deferred Acceptance Procedure with Soft Bounds in Ehlers et al. (2011), where there are no upper quotas. See also Kominers and Sönmez (2012) and Westkamp (2012) for more generalizations.}, expands the access that some students have to every school (being, thus, an affirmative action mechanism). This is done by reserving certain seats in each school for certain types of students, but converting them into regular ones when those are not claimed by those students. Although this mechanism may help minorities obtain seats in more competitive schools, it in a sense outsources to their choices the responsibility of obtaining diverse school cohorts.

In order to evaluate the assignments in terms of the distribution of students across schools, we define two classes of assignments that represent the two extremes: one in which students are completely segregated by their types (assignments that maximize segregation) and one in which the distribution of types in the set of students in each school is identical to the distribution of the population as a whole (assignments that minimize segregation). Although it is not necessarily the case that those are the designer’s diversity objectives, we believe that
the ability of the mechanism to attain such objective is a good measure of how successful it is for general objectives. We now define those formally and give simple examples of both:

**Definition 6.** An assignment $\mu$ maximizes segregation if for every school $c \in C$, $s, s' \in \mu(c) \implies \tau(s) = \tau(s')$.

**Definition 7.** Diversity objectives mirror the population distribution if for every $c \in C$ and $t \in T$, $q^t_c = \left\lceil \frac{q_c\cdot|S^t(S)|}{|S|} \right\rceil$. An assignment $\mu$ minimizes segregation if $\mu$ fully implements diversity when the diversity objectives mirror the population distribution.

**Example 1.** Suppose that $C = \{c_1, c_2\}$, $S = \{s_1, s_2, s_3, s_4\}$, $T = \{t_1, t_2\}$, $S^{t_1}(S) = \{s_1, s_2\}$ and $S^{t_2}(S) = \{s_3, s_4\}$. The assignment $\mu$, where $\mu(c_1) = \{s_1, s_2\}$ and $\mu(c_2) = \{s_3, s_4\}$ maximizes segregation, while the assignment $\mu'$, where $\mu'(c_1) = \{s_1, s_3\}$ and $\mu(c_2) = \{s_2, s_4\}$, minimizes segregation.

We denote the partition of the students by type by $S = S_1 \cup \cdots \cup S_k$, where for every $i \in \{1, \ldots, k\}$ and $s \in S_i$, $\tau(s) = t_i$. We consider a simplified configuration in which the number of students of each type is the same, that is, $|S_i| = |S_j|$ for all $i, j$, every school has the same capacity $q$ and the number of seats in schools equals the number of students ($|S| = q|C|$). In order to avoid issues related to fractional values throughout the analysis we will, for any given value of $k$ (the number of types of students), assume that the number of students is such that $|S| = n_1n_2k^2$ and that the number of schools is such that $|C| = n_2k$, for some $n_1, n_2 \in \mathbb{N}$. As a result, $q = n_1k$ and for every $i$, $|S_i| = n_1n_2k$ and $\frac{q|S_i|}{|S|} = n_1$.

Our first result shows that the DDA mechanism generates assignments that minimize segregation regardless of students’ preference profiles or schools’ priorities.

**Theorem 2.** Every assignment generated by the DDA mechanism minimizes segregation when the diversity objectives mirror the population distribution.

Theorem 2 is the strongest and most general result in this section, and shows that the DDA mechanism is typically effective in the task for which it was designed. For the other mechanisms the results are not as general, and some restrictions on students’ preferences and/or schools’ priorities are necessary in order to obtain positive results. Regarding students’ preferences, we will consider different scenarios. The first scenario below is one in which students of each type have an exclusive set of schools that are preferred by those students to all other schools. The preferences among those schools or the preferences among all other schools are unrestricted. This class of preferences accommodates, among others, a situation that is commonly observed: students (and their parents) have a preference for schools that have, historically, a significant proportion of students of their own type.
Scenario 1. (Favorite schools for each type) There is a partition of schools \( C = C_1 \cup \cdots \cup C_k \) where \( |C_1| = \cdots = |C_k| \) and for every student \( s_i \in S \) and \( j \neq i \) it follows that \( c_i \in C_i \) and \( c_j \in C_j \) implies \( c_i > s_i c_j \).

Now we define a scenario in which the set of schools can be partitioned such that the preference among schools in different partitions is perfectly correlated across students, but not between schools in them. This is common when, for example, schools in wealthier neighborhoods are perceived as being better than those in less wealthy neighborhoods.

Scenario 2. (Tiered schools) There is a partition of schools \( C = C_1 \cup \cdots \cup C_k \) where \( |C_1| = \cdots = |C_k| \) and for every student \( s \in S \), schools \( c_i \in C_i \) and \( c_{i+1} \in C_{i+1} \), student \( s \)'s preferences are such that \( c_i > s c_{i+1} \).

Finally, we consider a scenario in which all students rank the schools using the same ranking.

Scenario 3. (Common preferences) For every \( s, s' \in S \) and \( c, c' \in C \), \( c > s c' \iff c > s' c' \)

The following property connects DAMR outcomes with the set of fair with diversity assignments:

**Lemma 6.** Every assignment \( \mu \) generated by the DAMR mechanism is fair with diversity.

**Proof.** By Theorem 6.8 in Roth and Sotomayor (1992) and Lemma 1, \( \mu \) is stable (and thus pairwise stable). Thus by Lemma 3 \( \mu \) is fair with diversity. \( \square \)

Our first results consider the application of the DAMR mechanism in those different scenarios:

**Proposition 5.** Every assignment generated by the DAMR mechanism maximizes segregation in Scenario 1.

It is easy to see that the result in Proposition 5 applies not only to the DAMR mechanism, but to any mechanism that uses the student-proposing deferred acceptance procedure when the school’s choice function satisfies the substitutability condition and the law of aggregate demand.

**Proposition 6.** Let \( \mu \) be an assignment generated by the DAMR mechanism when the diversity objectives mirror the population distribution, and let \( C^* \subseteq C \) be the set of schools such that if \( c^* \in C^* \), \( \mu (c^*) \) enables diversity in \( c^* \). Then in Scenario 2 \( \frac{|C^*|}{|C|} \geq \frac{k-1}{k} \).
It is important to observe that the result in proposition 6 does not say that as the number of students or schools grows the proportion of schools which enable diversity converges to 1, since \( k \) is the number of student types which have some diversity objective associated with it. In fact, the lower-bound in the proposition may be binding. As a result, in a situation where there are only two types of students (minority and majority), for example, half of the schools may constitute a totally segregated subset of them, as shown in the example below.

**Example 2.** Consider the following school choice with diversity problem:

\[
S = \{s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8\} \\
T = \{t_1, t_2\} \\
S^{t_1}(S) = \{s_1, s_2, s_3, s_4\} \\
S^{t_2}(S) = \{s_5, s_6, s_7, s_8\} \\
>_{s_1} : c_1, c_2, c_3, c_4 \\
>_{s_2} : c_2, c_1, c_3, c_4 \\
>_{s_3} : c_1, c_2, c_3, c_4 \\
>_{s_4} : c_2, c_1, c_3, c_4 \\
>_{s_5} : c_1, c_2, c_3, c_4 \\
>_{s_6} : c_2, c_1, c_3, c_4 \\
>_{s_7} : c_1, c_2, c_4, c_3 \\
>_{s_8} : c_2, c_1, c_4, c_3
\]

Capacities are \( q_c = 2 \), diversity objectives are \( q_{t_i}^{c_j} = 1 \), for all \( i \in \{1, 2\} \) and \( j \in \{1, 2, 3, 4\} \). The assignment generated by the DAMR mechanism is \( \mu \), as follows:

\[
\mu = \begin{pmatrix}
c_1 & c_2 & c_3 & c_4 \\
s_1, s_5 & s_2, s_6 & s_3, s_4 & s_7, s_8
\end{pmatrix}
\]

Notice that both \( \mu(c_1) \) and \( \mu(c_2) \) enable diversity in those schools but the remaining population is segregated: to school \( c_3 \) are assigned only students of type \( t_1 \) and to school \( c_4 \) only students of type \( t_2 \).

**Proposition 7.** Every assignment generated by the DAMR mechanism minimizes segregation in scenario 3.

For the purpose of analyzing outcomes when the TTC mechanism is used, only scenario 1 allows for a result without assumptions on schools’ priorities:
Proposition 8. Every assignment generated by the TTC mechanism maximizes segregation in Scenario 1.

For the analysis of the outcomes of TTC and NSA, some restrictions on the schools’ priorities have to be made. A commonly used criterion is to give higher priority in a school $c$ for students that live nearby school $c$. We combine that priority rule with the assumption of geographic segregation, defined below.

Assumption 1. (Geographic segregation) There are $k$ neighborhoods and every student $s_i \in S_i$ lives in neighborhood $i$. The set of schools $C$ can also be partitioned as $C = C_1 \cup \cdots \cup C_k$, where $C_i$ is the set of schools in neighborhood $i$, $|C_i| = |C_j|$ for every $i, j$ and students have priority in their neighborhoods’ schools. That is, for all $s \in S_i$, $s' \notin S_i$ and $c \in C_i$, $s >_c s'$.

Proposition 9. Every assignment generated by the TTC mechanism maximizes segregation in scenarios 2 and 3 under geographic segregation.

Finally, it is easy to see that under geographic segregation, any mechanism that restricts students’ choices to those schools in their neighborhoods will lead to a situation of complete segregation.

Proposition 10. Every assignment generated by the NSA mechanism maximizes segregation under geographic segregation.

The proof for Proposition 10 is trivial and therefore omitted. Table 1 summarizes the results of the comparative analysis. For each mechanism, it shows under which conditions we have shown that each of the properties holds for an arbitrary instance of a school choice with diversity problem that satisfies the simplified configuration, as described at the beginning of this section and used throughout the analysis. If the cell is empty, then the property is not satisfied for some instances but may be satisfied for others.

5 Incentives

A desirable property for a mechanism is that of strategy-proofness. A mechanism is strategy-proof if truth-telling is a weak dominant strategy for the game induced by the mechanism in its participants (in this case, students) where the strategies are the stated preferences over schools. Unfortunately, this is not the case of the DDA mechanism, as shown by the following example.

---

17This assumption is consistent both with the separation of the city in zones and with walking-distance priority when students are eligible for such priority only in one school each.
<table>
<thead>
<tr>
<th></th>
<th>DDA</th>
<th>DAMR</th>
<th>TTC</th>
<th>NSA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fair with Diversity</strong></td>
<td>Always</td>
<td>Always</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
| **Implements Diversity**       | Always| Scenario 2<sup>b</sup>
|                                |       | Scenario 3 | —     | —     |
| **Minimizes Segregation**      | Always| Scenario 2<sup>b</sup>
|                                |       | Scenario 3 | —     | —     |
| **Maximizes Segregation**      | Never | Scenario 1
|                                |       | Scenario 2<sup>a</sup>
|                                |       | Scenario 3<sup>a</sup> | Always<sup>a</sup> |

a: Under Geographic Segregation
b: For a subset of the schools

**Scenario 1**: Favorite schools for each type.
**Scenario 2**: Tiered schools.
**Scenario 3**: Common preferences.

Table 1: Summary of the comparative analysis results

**Example 3.** Consider the following school choice with diversity problem:

\[
S = \{s_1, s_2\}
\]

\[
T = \{t_1\} \quad C = \{c_1, c_2\}
\]

\[
S^{t_1}(S) = \{s_1, s_2\}
\]

\[
>_{s_1}: c_2 <_{c_1} s_1 <_{s_2} c_1 <_{c_2} s_2 <_{s_1} c_1 <_{c_2} s_2 <_{s_1} c_2 <_{c_1} s_2 <_{c_2} s_1
\]

Capacities are \(q_{c_1} = q_{c_2} = 1\), diversity objectives are \(q^{T}_{c_1} = (q^{T}_{c_1}) = (0)\) and \(q^{T}_{c_2} = (q^{T}_{c_2}) = (0)\). Consider the following assignments:

\[
\mu = \begin{pmatrix}
c_1 & c_2 \\
s_1 & s_2
\end{pmatrix} \quad \mu' = \begin{pmatrix}
c_1 & c_2 \\
s_2 & s_1
\end{pmatrix}
\]

If students truthfully submit their preferences, the outcome of the DDA mechanism is \(\mu\). Now suppose that student \(s_1\) manipulates its preference and submits \(>_{s_1}'\), where \(c_2 >_{s_1}' c_1\) \(\not\supset_{s_1}' c_1\), that is, under \(>_{s_1}'\) school \(c_1\) is unacceptable. In this case the outcome of the DDA mechanism will be \(\mu'\), under which \(s_1\) is assigned to school \(c_2\). Thus, by misrepresenting her preferences, \(s_1\) is assigned to a more preferred school.

Not being strategy-proof, however, is a property not only of the DDA mechanism, but of any mechanism that implements diversity, as we show in the theorem below.

**Theorem 3.** There is no mechanism that implements diversity and is strategy-proof

22
There are evidences that suggest, however, that successful manipulations of stable mechanisms are rare in the presence of many participants or under low information environments. Roth and Peranson (1999) show, in an empirical study, that only about 0.01% of doctors would be able to successfully manipulate the mechanism for the National Resident Matching Program, which is a non-strategy-proof stable mechanism, as the DDA. Theoretical work on big markets also add to these evidences. Immorlica and Mahdian (2005) and Kojima and Pathak (2009) show that under certain regularity assumptions on other agents’ preferences, the number of players that have profitable deviations converges to zero as the number of participants grows in marriage markets (one-to-one matching) and college admissions (many-to-one matching) when colleges are the ones manipulating their priorities or capacities.

In the following sections we give two arguments on why we shouldn’t expect students to manipulate their preferences when the DDA mechanism is used. First, we show that the expected benefits of manipulating the DDA mechanism are reduced to an arbitrarily small value as the number of students grows (section 5.1). Then, we show that when students have low information over each others’ preferences and priorities, truth-telling maximizes expected utility (section 5.2).

5.1 Large markets

The concept that we will use for incentives in large markets is that of Strategy-proofness in the Large (SP-L), introduced by Azevedo and Budish (2013). A mechanism is SP-L if for any student, full-support iid distribution over other students’ reports and $\varepsilon > 0$, there is a large enough market such that the student maximizes her expected utility to within $\varepsilon$ by reporting her preferences truthfully. One of the reasons why SP-L can be considered a good alternative to strategy-proofness in large markets is that when classifying existing non-strategy-proof mechanisms in the literature, those who were not SP-L coincided with those with explicit empirical evidence that agents were strategically manipulating their preferences, as for example the well-known Boston Mechanism (Abdulkadiroğlu et al., 2006). Those which were classified as SP-L, such as the Gale-Shapley deferred acceptance mechanism, for example, however, have been shown to have approximate incentived for truth-telling in large markets (Immorlica and Mahdian, 2005; Kojima and Pathak, 2009).

Our definition of the formal model for the large market extension of the DDA mechanism is based on a similar result in Troyan and Fragiadakis (2013), in which they show that their proposed EDQDA mechanism is SP-L. Differently from their case, however, no assumption on a consistency between diversity objectives and the distribution of types of the population
is necessary here.

For the following results we will consider a sequence of economies indexed by \( n \in \mathbb{N} \), with \( S^n \) corresponding to the set of students, where \( |S^n| = n \). The set of student types \( T \) and of schools \( C \) is fixed for all \( n \), but the number of seats in each school may grow as \( n \) increases. Students’ types are defined, for each \( n \), by a function \( \tau^n : S^n \to T \). There is a finite set of priority classes \( Z = \{1, \ldots, |Z|\} \) and, for each school \( c \in C \), a partition of the set of student into those classes: \( S^n = S^{c,n}_1 \cup \cdots \cup S^{c,n}_{|Z|} \). For each student \( s \), let \( z_s \in Z^{|C|} \) denote the vector of priority classes with which student \( s \) is associated with at each school in \( C \). School \( c \)’s priorities between students \( s, s' \in S^n \) follow the order of that partition in the sense that if \( s \in S^{c,n}_i, s' \in S^{c,n}_j \) and \( i < j \) then \( s >_c s' \iff i < j \), for all \( n \). For a given type-priority-class pair \((t, z)\), the number of students of each type-priority-class pair, denoted \( n_{(t,z)} \) grows according to some fixed sequence, so that \( n_{(t,z)} \to \infty \) as \( n \to \infty \) for any \((t, z) \in T \times Z^{|C|}\).

Since the set of schools is fixed, there is a finite set of preference types \( A \). Each type \( a \in A \) has associated with it a von Neumann-Morgenstern expected utility function over lotteries over schools \( u_a : \Delta C \to [0, 1] \). The set of preference types \( A \) is such that for each preference ranking \( >_i \) over the elements of \( C \cup \{\emptyset\} \) there is a type \( a_i \in A \) such that \( u_{a_i} (\cdot) \) represents the ordinal preferences in \( >_i \) over degenerate lotteries over \( C \cup \{\emptyset\} \). Define group types as the set \( G = T \times A \times Z^{|C|} \). Denote the group type of a student \( s \) by \( g_s \in G \). Given the setup for the economies, we can proceed to the key definitions that are used in this section.

**Definition 8.** Fix a set of schools \( C \), a sequence of capacity vectors \((q^n)_n\) and of diversity objectives \((q^n)_n\). A school choice with diversity mechanism \( \{(\varphi^n)_n, G\} \) is a sequence of allocation functions \( \varphi^n : G^n \to \Delta (C^n) \) such that for every \( n \in \mathbb{N} \) and \( g \in G^n \), every element in the support of \( \varphi^n (g) \) is feasible with respect to \( q^n \).

Denote by \( \varphi^n_s (g_s, g_{-s}) \) the marginal distribution of \( \varphi^n (g_s, g_{-s}) \) in student \( s \)’s dimension. We can now define, from the perspective of a student \( s \), the individual allocation function that is induced by the mechanism and a distribution with full support over preference types \( m \in \Delta A \), given that her type is \( \tau^n (s) \) and her priority class is \( z_s \):

\[
\phi^n_s (a_s, m) = \sum_{g_{-s} \in C^{n-1}} \varphi^n ((\tau^n (s), a_s, z_s), g_{-s}) \cdot Pr (g_{-s} | a_{-s} \sim iid (m))
\]

---

18For simplicity of notation in this section we represent by \( \emptyset \) the possibility of remaining unmatched to any school.

19Formally, let \( q_n = (q^n_1, \ldots, q^n_c) \). An element \((\phi_1, \ldots, \phi_n) \in C^n \) in the support of \( \varphi^n (g) \) is feasible with respect to \( q_n \) if for every \( c \in C \), \( \sum_{j=1}^{n} 1_{\phi_j = c} \leq q^n_c \).
Notice that the values of $\tau^n(s)$ and $z_s$ are not arguments of the function, as opposed to $a_s$. This is because students are assumed to be able to manipulate their preference reports but not their types or priority classes. Moreover, by the definition of each economy $n$, $z_s$ and $t_s$ are fixed and therefore $g_s$ depends only on the realization of $a_s$. The definition of strategy-proofness in the large is therefore made in terms of manipulation of the value of $a_s$ for a given $(\tau^n(s), z_s)$:

**Definition 9.** (Azevedo and Budish, 2013) A school choice with diversity mechanism $\{(\varphi^n)_n, G\}$ is **strategy-proof in the large (SP-L)** if, for any $\varepsilon > 0$ and any $m \in \Delta A$, there exists $n_0$ such that for all $n \geq n_0$, all $(t, z) \in T \times Z$ and all $a, a' \in A$:

$$u_a[\phi^n_{(t,z)}(a, m)] \geq u_a[\phi^n_{(t,z)}(a', m)] - \varepsilon$$

Consider now the following version of the DDA mechanism, defined for each economy $n$. In the first stage, each student $s$ submits her rankings $>_s$ over $C \cup \{\emptyset\}$. In the second stage, the lottery vector $\ell \in [0, 1]^n$ is drawn uniformly at random and for each school $c \in C$, the priorities over $S^n$ for each school are constructed using the following procedure. Priorities between students in different priority classes follow the procedure given in their definitions above. For students in the same priority classes, higher lottery numbers imply higher priorities. Therefore, if $s, s' \in S^{c,n}_i$ then $s >_c s' \iff \ell_s > \ell_{s'}$. Given students’ preferences $>_S$, schools’ priorities $>_C$ and economy $n$ as described, the DDA mechanism is used to produce a school assignment, as described in section 3.

Let $\mu^n((g_i, \ell_i)_{i \in S^n})$ be the matching generated by the DDA mechanism for economy $n$, vector of group types $g \in G^n$ and lottery vector $\ell \in [0, 1]^n$. Define the function $\mathcal{M}^n(g)$ as follows:

$$\mathcal{M}^n(g) = \int_{\ell \in [0,1]^n} \mu^n((g_i, \ell_i)_{i \in S^n}) \, d\ell$$

We can now proceed to the main result:

**Theorem 4.** The school choice with diversity mechanism $\{(\mathcal{M}^n)_n, G\}$ is strategy-proof in the large.

### 5.2 Low information scenario

In this part we will consider a scenario in which students have low information regarding other students’ preferences and over schools’ priorities, denoted **symmetric information environments**. Roth and Rothblum (1999) introduced the concept of symmetric information
in matching mechanisms as a way to evaluate incentives in the presence of uncertainty on the part of students on other students’ and schools’ preferences, priorities and/or strategies. They show that in symmetric information environments, a strategy that reverts the ranking of two or more schools with respect to the true preference is stochastically dominated by truthfully submitting their preference ranking. While Roth and Rothblum (1999) considers the school-proposing deferred acceptance mechanism, Ehlers (2008) generalizes their result by giving sufficient conditions on any mechanism for those results to hold true. Most of the notation and the results below are based on the latter.\textsuperscript{20}

For the following arguments, some new definitions are necessary. Let $\mathcal{P}_s$ denote the set of all strict preference relations of student $s$ over $C$, $\mathcal{P}_c^1$ denote the set of all strict priority orderings of school $c$ over $S$, $\mathcal{P}_c^2 \equiv \mathbb{N}$ denote all possible school capacities, $\mathcal{P}_c^3$ denote all possible diversity objectives for school $c$ and $\mathcal{P}_c^4 \equiv \mathcal{P}_c^1 \times \mathcal{P}_c^2 \times \mathcal{P}_c^3$. Given a student $s$, let $\mathcal{P}_{-s} \equiv \prod_{v \not= s} \mathcal{P}_v$. A random profile is a probability distribution $\tilde{\succ}_{-s}$ over $\mathcal{P}_{-s}$. In a scenario in which student $s$ has uncertainty over $\mathcal{P}_{-s}$, $\tilde{\succ}_{-s}$ will be interpreted as student $s$’s beliefs over the stated preferences and priorities submitted by other students and schools.

Given a student $s \in S$, $\succ_s \in \mathcal{P}_s$ and $c, c' \in C$, let $\succ^c_{s} \succ^c_{s}$ denote the ranking that exchanges the positions of $c$ and $c'$ in $\succ_s$ and leave other positions in $\succ_s$ unchanged. Given a profile $\succ_{-s} \in \mathcal{P}_{-s}$, denote $\succ^c_{s} \succ^c_{s}$ the profile in which each student in $S \setminus \{s\}$ exchanges the positions of $c$ and $c'$ in her preference ranking, schools $c$ and $c'$ exchange their priority rankings, capacities and diversity objectives, and the other schools’ priorities, capacities and diversity objectives remain unchanged. Finally, if $\succ \succ_{s} \succ_{s}$, denote $\succ^c_{s} \succ^c_{s}$.

**Definition 10.** Given $s \in S$ and $c, c' \in C$, student $s$’s beliefs $\tilde{\succ}_{-s}$ are $\{c, c'\}$-symmetric if for every profile $\succ_{-s}$, both $\succ_{-s}$ and $\succ^c_{s} \succ^c_{s}$ are equally probable under $\tilde{\succ}_{-s}$, that is, $\Pr\{\tilde{\succ}_{-s} \succ_{-s}\} = \Pr\{\tilde{\succ}_{-s} \succ^c_{s} \succ^c_{s}\}$. A student $s$’s beliefs are $C$-symmetric if they are $\{c, c'\}$-symmetric for any $c, c' \in C$.

Notice that symmetric beliefs don’t require students’ preferences being independent, allowing for correlated preferences across them, for example. Moreover, although the definition of symmetric beliefs includes uncertainty about how the diversity objectives differ across schools, if those are common across schools then no uncertainty about those is necessary. A mechanism is a function $\varphi : \mathcal{P} \rightarrow \mathcal{M}$. We will denote $\varphi[\succ_s, \succ_{-s}]\{s\}$ the assignment of

\textsuperscript{20}For other examples in the school choice literature using this approach, see Erdil and Ergin (2008) and Kesten (2010).

\textsuperscript{21}We are assuming that students cannot make schools unacceptable. This is consistent with the objective of assigning every student to a school.
student $s$ in mechanism $\varphi$ when she submits the preference $>_s \in \mathcal{P}_s$ and other students and schools submit $>_s \in \mathcal{P}_s$.

A random assignment $\tilde{\mu}$ is a probability distribution over the set of assignments $\mathcal{M}$, and therefore $\tilde{\mu}(s)$ is the distribution over student $s$’s set of assignments. Given a mechanism $\varphi$ and $>_s \in \mathcal{P}_s$, each random profile $\tilde{>_s}$ induces a random assignment $\varphi[>_s, \tilde{>_s}]$, where:

$$\Pr \{ \varphi[>_s, \tilde{>_s}] = \mu \} = \sum_{>_s \in \mathcal{P}_s: \varphi[>_s, >_s] = \mu} \Pr \{ \tilde{>_s} = >_s \}$$

Given a student $s \in S$, $>_s, >'_s, >''_s \in \mathcal{P}_s$ and a random profile $\tilde{>_s}$. Let $>_s$ be student $s$’s true preference over $C$. We say that strategy $>_s$ stochastically dominates strategy $>''_s$ under beliefs $\tilde{>_s}$, denoted by $\varphi[>_s, \tilde{>_s}] (s) >_s \varphi[>''_s, \tilde{>_s}] (s)$, if for all $v \in C \cup \{s\}$:

$$\Pr \{ \varphi[>_s, \tilde{>_s}] (s) >_s v \} \geq \Pr \{ \varphi[>''_s, \tilde{>_s}] (s) >_s v \}$$

It seems clear that when choosing between $>_s$ and $>'_s$ a student that has the preferences $>_s$ would choose $>'_s$. In fact, for any von Neuman-Morgenstern utility function $u_s(\cdot)$ that represents $>_s$, the following is true:

$$E_{\tilde{>_s}} [u_s(\varphi[>_s, \tilde{>_s}] (s))] \geq E_{\tilde{>_s}} [u_s(\varphi[>''_s, \tilde{>_s}] (s))]$$

Majumdar and Sen (2004) introduced the class of Ordinally Bayesian Incentive Compatible mechanisms, as an alternative for strategy-proofness that considers a specific belief over other agents’ types.

**Definition 11.** (Majumdar and Sen, 2004) A mechanism is **Ordinally Bayesian Incentive Compatible (OBIC)** with respect to the belief $\tilde{\succ}$ if for all student $s \in S$, all $>_s, >'_s \in \mathcal{P}_s$ and every von Neuman-Morgenstern utility function $u_s(\cdot)$ that represents $>_s$, we have:

$$E_{\tilde{>_s}} [u_s(\varphi[>_s, \tilde{>_s}] (s))] \geq E_{\tilde{>_s}} [u_s(\varphi[>'_s, \tilde{>_s}] (s))]$$

In other words, a mechanism is OBIC with respect to the belief $\tilde{\succ}$ if a student never expects to be better off, with respect to any utility function that represents her ordinal preferences, by misrepresenting her preferences, on the assumption that other students follow the truth-telling strategy. We can now proceed to our first result regarding the incentives induced by the DDA mechanism under symmetric beliefs:
Proposition 11. Any strategy that changes the true ranking of the schools while using the DDA mechanism is stochastically dominated by a strategy that preserves the true ranking of the schools for a student with $C$-symmetric beliefs.

As a consequence of proposition 11, if every student have $C$-symmetric beliefs and other students are truthful, then truth-telling maximizes the students’ expected utility for any set of von Neuman-Morgensten utility functions that represent the students’ preferences over schools.

Theorem 5. The DDA mechanism is OBIC with respect to $C$-symmetric beliefs.

The result in Theorem 5 comes immediately from applying the result in proposition 11 to each student.

6 Conclusion

This paper proposed a school choice mechanism that incorporates diversity objectives both as a distributional concern and as an element of fairness. Whereas most of the literature on the subject focuses on giving the students the possibility of, through their preferences, obtain more diverse school cohorts, very few mechanisms make an explicit attempt to enforce them to some extent. The proposed DDA mechanism generates an assignment that is as close as possible to the distribution implied by the diversity objectives while requiring that such assignment satisfies a well-defined fairness criterion.

The analysis presented in this paper focused on the aspects of fairness and the distribution of student types across schools. Both requirements may have negative welfare implications. An evaluation of these, in general and in relevant scenarios, could complement the ones presented in this paper so that gains and losses can be compared in an objective way by policy makers.
References


Aygün, Orhan and Inacio Bo (2013), “College admissions with multidimensional reserves: the brazilian affirmative action case.” Mimeo. 1

Azevedo, Eduardo and Eric Budish (2013), “Strategyproofness in the large.” Working Paper, University of Chicago. 5.1, 9, 6, 12


Coldron, John, Emily Tanner, Steven Finch, Lucy Shipton, Claire Wolstenholme, Ben Willis, Sean Demack, and Bernadette Stiell (2008), “Secondary school admissions.” Technical report, Sheffield Hallam University and National Centre for Social Research. 1


Matos, Mauricio dos Santos, Selma Garrido Pimenta, Maria Isabel de Almeida, and Maria Amélia de Campos Oliveira (2012), “The impact of the social inclusion program of the university of são paulo on the access of public school students to free public higher education.” *Revista Brasileira de Estudos Pedagógicos*, 93, 720–742. 1


Roth, Alvin E and Marilda A Oliveira Sotomayor (1992), *Two-sided matching: A study in game-theoretic modeling and analysis*. 18, Cambridge University Press. 4, 6


Welch, F. and Audrey Light (1987), New evidence on school desegregation. Clearinghouse publication, United States Commission on Civil Rights. 2

Appendix

Proofs

Lemma 1

Suppose not, and consider the notation used in the definition of the choice function above. Then there exists $S' \subset S$ and $s_i, s_j \notin S'$ where $s_i \notin \mathbb{C}_c(S' \cup \{s_i\})$ and $s_i \in \mathbb{C}_c(S' \cup \{s_i, s_j\})$. For the rest of this proof we will assume, without loss of generality, that $s_i, s_j$ and all students in $S'$ are acceptable for school $c$. Let $t_i = \tau(s_i)$, $t_j = \tau(s_j)$ and define $\overline{S}_c(s, I) \equiv \{s' \in I : s' \succ_c s\}$, that is, $\overline{S}_c(s, I)$ is set of students in $I$ that have a higher priority in school $c$ than student $s$. For simplicity of notation, denote $\overline{S}_1 \equiv \overline{S}_c(s, S' \cup \{s_i\})$ and $\overline{S}_2 \equiv \overline{S}_c(s, S' \cup \{s_i, s_j\})$. Since $s_i \notin \mathbb{C}_c(S' \cup \{s_i\})$, $|\overline{S}_1| \geq q'_c$. Moreover, let $q^*$ be the number of students accepted in the final step of the procedure for $\mathbb{C}_c(S' \cup \{s_i\})$.

Since for any $t$ it is true that $S' \supseteq S' \cup \{s_i\}$, it easily follows that $|\overline{S}_1| \leq \overline{S}_2$. That is, the set of students of type $t_i$ that have higher priority than $s_i$ in school $c$ in $S' \cup \{s_i\}$ is a superset of those in $S' \cup \{s_i, s_j\}$ and thus it follows that $|\overline{S}_2| \geq |\overline{S}_1| \geq q'_c$. As a consequence, student $s_i$ is not accepted from $S' \cup \{s_i, s_j\}$ between steps 1 and $t$.

It must be then that $s_i$ is accepted in the final step of the procedure. We will consider the three circumstances under which $s_j$ could have been accepted. Suppose first, that $s_j$ is accepted in step $j$ (the one associated with type $t_j$). Then either $\Psi_{t_j}(S' \cup \{s_i, s_j\}) = \Psi_j(S' \cup \{s_i\}) \cup \{s_j\}$ or $\Psi_{t_j}(S' \cup \{s_i, s_j\}) = (\Psi_{t_j}(S' \cup \{s_i\}) \setminus \{s_k\}) \cup \{s_j\}$ for some $s_k \in S^{t_j}(S' \cup \{s_i\})$. In the former case:

$\overline{S}_c(s_i, S' \cup \{s_i, s_j\}) \setminus \Psi_{t_k}(S' \cup \{s_i, s_j\}) = \overline{S}_c(s_i, S' \cup \{s_i\}) \setminus \Psi_{t_k}(S' \cup \{s_i\})$

That is, the set of students that were not yet accepted by the end of step $k$ that have higher priority in school $c$ than $s_i$ is the same as in $S' \cup \{s_i, s_j\}$ and so $s_i$ is not accepted in the final step. In the latter case:

$\overline{S}_c(s_i, S' \cup \{s_i, s_j\}) \setminus \Psi_{t_k}(S' \cup \{s_i, s_j\}) \supseteq \overline{S}_c(s_i, S' \cup \{s_i\}) \setminus \Psi_{t_k}(S' \cup \{s_i\})$

The procedure will accept $q^*$ students from $S^* \overline{S}_c(s_i, S' \cup \{s_i, s_j\}) \setminus \Psi_{t_k}(S' \cup \{s_i, s_j\})$ in the final step and thus $s_i$ will again not be accepted. Finally, if $s_j$ isn’t accepted in step $j$, then:

$\overline{S}_c(s_i, S' \cup \{s_i, s_j\}) \setminus \Psi_{t_k}(S' \cup \{s_i, s_j\}) \supseteq \overline{S}_c(s_i, S' \cup \{s_i\}) \setminus \Psi_{t_k}(S' \cup \{s_i\})$

33
The procedure will accept $q^*$ students from $S''$ in the final step and thus $s_i$ is not accepted. Contradiction with $s_i \in \mathbb{C}_c(S' \cup \{s_i, s_j\})$.

**Lemma 3**

If $\mu$ is pairwise stable then $\mu$ is fair with diversity.

First, notice that $\mu$ is individually rational by the definition of pairwise stability and the step 0 that eliminates all unacceptable students in $\mathbb{C}_c$. Suppose that $\mu$ is not fair with diversity. Since $\mu$ is non-wasteful it must then be that a students justifiably demands a seat in a school $c$, which implies that $\mu(s) \neq c$ and $c \succ s \mu(s)$. By pairwise stability, $s \notin \mathbb{C}_c(\mu(c) \cup \{s\})$. By non-wastefulness, $|\mu(c)| = q_c$ and thus while obtaining $\mathbb{C}_c(\mu(c) \cup \{s\})$, the procedure only finished after the final step. We will now show that every condition in which a student justifiably demands a seat will lead to a contradiction:

$$t = \tau(s) \text{ and } |S'(\mu(c))| < q_c^1.$$ But then $|S'(\mu(c)) \cup \{s\}| \leq q_c^1$ and $s$ is accepted at the step of the procedure for $\mathbb{C}_c$ associated with $t$. Contradiction with $s \notin \mathbb{C}_c(\mu(c) \cup \{s\})$.

There is a student $s' \in \mu(c)$ such that $\tau(s') = \tau(s)$ and $s \succ_c s'$. Let $t = \tau(s') = \tau(s)$. Since $s \notin \mathbb{C}_c(\mu(c) \cup \{s\})$ and $s \succ_c s'$, $s'$ isn’t accepted on the step associated with $t$ of the procedure, otherwise $s$ would also be accepted. The same holds for the final step. But then $s' \notin \mathbb{C}_c(\mu(c) \cup \{s\})$, which implies that $|\mathbb{C}_c(\mu(c) \cup \{s\})| < q_c$, contradicting lemma 2.

There is $t \in T$ and $s' \in S'(\mu(c))$ such that $|S'(\mu(c))| > q_c^1$ and $s \succ_c s'$. Using the same notation used in the description of the procedure for $\mathbb{C}_c$, $\Psi_t(\mu(c))$ is the set of students in $S'(\mu(c))$ accepted during the step associated with $t$ in $\mathbb{C}_c(\mu(c))$ and denote, additionally, $\Psi_s(\mu(c))$ be the set of students accepted during the final step of $\mathbb{C}_c(\mu(c))$. Since $|S'(\mu(c))| > q_c^1$, $|S'(\Psi_s(\mu(c)))| > 0$. By the description of the step associated with $t$ in the procedure for $\mathbb{C}_c$, $\Psi_t(\mu(c))$ contains the top $q_c^1$ students in $S'(\mu(c))$ with respect to $\succ_c$ and thus for any $s_i \in \Psi_t(\mu(c))$ and $s_j \in S'(\Psi_s(\mu(c)))$, $s_i \succ_c s_j$. Since $s \succ_c s'$, if either $s' \in \Psi_t(\mu(c))$ or $s' \in S'(\Psi_s(\mu(c)))$, $s \succ_c s''$ for some $s'' \in S'(\Psi_s(\mu(c)))$ and thus since $s \notin \mathbb{C}_c(\mu(c) \cup \{s\})$, $s'' \notin \mathbb{C}_c(\mu(c) \cup \{s\})$, implying $|\mathbb{C}_c(\mu(c) \cup \{s\})| < q_c$, once again a contradiction of lemma 2.

If $\mu$ is fair with diversity then $\mu$ is pairwise stable.

Suppose that $\mu$ is fair with diversity but that it is not pairwise stable. Then $\mu$ is blocked by a student, a school or a student-school pair. We’ll examine each possibility:

$\mu$ is blocked by a student, that is, $s \succ s \mu(s)$. But then $\mu$ is not individually rational. Contradiction with $\mu$ being fair with diversity.

$\mu$ is blocked by a school $s$, that is, $\mu(c) \neq \mathbb{C}_c(\mu(c))$. Since $\mu(c) \subseteq \mathbb{C}_c(\mu(c))$ and
\[\mu (c) \neq C_c (\mu (c)), \mu (c) \subset C_c (\mu (c))\] and thus \(|\mu (c)| < q_c\). Since \(\mu\) is fair with diversity, all students in \(\mu (c)\) are acceptable to \(c\) and by Lemma 2 \(\mu (c) = C_c (\mu (c))\), which is a contradiction.

\(\mu\) is blocked by a student-school pair \((s, c)\), that is, \(\mu (s) \neq c, c >_s \mu (s)\) and \(s \in C_c (\mu (c) \cup \{s\})\). Since \(\mu\) is fair with diversity, \(|\mu (c)\) = \(q_c\) and \(s\) doesn’t justifiably demands a seat in \(c\). We will show that \(s\) couldn’t be accepted in any step of the procedure for \(C_c (\mu (c) \cup \{s\})\). Since \(|\mu (c) \cup \{s\}| > q_c\), \(s\) cannot be accepted in step 0 of the procedure for \(C_c\). Let \(t = \tau (s)\). In order for \(s\) to be accepted in the step associated with \(t\), then either \(|S^t (\mu (c))| < q_c\) or there is a student \(s' \in S^t (\mu (c))\) such that \(s >_c s'\). Both possibilities contradict \(\mu\) being fair with diversity, more specifically the first two items in the definition of when a student justifiably demands a seat in a school. It must then be that \(s\) is accepted in the final step and thus there is a student \(s' \in \mu (c)\) that is accepted in the final step that will be replaced by \(s\). This can only be true if \(s >_c s'\). Notice that all students accepted in the final step are in “excess” of their diversity objectives, that is, if \(t' = \tau (s')\), \(|A^{t'} (\mu (c))| > q_c\).

But then we have a contradiction with the third item in the definition cited above.

**Theorem 2**

We will split the analysis for each type and show that during the deferred acceptance process, each school will be accepted by \(n_1\) students of each type, leading thus to an assignment that fully implements diversity. Suppose, for contradiction, that schools’ diversity objectives mirror the population distribution (so that for every \(t_i \in T\) \(q_{c_i} = n_1\)) but the assignment \(\mu\) generated by the DDA mechanism doesn’t minimize segregation. Then there is a school \(c \in C\) and a type \(t_i \in T\) such that \(|S^{t_i} (\mu (c))| < n_1\). It must then be that at some step \(\ell\) in the deferred acceptance procedure of the DDA mechanism the set of students of type \(t_i\) that rejected school \(c\) has more than \((n_1 - 1) n_2 k\) elements. Without loss of generality, let \(\ell\) be the earliest step at which this happens to some school for any type (\(c\) may or may not be the only school for which that happens during step \(\ell\)). Since students consider all schools acceptable, all those students rejected \(c\) because another school proposed simultaneously. By Proposition 2 in Roth (1984), offers made by schools during the deferred acceptance procedure remain open. That is, if a student receives an offer during a step of the procedure, it may change its assignment over time, but will not become unmatched at any subsequent step. Therefore, those students who rejected \(c\) are assigned to other schools. But then at least one school \(c' \in C\), with \(c' \neq c\) proposed to more than \(n_1\) students of type \(t_i\) during step \(\ell\). This can only happen if some student of type \(t_i\) is accepted during the final step of the procedure for
the choice function $C_{c'}$. This implies that at the step associated with some type $t_j \neq t_i$ the number of students of type $t_j$ that rejected $c'$ at some step earlier than $\ell$ is greater than $(n_1 - 1)n_2k$. Contradiction with $\ell$ being the earliest step at which this happened.

**Proposition 5**

In order to show this, it is sufficient to show that, for any given $i$, every student $s_i \in S_i$ is accepted by some school in $C_i$. Now suppose that there is a student $s_i \in S_i$ that is not assigned to any school in $C_i$. Since all schools in $C_i$ are preferred by $s_i$ to any other schools, this implies that $s_i$ was rejected by all schools in $C_i$. Since every student is acceptable by all schools and all students in $S_i$ have the same type, every school $c$ will simply accept the top $q_c$ students according to $\mu(c)$ or all students in case the number of those pointing to $c$ is lower than $q_c$. Thus, the only way in which a student is rejected by a school is if that school has already accepted $q_c$ students. Notice that since $C_c$ satisfies the law of aggregate demand (see the proof of Theorem 3), by the end of each step of the procedure the number of accepted students in each school never decreases. Thus, by the end of step 1, since $s_i$ was rejected by her first choice, at least $q_c$ students in $S_i$ were accepted by schools in $C_1$. By the end of step 2, at least $2q_c$ students are accepted, since the school mentioned in step 1 still accepts $q_c$ students by the end of step 2, and the second-best school for $s_i$ also accepts $q_c$ students. By repeating the argument, by the end of step $|C_i|$, at least $|C_i|q_c$ students were accepted in the first $|C_i|$ steps. But notice that during the first $|C_i|$ steps, students in $S_j$ could have only pointed to schools in $C_j$, for any $j \in \{1, \ldots, k\}$. Thus there is at least $|C_i|q_c + 1$ students in $S_i$, which is a contradiction with $\sum_{c \in C_i} q_c = |S_i|$. Therefore, students in $S_i$ will all be assigned to schools in $C_i$ and thus $\mu(s) \in C_i \implies s \in S_i$.

**Proposition 6**

Let $\mu$ be the assignment generated by the DAMR mechanism. By Lemma 6, $\mu$ is fair with diversity. Therefore $\mu$ is non-wasteful and thus every school is assigned $q_c$ students and every student is assigned to a school. We will now show that when diversity objectives mirror the population distribution, $\mu(c)$ enables diversity in every school $c \in C_1 \cup \cdots \cup C_{k-1}$. We will prove by induction in the sets $C_1, \ldots, C_{k-1}$ when $k > 1$ since the case $k = 1$ is trivial.

**Step 1**: we want to show that for every school $c \in C_1$, $\mu(c)$ enables diversity at $c$.

In order to do so, we must first show that there are at least $n_1n_2$ students of each type $t \in T$ in $\bigcup_{c \in C_1} S^t(\mu(c))$. Suppose not. Then there is at least one school $c' \in C_1$ such that $|S^t(\mu(c'))| < n_1$. Since $|S^t(S)| = n_1n_2k$ and $k > 1$, then there is a student $s'$ such that
τ (s') = t and µ (s') ≠ C1. But then c' > µ (s') and s' justifiably demands a seat in c', which implies that µ isn’t fair with diversity and thus we have a contradiction. Moreover, since there are at least n1n2 students of each type t in \( \bigcup_{c \in C_1} S^t (\mu (c)) \), there are at least n1n2k students in \( \bigcup_{c \in C_1} S^t (\mu (c)) \). Since \( q_c = n_1k \) and for any i, \( |C_i| = n_2 \) it follows that there are exactly n1n2 students of each type t in \( \bigcup_{c \in C_1} S^t (\mu (c)) \).

Suppose now that there is a school c' ∈ C1 such that µ (c) doesn’t enable diversity at c'. Then there is a type t ∈ T such that \( |S^t (\mu (c'))| < n_1 \). By the result above we know that there is a student s of type t such that \( \mu (s) \notin C_1 \). By assumption on preferences, c' > µ (s), implying that s justifiably demands a seat in c', a contradiction.

**Step k*: by induction assumption, for every i ∈ \{1, \ldots, k^* - 1\} and c ∈ Ci, µ (c) enables diversity in c. The proof follows the same argument as for step 1, with the difference that in the instances in which a student s justifiably demands a seat in some school c' ∈ Ck*, that student is assigned in µ to some school in Ck', where k' > k*, implying that c' > µ (s).

Notice, however, that when k* = k this argument cannot be made anymore, that is, if there is a school c' ∈ Ck such that \( |S^t (\mu (c'))| < n_1 \), there may not exist a student s of type t such that c' > µ (s).

**Proposition 7**

We will show by induction on the steps of the DAMR mechanism that for every school c_i ∈ C, µ (c_i) enables diversity in c_i when the diversity objectives mirror the population distribution (that is, for every t ∈ T and \( q_{c_i}^t = n_1 \)). Given that all students share the same preference ranking, the student-proposing deferred acceptance procedure consists in, at each step k*, all students that were not yet accepted by a school propose to school c_k*. Denote \( S^0 = S \). Therefore, for every t ∈ T, \( |S^t (S^0)| = n_1n_2k \).

**Step 1:** During step 1, every student in \( S^0 \) proposes to school c_1. School c_1 will accept the students in C_c (S^0). Since k ≥ 0, for every t ∈ T, \( |S^t (S^0)| \geq n_1 \) and thus by Proposition 4, C_c (S^0) (or µ (c_1)) enables diversity in c_1. Denote \( S^1 \equiv \{ s \in S^{i-1} : s \notin C_c (S^{i-1}) \} \). That is, \( S^1 \) is the set of students rejected by school c_1 during step 1.

**Step k*: By induction assumption, in every step i for \( i \in \{1, \ldots, k^* - 1\} \), µ (c_i) enables diversity in c_i. As a result, there are \( n_1 \) students of each type assigned to each school c_1, \ldots, c_{k^* - 1}. Therefore, for every t ∈ T, \( |S^t (S^{k^* - 1})| = n_1(n_2k - k^* + 1) \). Following the deferred acceptance procedure, all students that were rejected in step k* - 1 will propose to school c_k*. Since \( |C| = n_2k \), for any k* ≤ |C| and t we have \( |S^t (S^{k^* - 1})| \geq n_1 \). Therefore, by Proposition 4, C_c (S^{k^* - 1}) (or µ (c_k*)) enables diversity in c_k*. 

37
Proposition 8

Notice that in each step of the TTC mechanism, students are only assigned to schools that they are pointing to, and that since for all \( i, j \ |C_i| = |C_j| \) and \( C = C_1 \cup \cdots \cup C_k \), \( \sum_{c \in C_i} q_c = |S_i| \) (or simply \( |C_i| q = |S_i| \)). Fix an \( i \in \{1, \ldots, k\} \). Since for every student \( s_i \in S_i \), \( c_i \in C_i \) and \( c_j \in C_j \) implies \( c_i > s_i, c_j \), students in \( S_i \) will point to schools in \( C_i \) until there are no schools in \( C_i \) with seats available. Let \( \mu \) be the assignment generated by the TTC mechanism. We will show, by induction on the steps of the TTC procedure, that \( \mu (s) \in C_i \implies s \in S_i \).

**Step 1:** For every \( i \), students in \( S_i \) point only to schools in \( C_i \). Thus, students in \( S_i \) can only be assigned to schools in \( C_i \) in step 1.

**Step \( k^* \):** By induction, for every \( i \) and student \( s \in S \) assigned until step \( k^* - 1 \), \( \mu (s) \in C_i \implies s \in S_i \). Now fix an \( i \) and consider the schools in \( C_i \) and the students in \( S_i \). If there are no seats left in any school in \( C_i \), by induction all those are occupied by students in \( S_i \). But since \( \sum_{c \in C_i} q_c = |S_i| \), every student in \( S_i \) is assigned to schools and are not available for cycles in the procedure anymore. Thus, at step \( k^* \), students in \( S_i \) are either already assigned to schools in \( C_i \) or are pointing to schools in \( C_i \). Therefore, at step \( k^* \) they can only be assigned to schools in \( C_i \).

Proposition 9

We will show, by induction, two facts about \( \mu \) at the end of each step of the procedure: one is that if there is some \( c_i \in C_i \) where \( |\mu (c_i)| < q_c \), then for all \( j \) such that \( i < j \leq k \), \( c_j \in C_j \implies \mu (c_j) = \emptyset \). The other is that if \( c \in C_{k'} \) and \( s \in \mu (c) \) then \( s \in S_{k'} \).

**Scenario 2:**

**Step 1:** During step 1, every student points to some school in \( C_1 \), and every school in \( C_1 \) points to some student in \( S_1 \) (since schools in \( S_1 \) give priority to students in \( S_1 \)). Therefore, any cycle involves only students in \( S_1 \) and schools in \( C_1 \) and by the end of this step, all students assigned to a school are in \( S_1 \), and every school receiving a student is in \( C_1 \). Thus, for all \( j > 1, c_j \in C_j \implies \mu (c_j) = \emptyset \).

**Step \( k \):** By induction assumption, by the end of step \( k^* - 1 \) there is a \( j \in \{1, \ldots, k\} \) such that, for every \( j < j' \leq k \) and \( c \in C_{j'} \), \( \mu (c_{j'}) = \emptyset \) and for all \( j'' \) such that \( 1 \leq j'' < j \) and \( c_{j''} \in C_{j''} \) we have \( |\mu (c_{j''})| = q_c, \mu (c_{j''}) \subseteq S_{j''} \) and since \( \sum_{c \in C_{j''}} q_c = |S_{j''}|, \cup_{c_{j''} \in C_{j''}} \mu (c_{j''}) = S_{j''} \). Thus, during step \( k^* \) there are no schools with seats available in the set \( C_1 \cup \cdots \cup C_{j-1} \) but there is at least one school with at least one seat available in \( C_j \). Since every student prefers any school in \( C_j \) to any other school with seats available, every student that wasn’t yet assigned to a school points to a school in \( C_j \). It is also the case that by the end of step
the number of students in $S_j$ that are not yet assigned to a school equals the number of seats available in schools in $C_j$. Since schools in $C_j$ give higher priority to students in $S_j$, then in step $k^*$ every school in $C_j$ points to students in $S_j$. Therefore, any cycle involves only students in $S_j$ and schools in $C_j$ and thus by the end of this step, all students assigned to a school are in $S_j$, and every school receiving a student is in $C_j$.

Since all students are acceptable to all schools and all schools are acceptable to all students, Pareto efficiency of the TTC implies that every student is matched to a school in the assignment $\mu$ generated by the end of the process. Therefore, for any $i$ and every school $c_i \in C_i, s \in \mu(c_i) \implies s \in S_i$ and thus $\mu$ maximizes segregation.

**Scenario 3:**

First, since TTC is efficient, every student is assigned to a school. We will then show that, for any $i$, $\mu(s) \in C_i \implies s \in S_i$, which implies that $\mu$ maximizes segregation. By assumption, every student has the same preferences over the schools in $C$. We will prove the statement by induction in the steps of the TTC procedure.

**Step 1**: at the beginning of step 1, every school and every student is available. Let $c_1 \in C_j$ for some $j$. Since $c_1$ has seats available and it is the first option for every student, every student points to $c_1$. Since all students are available, that includes students in $S_j$ and since $c_1$ gives priority to students in $S_j$, $c_1$ points to some student in $S_j$, forming the unique cycle, between $c_1 \in C_j$ and $s_j \in S_j$. Therefore, $\mu(s_j) \in C_j$ and $s_j \in S_j$.

**Step $k^*$**: let $j \in \{1, \ldots, |C|\}$ be the lowest value for which school $c_j$ is still available in the procedure. This implies that $c_j$ is the most preferred school among those still available and thus every student points to $c_j$. Since by induction assumption for any $i$, $\mu(s) \in C_i \implies s \in S_i$ for those $s$ that were removed from the procedure in earlier steps. Therefore, the number of students still available in $S_j$ equals the number of seats available in schools in $C_j$. Thus at step $k^*$, school $c_j$ points to some student in $s_j \in S_j$, forming the unique cycle, between $c_j \in C_j$ and $s_j \in S_j$. Therefore, $\mu(s_j) \in C_j$ and $s_j \in S_j$.

**Theorem 3**

We first show that $C_c$ satisfies the law of aggregate demand, which is a property of a choice function where $S'' \subset S' \subset S$ implies $|C(S')| \geq |C(S'')|$. To see that, notice first that the addition of unacceptable students has no effect on the number of students accepted and thus the analysis can focus only on sets of acceptable students. If $|S'| < q_c$ then all students in $S'$ or $S''$ are accepted and thus $|C_c(S'')| < |C_c(S')| \leq q_c$. If $|S'| \geq q_c$ then $|C_c(S')| = q_c$, and since $C_c$ never accepts more than $q_c$ students, the property is satisfied. By Proposition
In Sönmez (2013) and Proposition 6.4 in Roth and Sotomayor (1992), the mechanism that yields the student-optimal stable matching is the only mechanism that is pairwise stable and is strategy-proof.

We now show, through an example, that the student-optimal stable assignment may not implement diversity, finishing the proof.

\[ S = \{s_1, s_2\}, \quad T = \{t_1\}, \quad S^{t_1}(S) = \{s_1\}, \quad C = \{c_1, c_2\} \]

\[ >_{c_1}: s_1 \ s_2 \]
\[ >_{c_2}: s_2 \ s_1 \]
\[ >_{s_1}: c_2 \ c_1 \]
\[ >_{s_2}: c_1 \ c_2 \]

\[ q_{c_1} = q_{c_2} = 1, \quad q^{T}_{c_1} = (q^{T}_{c_1}) = (1) \text{ and } q^{T}_{c_2} = (q^{T}_{c_2}) = (0) \]

There are two fair with diversity assignments, \( \mu \) and \( \mu' \), where \( \mu(c_1) = \{s_2\}, \mu(c_2) = \{s_1\}, \mu'(c_1) = \{s_1\} \text{ and } \mu'(c_2) = \{s_2\} \). The assignment \( \mu' \) implements diversity, since \( \mu \not\succ \mu' \). The student-optimal stable assignment, however, is \( \mu \), and it is easy to see that \( \mu' \succ \mu \) and thus \( \mu \) doesn’t implement diversity.

**Theorem 4**

The proof of this theorem consists of showing that the DDA mechanism, as defined for the sequence of economies \( n \in \mathbb{N} \) in section 5.1, is semi-anonymous and envy-free but for tie-breaking, which Azevedo and Budish (2013) show to be a sufficient condition for \( \{(\mathcal{M}^n)_n, G\} \) to be SP-L.

Agents (students) belong to groups \( h \) in a finite set \( H \). A semi-anonymous mechanism is defined as \( \{(\Psi^n)_n, (G_h)_{h \in H}\} \), where the \( G_h \) are the sets of actions available to each subgroup \( h \), and

\[ G = G_{h_1} \cup \cdots \cup G_{h_{|H|}} \]

is the set of actions. The \( (\Psi^n)_n \) are functions

\[ \Psi^n : G^n \to \Delta(X^n_0) \]

where \( X^n_0 \) are feasible allocations for the economy \( n \). Agents in subgroup \( h \) are restricted to play strategies in \( G_h \). It is easy to see that the DDA mechanism defined in section 5.1 is semi-anonymous. A subgroup is a combination of student type and priority class, so that \( H = T \times Z^{|C|} \). The elements of each sets of actions \( G_h \) differ, therefore, only in the preference
types. Since students are not able to manipulate their types or priority classes, a student in subgroup $h$ is restricted to play strategies in $G_h$, and therefore DDA is semi-anonymous in this setting.

The property that Azevedo and Budish (2013) show being sufficient for a semi-anonymous mechanism to be SP-L is the following:

**Definition 12.** (Azevedo and Budish, 2013) A direct semi-anonymous mechanism $\{\Psi^n\}_n : (G_h)_{h \in H}$ is envy-free but for tie-breaking if for each $n$ there exists a function $x^n : (G \times [0, 1])^n \to \Delta (X^n_h)$, symmetric over its coordinates, such that

$$\Psi^n(g) = \int_{l \in [0,1]^n} x^n((g_i, \ell_i)_{i \in S^n}) \, dl$$

and, for all $s, s', n, g$ and $l$, if $\ell_s \geq \ell_{s'}$ and if $g_s$ and $g_{s'}$ belong to the same subgroup, then

$$u_{g_s}[x^n_s((g_i, \ell_i)_{i \in S^n})] \geq u_{g_s}[x^n_{s'}((g_i, \ell_i)_{i \in S^n})].$$

In other words, if students $s$ and $s'$ belong to the same subgroup and $\ell_s \geq \ell_{s'}$, then $s$ doesn’t envy the outcome of student $s'$. Notice that the matching function $\mu^n((g_i, \ell_i)_{i \in S^n})$ described in section 5.1 is symmetric over its coordinates: if the characteristics group $g$ and the lottery numbers (which together completely characterize the information about a student used by the mechanism) of two students are switched, it is easy to see that the outcome of the DDA mechanism will be the same except for the assignment of those two students, which will be switched.

Suppose now, for contradiction, that there are two students $s$ and $s'$ such that $\ell_s \geq \ell_{s'}$, $g_s = g_s'$ but $u_{g_s}[x^n_s((g_i, \ell_i)_{i \in S^n})] < u_{g_s}[x^n_{s'}((g_i, \ell_i)_{i \in S^n})]$. Let $c = x^n_s((g_i, \ell_i)_{i \in S^n})$ and $c' = x^n_{s'}((g_i, \ell_i)_{i \in S^n})$. In terms of student $s$’s preferences, $c' >_s c$. Since $s$ and $s'$ are in the same subgroup and $\ell_s \geq \ell_{s'}$, then $s >_{c'} s'$, and $\tau^n(s) = \tau^n(s')$. But this implies that student $s$ justifiably demands a seat in school $c'$, as in definition 2, which is a contradiction with Theorem 1.

When drawing the lottery numbers vector $\ell$ uniformly at random as described in section 5.1 we conclude, therefore, that the induced random mechanism $\{M^n\}_n : (G_h)_{h \in H}$ is envy-free but for tie-breaking and SP-L.

**Proposition 11**

The proof is based on the two conditions that are sufficient for a mechanism to be immune to strategic behavior based on switching the order of schools in the submitted ranking by
students with symmetric beliefs described in Ehlers (2008): Anonymity and Positive Association. Although the original theorem considers a situation in which schools are described by having only a strict ranking of students and a capacity, it is easy to verify that the results hold with the addition of the diversity objectives as well since our definition of a profile $\succ_s^{c\leftrightarrow c'}$ we consider that schools $c$ and $c'$ exchange their diversity objectives as well.

Given an assignment $\mu \in \mathcal{M}$ and $c, c' \in C$, let $\mu^{c\leftrightarrow c'}$ denote the assignment under which $\mu^{c\leftrightarrow c'}(c) = \mu(c')$ and $\mu^{c\leftrightarrow c'}(c') = \mu(c)$ and for all $c'' \in C \setminus \{c, c'\}$, $\mu^{c\leftrightarrow c'}(c'') = \mu(c'')$. A mechanism satisfies Anonymity if for all $s \in \mathcal{P}$, all $\mu \in \mathcal{M}$ and all $c, c' \in C$, if $\varphi(\succ) = \mu$, then $\varphi(\succ^{c\leftrightarrow c'}) = \mu^{c\leftrightarrow c'}$. A mechanism $\varphi$ satisfies Positive Association if for all $s \in \mathcal{P}$, all $s \in S$ and $c, c' \in C$, if $\varphi(\succ)(s) = c$ and $c' \succ_s c$, then $\varphi(\succ^{c\leftrightarrow c'}, \succ^{c\leftrightarrow c'}) = c$.

Since what identifies a school for the purpose of the DDA mechanism is its priorities, capacity and diversity objectives, it is easy to see that DDA satisfies Anonymity.

Denote the DDA mechanism by $\varphi^{DDA}$. We will show that if for some $s \in S$ and $c_i, c_j \in C$ where $c_i \succ_s c_j$, if $\varphi^{DDA}(\succ)(s) = c_j$ then $\varphi^{DDA}(\succ^{c_i\leftrightarrow c_j}, \succ, \succ) = \varphi^{DDA}(\succ)$, implying that DDA satisfies Positive Association. To do so, it is sufficient to show that at any step of the procedure in which $\succ_s$ is used there is no change in its outcome when compared with when $\succ^{c_i\leftrightarrow c_j}$ is used instead.

We will use below two facts: (i) By Proposition 2 in Roth (1984), if at some step $t - 1$ of the DDA mechanism a student $s \in S$ receives a proposal from a school $c \in C$, during step $t$ student $s$ will still receive a proposal from school $c$ unless she rejected $c$. (ii) The only circumstance in which a student rejects a proposal is when she receives another proposal from a school with a higher rank with respect to the submitted preferences.

Now consider, without loss of generality, that $\succ_s$ orders schools as follows:

$$c_1 \succ_s c_2 \succ_s \cdots \succ_s c_i \succ_s \cdots \succ_s c_j \succ_s c_{j+1} \succ_s \cdots \succ_s c_{|C|}$$

By facts (i) and (ii) and $\varphi^{DDA}(\succ)(s) = c_j$, during the procedure for $\varphi^{DDA}(\succ)$ there was no step during which student $s$ received a proposal from some school $c_\ell$ for $\ell < j$. As a consequence, the only part of $\succ_s$ that may have been used to determine $\varphi^{DDA}(\succ)$ is that $c_j \succ_s c_{j+1} \succ_s \cdots \succ_s c_{|C|}$. But it is also true that $c_j \succ^{c_i\leftrightarrow c_j} c_{j+1} \succ^{c_i\leftrightarrow c_j} \cdots \succ^{c_i\leftrightarrow c_j} c_{|C|}$, and so therefore $\varphi^{DDA}(\succ^{c_i\leftrightarrow c_j}, \succ, \succ) = \varphi^{DDA}(\succ)$ and thus DDA satisfies Positive Association.
Specification of the alternative mechanisms

Top Trading Cycles (TTC)

The following procedure is described in Abdulkadiroğlu and Sönmez (2003).

**Step 1**: Assign a counter for each school which keeps track of how many seats are still available at the school. Initially set the counters equal to the capacities of the schools. Each student points to his favorite school under her announced preferences. Each school points to the student who has the highest priority for the school. Since the number of students and schools are finite, there is at least one cycle. (A cycle is an ordered list of distinct schools and distinct students \((s_1, c_1, s_2, \ldots, s_k, c_k)\) where \(s_1\) points to \(c_1\), \(c_1\) points to \(s_2\), \ldots, \(s_k\) points to \(c_k\), \(c_k\) points to \(s_1\).) Moreover, each school can be part of at most one cycle. Similarly, each student can be part of at most one cycle. Every student in a cycle is assigned a seat at the school she points to and is removed. The counter of each school in a cycle is reduced by one and if it reduces to zero, the school is also removed. Counters of all other schools stay put.

In general, at step \(k\): each remaining student points to her favorite school among the remaining schools and each remaining school points to the student with highest priority among the remaining students. There is at least one cycle. Every student in a cycle is assigned a seat at the school that she points to and is removed. The counter of each school in a cycle is reduced by one and if it reduces to zero the school is also removed. Counters of all other schools stay put.

The procedure terminates when all students are assigned a seat.

Deferred Acceptance Minority Reserves (DAMR)

The extension of the DAMR mechanism in Hafalir et al. (2013) for multiple types consists in applying the student-proposing deferred acceptance mechanism when schools’ choice function is \(C_e\), presented in section 3. More specifically:

**Step 1**: Start with the matching in which no student is matched. Each student \(s\) applies to her first-choice school. Let \(S_c^1\) be the set of students that applied to school \(c\). Each school \(c\) accepts all students in \(C_e\left(S_c^1\right)\) and rejects the rest, if any.

43
**Step $k^*$**: Start with the tentative matching obtained at the end of step $k^* - 1$. Each student $s$ who got rejected at step $k - 1$ applies to her next-choice school. Each school $c$ considers the new applicants ($S^k_c$) and students admitted tentatively at step $k^* - 1$ ($C_c (S^{k^* - 1}_c)$). Each school $c$ accepts all students in $C_c (S^{k^* - 1}_c) \cup S^k_c$ and rejects the rest, if any. If there are no rejections, then stop.

The procedure terminates when no rejection occurs and the tentative matching at that step is finalized. Since no student reapplies to a school that has rejected her and at least one rejection occurs in each step, the procedure stops in finite time.