Fair Implementation of Diversity in School Choice*

Inácio Bó†

September, 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 are introducing 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 we 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 optimally approximates 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 and IRC conditions 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 also show analytically that the mechanism that we provide has a general ability to satisfy those objectives, as opposed to some currently proposed mechanisms, which may yield segregated assignments.

JEL classification: C78, D63, D78, D82
Keywords: Mechanism Design, Matching, School Choice, Affirmative Action, Diversity.

*In a version of their paper published in September 2014, Ehlers et al. (2014) show simultaneous and independent work presenting some of the properties of the school-proposing deferred acceptance mechanism described here.

†Address: WZB Berlin Social Science Center, Reichpietschufer 50, D-10785 Berlin, Germany; website: http://www.inaciobo.com; e-mail: inacio.bo@wzb.eu; phone: +49 (0)30-25491-292.

‡An earlier version of this paper was made available in November 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 to choose which school to apply to the forced reallocation of students.[2] More recent such efforts include Brazil's racial and income-based public university reserves (Aygün and Bó, 2013) and the attempt to increase religious diversity in British schools (Coldron et al., 2008).

Most policies used to achieve that objective consist of 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 Ciotti (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 district 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

[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).

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 otherwise available to everyone.[3] Examples of affirmative action mechanisms include artificial increases on exam scores for students from public schools in university admissions (Matos et al., 2012) and giving higher priority to racial minorities for a number of seats in schools.[4] Another class of mechanisms takes 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 admission – which is in many cases determined by student performance in tests and high-school grades – affirmative action mechanisms could increase the diversity of cohorts by improving the access of minority students to more competitive universities. In the case of school choice, however, that it 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 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 certain scenarios these 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 desired 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 same-type 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

[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. (2014).

[5]See Ehlers et al. (2014) and Troyan and Fragiadakis (2013).

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 if 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 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 other diversity implementation mechanisms in the literature is the fact that it doesn't use the diversity objectives as a binary objective (that is either satisfied or not). It instead *induces an ordering* of the the satisfaction of those objectives across assignments, 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, one key aspect of this paper is our use of 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 some technical conditions, namely Substitutability and Irrelevance of Rejected Contracts (IRC). 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 school-optimality of the stable allocations selected *induces the desired properties on the allocation.*

[6] A similar definition of fairness is presented in the context of an affirmative action mechanism in Ehlers et al. (2014).

While the use of the school-proposing version of the deferred acceptance procedure allows the use of the school-optimality property to select among stable assignments with some degree of arbitrariness, the student-proposing deferred acceptance procedure is widely known for being strategy-proof for the students and having desirable 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.

We complement our paper with an analytic evaluation of how the outcome of our mechanism compares to the use of student-proposing deferred acceptance affirmative action mechanisms. 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 alternative may lead to highly 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 (School-Proposing Diversity, or SPDiv) that generates an assignment which satisfies, in a well-defined maximal way, those diversity objectives without being wasteful or compromising fairness (Theorem 1). The mechanism we propose may be vulnerable to strategic manipulation by students (Example 3). However, there is no mechanism that both implements diversity with fairness and is strategy-proof (Theorem 3). Moreover, we show that when the number of students is large, gains from manipulating their preferences disappear (Theorem 4).

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 the SPDiv and those generated by the student-proposing affirnative action mechanisms under some symmetric population distributions. We show that the SPDiv mechanism is able to minimize segregation regardless of schools' priorities or students' preferences (Theorem 2). We show that, on the other hand, student-proposing affirnative action mechanisms yield segregated assignments for some familiar preference profiles (propositions 5 and 6).

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.[7]

The incentive and welfare properties of both mechanisms come into play in the context of college admissions in Balinski and Sönmez (1999). In their model, colleges are not considered strategic agents but their seats are simple objects to be consumed by the students. In this scenario there is no need for strategic or welfare considerations on the part of colleges. Moreover, they show that when colleges preferences are based on exogenous 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) use the SPDA associated with a maximum quota for certain types. After Kojima (2012) showed that maximal quotas may hurt every minority student, the focus shifted to minority reserves used, for example, in Ehlers et al. (2014), 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.

When students have strict preferences, however, diversity objectives and student welfare (when measured in terms of these preferences) may go in opposite directions. As a result, unless student 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 and IRC, 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.

Though 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. (2014) present, 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.[8] 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

[7] Whereas 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).

[8] This statement and the one below in this section refer to the "hard-bounds" mechanism in Ehlers et al. (2014).

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. (2014) 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[9] and has good incentives properties in large markets.

The paper proceeds as follows. Section 2 introduces the model and the basic definitions of fairness and implementation of diversity. Section 3 presents the SPDiv mechanism and its general properties. Section 4 presents the analytical results of the outcomes generated by the SPDiv mechanism and student-proposing affirmative action mechanisms. Section 5 discusses the incentives in the game induced by the SPDiv mechanism in large markets. Section 6 concludes. Proofs omitted from the main text can be found in the Appendix.

2 Model

A **school choice with diversity problem** consists of a tuple $\langle S, C, T, \tau, q, \underline{q}, \succ_S, \succ_C \rangle$:

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, $S^t(I) = \{s \in I : t = \tau(s)\}$.
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 $q_c^T = (q_c^{t_1}, \ldots, q_c^{t_k})$ of **diversity objectives**, where q_c^t is the minimum desired number of students with type t at school c, where $\sum_{t \in T} q_c^t \leqslant q_c$. Let $\underline{q} = (q_{c_1}^T, \ldots)$.
7. Students' **preference profile** $\succ_S = (\succ_{s_1}, \ldots, \succ_{s_{|S|}})$, where $\succ_s$ is a strict ranking over $C \cup \{s\}$, where s represents remaining unmatched to any school. If $s \succ_s c$, school c is deemed *unacceptable* to student s.
8. Schools' **priority profile** $\succ_C = (\succ_{c_1}, \ldots, \succ_{c_{|C|}})$, which is a collection of complete and strict rankings over the students in $S \cup \varnothing$. If $\varnothing \succ_c s$, student s is deemed *unacceptable* to school c.

[9] More specifically, every assignment that is fair in our model is fair in theirs, but not vice-versa.

An **assignment** μ is a function from $C \cup S$ to subsets of $C \cup S \cup \{\varnothing\}$ such that:

- $\mu(s) \in C \cup \{s\}$ and $|\mu(s)| = 1$ for every student s[10]
- $|\mu(c)| \leqslant q_c$ 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 μ, and for a school c, $\mu(c)$ is the set of students that are assigned to school c under μ. For a given school choice with diversity problem, we will denote by $\mathcal{M}$ the set of all assignments. A set of students $I \subseteq S$ **enables diversity at school** c if for all $t \in T$, $|S^t(I)| \geqslant q_c^t$. An assignment μ **fully implements diversity** if for every school c, $\mu(c)$ enables diversity at c. If there is an assignment $\mu^* \in \mathcal{M}$ where μ^* 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** μ if $|\mu(c)| < q_c$ and $c \succ_s \mu(s)$. An assignment μ is **non-wasteful** if no student justifiably claims an empty seat at some school. An assignment is **individually rational** if for every student s, $\mu(s) \succ_s s$ and for every school c and every student $s' \in \mu(c)$, $s' \succ_c \varnothing$. A traditionally desirable condition for an assignment to satisfy is that of having no student that justifiably envies another. We define that formally, using the notion of fairness in Balinski and Sönmez (1999):

Definition 1. A student s **justifiably envies** student s' under the assignment μ, where $c = \mu(s')$, if and only if $c \succ_s \mu(s)$ and $s \succ_c s'$. An assignment μ satisfies **no justified envy** if no student justifiably envies another under μ. An assignment μ 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.[11] 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:

$$\begin{aligned} S &= \{s_1, s_2\} & & \\ T &= \{t_1\} & C &= \{c_1, c_2\} \\ S^{t_1}(S) &= \{s_1\} & & \\ \succ_{s_1}&: c_2\ c_1 & \succ_{c_1}&: s_2\ s_1 \\ \succ_{s_2}&: c_2\ c_1 & \succ_{c_2}&: s_2\ s_1 \end{aligned}$$

[10] We will abuse notation and consider $\mu(s)$ as an element of C, instead of a set with an element of C.

[11] 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 another student with lower priority.

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

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

Diversity objectives are feasible, since the assignment μ' fully implements diversity. The unique fair assignment is μ, 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 students with type t whenever the diversity objective associated with t is not yet satisfied, μ' 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 $c \succ_s \mu(s)$ and either:

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

An assignment is **fair with diversity** if it is individually rational, non-wasteful, and if no student justifiably demands a seat in any school.[12]

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

[12]The reader may verify that in the example used in proposition 1 the unique fair with diversity assignment fully implements diversity.

Proposition 2. *If μ is fair and fully implements diversity, then μ is also fair with diversity.*

Proof. Since μ is non-wasteful, we only need to show that if μ 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 μ fully implements diversity, $|S^t(\mu(c))| \geqslant q_c^t$.
- Conditions 2 and 3: Since μ 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:

$$\begin{aligned} S &= \{s_1, s_2\} & & \\ T &= \{t_1\} & C &= \{c_1, c_2\} \\ S^{t_1}(S) &= \{s_1\} & & \\ \succ_{s_1}&: c_1\ c_2 & \succ_{c_1}&: s_1\ s_2 \\ \succ_{s_2}&: c_2\ c_1 & \succ_{c_2}&: s_2\ s_1 \end{aligned}$$

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

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

Diversity objectives are feasible, since the assignment μ' fully implements diversity. The unique fair with diversity assignment is μ, which doesn't fully implement diversity. □

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 $>^{\underline{q}}$ be the partial order over the set of assignments $\mathcal{M}$ such that $\mu' >^{\underline{q}} \mu$ if :

1. For all $c \in C$ and $t \in T$ such that $|S^t(\mu(c))| \leqslant q_c^t$, $|S^t(\mu'(c))| \geqslant |S^t(\mu(c))|$ and
2. There are $c' \in C$ and $t' \in T$ such that $|S^t(\mu(c'))| < q_{c'}^t$ and $|S^t(\mu'(c'))| > |S^t(\mu(c'))|$.

Denote $\mu' \not>^{\underline{q}} \mu$ if $\mu' >^{\underline{q}} \mu$ is false.

In other words, $\mu' >^{\underline{q}} \mu$ if μ' has less seats "reserved" for students of certain types occupied by students that are not of those types when compared to μ. Notice that if $\mu' >^{\underline{q}} \mu$ then it cannot be the case that both μ' and μ 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 μ assigns 40 minority students to c and μ' assigns 45, then $\mu' >^{\underline{q}} \mu$. However, If μ'' assigns 51 minority students to c and μ''' assigns 55, both $\mu'' \not>^{\underline{q}} \mu'''$ and $\mu''' \not>^{\underline{q}} \mu''$. Increasing the number of minority students after the diversity objective is satisfied doesn't make an assignment greater with respect to $>^{\underline{q}}$. 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 formally define what it means to implement diversity in this framework.

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

An assignment μ thus implements diversity if μ either fully implements diversity or μ does not fully implement diversity but there is no assignment that is fair with diversity and "further satisfies" some diversity objective in some school without jeopardizing another in the same or some other school.

3 The mechanism

We start by defining the choice function $\mathbb{C}_{\mathbf{c}} : 2^S \to 2^S$ at school c for the school choice with diversity problem $\left\langle S, C, T, \tau, q, \underline{q}, \succ_S, \succ_C \right\rangle$. Fix any $S' \subseteq S$ and let $I \subseteq S'$ be the set of students **acceptable to** c among S'. $\mathbb{C}_{\mathbf{c}}(S')$ is defined by the following procedure.

1. **Step 0**: If $|I| \leqslant q_c$, $\mathbb{C}_{\mathbf{c}}(S') = I$.

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

$$\vdots$$

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

$$\vdots$$

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

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 proposed in Hafalir et al. (2013). We now proceed to show some important properties of $\mathbb{C}_{\mathbf{c}}$.

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

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

In a recent paper, Aygün and Sönmez (2012) show that when schools' choices are primitives of the problem (as opposed to choice functions derived from preferences over sets of students), substitutability is not sufficient for guaranteeing the existence of stable matchings. A new condition, shown below, together with substitutability, suffices for the existence result.

Definition 6. A choice function C satisfies irrelevance of rejected contracts (IRC) if:

$$\forall I \subset S, \forall s \in S \backslash I \qquad s \notin C(I \cup \{s\}) \implies C(I) = C(I \cup \{s\})$$

Lemma 1. *The function $\mathbb{C}_{\mathbf{c}}$ satisfies the substitutability condition and IRC.*

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 $t \in T$, $S^t(I) \geqslant q_c^t$. Then $\mathbb{C}_{\mathbf{c}}(I)$* enables diversity at *$c$.*

Proof. Suppose not. Then there is at least one type $t \in T$ such that $|S^t(\mathbb{C}_{\mathbf{c}}(I))| < q_c^t$. Suppose first that $|I| < q_c$. Then the choice procedure would accept all students in I in step 0, which contradicts with $|S^t(I)| \geqslant q_c^t$, so it must be that $|I| \geqslant q_c$ and that the choice procedure will run until the final step. Notice, however, that in the step associated with type t, since $|S^t(I)| \geqslant q_c^t$, the top q_c^t students with respect to $\succ_c$ are accepted, which is a contradiction with $|S^t(\mathbb{C}_{\mathbf{c}}(I))| < q_c^t$. □

The proposition above shows that whenever the set of students available is such that there are subsets of them which enable diversity at c, $\mathbb{C}_{\mathbf{c}}$ selects one of them. As we show in section 4, however, this does not guarantee that the outcome of a deferred acceptance procedure fully implements diversity.

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

1. $\mathbb{C}_{\mathbf{c}}(I) = I'$.
2. $\mathbb{C}_{\mathbf{c}}(I \cup \{s\}) = \mathbb{C}_{\mathbf{c}}(I) \cup \{s\}$ for any $s \in S$ if s is acceptable to c.
3. $|\mathbb{C}_{\mathbf{c}}(I \cup J)| > |\mathbb{C}_{\mathbf{c}}(I)|$ for any $J \subset S$ such that for every $s \in J$, s is acceptable to c , $I \cap J \neq \varnothing$ and $I \neq J$.

The results presented in the lemma above come easily from the definition of the procedure for $\mathbb{C}_{\mathbf{c}}$, so we omit the proof.

An assignment μ is **blocked** by a student s if $s \succ_s \mu(s)$, and by a school c if $\mu(c) \neq \mathbb{C}_{\mathbf{c}}(\mu(c))$. Similarly, μ is **blocked by a student-school pair** (s,c) if $\mu(s) \neq c$, $c \succ_s \mu(s)$ and $s \in \mathbb{C}_{\mathbf{c}}(\mu(c) \cup \{s\})$. An assignment μ 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 μ is pairwise stable then μ 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 μ is fair with diversity if and only if μ is pairwise stable.*

The School-Proposing Diversity (**SPDiv**) mechanism that we propose consists of applying the school-proposing deferred acceptance procedure described in Roth (1984) using $\mathbb{C}_{\mathbf{c}}$ as the schools' choice function:

1. **Step 1:** Let $\mathbb{S}_c(1) = S$ for all $c \in C$

   (a) Each school c proposes to the students in $\mathbb{C}_{\mathbf{c}}(\mathbb{S}_c(1))$.

   (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. Let, for all $c \in C$, $R_c(1)$ be the set of students who rejected school c at this step.

   $\vdots$

2. **Step k:** Let $\mathbb{S}_c(k) = \mathbb{S}_c(k-1) \setminus R_c(k-1)$ for all $c \in C$.

   (a) Each school c proposes to the students in $\mathbb{C}_{\mathbf{c}}(\mathbb{S}_c(k))$.

   (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.

The procedure terminates at any step T in which no rejections are issued, and the resulting assignment μ is such that for every school c, $\mu(c) = \mathbb{C}_{\mathbf{c}}(\mathbb{S}_c(T))$ as defined above. Students that are not in the choice set of any school are left unmatched.

The following result extends a theorem in Roth (1984) for cases in which, as in this paper, choice functions are the primitives instead of preference relations:

Lemma 4. *Suppose that students' preferences are strict and that the schools' choice function satisfies the substitutability condition and IRC. Then the assignment μ^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 μ, $\mu^C(c) = \mathbb{C}_{\mathbf{c}}\left(\mu^C(c) \cup \mu(c)\right)$.*

Lemma 5. *Let μ and μ' be fair with diversity assignments. If for every $c \in C$, $\mu(c) = \mathbb{C}_{\mathbf{c}}(\mu(c) \cup \mu'(c))$ then $\mu' \not\succ^{q} \mu$.*

Proof. Suppose not. Then $\mu' \succ^{q} \mu$ and therefore there is $c \in C$ and $t \in T$ such that $|S^t(\mu(c))| < q_c^t$ and $|S^t(\mu'(c))| > |S^t(\mu(c))|$. By Lemma 3, both $\mu(c)$ and $\mu'(c)$ contain only acceptable students for c, $\mu(c) = \mathbb{C}_{\mathbf{c}}(\mu(c))$ and $\mu'(c) = \mathbb{C}_{\mathbf{c}}(\mu'(c))$.

If $|\mu(c)| < q_c$, then by Lemma 2 and the fact that $\mu(c) \neq \mu'(c)$, $|\mathbb{C}_{\mathbf{c}}(\mu(c) \cup \mu'(c))| > |\mathbb{C}_{\mathbf{c}}(\mu(c))|$. But then $|\mathbb{C}_{\mathbf{c}}(\mu(c) \cup \mu'(c))| > |\mu(c)|$ which implies $\mu(c) \neq \mathbb{C}_{\mathbf{c}}(\mu(c) \cup \mu'(c))$, a contradiction. Thus, $|\mu(c)| = q_c$ and the procedure for $\mathbb{C}_{\mathbf{c}}(\mu(c) \cup \mu'(c))$ 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 $|S^t(\mu(c))| < q_c^t$, student s' is accepted in the step associated with t in $\mathbb{C}_{\mathbf{c}}(\mu(c) \cup \mu'(c))$. As a consequence, $s' \in \mathbb{C}_{\mathbf{c}}(\mu(c) \cup \mu'(c))$ and thus $\mu(c) \neq \mathbb{C}_{\mathbf{c}}(\mu(c) \cup \mu'(c))$, a contradiction. □

Putting it all together we get our main result:

Theorem 1. *The SPDiv mechanism implements diversity.*

Proof. Let μ^C be the outcome of the SPDiv mechanism. By lemmas 1 and 4, μ^C is pairwise stable and thus, by Lemma 3, μ^C is fair with diversity. By lemmas 4 and 5, for any fair with diversity assignment μ', $\mu' \not\succ^{q} \mu^C$. As a consequence, μ^C implements diversity. □

4 Comparative Analysis

The purpose of the SPDiv mechanism is to attain a school assignment that is not only fair but approximates, as much as possible, the assignment to the diversity objectives. In this section we show how it compares to an afirmative action mechanism based on the student-proposing deferred acceptance procedure. We consider a generalization for multiple types of students of the Deferred Acceptance with Minority Reserves (DAMR) mechanism, proposed in Hafalir et al. (2013)[13]. That mechanism expands the access that some students have to every school (thus being 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, in a sense it 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

[13]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. (2014), where there are no upper quotas. See also Kominers and Sönmez (2012) and Westkamp (2013) for more generalizations.

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, the ability of the mechanism to attain such an 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 7. An assignment μ **maximizes segregation** if for every school $c \in C$, $s, s' \in \mu(c) \implies \tau(s) = \tau(s')$.

Definition 8. Diversity objectives **mirror the population distribution** if for every $c \in C$ and $t \in T$, $q_c^t = \left\lfloor \frac{q_c \left| S^t(S) \right|}{|S|} \right\rfloor$. An assignment μ **minimizes segregation** if μ 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 μ, where $\mu(c_1) = \{s_1, s_2\}$ and $\mu(c_2) = \{s_3, s_4\}$ *maximizes segregation*, while the assignment μ', 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 $|S| = n_1 n_2 k^2$ and that the number of schools is such that $|C| = n_2 k$, for some $n_1, n_2 \in \mathbb{N}$. As a result, $q = n_1 k$ and for every i, $|S_i| = n_1 n_2 k$ and $\frac{q|S_i|}{|S|} = n_1$.

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

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

Theorem 2 shows that the SPDiv 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 consider two 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 not specified. This class of preferences accommodates, among other things, 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 $s_i \in S_i$ and $j \neq i$ it follows that $c_i \in C_i$ and $c_j \in C_j$ implies $c_i \succ_{s_i} c_j$.

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 $C = C_1 \cup \cdots \cup C_a$, where $a \geqslant k$, where $|C_1| = \cdots = |C_a|$ 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 $c_i \succ_s c_{i+1}$.

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

Lemma 6. *Every assignment μ generated by the DAMR mechanism is fair with diversity.*

Proof. By Lemma 1 and Theorem 1 in Aygün and Sönmez (2012), μ is stable (and thus pairwise stable). Thus by Lemma 3 μ is fair with diversity. □

We now 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 μ 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 $\frac{|C^*|}{|C|} \geqslant \frac{a-1}{a}$.*

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:

$$
\begin{aligned}
S &= \{s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8\} \\
T &= \{t_1, t_2\} \qquad\qquad C = \{c_1, c_2, c_3, c_4\} \\
S^{t_1}(S) &= \{s_1, s_2, s_3, s_4\} \\
S^{t_2}(S) &= \{s_5, s_6, s_7, s_8\} \\
\succ_{s_1}&: c_1\ c_2\ c_3\ c_4 \qquad \succ_{c_1}: s_1\ s_2\ s_3\ s_4\ s_5\ s_6\ s_7\ s_8 \\
\succ_{s_2}&: c_2\ c_1\ c_3\ c_4 \qquad \succ_{c_2}: s_1\ s_2\ s_3\ s_4\ s_5\ s_6\ s_7\ s_8 \\
\succ_{s_3}&: c_1\ c_2\ c_3\ c_4 \qquad \succ_{c_3}: s_1\ s_2\ s_3\ s_4\ s_5\ s_6\ s_7\ s_8 \\
\succ_{s_4}&: c_2\ c_1\ c_3\ c_4 \qquad \succ_{c_4}: s_1\ s_2\ s_3\ s_4\ s_5\ s_6\ s_7\ s_8 \\
\succ_{s_5}&: c_1\ c_2\ c_3\ c_4 \\
\succ_{s_6}&: c_2\ c_1\ c_3\ c_4 \\
\succ_{s_7}&: c_1\ c_2\ c_4\ c_3 \\
\succ_{s_8}&: c_2\ c_1\ c_4\ c_3
\end{aligned}
$$

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

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

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

When $a = |C|$, that is, the set of schools is partitioned such that each partition has only one school, the preferences in scenario 2 are equivalent to a situation in which all students have the same preferences among schools. It is easy to see that proposition 6 leads to the following corollary:

Corollary 2. *Every assignment generated by the DAMR mechanism minimizes segregation when all students have the same preferences.*

The value of a in scenario 6 indicates, in a way, the degree of correlation among students' preferences over schools: the larger the value of a, the more similar they are. Proposition 6 shows, therefore, that the DAMR mechanism may be an adequate choice of mechanism in situations where students' preferences follow, for example, a widely known ranking, but less so when preferences are more heterogeneous.

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 SPDiv mechanism, as shown by the following example.

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

$$\begin{aligned} S &= \{s_1, s_2\} & \\ T &= \{t_1\} & C = \{c_1, c_2\} \\ S^{t_1}(S) &= \{s_1, s_2\} & \\ \succ_{s_1}&: c_2\ c_1 & \succ_{c_1}: s_1\ s_2 \\ \succ_{s_2}&: c_1\ c_2 & \succ_{c_2}: s_2\ s_1 \end{aligned}$$

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

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

If students truthfully submit their preferences, the outcome of the SPDiv mechanism is μ. Now suppose that student s_1 manipulates its preference and submits $\succ'_{s_1}$, where $c_2 \succ'_{s_1} \varnothing \succ'_{s_1} c_1$, that is, under $\succ'_{s_1}$ school c_1 is unacceptable. In this case the outcome of the SPDiv mechanism will be μ', 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 SPDiv 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*

There is evidence that suggests, however, that successful manipulations of stable mechanisms are rare in the presence of a large number of 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, like the SPDiv. Theoretical work on big markets also add to this evidence. 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 an argument for why we shouldn't expect students to manipulate their preferences when the SPDiv mechanism is used. We show that the expected benefits of manipulating the SPDiv mechanism are reduced to an arbitrarily small value as the number of students grows.

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 in 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 ε 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 incentives 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 SPDiv mechanism is based on a similar result in Troyan and Fragiadakis (2013), in which they show that their proposed EDQDA mechanism is SP-L. [14]

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: $S^n = S_1^{c,n} \cup \cdots \cup S_{|Z|}^{c,n}$. For each student s, let $z_s \in Z^{|C|}$ denote the vector of priority classes with which student s is associated 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 $s \in S_i^{c,n}$, $s' \in S_j^{c,n}$ and $i \neq j$ then $s \succ_c s' \iff i < j$, 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 $\succ_i$ over the elements of $C \cup \{\varnothing\}$ there is a type $a_i \in A$ such that u_{a_i} represents the ordinal preferences in $\succ_i$ over degenerate lotteries over $C \cup \{\varnothing\}$[15]. 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 9. Fix a set of schools C, a sequence of capacity vectors $(q_n)_{\mathbb{N}}$ and of corresponding diversity objectives $\left(\underline{q_n}\right)_{\mathbb{N}}$. A *school choice with diversity mechanism* $\{(\varphi^n)_{\mathbb{N}}, G\}$ is a sequence of allocation functions $\varphi^n : G^n \to \Delta\left(C^n\right)$ 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. [16]

Denote by $\varphi_s^n(g_s, g_{-s})$ 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

[14] Unlike their case, however, no assumption on a consistency between diversity objectives and the distribution of types of the population is necessary here.

[15] For simplicity of notation in this section we represent by $\varnothing$ the possibility of remaining unmatched to any school.

[16] Formally, let $q_n = \left(q_{c_1}^n, \ldots, q_{c_m}^n\right)$. 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 $c \in C$, $\sum_{i=1}^n \mathbf{1}_{\phi_i = c} \leqslant q_c^n$.

that is induced by the mechanism and a distribution with full support over preference types $m \in \overline{\Delta}A$, given that her type is $\tau^n(s)$ and her priority class is z_s:

$$\phi_s^n(a_s, m) = \sum_{g_{-s} \in G^{n-1}} \varphi_s^n\left(\left(\tau^n(s), a_s, z_s\right), g_{-s}\right) \cdot Pr\left(g_{-s} | a_{-s} \sim iid(m)\right)$$

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 the manipulation of the value of a_s for a given $(\tau^n(s), z_s)$:

Definition 10. (Azevedo and Budish, 2013) A school choice with diversity mechanism $\{(\varphi^n)_{\mathbb{N}}, G\}$ is **strategy-proof in the large (SP-L)** if, for any $\varepsilon > 0$ and any $m \in \overline{\Delta}A$, there exists n_0 such that for all $n \geqslant n_0$, all $(t, z) \in T \times Z$ and all $a, a' \in A$:

$$u_a\left[\phi_{(t,z)}^n(a, m)\right] \geqslant u_a\left[\phi_{(t,z)}^n(a', m)\right] - \varepsilon$$

Consider now the following version of the SPDiv mechanism, defined for each economy n. In the first stage, each student s submits her rankings $\succ_s$ over $C \cup \{\varnothing\}$. In the second stage, the lottery vector $\ell \in [0,1]^n$ is uniformly drawn 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 $s, s' \in S_i^{c,n}$ then $s \succ_c s' \iff \ell_s > \ell_{s'}$. Given students' preferences $\succ_{S^n}$, schools' priorities $\succ_C$ and economy n as described, the SPDiv mechanism is used to produce a school assignment, as described in section 3.

Let $\mu^n\left((g_i, \ell_i)_{i \in S^n}\right)$ be the matching generated by the SPDiv 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:

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

That is, $\mathcal{M}^n$ is the *school choice with diversity mechanism* that results from the procedure described above to assign schools to students. We can now proceed to the main result:

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

6 Conclusion

This paper proposes 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, through their preferences, the possibility to obtain more diverse school cohorts, very few mechanisms make an explicit attempt to enforce them to some extent. The proposed SPDiv mechanism generates an assignment that is as close as possible to the distribution implied by the diversity objectives while requiring that such an assignment satisfies a well-defined fairness criterion. This property is achieved by using the school-proposing deferred acceptance procedure when selecting a stable matching instead of the widely used student-proposing version. Analytical results show a general ability of the proposed mechanism for satisfying those objectives, as opposed to other mechanisms proposed in the literature.

References

Abdulkadiroğlu, Atila (2005), "College admissions with affirmative action." *International Journal of Game Theory*, 33, 535–549. 3, 7

Abdulkadiroğlu, Atila, Parag Pathak, Alvin E. Roth, and Tayfun Sönmez (2006), "Changing the boston school choice mechanism." Working Paper, National Bureau of Economic Research. 1.2, 5.1

Abdulkadiroğlu, Atila, Parag A Pathak, and Alvin E Roth (2005), "The new york city high school match." *The American Economic Review*, 95, 364–367. 1.2

Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth (2009), "Strategy-proofness versus efficiency in matching with indifferences: Redesigning the nyc high school match." *The American Economic Review*, 99, 1954–78. 1.2

Abdulkadiroğlu, Atila and Tayfun Sönmez (2003), "School choice: A mechanism design approach." *The American Economic Review*, 93, 729–747. 1, 3, 1.2, 2

Aygün, O. and T. Sönmez (2012), "Matching with contracts: The critical role of irrelevance of rejected contracts." Boston College working paper. 3, 4, A

Aygün, Orhan and Inácio Bó (2013), "College admissions with multidimensional reserves: the brazillian affirmative action case." Working paper, Boston College. 1

Azevedo, Eduardo and Eric Budish (2013), "Strategyproofness in the large." Working Paper, University of Chicago. 5.1, 10, A, 11

Balinski, Michel and Tayfun Sönmez (1999), "A tale of two mechanisms: student placement." *Journal of Economic theory*, 84, 73–94. 1.2, 2, 2

Ciotti, Paul (2001), "Money and school performance: Lessons from the kansas city desegregation experiment." In *School Reform: The Critical Issues*, 308–338, Hoover Institution Press. 1

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

Dubins, Lester E and David A Freedman (1981), "Machiavelli and the gale-shapley algorithm." *The American Mathematical Monthly*, 88, 485–494. 1.2, 7

Echenique, Federico and M. Bumin Yenmez (2012), "How to control controlled school choice." SS Working Paper 1366, Caltech. 1.2, 3, 13

Ehlers, Lars, Isa E Hafalir, M Bumin Yenmez, and Muhammed A Yildirim (2014), "School choice with controlled choice constraints: Hard bounds versus soft bounds." *Journal of Economic Theory*. ∗, 4, 5, 6, 1.2, 8, 13

Erdil, A and T Kumano (2012), "Prioritizing diversity in school choice." Working paper, Washington University. 1.2

Gale, D. and L. S. Shapley (1962), "College admissions and the stability of marriage." *The American Mathematical Monthly*, 69, 9–15. 1.2, 2

Gale, David and Marilda Sotomayor (1985), "Ms. machiavelli and the stable matching problem." *The American Mathematical Monthly*, 92, 261–268. 1.2

Hafalir, Isa E, M Bumin Yenmez, and Muhammed A Yildirim (2013), "Effective affirmative action in school choice." *Theoretical Economics*, 8, 325–363. 4, 1.2, 3, 4, 13, A

Hatfield, John William and Fuhito Kojima (2010), "Substitutes and stability for matching with contracts." *Journal of Economic Theory*, 145, 1704–1723. 7

Immorlica, Nicole and Mohammad Mahdian (2005), "Marriage, honesty, and stability." In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, SODA '05, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA. 5, 5.1

Kojima, Fuhito (2012), "School choice: Impossibilities for affirmative action." *Games and Economic Behavior*, 75, 685–693. 1.2

Kojima, Fuhito and Parag A. Pathak (2009), "Incentives and stability in large two-sided matching markets." *The American Economic Review*, 99, 608–627. 5, 5.1

Kominers, Scott Duke and Tayfun Sönmez (2012), "Designing for diversity: Matching with slot-specific priorities." *Boston College and University of Chicago working paper*. 1.2, 13

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

Ministerial Group on Public Order and Community Cohesion (2001), "Building cohesive communities: A report of the ministerial group on public order and community cohesion." Technical report. 1

Roth, Alvin E. (1984), "Stability and polarization of interests in job matching." *Econometrica*, 52, 47–57. 3, 3, A, A.1, A.1, A.2, A

Roth, Alvin E (1985), "The college admissions problem is not equivalent to the marriage problem." *Journal of economic Theory*, 36, 277–288. 1.2

Roth, Alvin E and Elliott Peranson (1999), "The redesign of the matching market for american physicians: Some engineering aspects of economic design." *The American Economic Review*, 89, 748–780. 5

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

Sönmez, Tayfun (2013), "Bidding for army career specialties: Improving the rotc branching mechanism." *Journal of Political Economy*, 121, 186–219. A

Troyan, Peter and Daniel Fragiadakis (2013), "Market design under distributional constraints: Diversity in school choice and other applications." Working paper, Stanford University. 5, 1.2, 5.1

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

Westkamp, Alexander (2013), "An analysis of the german university admissions system." *Economic Theory*, 53, 561–589. 13

Acknowledgements

I am grateful to Samson Alva, Orhan Aygün, Zhaochen He, Onur Kesten, Vikram Manjunath, Tayfun Sönmez, Bertan Turhan, Utku Ünver, and the seminar participants at Boston College, WZB Berlin Social Science Center and the Royal Economic Society Postgraduate Meeting for their helpful comments. All errors are mine, despite their efforts.

A Appendix

Proofs

Lemma 1

We first show that $\mathbb{C}_{\mathbf{c}}$ satisfies substitutability. 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 $s_i \notin \mathbb{C}_{\mathbf{c}}(S' \cup \{s_i\})$ and $s_i \in \mathbb{C}_{\mathbf{c}}(S' \cup \{s_i, s_j\})$. 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 $\overline{S_1^t} \equiv \overline{S_c}(s_i, S^t(S' \cup \{s_i\}))$ and $\overline{S_2^t} \equiv \overline{S_c}(s_i, S^t(S' \cup \{s_i, s_j\}))$. Since $s_i \notin \mathbb{C}_{\mathbf{c}}(S' \cup \{s_i\})$, $\left|\overline{S_1^{t_i}}\right| \geqslant q_c^t$. Moreover, let q^* be the number of students accepted in the final step of the procedure for $\mathbb{C}_{\mathbf{c}}(S' \cup \{s_i\})$.

Since for any t it is true that $S^t(S' \cup \{s_i\}) \subseteq S^t(S' \cup \{s_i, s_j\})$, it easily follows that $\overline{S_1^{t_i}} \subseteq \overline{S_2^{t_i}}$. 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 $\left|\overline{S_2^{t_i}}\right| \geqslant \left|\overline{S_1^{t_i}}\right| \geqslant q_c^{t_i}$. 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 be 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 $\Psi_{t_j}(S' \cup \{s_i, s_j\}) = \left(\Psi_{t_j}(S' \cup \{s_i\}) \setminus \{s_k\}\right) \cup \{s_j\}$ for some $s_k \in S^{t_j}(S' \cup \{s_i\})$. In the former case:\myCF

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

That is, the set of students that had not yet been 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:

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

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

We now show that $\mathbb{C}_{\mathbf{c}}$ satisfies the law of aggregate demand, which is a property of a choice function where $S'' \subset S' \subseteq S$ implies $|C(S')| \geqslant |C(S'')|$. To see that, first note 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 $|\mathbb{C}_{\mathbf{c}}(S'')| < |\mathbb{C}_{\mathbf{c}}(S')| \leqslant q_c$. If $|S'| \geqslant q_c$ then $|\mathbb{C}_{\mathbf{c}}(S')| = q_c$, and since $\mathbb{C}_{\mathbf{c}}$ never accepts more than q_c students, the property is satisfied.

As shown in Aygün and Sönmez (2012), if a choice function satisfies substitutability and the law of aggregate demand, it also satisfies IRC.

Lemma 3

If μ is pairwise stable then μ is fair with diversity.

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

$t = \tau(s)$ and $|S^t(\mu(c))| < q_c^t$. But then $|S^t(\mu(c) \cup \{s\})| \leqslant q_c^t$ and s is accepted at the step of the procedure for $\mathbb{C}_{\mathbf{c}}$ associated with t. Contradiction with $s \notin \mathbb{C}_{\mathbf{c}}(\mu(c) \cup \{s\})$.

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 $s \notin \mathbb{C}_{\mathbf{c}}(\mu(c) \cup \{s\})$ and $s \succ_c s'$, s' isn't accepted in the step associated with t of the procedure, otherwise s would also be accepted. The same holds for the final step. But then $s' \notin \mathbb{C}_{\mathbf{c}}(\mu(c) \cup \{s\})$, which implies that $|\mathbb{C}_{\mathbf{c}}(\mu(c) \cup \{s\})| < q_c$, contradicting lemma 2.

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

If μ is fair with diversity then μ is pairwise stable.

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

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

μ is blocked by a school s, that is, $\mu(c) \neq \mathbb{C}_{\mathbf{c}}(\mu(c))$. Since $\mu(c) \subseteq \mathbb{C}_{\mathbf{c}}(\mu(c))$ and

$\mu(c) \neq \mathbb{C}_{\mathbf{c}}(\mu(c))$, $\mu(c) \subset \mathbb{C}_{\mathbf{c}}(\mu(c))$ and thus $|\mu(c)| < q_c$. Since μ is fair with diversity, all students in $\mu(c)$ are acceptable to c and by Lemma 2 $\mu(c) = \mathbb{C}_{\mathbf{c}}(\mu(c))$, which is a contradiction.

μ is blocked by a student-school pair (s, c), that is, $\mu(s) \neq c$, $c \succ_s \mu(s)$ and $s \in \mathbb{C}_{\mathbf{c}}(\mu(c) \cup \{s\})$. Since μ is fair with diversity, $|\mu(c)| = q_c$ and s doesn't justifiably demand a seat in c. We will show that s couldn't be accepted in any step of the procedure for $\mathbb{C}_{\mathbf{c}}(\mu(c) \cup \{s\})$. Since $|\mu(c) \cup \{s\}| > q_c$, s cannot be accepted in step 0 of the procedure for $\mathbb{C}_{\mathbf{c}}$. Let $t = \tau(s)$. In order for s to be accepted in the step associated with t, then either $|S^t(\mu(c))| < q_c^t$ or there is a student $s' \in S^t(\mu(c))$ such that $s \succ_c s'$. Both possibilities contradict μ 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 $s \succ_c s'$. Notice that all students accepted in the final step are in "excess" of their diversity objectives, that is, if $t' = \tau(s')$, $\left|A^{t'}(\mu(c))\right| > q_c^{t'}$. But then we have a contradiction with the third item in the definition cited above.

Lemma 4

This proof consists of reproducing the steps in the same result given in Roth (1984) under the more general assumption that the primitives are choice functions that satisfy substitutability and IRC instead of choice functions derived from preferences over sets of students.

Proof. We make use of the following results in Roth (1984), which remain valid without assuming IRC:

Lemma A.1. *(Roth, 1984) Let $s^* \in S_1$ and $s^* \in \mathbb{C}_{\mathbf{c}}(S_1 \cup S_2)$. Then $s^* \in \mathbb{C}_{\mathbf{c}}(S_2 \cup \{s^*\})$.*

Proposition A.1. *(Roth, 1984)Offers remain open: for every school c, if $s \in \mathbb{C}_{\mathbf{c}}(\mathbb{S}_c(k-1))$ and is not rejected by student s in step $k-1$, then $s \in \mathbb{C}_{\mathbf{c}}(\mathbb{S}_c(k))$.*

Proposition A.2. *(Roth, 1984)Rejections are final: If s rejects school c at step k, then for any $p \geqslant k$ student s would reject another proposal from c. In other words, if $s(p)$ is the set of schools that propose to student s at step p, she would never choose c out of $\{\{s\} \cup s(p) \cup \{c\}\}$.*

For the next steps of the proof, however, in order to not have to assume that the choice functions are derived from strict preferences over sets of students, we derive the results by assuming that the choice function satisfies IRC. We will use the following lemma:

Lemma A.2. *Let C satisfy IRC and let $X \subseteq S$ and $Y = C(X)$. Then, for any $Z \subseteq X \backslash Y$, $C(Y \cup Z) = C(Y)$.*

Proof. Let $\overline{Z} \equiv X \backslash (Z \cup Y)$, $Z = \{z_1, z_2, \ldots, z_n\}$ and $\overline{Z} = \{\overline{z}_1, \overline{z}_2, \ldots, \overline{z}_m\}$. Then $C(X) = C\left(Y \cup \overline{Z} \cup Z\right) = C(Y \cup \{\overline{z}_1, \overline{z}_2, \ldots, \overline{z}_{m-1}, z_1, z_2, \ldots, z_n\} \cup \{\overline{z}_m\})$. Since $\overline{z}_m \notin C(X)$, by IRC $C\left(Y \cup \overline{Z} \cup Z\right) = C(Y \cup \{\overline{z}_1, \overline{z}_2, \ldots, \overline{z}_{m-1}, z_1, z_2, \ldots, z_n\})$. Repeating this step for each element of $\overline{Z}$, we get that $C(X) = C(Y \cup Z)$. And repeating now for the elements in Z, we get $C(Y \cup Z) = C(Y)$. □

Proposition A.3. *The outcome of the school-proposing deferred acceptance procedure above is (pairwise) stable.*

Proof. Let μ^C be the outcome of the procedure. First, note that for every student s, $\mu^C(s) \succeq_s s$, otherwise $\mu^C(s)$ would have been rejected by s. Now suppose that $\mathbb{C}_{\mathbf{c}}\left(\mu^C(c)\right) \neq \mu^C(c)$. Then there is s such that $s \in \mathbb{C}_{\mathbf{c}}(\mathbb{S}_c(T))$ but $s \notin \mathbb{C}_{\mathbf{c}}\left(\mu^C(c)\right)$. But since $\mu^C(c) \subseteq \mathbb{S}_c(T)$ this would violate substitutability of $\mathbb{C}_{\mathbf{c}}$, and therefore $\mathbb{C}_{\mathbf{c}}\left(\mu^C(c)\right) = \mu^C(c)$. Now suppose that there is a student s and school c such that $c \succ_s \mu^C(s)$ and $s \in \mathbb{C}_{\mathbf{c}}\left(\mu^C(c) \cup \{s\}\right)$. By the assumption and propositions 1 and 2, student s didn't reject any proposal from school c, and therefore $s \in \mathbb{S}_c(T)$. Denote $R^T \equiv \mathbb{S}_c(T) \setminus (\mathbb{C}_{\mathbf{c}}(\mathbb{S}_c(T)) \cup \{s\})$. Then $\mathbb{S}_c(T) = \mu^C(c) \cup R^T \cup \{s\}$, and therefore $\mu^C(c) = \mathbb{C}_{\mathbf{c}}\left(\mu^C(c) \cup R^T \cup \{s\}\right) = \mathbb{C}_{\mathbf{c}}\left(\mu^C(c)\right)$. Since students in R^T are rejected, by IRC $\mathbb{C}_{\mathbf{c}}\left(\mu^C(c) \cup \{s\}\right) = \mathbb{C}_{\mathbf{c}}\left(\mu^C(c)\right)$. Contradiction with $s \in \mathbb{C}_{\mathbf{c}}\left(\mu^C(c) \cup \{s\}\right)$. □

Proposition A.4. *The outcome of the procedure above, μ^C, is school-optimal in the sense that, for each school c and every stable outcome μ, $\mu^C(c) = C_c\left(\mu^C(c) \cup \mu(c)\right)$.*

Proof. Suppose not. Then there exists a stable outcome μ^* and a school c such that $\mu^C(c) \neq \mathbb{C}_{\mathbf{c}}\left(\mu^C(c) \cup \mu^*(c)\right)$. We first show that there is at least one student $s \in \mathbb{C}_{\mathbf{c}}\left(\mu^C(c) \cup \mu^*(c)\right)$ such that $s \notin \mu^C(c)$ and $s \in \mu^*(c)$. If that isn't the case, since $\mu^C(c) \neq \mathbb{C}_{\mathbf{c}}\left(\mu^C(c) \cup \mu^*(c)\right)$, it must be that $\mathbb{C}_{\mathbf{c}}\left(\mu^C(c) \cup \mu^*(c)\right) \subsetneq \mu^C(c)$. But then all students in $\mu^*(c)$ are rejected and by Lemma A.2 $\mathbb{C}_{\mathbf{c}}\left(\mu^C(c) \cup \mu^*(c)\right) = \mathbb{C}_{\mathbf{c}}\left(\mu^C(c)\right) = \mu^C(c)$, where the second equality is proved in proposition A.3 and therefore constitutes a contradiction.

Next, we will show that no student will reject an *achievable* school during the deferred acceptance procedure. We say that school c is *achievable* to student s if there exists a stable matching μ' where $\mu'(s) = c$. The proof is by induction on the steps of the deferred-acceptance procedure. By induction assumption, up to step $k-1$ no student rejected any achievable school. Now suppose that student s rejects school c, which is achievable for her, in favor of another school c'. It must then be that $c' \succ_s c$. Moreover, since school c is achievable to s, there exists at least on stable matching μ' in which $\mu'(s) = c$. Since up to step $k-1$ no student rejected an achievable school, no student in $S \backslash \mathbb{S}_{c'}(k)$ can be in $\mu'(c')$ and therefore $\mu'(c') \subset \mathbb{S}_{c'}(k)$. But since student s rejected school c in favor of c' at step k, it must be that $s \in \mathbb{C}_{c'}(\mathbb{S}_{c'}(k))$. Since μ' is stable, $\mu'(c) = \mathbb{C}_{c'}(\mu'(c))$. We have, therefore, $s \notin \mathbb{C}_{c'}(\mu'(c))$, $\mu'(c') \subset \mathbb{S}_{c'}(k)$ and $s \in \mathbb{C}_{c'}(\mathbb{S}_{c'}(k))$, which is a contradiction with $\mathbb{C}_{c'}$ satisfying substitutability.

Given that, it must then be that all students who have c as achievable remain available until the last step of the algorithm and therefore $\mu^*(c) \subset \mathbb{S}_c(T)$. Since $\mu^C(c) = \mathbb{C}_{\mathbf{c}}(\mathbb{S}_c(T)) \subset \mathbb{S}_c(T)$, we can rewrite this as:

$$\mu^C(c) = \mathbb{C}_{\mathbf{c}}\left(\mu^C(c) \cup \mu^*(c) \cup \left(\mu^*(c) \backslash \mu^C(c)\right) \cup \left(\mathbb{S}_c(T) \backslash \mu^C(c)\right)\right)$$

By Lemma A.2:

$$\mathbb{C}_{\mathbf{c}}\left(\mu^C(c) \cup \mu^*(c) \cup \left(\mu^*(c) \backslash \mu^C(c)\right) \cup \left(\mathbb{S}_c(T) \backslash \mu^C(c)\right)\right) = \mathbb{C}_{\mathbf{c}}\left(\mu^C(c) \cup \mu^*(c)\right)$$

That is, $\mu^C(c) = \mathbb{C}_{\mathbf{c}}\left(\mu^C(c) \cup \mu^*(c)\right)$ which is a contradiction with our initial assumption. □

□

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, thus leading to an assignment that fully implements diversity. Suppose, for contradiction, that schools' diversity objectives mirror the population distribution (so that for every $t_i \in T$ $q_c^{t_i} = n_1$) but the assignment μ generated by the SPDiv 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 ℓ in the deferred acceptance procedure of the SPDiv 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 ℓ 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 ℓ.) 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 ℓ. 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 ℓ is greater than $(n_1 - 1)\, n_2 k$. Contradiction with ℓ 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 $\succ_c$ 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 $\mathbb{C}_{\mathbf{c}}$ 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_i'. By the end of step 2, at least $2q_c$ students are accepted, since the school mentioned in step 1 will still accept 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 μ be the assignment generated by the DAMR mechanism. By Lemma 6, μ is fair with diversity. Therefore μ 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 $c \in C_1 \cup \cdots \cup C_{a-1}$. We will prove by induction in the sets $C_1, \ldots, C_{a-1}$ when $t > 1$ since the case $a = k = 1$ is trivial.

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

Let n_3 be the integer such that $n_3 = |C_1| = \cdots = |C_a|$. Since $|C| = n_2 k$, $n_3 a = n_2 k$. And since $a \geqslant k$, $n_3 \leqslant n_2$. We must first show that there are at least $n_1 n_3$ 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 $|S^t(S)| = n_1 n_2 k \geqslant n_1 n_3 k$ and $k > 1$, then there is a student s' such that $\tau(s') = t$ and $\mu(s') \notin C_1$. But then $c' \succ_{s'} \mu(s')$ 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 $n_1 n_3$ students of each type t in $\bigcup_{c \in C_1} S^t(\mu(c))$, there are at least $n_1 n_3 k$ students in $\bigcup_{t \in T} \bigcup_{c \in C_1} S^t(\mu(c))$. Since $q_c = n_1 k$ and for any i, $|C_i| = n_3$ it follows that there are *exactly* $n_1 n_3$ students of each type t in $\bigcup_{c \in C_1} S^t(\mu(c))$.

Suppose now that there is a school $c' \in C_1$ such that $\mu(c)$ doesn't enable diversity at c'. Then there is a type $t \in T$ 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, $c' \succ_s \mu(s)$, implying that s justifiably demands a seat in c', a contradiction.

Step k^*: by induction assumption, for every $i \in \{1, \ldots, k^* - 1\}$ and $c \in C_i$, $\mu(c)$ 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 $c' \in C_{k^*}$, that student is assigned in μ to some school in $C_{k'}$, where $k' > k^*$, implying that $c' \succ_s \mu(s)$.

Note, however, that when $k^* = a$ this argument cannot be made any longer, that is, if there is a school $c' \in C_k$ such that $|S^t(\mu(c'))| < n_1$, there may not exist a student s of type t such that $c' \succ_s \mu(s)$.

Theorem 3

Remember that the choice function $\mathbb{C}_{\mathbf{c}}$ satisfies the law of aggregate demand (shown in the proof of Lemma 1.) By proposition 4 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.

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

$\succ_{c_1}$: s_1 s_2
$\succ_{c_2}$: s_2 s_1
$\succ_{s_1}$: c_2 c_1
$\succ_{s_2}$: c_1 c_2
$q_{c_1} = q_{c_2} = 1$, $q_{c_1}^T = \left(q_{c_1}^{t_1}\right) = (1)$ and $q_{c_2}^T = \left(q_{c_2}^{t_1}\right) = (0)$

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

Theorem 4

The proof of this theorem consists of showing that the SPDiv 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 $\{(\mathcal{M}^n)_{\mathbb{N}}, G\}$ to be SP-L.

Agents (students) belong to groups h in a finite set H. A semi-anonymous mechanism is defined as $\{(\Psi^n)_{\mathbb{N}}, (G_h)_{h\in H}\}$, 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 $(\Psi^n)_{\mathbb{N}}$ are functions

$$\Psi^n : G^n \to \Delta(X_0^n)$$

where X_0^n are feasible allocations for the economy n. Agents in subgroup h are restricted to playing strategies in G_h. It is easy to see that the SPDiv 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 playing strategies in G_h, and therefore SPDiv is semi-anonymous in this setting.

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

Definition 11. (Azevedo and Budish, 2013) A direct semi-anonymous mechanism $\{(\Psi^n)_{\mathbb{N}}, (G_h)_{h\in H}\}$ is **envy-free but for tie-breaking** if for each n there exists a function $x^n : (G \times [0,1])^n \to \Delta(X_0^n)$, symmetric over its coordinates, such that

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

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

$$u_{g_s}\left[x_s^n\left((g_i, \ell_i)_{i\in S^n}\right)\right] \geqslant u_{g_s}\left[x_{s'}^n\left((g_i, \ell_i)_{i\in S^n}\right)\right]$$

In other words, if students s and s' belong to the same subgroup and $\ell_s \geqslant \ell_{s'}$, then s doesn't envy the outcome of student s'. Notice that the matching function $\mu^n\left((g_i,\ell_i)_{i\in S^n}\right)$ 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 SPDiv 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 $\ell_s \geqslant \ell_{s'}$, $g_s = g_{s'}$ but $u_{g_s}\left[x_s^n\left((g_i,\ell_i)_{i\in S^n}\right)\right] < u_{g_s}\left[x_{s'}^n\left((g_i,\ell_i)_{i\in S^n}\right)\right]$. Let $c = x_s^n\left((g_i,\ell_i)_{i\in S^n}\right)$ and $c' = x_{s'}^n\left((g_i,\ell_i)_{i\in S^n}\right)$. In terms of student s's preferences, $c' \succ_s c$. Since s and s' are in the same subgroup and $\ell_s \geqslant \ell_{s'}$, then $s \succ_{c'} s'$, 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 ℓ uniformly at random as described in section 5.1 we conclude, therefore, that the induced random mechanism $\{(\mathcal{M}^n)_{\mathbb{N}}, (G_h)_{h\in H}\}$ is *envy-free but for tie-breaking* and *SP-L*.

Specification of the alternative mechanism

Deferred Acceptance Minority Reserves (DAMR)

The extension of the DAMR mechanism in Hafalir et al. (2013) for multiple types consists of applying the student-proposing deferred acceptance mechanism when schools' choice function is $\mathbb{C}_{\mathbf{c}}$, 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 $\mathbb{C}_{\mathbf{c}}(S_c^1)$ and rejects the rest, if any.

$\vdots$

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_c^k) and students admitted tentatively at step k^*-1 $(\mathbb{C}_{\mathbf{c}}\left(S_c^{k^*-1}\right))$. Each school c accepts all students in $\mathbb{C}_{\mathbf{c}}\left(\mathbb{C}_{\mathbf{c}}\left(S_c^{k^*-1}\right)\cup S_c^{k^*}\right)$ 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.