EFFECTIVE AFFIRMATIVE ACTION IN SCHOOL CHOICE

ISA E. HAFALIR, M. BUMIN YENMEZ, AND MUHAMMED A. YILDIRIM

Abstract. The prevalent affirmative action policy in school choice limits the number of majority students to increase minority representation. Nevertheless, this policy may be detrimental to minority students. Instead, we introduce an alternative policy that gives preferential treatment to minorities. We compare the welfare effects of these policies under the deferred acceptance and the top trading cycles algorithms. The deferred acceptance algorithm with preferential treatment Pareto dominates the one with majority quotas. Under simulations, minorities are on average better off with the preferential treatment while adverse effects on majorities are mitigated. Our main results carry over to multiple student types.

1. Introduction

Affirmative action is a popular albeit controversial scheme that is implemented to close socio-economic gaps that exist between groups as a result of historic discrimination. To this end, it involves policies designed to increase the representation of some groups in public areas such as employment, education, and business contracting. To determine the underrepresented groups, it may take into consideration factors including race, religion, color, sex, caste, neighborhood, economic status, or national origin.

Since its introduction in the Unites States, affirmative action has been a source of debate in philosophy, law, and economics.[1] Holzer and Neumark (2000) provide a review of the economics literature and assess affirmative action from an efficiency-performance perspective. They point out that there is no consensus on whether affirmative action policies result in efficiency gains or losses. They conclude that affirmative action seems to offer significant redistribution of welfare toward women and minorities with relatively small efficiency consequences. Loury and Garman (1993) and Arcidiacono (2005) consider the effects of affirmative action in higher education and elucidate the importance of these policies for the

Date: April 26, 2011.

Hafalir and Yenmez are affiliated with the Tepper School of Business, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213; Yenmez is also with Microsoft Research New England, 1 Memorial Drive, Cambridge, MA 02143; Yildirim is with Center for International Development, Harvard University, 79 John F. Kennedy Street, Cambridge, MA 02138. Emails: isaemin@cmu.edu, byenmez@andrew.cmu.edu, and muhammed_yildirim@hks.harvard.edu.

[1]In fact, affirmative action is ubiquitous: various forms of it are present all over the world. See Sowell (2004) for a review of affirmative action in different countries.

decision making process of minority students. More recently, Bertrand, Hanna and Mullainathan (2010) examine an affirmative action program for lower-caste groups in Indian engineering colleges. They show empirically that affirmative action benefits the financially disadvantaged.[2]

Although affirmative action can be seen in many public areas, we study it in the context of public school admissions, where the goal is to maintain a racial and ethnic balance at schools by giving underrepresented groups (usually minorities) higher chances to attend better schools. Many of the minorities who are targets of affirmative action policies live together in isolated, economically challenged neighborhoods that lack good schools. The better schools tend to be located in the wealthier neighborhoods, increasing the chances of wealthier students, who are often majorities, to attend those schools. To circumvent this shortcoming, some school districts employ affirmative action policies that impose racial quotas (e.g., historically in Seattle (WA), Louisville (KY), Minneapolis (MN), and White Plains (NY)). On the other hand, some school districts employ affirmative action because of court orders enforcing desegregation (e.g., historically in Boston (MA), St. Louis (MO), and Kansas City (MO)).

It is of great importance that we understand the social and economic effects of affirmative action policies. Unfortunately, the consequences of these policies receive surprisingly little attention (Sowell, 2004). In very general settings, it is nearly impossible to assess the overall efficiency or welfare effects of affirmative action policies (Holzer and Neumark, 2000). There are numerous reasons for this. To name just a few, markets may be decentralized, or the admission and affirmative action policies may not be clear. By contrast, public school admissions are increasingly handled in a centralized manner where students submit an ordered preference list of schools and school priorities are fixed by school policies, the so-called *school choice problem*. Moreover, the affirmative action policies in the school choice setting are transparent, making it close to an ideal environment for studying the welfare effects of these policies.

We build on the work of Abdulkadiroğlu and Sönmez (2003), who approach the school choice problem from a mechanism-design perspective.[3] They illustrate that some popular mechanisms used in practice have shortcomings and propose as alternatives two celebrated algorithms, the *student-proposing deferred acceptance algorithm* (DA) and the *top trading*

[2]In India, affirmative action policies have been used since the 1930s and there is an intense debate over them. In May 2006, the government announced a plan to extend reservations of low-caste groups in universities, which resulted in massive protests (http://www.time.com/time/world/article/0,8599,1198102,00.html). For a comparison of affirmative action in the United States and India, see Deshpande (2006).

[3]There is a large literature on matching theory and its applications to real-life markets including school choice. We refer the reader to Roth and Sotomayor (1990) for background reading in matching, and three excellent reviews for recent applications (Roth, 2008; Sönmez and Ünver, 2009; Pathak, 2011).

cycles algorithm (TTC).[4] DA, introduced by Gale and Shapley (1962), produces *stable* outcomes and assigns the best outcome among all stable outcomes to one side of the market and the worst to the other side. In the school choice context, stability requires that each student prefers her assignment to her outside option and that there is no school-student pair (c, s) such that s prefers c to her assignment and that either c has an empty seat or that there exists a student assigned to c who has a lower priority at c than s.[5] Moreover, the student-proposing deferred acceptance algorithm is *group strategy-proof*, i.e., there exists no group of students who can jointly submit a modified preference list and all be better off (Dubins and Freedman, 1981). TTC was first studied by Shapley and Scarf (1974), who attribute it to David Gale. TTC is *Pareto efficient*, hence one cannot make any student better off without hurting others. Moreover, it is also group strategy-proof (Bird, 1984). Therefore, the choice between these two algorithms boils down to whether one prefers Pareto efficiency or stability. If a school district puts more weight on Pareto efficiency, then they should implement TTC; if they do not want to violate stability, then DA is the right choice.[6]

Abdulkadiroğlu and Sönmez (2003) also model a simple affirmative action policy with racial and ethnic quotas, the so-called *controlled choice problem*, and show that modified versions of the two aforementioned mechanisms maintain their desirable properties. Subsequently, Abdulkadiroğlu (2005) considers the same affirmative action policies where schools can have more general preferences rather than having an ordered list. He shows that controlled choice constraints induce substitutable preferences and that the student-proposing deferred acceptance algorithm is strategy-proof for students if school preferences satisfy a responsiveness condition.

In a recent paper, Kojima (2010) investigates the consequences of affirmative action on students' welfare. In a setup where there are two student types (minority and majority) and schools have quotas for majority students, he shows that affirmative action policies may hurt minority students, the purported beneficiaries. To be more explicit, he finds examples in which all minority students are made worse off because of the affirmative action policies for both the student-proposing deferred acceptance algorithm and the top trading cycles algorithm, and he concludes that caution should be exercised when applying such policies.

[4]Kesten (2006) shows that these two mechanisms are the same if and only if school priorities are acyclic. Acyclicity is a strong condition and usually not satisfied. Haeringer and Klijn (2009) study a preference revelation game when students can submit limited lists and show that both mechanisms may have equilibria that produce *unstable* matchings.

[5]The second property is also called *no justified-envy* in the school choice context.

[6]Kesten (2010) recognizes the efficiency loss caused by DA and proposes a modified algorithm where students give up their priorities in certain schools to correct for the loss. Similarly, Erdil and Ergin (2008) introduce a new mechanism to improve the welfare losses created by random breaking of ties in priorities caused by DA. In contrast, Kesten and Ünver (2009) approach the problem from an *ex-ante* perspective instead of randomly breaking ties.

The reason that a quota for majority students can have adverse effects on minority students is simple. Consider a situation in which a school c is mostly desired by majorities. Then having a majority quota for c decreases the number of majority students that can be assigned to c even if there are empty seats.[7] This, in turn, increases the competition for other schools and thus can make the minority students worse off. In this paper, we circumvent this inefficiency caused by majority student quotas by offering *minority student reserves*. More specifically, schools assign minority reserves such that if the number of minority students in a school is less than its minority reserves, then any minority is preferred to any majority in that school. If there are not enough minority students to fill the reserves, majority students can be admitted to fill up that school's reserved seats. Minority reserves can also be interpreted as majority quotas, but with a big difference: the number of majority students can be more than its allotted share, which is the capacity of the school less the minority reserves, as long as there are no minority students who veto this match.[8] To study the effects of affirmative action with minority reserves policies in the school choice context, we first adapt the deferred acceptance and the top trading cycles algorithms to our model and then prove that each algorithm preserves its aforementioned desirable properties.

Recent papers by Ehlers (2010) and Abdulkadiroğlu (2010*a*) consider a more general controlled choice problem in which each school has hard floors and ceilings for multiple types of students. We also consider this environment, but instead of considering these floors and ceilings as hard bounds, we regard them as regulatory boundaries of preferential treatment, i.e., soft bounds. When there are two student types, say, minority and majority, the affirmative action with majority quotas is a special case of the hard-bound policy, whereas the affirmative action with minority reserves is an instance of soft bounds. We also study both of these generalized policies under the deferred acceptance and the top trading cycles algorithms.

Our main results are as follows: First, we show that for any stable matching under the *affirmative action with majority quotas* policy, there exists a stable matching under the corresponding *affirmative action with minority reserves* policy that is better for all students (Theorem 1). Next, we prove that the student-proposing deferred acceptance algorithm (DA) with minority reserves is never strictly Pareto dominated by DA with *no affirmative*

[7]In fact this is not only a theoretical possibility but also a reality. A parent in Louisville (KY) sued a school district exactly because of just such a situation: "There was room at the school. There were plenty of empty seats. This was a racial quota" (http://abcnews.go.com/Politics/SupremeCourt/story?id=2693451).

[8]Another way of interpreting minority reserves is as "soft" quotas. Jencks (1992) suggests the use of soft quotas in a general context where he defines soft quotas to be targets that institutions try to reach but inevitably may fail. However, Fryer (2009) states that when the auditors have imperfect information about the hiring or admission process, soft quotas or goals become hard quotas. But in the school choice context, the process is transparent and priorities of both students and schools can be accessed by an auditor. Hence, implementation of soft quotas would not lead to hard quotas in the school choice problem.

action for minority students (Theorem 2).[9] On the other hand, we show by an example that it can be weakly Pareto dominated by minority students (Example 1). Still, we prove that if all schools and all students have the same priorities/preferences, then there exists a unique stable matching under each policy and the stable matchings under minority reserves and majority quotas Pareto dominate the stable matching under no affirmative action for minority students (Proposition 2). Furthermore, we prove that if minority reserves for all schools are greater than the number of minority students assigned to those schools in DA with no affirmative action, then DA with minority reserves Pareto dominates DA with no affirmative action (Proposition 4).[10]

We then analyze the performance of these three policies in the top trading cycles algorithm (TTC). Similar to our result for the deferred acceptance algorithm, we find that TTC with minority reserves is never strictly Pareto dominated by TTC with no affirmative action for minority students (Theorem 3). However, it turns out that there is no Pareto dominance relationship between TTC with minority reserves and majority quotas, and TTC with minority reserves and no affirmative action (Proposition 7).

On top of our theoretical results, we devise computer simulations to quantify the differences between outcomes of the aforementioned affirmative action policies by examining how much better/worse off both minorities and majorities are in comparison to other policies. In our simulations, we allow for correlations between student preferences over schools and correlations between school priorities over students. The simulations indicate that, on average, (i) minority reserves make minorities better off (but can also make majorities worse off) than no affirmative action, in both DA and TTC; (ii) DA with minority reserves not only Pareto dominates DA with majority quotas, but also benefits both minorities and majorities significantly; (iii) majority quota-based mechanisms are very sensitive to quota size, especially for majority welfare, whereas minority reserve-based mechanisms moderate the adverse effects of affirmative action policies on majorities; (iv) TTC with minority reserves results in better matchings for minorities than TTC with majority quotas; and (v) students on average prefer TTC over DA for all affirmative action policies.

In the general case with multiple student types, a stable matching may not exist with hard floor and ceiling bounds (Ehlers, 2010) . On the contrary, DA with soft bounds always yields a stable matching, which Pareto dominates any stable matching with hard bounds if it exists (Theorem 1′). Moreover, increasing the floor or ceiling of any student type cannot

[9]This is in contrast with the result of Kojima (2010) that all minorities can be hurt by an affirmative action policy with majority quotas.

[10]This result shows the importance of choosing minority reserves carefully. Although minorities can be made weakly worse off by affirmative action, if the school districts use past data to figure out what the matching would be without affirmative action, then by making sure that schools have at least that much reserve for minority students, they can guarantee that all minority students would be made better off by minority reserves.

make every member of this group worse off under DA or TTC with soft bounds (Theorem $2'$ and $3'$).

The rest of the paper is organized as follows. Section 2 sets up the model and introduces formal definitions of different affirmative action policies. Section 3 defines the deferred acceptance algorithm with minority reserves and compares outcomes of the algorithm under different policies. Similarly, Section 4 adapts the top trading cycles algorithm to minority reserves. Section 5 describes our simulation model and presents the simulation results. Section 6 extends our results to more general affirmative action policies with multiple student types, floors, and ceilings. Section 7 concludes. All proofs are in the Appendix.

2. **Model**

Let S and C be finite and disjoint sets of students and schools. For each student $s \in S$, $\succ_s$ is a strict preference relation over $C \cup \{s\}$ where s denotes the outside option.[11] School c is ***acceptable*** to student s if $c \succ_s s$. The list of preferences for a group of students S' is denoted by $\succ_{S'} \equiv (\succ_s)_{s \in S'}$. For each school $c \in C$, $\succ_c$ is a strict priority order over S. Following Kojima (2010), students can be one of two types, minority or majority. The set of minority students is denoted by S^m, and the set of majority students is denoted by S^M, so $S = S^m \cup S^M$. For all $c \in C$, q_c is the capacity of c or the number of seats in c. There are enough seats for all students, so $\sum_{c \in C} q_c \geq |S|$. The vector of capacities is denoted by q. A ***school choice problem*** or simply a ***problem*** is a quadruple $\langle C, S, (\succ_i)_{i \in C \cup S}, (q_c)_{c \in C} \rangle$.

A ***matching*** μ is a mapping from $C \cup S$ to the subsets of $C \cup S$ such that

(1) $\mu(s) \in C \cup \{s\}$ for every $s \in S$,
(2) $\mu(c) \subseteq S$ and $|\mu(c)| \leq q_c$ for every $c \in C$, and
(3) $\mu(s) = c$ if and only if $s \in \mu(c)$ for every $c \in C$ and $s \in S$.

A matching μ ***Pareto dominates*** matching ν if $\mu(s) \succeq_s \nu(s)$ for all $s \in S$ and $\mu(s) \succ_s \nu(s)$ for at least one $s \in S$. A matching is ***Pareto efficient*** if it is not Pareto dominated by another matching. Affirmative action policies are implemented to improve the matches of minorities, sometimes at the expense of majorities. Therefore, we also need an efficiency concept to analyze the welfare of minority students. A matching μ ***Pareto dominates*** matching ν ***for minorities*** if $\mu(s) \succeq_s \nu(s)$ for all $s \in S^m$ and $\mu(s) \succ_s \nu(s)$ for at least one $s \in S^m$. A matching is ***Pareto efficient for minorities*** if it is not Pareto dominated for minorities by another matching.

A matching is ***stable*** if it is individually rational and does not have a blocking pair. ***Individual rationality*** is the same regardless of the affirmative action policy employed and can be defined as $\mu(s) \succeq_s s$, for all $s \in S$. However, whether a pair (c, s) can block a

[11] This could be attending a private school or being home-schooled.

matching or not depends on the affirmative action policy. Below, we define three different affirmative action policies; and for each one we also consider the notion of blocking.

The first affirmative action policy is really the absence of one, or ***no affirmative action***. To be more explicit, schools do not discriminate students based on their types. Therefore, a matching μ does not have a blocking pair if for all $c \succ_s \mu(s)$, we have $|\mu(c)| = q_c$ and $s' \succ_c s$ for all $s' \in \mu(c)$.

The second affirmative action policy is called ***affirmative action with majority quotas*** or simply ***majority quotas***. It is implemented by prohibiting schools to admit more than a certain number of majority students. That is, given a vector of majority quotas $q^M \equiv (q_c^M)_{c\in C}$, a matching μ is feasible with majority quotas if for all c, $|\mu(c) \cap S^M| \leq q_c^M$. Moreover, a matching μ does not have a blocking pair, if for all $c \succ_s \mu(s)$ we have either (i) $|\mu(c)| = q_c$ and $s' \succ_c s$ for all $s' \in \mu(c)$, or (ii) $s \in S^M$, $s' \succ_c s$ for all $s' \in \mu(c) \cap S^M$ and $|\mu(c) \cap S^M| = q_c^M$.

These quotas can make not only the majority students worse off but also the minority students (Kojima, 2010). To avoid this inefficiency, we introduce a new affirmative action policy, which gives priority to minority students up to the reserve numbers. More specifically, each school c is assigned a minority reserve r_c^m such that if the number of minority students admitted to c is less than r_c^m, then any minority applicant is preferred to any majority applicant in c. The vector of minority reserves is denoted by r^m.

Hence, the last affirmative action policy is called ***affirmative action with minority reserves*** or simply ***minority reserves***. For minority reserves, a matching μ does not have a blocking pair if for all $c \succ_s \mu(s)$ then $|\mu(c)| = q_c$ and either (i) $s' \succ_c s$ for all $s' \in \mu(c)$ where either " $s \in S^m$ and $|\mu(c) \cap S^m| \geq r_c^m$ " or " $s \in S^M$ and $|\mu(c) \cap S^m| > r_c^m$ " or (ii) $s \in S^M$, $s' \succ_c s$ for all $s' \in \mu(c) \cap S^M$ and $|\mu(c) \cap S^m| \leq r_c^m$.

Condition (i) describes a situation where (c, s) does not form a blocking pair because c has filled its minority reserves, and c prefers all students in c to s. Whereas in condition (ii) blocking does not happen because s is a majority student, the number of minority students in c does not exceed minority reserves and c prefers all majority students in c to s. Note that in the latter case there can be a minority student s' assigned to c such that c prefers s to s'. If c had no affirmative action, then (c, s) would form a blocking pair.

Throughout the paper we perform welfare comparisons between these affirmative action policies. Whenever we compare the effects of minority reserves r^m and majority reserves q^M, we assume that $r^m + q^M = q$.

A ***matching mechanism*** ϕ (or ***algorithm***) is a mapping from school choice problems into matchings. In a school choice problem $\langle C, S, (\succ_i)_{i\in C\cup S}, (q_c)_{c\in C}\rangle$, we assume that everything is known except $(\succ_s)_{s\in S}$.[12] Therefore, students are the only strategic agents in the

[12]The priority orders of schools are usually determined by a public rule.

problem and can manipulate the mechanism by misreporting their preferences. When other components of the problem are clear, we represent the problem just by $\succ_S$ and the outcome of the mechanism by $\phi(\succ_S)$.

A matching mechanism ϕ is ***strategy-proof*** if for each student s, for any $\succ_S$ there exists no $\succ'_s$ such that $\phi_s(\succ'_s, \succ_{S\setminus\{s\}}) \succ_s \phi_s(\succ_S)$. If a mechanism is strategy-proof, each student finds it optimal to report her preferences truthfully regardless of the preferences of other agents. Similarly, a matching mechanism ϕ is ***group strategy-proof*** if for any group of students $\hat{S} \subseteq S$, for any $\succ_S$ there exists no $\succ'_{\hat{S}}$ such that $\phi_s(\succ'_{\hat{S}}, \succ_{S\setminus\hat{S}}) \succ_s \phi_s(\succ_S)$ for all $s \in \hat{S}$. If a mechanism is group strategy-proof then there exists no group of students who can jointly change their preference profiles to make each student in the group better off.[13] A matching mechanism ϕ is ***Pareto efficient*** if $\phi(\succ_S)$ is Pareto efficient for all $\succ_S$. Finally, a matching mechanism ϕ ***Pareto dominates*** another matching mechanism ψ if for all $\succ_S$ either $\phi(\succ_S) = \psi(\succ_S)$ or $\phi(\succ_S)$ Pareto dominates $\psi(\succ_S)$.

3. Deferred Acceptance Algorithm with Minority Reserves

We adapt the student-proposing deferred acceptance algorithm to our setup when schools have minority reserves.

Step 1: Start with the matching in which no student is matched. Each student s applies to her first-choice school. Each school c first accepts as many as r_c^m minority applicants with the highest priorities if there are enough minority applicants. Then it accepts applicants with the highest priorities from the remaining applicants until its capacity is filled or the applicants are exhausted. The rest of the applicants, if any remain, are rejected by c.

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 and students admitted tentatively from the previous steps. Among these students, school c first accepts as many as r_c^m minority students with the highest priorities if there are enough minority students. Then it accepts students with the highest priorities from the remaining students. The rest of the students, if any remain, are rejected by c. If there are no rejections, then stop.

The algorithm 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 algorithm stops in finite time.

We first show that the above algorithm produces a stable matching that assigns each student to the best outcome among all stable matching outcomes, and is group strategy-proof for students.

[13] Ergin (2002) studies a stronger version of group strategy-proofness.

Proposition 1. *The student-proposing deferred acceptance algorithm with minority reserves produces a stable matching that assigns the best outcome among the set of stable matching outcomes for each student and is group strategy-proof.*

In the proof we show that an equivalent way to implement the deferred acceptance algorithm with minority reserves is to first create a new matching problem with no affirmative action and then apply the original deferred acceptance algorithm to this market. The new problem is created by replicating a school c with minority reserves r_m^c, capacity q_c, and priorities $\succ_c$ by two schools c^1 ("original") with capacity $q_c - r_c^m$ and priorities $\succ_c$; and c^2 ("minority favoring") with capacity r_c^m and priorities $\succ'_c$ where:

$$s \succ'_c s' \iff \begin{cases} s \in S^m & \text{and } s' \in S^M \\ s, s' \in S^m & \text{and } s \succ_c s' \\ s, s' \in S^M & \text{and } s \succ_c s' \end{cases}.$$

For each student s we replace school c with its copies in the same order to get the new preference $\succ'_s$. For example, if $c_1 \succ_s c_2$ then $c_1^2 \succ'_s c_1^1 \succ'_s c_2^2 \succ'_s c_2^1$. Less formally, each student keeps the relative rankings of schools the same and prefers the minority-favoring schools over the originals. Therefore, the student-proposing deferred acceptance algorithm with minority reserves preserves the properties of the original one.

Next, we show that for any stable matching under majority quotas, there exists a stable matching under the corresponding minority reserves which Pareto dominates it.

Theorem 1. *Consider majority quotas q^M and minority reserves r^m such that $r^m = q - q^M$. Take a matching μ that is stable under majority quotas q^M. Then, either μ is stable under minority reserves r^m or there exists a matching which is stable under minority reserves r^m that Pareto dominates μ.*

If μ is stable under minority reserves, then there is nothing to prove. Otherwise, that is, if μ is not stable under minority reserves, then there exists a blocking pair (c, s) such that s is a majority student and c has not filled its capacity yet. Whenever there is school c with empty seats that a student prefers to her current assignment, we execute the following ***improvement algorithm***.

Step 1: For college c defined above, find $S^1 \equiv \{s \in S : c \succ_s \mu(s)\}$. Among the students in S^1 match the best students according to $\succ_c$ up to the capacity. Define μ_1 to be the new matching.

Step k: If there is no school with an empty seat that a student prefers to her match in μ_{k-1}, then stop. Otherwise consider one such school, say c_k. Let $S^k \equiv \{s \in S : c_k \succ_s \mu_{k-1}(s)\}$. Among the students in S^k first match the most-preferred minority students

according to $\succ_{c_k}$ to c_k until the minority reserves are filled or minority students are exhausted. Then match the best students according to $\succ_{c_k}$ if there are more seats and students available. Define μ_k to be the new matching.

The algorithm ends in a finite number of steps since it improves the match of at least one student at every step of the algorithm.[14]

The improvement algorithm produces a stable matching under minority reserves (see the Appendix for the proof) because the starting point is a stable matching under majority quotas. If it starts from an arbitrary matching, then it does not produce a stable matching. Surprisingly, if it starts from the matching in which no agent is previously matched, then it proceeds like the school-proposing deferred acceptance algorithm with the exception that offers are made randomly. Since the order of proposals does not change the outcome of the deferred acceptance algorithm (McVitie and Wilson, 1970), the improvement algorithm starting from the matching in which no agent is matched produces the same outcome as the school-proposing deferred acceptance algorithm.

Remark 1. Theorem 1 is equivalent to the statement that the student-proposing deferred acceptance algorithm with minority reserves Pareto dominates the algorithm with majority quotas. To see this, note that for each affirmative action policy, the student-optimal stable matching Pareto dominates any other stable matching. Therefore, Pareto domination relationship in Theorem 1 holds if and only if it holds for the student-optimal stable matchings under the corresponding policies.

Kojima (2010) shows that using majority quotas may hurt all minority students in some settings. Specifically, in Theorem 1 of his paper, he gives an example in which the only minority student is made strictly worse off by implementing majority quotas. We next show that this is not possible with minority reserves.

Theorem 2. *Consider minority reserves r^m. Let μ^r and μ be the matchings produced by the student-proposing deferred acceptance algorithm with or without minority reserves r^m, respectively, for a given preference profile. Then there exists at least one minority student s such that $\mu^r(s) \succeq_s \mu(s)$.*

The outline of the proof is as follows. Suppose, for contradiction, that $\mu(s) \succ_s \mu^r(s)$ for all $s \in S^m$. If each minority student reports $\mu^r(s)$ as the only acceptable school, then $\mu(s)$ can be shown to be stable with minority reserves r^m. Since the student-proposing deferred acceptance algorithm with minority reserves is student optimal (Proposition 1), $\mu^r(s) \succeq_s \mu(s)$ for all $s \in S^m$. This contradicts the fact that the algorithm is group strategy-proof (Proposition 1).

[14]This algorithm is reminiscent of the algorithms that utilize *stable improvement cycles*; see Erdil and Ergin (2008); Abdulkadiroğlu (2010*b*).

On the other hand, the example below shows that imposing minority reserves can make some minorities worse off while leaving some of them indifferent.

Example 1. Consider the following problem: $C = \{c_1, c_2, c_3\}$, $S^M = \{s_1\}$, and $S^m = \{s_2, s_3\}$. All schools have a capacity of one, $q = (1, 1, 1)$. Students' preferences are

$$\succ_{s_1}: c_1 \ \succ c_3 \ \succ c_2, \ \ \succ_{s_2}: c_3 \ \succ c_1 \ \succ c_2, \ \ \succ_{s_3}: c_1 \ \succ c_2 \ \succ c_3.$$

All schools have the same preferences: $s_1 \succ_c s_2 \succ_c s_2 \succ_c s_3$ for all $c \in C$.

Minority reserves are given by $r^m = (0, 0, 0)$. In this case, the unique stable matching, which is also the outcome of the deferred acceptance algorithm, is

$$\mu(c_1) = s_1, \ \mu(c_2) = s_3, \ \mu(c_3) = s_2.$$

However, when minority reserves are $r^m = (1, 0, 0)$, then the unique stable matching, which is also the outcome of the deferred acceptance algorithm, is

$$\mu'(c_1) = s_2, \ \mu'(c_2) = s_3, \ \mu'(c_3) = s_1.$$

With minority reserves, s_1 gets rejected from c_1 because of the presence of minority reserves at the first step of the algorithm. Then, s_1 applies to c_3 and c_3 rejects s_2 in return. Next, s_2 applies to c_1 and c_1 rejects s_3. Finally, s_3 applies to c_2, which accepts her. Therefore, the introduction of minority reserves creates a rejection chain that makes some minority students worse off. Hence an increase in the minority reserves of c_1 makes s_2 worse off and s_3 indifferent.

Example 1 shows that, in general, having minority reserves does not necessarily improve the outcome for minorities without making further assumptions about minority preferences and/or reserve sizes. In next two subsections we provide two positive results which guarantee that minorities are better off with minority reserves policies. The first one is obtained by considering common preferences/priorities of students/schools, whereas the second one is obtained by considering endogenous reserves.

3.1. **Common Preferences/Priorities.** In reality, students' preferences over schools are correlated, and similarly, schools' priorities are correlated over students. In the next proposition, we consider the extreme case where students have the same preferences over schools and schools have the same priorities over students, and we show that the student-proposing deferred acceptance algorithm with minority reserves Pareto dominates those with no affirmative action and with majority quotas.

Proposition 2. *Consider majority quotas q^M and minority reserves r^m such that $r^m = q - q^M$. If students have the same preferences over schools, and schools have the same priority orders over students, then each affirmative action policy results in a unique stable*

matching. Let μ, μ^r, and μ^q be the stable matchings with no affirmative action, minority reserves r^m, and majority quotas q^M respectively for a given preference profile. Then $\mu^r(s) = \mu^q(s) \succeq_s \mu(s)$ for any $s \in S^m$ and $\mu^r(s) \succeq_s \mu^q(s)$ for any $s \in S^M$.

With each affirmative action policy the unique stable matching can be attained by a serial dictatorship of schools: Each school chooses the best students, taking affirmative action policies into account. Since both affirmative action policies favor minorities in the same way when schools are overdemanded, minorities are matched to the same schools with minority reserves and majority quotas. Also, matches of the minority students are at least as good as the schools they are matched with under no affirmative action. The stable matchings with minority reserves and majority quotas can only differ for majority students. This happens when minority students are exhausted at some step of the serial dictatorship. After this step with minority reserves more majority students can be admitted than can be with majority quotas. And this makes majority students better off.

3.2. **Endogenous Reserves.** In the absence of assumptions about agents' preferences and priorities, we can only guarantee that at least one minority student is not going to be worse off in the student-proposing deferred acceptance algorithm if colleges set minority reserves exogenously. However, we now argue that if the reserves are chosen endogenously, all minority students can be made better off. More specifically, (i) if all schools' reserves are smaller than the number of minority students assigned to those schools in a stable matching under no affirmative action, then that stable matching remains stable under minority reserves, and (ii) if all schools' reserves are greater than the number of minority students assigned to those schools in a stable matching under no affirmative action, say μ, then there exists a stable matching under minority reserves that Pareto dominates μ.

Proposition 3. *Suppose that μ is a stable matching under no affirmative action. Let r_c^m be such that $r_c^m \leq |\mu(c) \cap S^m|$ for all c. Then μ is a stable matching under minority reserves r^m.*

The intuition behind this result is simple. Since minority reserves are already filled in each school with μ, if there is any blocking pair (c, s) for μ under minority reserves, then it would also block μ under no affirmative action. On the other hand, if the minority reserves are not filled, then there could be a blocking pair under minority reserves with a minority student since minority reserves give preferential treatment to minorities until they are filled. For this case we establish the following.

Proposition 4. *Suppose that μ is a stable matching under no affirmative action. Let r_c^m be such that $r_c^m \geq |\mu(c) \cap S^m|$ for all c. Then either μ is stable under minority reserves r^m or there exists a stable matching under minority reserves r^m that Pareto dominates μ for minorities.*

If minority reserves exceed the number of minority students in μ, then there could be blocking pairs with minority students. In that case, we apply an algorithm to improve the matches of the minorities by possibly rejecting majorities. At each step of the algorithm we take a school c that is preferred by a minority student to her match. Among such minorities, c admits the most-preferred ones until all minorities are exhausted or minority reserves are filled. Then c rejects an equal number of majority students. At the end of this procedure there exists no school with non-filled minority reserves for which there exists a minority student who prefers that school to her current matching. However, there may be unmatched majority students and schools with empty seats. Then, we apply the improvement algorithm introduced after Theorem 1 to get a stable matching. The details of the proof are given in the Appendix.

As a corollary to Propositions 3 and 4, we have the following:

Corollary 1. *Suppose that μ^r and μ are the matchings produced by the student-proposing deferred acceptance algorithms for a given preference profile with or without minority reserves r^m, respectively, where either $r_c^m \leq |\mu(c) \cap S^m|$ for all c or $r_c^m \geq |\mu(c) \cap S^m|$ for all c. Then μ^r Pareto dominates μ for minorities.*

Therefore, if minority reserves are chosen endogenously by calculating the number of minority students admitted in a stable matching μ under no affirmative action, then the student-optimal stable matching under minority reserves Pareto dominates μ for minorities.

Remark 2. If we set minority reserves to be the capacities for all schools ($r^m = q$) then the student-proposing deferred acceptance algorithm with minority reserves Pareto dominates the student-proposing deferred acceptance algorithm for minorities. This is an exogenous affirmative action policy that makes all minorities better off.

4. **Top Trading Cycles Algorithm with Minority Reserves**

We adapt the top trading cycles algorithm to our setup when schools have minority reserves.

Step 1: Start with the matching in which no agent is matched. If a school has minority reserves then it points to its most preferred minority student; otherwise it points to the most preferred student. Each student points to the most preferred school if there is an acceptable school, and otherwise points to herself. There exists at least one cycle. Each student in any of the cycles is matched to the school she is pointing to (if she is pointing to herself, then she gets her outside option). All students in the cycles and schools that have filled their capacities are removed. If there is no unmatched student, then stop.

Step k: If a school has not filled its minority reserves, then it points to the most preferred minority student if there is any minority student left. Otherwise, it points to the most preferred student. Each student points to the most preferred school if there is an acceptable school, and otherwise points to herself. There exists at least one cycle. Each student in any of the cycles is matched to the school she is pointing to (if she is pointing to herself, then she gets her outside option). All students in the cycles and schools that have filled their capacities are removed. If there is no unmatched student, then stop.

The algorithm terminates in a finite number of steps since there is at least one student matched and removed in any step of the algorithm.

If a school has minority reserves, then it points to minorities until the reserves are filled. Therefore, having minority reserves empowers minorities by facilitating cycles that are otherwise impossible. On the other hand, even if the school points to minority students, it may receive majority students in some cycles.

Proposition 5. *The top trading cycles algorithm with minority reserves is Pareto efficient and group strategy-proof.*

For Pareto efficiency, note that at each step of the algorithm students point to the school they like the most with empty seats. Therefore, any student who is matched at a particular step cannot be made better off without making students who are matched before her worse off. Hence, the algorithm is Pareto efficient. In contrast, the top trading cycles with majority quotas is only *constrained efficient* since quotas add extra feasibility constraints (Abdulkadiroğlu and Sönmez, 2003).

The intuition for group strategy-proofness is as follows. Suppose for contradiction that there exists a group of agents, say $\hat{S}$, who can jointly deviate from their original preferences to obtain better matches for all of them. Consider one of the students, say $\hat{s}$, in $\hat{S}$ who is matched the earliest in the algorithm under truthful reporting among $\hat{S}$. Since none of the students who are matched before $\hat{s}$ are changing their preferences, all of the cycles are going to be the same under the new preference profile and the original one. Therefore, all of the schools that $\hat{s}$ like more than her match under the original preference profile will have filled their capacities under the new preference profile as well. This gives a contradiction.

Next, we compare the top trading cycles algorithm with minority reserves to that with no affirmative action.

Theorem 3. *Suppose that ψ^r and ψ are the matchings produced by the top trading cycles algorithm with or without minority reserves r^m for a given preference profile. Then there exists $s \in S^m$ such that $\psi^r(s) \succeq_s \psi(s)$.*

The proof is by induction on the number of agents. If there exists a minority among the set of students who is matched at the first step of ψ^r, then we are done since she will be matched to her top-choice school. Otherwise, all students matched at the first step of ψ^r, say $\hat{S}$, are majority students. Therefore, all schools, say $\hat{C}$, who are matched at this step must have zero minority reserves. Moreover, in the first step of ψ we see the same matchings. Now, we can look at a smaller problem with $\hat{S}$ removed and the capacities of schools in $\hat{C}$ reduced by one. Both ψ^r and ψ produce the same matching in the smaller problem with what they produce in the larger one. The conclusion follows from this induction argument.

Theorem 3 only tells us that we cannot make all minority students worse off by having minority reserves.[15] However, if each school sets a positive minority reserve size then we obtain a stronger result and guarantee that at least some minority students will be matched with their top-choice schools.

Proposition 6. *Suppose that $r_c^m \geq 1$ for all $c \in C$. Then there exists a minority student who is matched with her top choice school in the top trading cycles algorithm with minority reserves r^m.*

Under this assumption, all schools point to minorities in the first step of the algorithm, so all cycles in this step consist of schools and minority students. These minorities are then matched to their top-choice schools.

It turns out that the top trading cycles algorithm with minority reserves does not Pareto dominate the top trading cycles algorithm with or without majority quotas for minorities. Similarly, the top trading cycles algorithm with or without majority quotas does not Pareto dominate that with minority reserves for minorities.

Proposition 7. *Consider majority quotas q^M and minority reserves r^m such that $r^m = q - q^M$. There exists no Pareto dominance relationship for minorities between the top trading cycles algorithm with minority reserves r^m and the top trading cycles algorithm with or without majority quotas q^M.*

For each pair of mechanisms, we show an example in the Appendix for which one mechanism outcome Pareto dominates the outcome of the other mechanism.

Next, we provide an example in which although all seats are reserved for minorities, there are some minorities who are worse off (than they would be with no affirmative action) because of the minority reserves. This is in contrast with our result for the student-proposing deferred acceptance algorithm (Remark 2).

[15]The corresponding result does not hold for majority quotas. Consider the following example: $C = \{c_1, c_2\}$, $S^M = \{s_2\}$, and $S^m = \{s_1\}$. All schools have a capacity of one, $q = (1, 1)$. Preferences and priorities are given as follows: $c_1 \succ_{s_1} c_2$, $c_2 \succ_{s_2} c_1$, $s_2 \succ_{c_1} s_1$, and $s_1 \succ_{c_2} s_2$. With no affirmative action, both students get their top choices in top trading cycles algorithm. Now consider majority quotas $q^M = (1, 0)$. Then in the top trading cycles algorithm with majority quotas, both students get their second choices, making the only minority student worse-off.

Example 2. Consider the following problem: $C = \{c_1, c_2, c_3\}$, $S^M = \{s_3\}$, and $S^m = \{s_1, s_2\}$. All schools have a capacity of one, $q = (1, 1, 1)$. Students' preferences are

$$\succ_{s_1}: c_2, \quad \succ_{s_2}: c_2 \ \succ c_1 \ \succ c_3, \quad \succ_{s_3}: c_3.$$

Schools' priorities are

$$\succ_{c_1}: s_2, \quad \succ_{c_2}: s_3 \ \succ s_1 \ \succ s_2, \quad \succ_{c_3}: s_1.$$

When minority reserves are $r^m = (0, 0, 0)$, the outcome of the top trading cycles algorithm is

$$\mu(c_1) = s_1, \ \mu(c_2) = s_2, \ \mu(c_3) = s_3.$$

However, when minority reserves are given by $r^m = (1, 1, 1)$, the outcome of the top trading cycles algorithm is given by

$$\mu'(c_1) = s_2,$$
$$\mu'(c_2) = s_1,$$
$$\mu'(c_3) = s_3.$$

Therefore, in this example, one of the minorities (s_2) is worse off because of a minority reserves policy with $r^m = q$.

5. **Simulations**

Our theoretical results show that the *student-proposing deferred acceptance algorithm (DA)* with minority reserves (DA_{MiR}) Pareto dominates *DA* with majority quotas (DA_{MaQ}) (Theorem 1), and is not strictly Pareto dominated by *DA* with no affirmative action (DA_{NAA}) for minority students (Theorem 2). Such Pareto dominance statements cannot be made in-between the *top trading cycles algorithms* (*TTC*) employing minority reserves (TTC_{MiR}), majority quotas (TTC_{MaQ}), or no quotas (TTC_{NAA}). Nevertheless, it is important to quantify how much better/worse each policy makes minorities compared to other policies. Furthermore, it is ultimately desirable to increase the representation of minorities without imposing severe effects on the majority welfare. Therefore, how many majorities improve in their matches and how many drop should also be taken into account while determining which policy to use.

To this end, we devised computer simulations to quantify the differences between outcomes of the aforementioned policies by examining how much better/worse off both minorities and majorities are in comparison to other policies.[16] We defined utility functions for students and schools to get strict preference relations over schools and students, respectively. In real-life school choice problems, some schools are in greater overall demand than others. To

[16]Similar experiments have been employed in the school choice literature; see Chen and Sönmez (2006); Erdil and Ergin (2008).

reflect this phenomenon we allowed for correlations between student preferences. Conversely, schools might also have correlated preferences over students. For instance, in many districts or countries, there are centralized exams that are integral to the school admissions process. Our school utility function takes into account the presence of such correlations.

Suppose there are n students and m schools in the district in which we want to match students to schools. Students are denoted by $s_1, \ldots, s_n$ and schools by $c_1, \ldots, c_m$. Proportion p of the students are minorities. Each school has M seats and allocates proportion r of their seats as minority reserves or proportion $1-r$ as majority quotas. Let Z denote i.i.d normally distributed random variables with zero mean and variance one. We define $Z(c_j)$ $[Z(s_j)]$ to reflect the overall preference of students [schools] for a particular school c_j [a particular student s_j], whereas $Z_{s_i}(c_j)$ $[Z_{c_i}(s_j)]$ is the student- [school-] specific preference distribution over the schools [students]. Initially, we did not assume any differences in terms of preferences between minorities and majorities except the reserve or quota allocations. We can formalize the utility function for student s_i and school c_j as:

$$U_{s_i}(c_j) = \alpha Z(c_j) + (1-\alpha) Z_{s_i}(c_j),$$
$$U_{c_j}(s_i) = \theta Z(s_i) + (1-\theta) Z_{c_j}(s_i),$$

where $\alpha,\ \theta \in [0,1]$ are fixed parameters that set the correlation levels between student preferences and school priorities, respectively.

For each simulation experiment, we set the parameters (n, m, p, M, r, α, θ) and randomly generated the utility functions. We defined the preference order for each student s_i for all pairs of schools (c_j, $c_{j'}$) by using the relation: $c_j \succ_{s_i} c_{j'} \iff U_{s_i}(c_j) > U_{s_i}(c_{j'})$ $\forall j, j' \in 1, \ldots, m$. Similarly, a school's priority order can be determined by comparing utility levels for each student pair $(s_i,\ s_{i'})$: $s_i \succ_{c_j} s_{i'} \iff U_{c_j}(s_i) > U_{c_j}(s_{i'})$, $\forall i, i' \in 1, \ldots, n$. For each set of parameters, we performed 100 simulations to capture representative behavior of different matching models. We implemented all six matching algorithms in PERL and ran more than five million simulations in total to sample throughout the parameter space.

In our first set of simulations, we set the number of students to $n = 1000$, number of schools to $m = 20$, each school size to $M = 50$, and proportion of minority students to $p = 20\%$[17] and varied minority reserve ratio r,[18] α, and θ. Note that the expected ratio of minority students assigned by DA_{NAA} and TTC_{NAA} is equal to p (20%) in each school.

[17]For robustness check, we tried different values for these variables, e.g., $n = 5000, 1200$, $m = 50$, $p = 15\%$, or used variable reserve sizes for each school, and our conclusions were not affected. Moreover, we also ran simulations for $n < m \times M$ and $n > m \times M$, but we did not observe any qualitative changes.

[18]We will only specify minority reserves from this point on; the corresponding majority quotas will be set to $1 - r$.

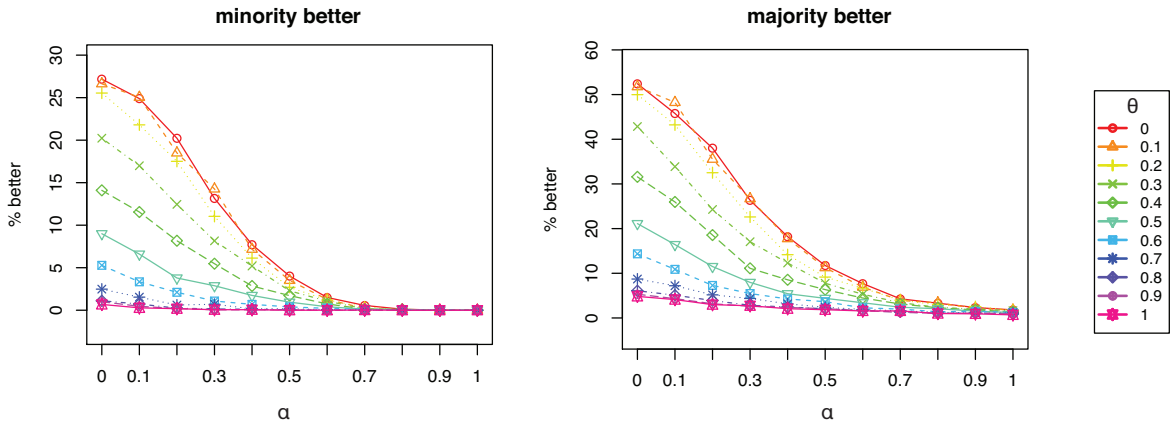


FIGURE 1. *Percent of minorities and majorities who are better off under DA_{MiR} than under DA_{MaQ}.*

Initially, we set $r = 20\%$ and changed α and θ from 0 to 1 in steps of 0.1. We first compared DA_{MiR} to DA_{MaQ}. As a sanity check, our simulations confirmed the Pareto dominance of DA_{MiR} over DA_{MaQ} (Supplementary Figure 1.2).[19] For small values of α and θ, as the level of correlation between school (student) priorities (preferences) increases, the ratio of minority and majority students who are better off under DA_{MiR} decreases (Figure 1). *When neither student preferences nor school priorities are correlated with each other (i.e., $\alpha, \theta = 0$) 27% of minorities and 52% of majorities are better off under DA_{MiR}.* When either school priorities or student preferences are perfectly correlated, both methods give rise to the same assignments for minorities. Under the same settings, DA_{MiR} increases the match quality of $5 - 40\%$ of minorities, while making $3 - 10\%$ of majorities and 0.4% of minorities worse off than under DA_{NAA} (Figure 2). When both $\alpha = \theta = 100\%$, we corroborate the results of Proposition 2, with 40% of minorities being better of under affirmative action policies. Overall, DA_{MaQ} makes $5-40\%$ of minorities better off, while decreasing the match quality of $5 - 60\%$ of majorities compared to DA_{NAA} (Supplementary Figure 1.3). *Most surprisingly, for low levels of α and θ, $\sim$ 20% of minorities are worse off under DA_{MaQ} than under DA_{NAA}, corroborating that the observations of Kojima (2010) are not peculiarities.*

The differences between matches of minorities under different TTC algorithms are almost exclusively α dependent, showing the power bestowed to students by TTC algorithms (Supplementary Figures 1.4-6). TTC_{MiR} increases the match qualities of minority students compared to both TTC_{MaQ} and TTC_{NAA}, more significantly when α is not close to 1. When $\alpha \approx 1$, TTC_{MaQ} makes minorities better off compared to TTC_{NAA} because the probability of reciprocality between choices of students and schools increases, thereby creating cycles and better matches for minority students.

[19] All of the supplementary figures are available in a separate document.

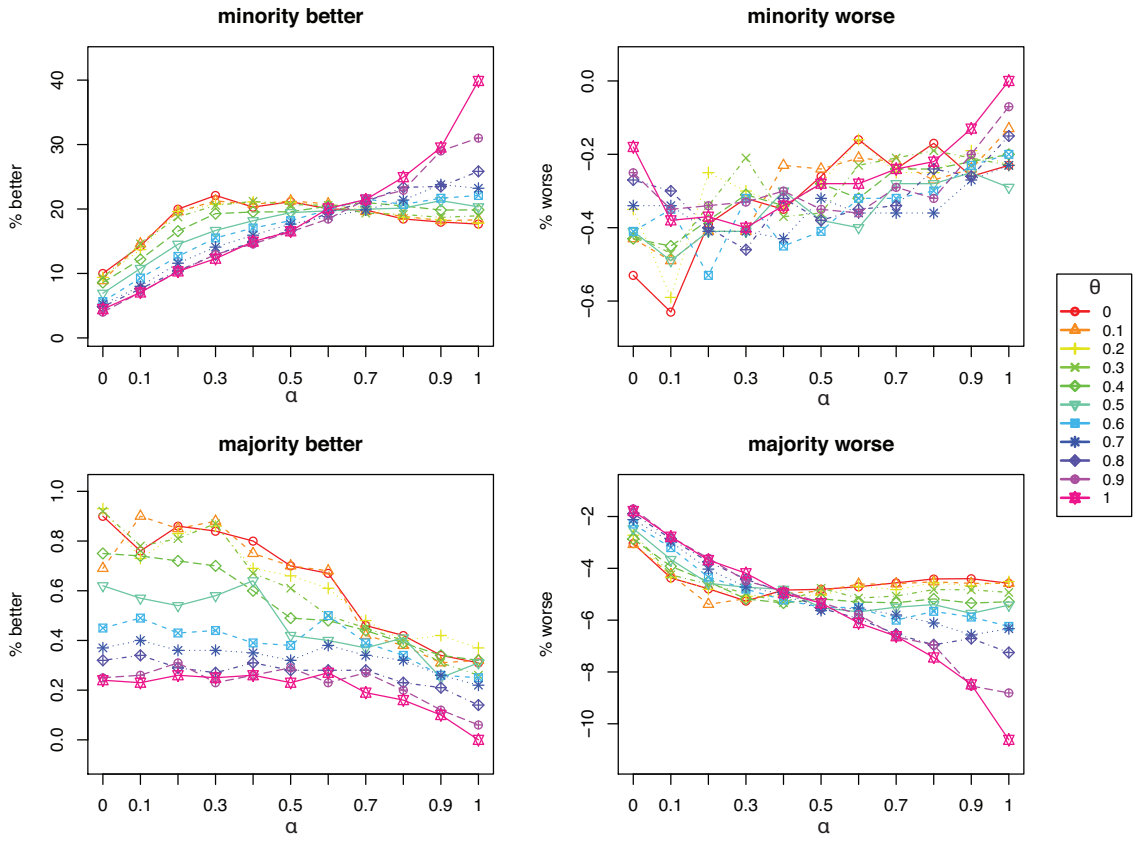


FIGURE 2. *Percent of minorities and majorities who are better/worse of under* DA_{MiR} *than under* DA_{NAA}.

After assessing the effects of α and θ parameters with respect to each other, we set them equal and varied them simultaneously from 0 to 1 in steps of 0.1 while changing the minority reserve ratio, r, from 0% to 32% in steps of 4%. When r is larger than 20%, which is the ratio of minority students in our simulations, all matching algorithms with minority reserves result in better matches for minority students (Supplementary Figures 2.1-6). But DA_{MaQ} made $50-90\%$ of majorities worse compared to DA_{NAA} (Supplementary Figure 2.3), whereas DA_{MiR} only affects majorities adversely when both school priorities and student preferences are highly correlated (Supplementary Figure 2.1). In this regime, minorities are placed at the schools that majorities desire, shifting most majorities to less attractive schools. For lower α and θ values, DA_{MiR} makes $50-90\%$ of majorities better off compared to DA_{MaQ}. Similar results hold for top trading cycles algorithms as well (Supplementary Figure 2.4-6). Specifically, TTC_{MiR} improves the welfare of majorities significantly. *These results suggest for both DA and TTC, matchings with majority quotas are very sensitive toward the levels of quotas and become very costly for welfare of the society, whereas allocations with minority reserves moderate the effects of higher reserves on majority students.*

In real-life situations, affirmative action policies are directed towards groups who tend to be left behind for variety of reasons. For instance, there might be observed differences between the exam scores or other academic achievements between majorities and minorities, which can be reflected in our model by changing the school priorities on minority students. To this end, we introduced a new variable, $\Delta \leq 0$, as average overall shared preference toward minority students. We can define an updated school utility function for minority students as:

$$U_{c_j}(s_i^{minority}) = \theta N_{\Delta,1}(s_i^{minority}) + (1-\theta) Z_{c_j}(s_i^{minority}),$$

where $N_{\Delta,1}$ is a normal distribution with mean Δ and variance 1.

With this new utility function in hand, we first checked the effects of Δ and its interactions with α and minority reserve, r, parameters. Initially, we varied Δ from 0 to -2 with steps of 0.2,[20] α from 0 to 1 in steps of 0.1 and set minority reserve at $r = 20\%$ and $\theta = 0.5$. As the correlation between student preferences increases, all affirmative action policies increased match qualities of minorities up to 80% for smaller Δ values compared to their no affirmative action counterparts (Supplementary Figures 3.1-6). Moreover, as Δ decreases from 0 to -2, the amount of improvements under DA_{MiR} compared to DA_{MaQ} decreases for both minorities and majorities (Supplementary Figure 3.2).

Next, we set $\alpha = \theta = 50\%$ and analyzed the interaction between Δ and minority reserve size r. With decreasing values of Δ, affirmative action policies make minorities better off more dramatically (Supplementary Figures 4.1-6). For lower values of Δ, positive effects of affirmative action promoting policies for minorities can be observed for lower minority reserve sizes. These lower minority reserve sizes coincide with the expected number of minority students being assigned to better schools under no affirmative action policies, corroborating the result of Proposition 4 and showing the importance of selecting appropriate reserve sizes.

Lastly, we compared the student-proposing deferred acceptance algorithms to the top trading cycles algorithms. Overall, the ratio of students who are better off to worse off under TTC_{NAA} compared to DA_{NAA} is around four, validating the notion that TTC based algorithms improve the overall social welfare of students (Supplementary Figure 1.7). For affirmative action policies, we also see that TTC based algorithms benefit a larger ratio of both minorities and majorities, albeit not as much as the increase seen in the no affirmative action counterpart (Supplementary Figures 1.8-9).

[20] $\Delta = -2$ corresponds to the case where average utility of minorities are 2 standard deviations lower than the average utility of majorities.

6. **Generalization: Controlled School Choice with Soft Bounds**

In this section, we consider a more general model of controlled school choice with multiple student types. Such affirmative action policies are readily adapted by various school districts. For example, the Minneapolis school district requires the racial composition of each school to be close to that of the school district. To reflect this variety, we introduce an arbitrary number of student types with a ceiling and a floor bound for each type-school pair, as in Abdulkadiroğlu (2010*a*) and Ehlers (2010). However, instead of considering these floors and ceilings as feasibility constraints, we view them as priority levels.

The basic parameters $\langle C, S, (\succ_i)_{i\in C\cup S}, (q_c)_{c\in C}\rangle$ are the same as in the model presented in Section 2. In addition, following Ehlers (2010), we define a ***type space*** $T = \{t_1, \ldots, t_k\}$ where t denotes a typical type. A ***type function*** $\tau : S \to T$ is given where $\tau(s)$ is the type of student s. Let S^t denote the set of students of type t, $S^t \equiv \{s \in S : \tau(s) = t\}$. For each school c, two vectors of type-specific bounds are considered: $\overline{q}_c = (\overline{q}_c^t)_{t\in T}$ and $\underline{q}_c = (\underline{q}_c^t)_{t\in T}$ such that $\underline{q}_c^t \leq \overline{q}_c^t \leq q_c$ for all $t \in T$ and $\sum_{t\in T} \underline{q}_c^t \leq q_c \leq \sum_{t\in T} \overline{q}_c^t$. For student type t, $\underline{q}_c^t$ is called the ***floor***, and $\overline{q}_c^t$ is called the ***ceiling*** in school c. Let $\overline{q} \equiv (\overline{q}_c)_{c\in C}$ and $\underline{q} \equiv (\underline{q}_c)_{c\in C}$. We call this problem ***controlled school choice with soft bounds***.

In Abdulkadiroğlu (2010*a*) and Ehlers (2010), a matching is not feasible if it assigns less than $\underline{q}_c^t$ or more than $\overline{q}_c^t$ number of type t students to school c. Some school districts administer these bounds as hard bounds, so a theoretical analysis of such policies is important. However, applications of these hard bounds are quite paternalistic in the sense that matchings can be forced *despite* the student preferences. That is, with this specification school districts end up not allowing students to take some available seats, even if there are no physical constraints. On the other hand, one can also view these bounds as *soft bounds*. To be more explicit, school districts may adapt a dynamic priority structure, giving highest priority to student types who do not fill their floors, and least priority to student types who fill their ceilings. Yet, schools can still admit fewer students than their floor or more than their ceiling so long as students with higher priorities do not veto this match. In other words, on the soft bounds view, affirmative action policies promote the desired balancing at schools, only when student preferences allow them to do so.

In controlled school choice with soft bounds, all matchings that assign at most q_c number of students to school c is ***feasible***. A matching μ is ***stable under soft bounds*** if it is individually rational and does not have a blocking pair. Individual rationality is the same as before. A matching μ does not have a blocking pair if for all $c \succ_s \mu(s)$ where $\tau(s) = t$, we have $|\mu(c)| = q_c$, $s' \succ_c s$ for all $s' \in \mu(c) \cap S^t$, and either

(i) $|\mu(c) \cap S^t| \geq \overline{q}_c^t$ and $s' \succ_c s$ for all $s' \in \mu(c)$ such that $|\mu(c) \cap S^{\tau(s')}| > \overline{q}_c^{\tau(s')}$, or
(ii) $\overline{q}_c^t > |\mu(c) \cap S^t| \geq \underline{q}_c^t$, and
  (a) $|\mu(c) \cap S^{t'}| \leq \overline{q}_c^{t'}$ for all $t' \in T \setminus \{t\}$, and

(b) $s' \succ_c s$ for all $s' \in \mu(c)$ such that $\overline{q}_c^{\tau(s')} \geq |\mu(c) \cap S^{\tau(s')}| > \underline{q}_c^{\tau(s')}$.

Less formally, if a student s of type t cannot form a blocking pair with a favorable school c, then c has filled its capacity and all students of type t matched with c are preferred to s by c. Moreover, either c has admitted more than its ceiling of type t students, and all students with types exceeding their ceilings are preferred to s; or c has admitted more than its floor, but not more than its ceiling of type t students, there are no students with types exceeding their ceilings, and all students with types exceeding their floors are preferred to s. Note that we do not have to consider the case when type t students have not filled their floor, because in such a situation there will be a student type that has exceeded its floor.

Affirmative action with minority reserves is a special case of controlled school choice with soft bounds where there are two student types, majorities (type t_1) and minorities (type t_2). To show the equivalence, we specify $(\underline{q}_c, \overline{q}_c)$ for each school c as follows: $\underline{q}_c^{t_1} = 0$, $\overline{q}_c^{t_1} = q_c$, $\underline{q}_c^{t_2} = r_c^m$, and $\overline{q}_c^{t_1} = q_c$. In this specification, for each school the minority floors are set to the minority reserves, and the majority floors are zero. Moreover, both ceilings are equal to the school quotas.

On the other hand, ***stability under hard bounds*** can be defined similarly with the following no blocking definition. A matching μ does not have a blocking pair if for all $c \succ_s \mu(s)$ student s can be hired by either taking an empty seat in c or by firing $s' \in \mu(c)$ such that $s \succ_c s'$ without violating the feasibility constraints in school c.[21]

Consider any school c. Given its priority ranking $\succ_c$, quota q_c, floors and ceilings $\underline{q}_c, \overline{q}_c$, we define the choice function of school c for an available set of students $\tilde{S}$ by $Ch_c(\tilde{S})$. On top of the priority order, the soft bounds impose on school c additional preference in the sense that student types who do not fill floors are preferred over those who fill their floors, and students who do not fill ceilings are preferred over those who fill their ceilings.

To define the choice function more formally, let $Ch_c(\tilde{S}, q_c, (q_c^t)_{t\in T})$ be the subset of $\tilde{S}$ that includes the highest ranked students such that there are no more than q_c students in total and q_c^t students of type t. In addition, let $Ch_c^{(1)}(\tilde{S}) \equiv Ch_c(\tilde{S}, \sum_{t\in T} \underline{q}_c^t, (\underline{q}_c^t)_{t\in T})$, $Ch_c^{(2)}(\tilde{S}) \equiv Ch_c(\tilde{S} \setminus Ch_c^{(1)}(\tilde{S}), q_c - |Ch_c^{(1)}(\tilde{S})|, (\overline{q}_c^t - \underline{q}_c^t)_{t\in T})$, and $Ch_c^{(3)}(\tilde{S}) \equiv Ch_c(\tilde{S} \setminus (Ch_c^{(1)}(\tilde{S}) \cup Ch_c^{(2)}(\tilde{S})), q_c - |Ch_c^{(1)}(\tilde{S})| - |Ch_c^{(2)}(\tilde{S})|, q_c - \overline{q}_c^t)$. Intuitively, $Ch_c^{(1)}(\tilde{S})$ is the set of students chosen with the highest priorities among $\tilde{S}$ without exceeding the floor of each student type, $Ch_c^{(2)}(\tilde{S})$ is the set of remaining students chosen from $\tilde{S}$ with the highest priorities without exceeding the ceiling, and $Ch_c^{(3)}(\tilde{S})$ is the set of students chosen above the ceiling. Finally, $Ch_c(\tilde{S}) \equiv Ch_c^{(1)}(\tilde{S}) \cup Ch_c^{(2)}(\tilde{S}) \cup Ch_c^{(3)}(\tilde{S})$.

[21] In contrast to Ehlers (2010), we do not require that s' be matched such that the feasibility constraints for all schools are satisfied.

6.1. **Deferred Acceptance Algorithm with Soft Bounds.** With hard bounds, no stable matching exists even if the notion of blocking is restricted so that schools can only reject a student who has the same type as the newly admitted one (Ehlers, 2010). However, with soft bounds we recover the existence of stable matchings. To show this, we consider the student-proposing deferred acceptance algorithm with soft bounds, defined as follows.

Step 1: Start with no student matched. Each student s applies to her first-choice school. Let $S_{c,1}$ denote the set of students who applied to school c. School c accepts the students in $Ch_c(S_{c,1})$ and rejects the rest.

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. Let $S_{c,k}$ denote the set of students who either were tentatively matched to c at the end of step $k-1$, or applied to school c in this step. Each school accepts the students in $Ch(S_{c,k})$ and rejects the rest. If there are no rejections, then stop.

The algorithm terminates when there are no new applications. At each step of the algorithm, there is at least one student rejected. Hence, the algorithm ends in finite time.

Proposition 1′. *The student-proposing deferred acceptance algorithm with soft bounds produces a stable matching that assigns the best outcome among the set of stable matching outcomes for each student and is group strategy-proof.*

It is easy to see that Ch_c is *substitutable* for each c. Ch_c satisfies **substitutability** if for any group of students $\tilde{S}$ that contains students s and s', $s \in Ch_c(\tilde{S})$ implies $s \in Ch_c(\tilde{S} \setminus \{s'\})$. To show substitutability, note that if $s \in Ch_c^{(1)}(\tilde{S})$, then $s \in Ch_c^{(1)}(\tilde{S} \setminus \{s'\})$. Otherwise, if $s \in Ch_c^{(i)}(\tilde{S})$ for $i = 2$ or $i = 3$, then $s \in Ch_c^{(i)}(\tilde{S} \setminus \{s'\}) \cup Ch_c^{(i-1)}(\tilde{S} \setminus \{s'\})$.

The first part of the proposition follows from Theorem 6.8 in Roth and Sotomayor (1990) since preferences are substitutable. The second part follows directly as an application of either Martínez et al. (2004), since the preferences are substitutable and *quota q-separable*, or Hatfield and Kojima (2009), since the preferences satisfy *the law of aggregate demand* and substitutability.

For the special case of minority reserves, $Ch_c(S)$ is equal to the r_c^m highest ranked minorities if enough minority students are present, and then the highest ranked students from the rest until reaching the capacity. This is exactly what school c chooses in the deferred acceptance algorithm with minority reserves, so the deferred acceptance algorithms proceed exactly in the same way with minority reserves and with the corresponding soft bounds.

In addition, we also establish the following results.

Theorem 1′. *Suppose that μ is a stable matching under hard bounds. Then either μ is stable under soft bounds or there exists a matching which is stable under soft bounds that Pareto dominates μ.*

Theorem 2′. *Increasing the floor and/or ceiling of a particular type t while keeping other bounds fixed cannot make all type t students worse off in the deferred acceptance algorithm with soft bounds.*

The proofs of these theorems are exactly like their counterparts in Section 3 and so they are omitted.

6.2. **Top Trading Cycles Algorithm with Soft Bounds.** Similarly, we adapt the top trading cycles algorithm to the general environment as follows.

Step k: Start with the matching at the end of step $k-1$ (step 0 matching is the one in which no agent is matched). Each student points to the most preferred school if there is an acceptable school, and otherwise points to herself. Each school points to the student with the highest priority whose types have not filled their floors, if such students remain. Otherwise, it points to the student with the highest priority whose types have not filled their ceilings, if such students remain. Otherwise, it points to the student with the highest priority. There exists at least one cycle. Each student in any of the cycles is matched to the school she is pointing to. All students in the cycles and schools that have filled their capacities are removed. If there is no unmatched student, then stop.

This algorithm ends in finite time since there is at least one student matched at every step of the algorithm.

For the special case of minority reserves, schools give precedence to minorities until they are matched with minorities up to the reserves. Therefore, the same cycles are executed at each step of the algorithm with minority reserves or with the corresponding soft quotas.

Proposition 5′. *The top trading cycles algorithm with soft bounds is Pareto efficient and group strategy-proof.*

Theorem 3′. *Increasing the floor and/or ceiling of a particular type t while keeping other bounds fixed cannot make all type t students worse off in the top trading cycles algorithm with soft bounds.*

The proofs of these results are exactly like their counterparts in Section 4 and so they are omitted.

7. Conclusion

In recent years, admissions processes of public schools have been substantially improved by implementing market-design-rooted mechanisms (Abdulkadiroğlu et al., 2005; Abdulkadiroğlu, Pathak and Roth, 2005). One of the key ingredients in the admissions process is the presence or absence of affirmative action policies in many school districts. A common

affirmative action policy sets *quotas* for different types of students in order to increase representation of minorities. Yet, Kojima (2010) shows that majority quotas may have adverse effects on minorities. In this paper, we introduced an innovative affirmative action policy that gives preferential treatment to minorities by setting *reserves* for them: Schools choose any minority over any majority if the reserve is not filled, but if no minority would like to take that reserved seat, schools can admit more majorities.

First, we showed that the student-optimal stable matching under the affirmative action policy with majority quotas is Pareto dominated by the student-optimal stable matching under the corresponding affirmative action policy with minority reserves. Therefore, having minority reserves in a way corrects for the efficiency loss due to limiting majority attendance in a school even if minorities do not want to attend that particular school. However, having minority reserves can make some minorities worse off, although minority students cannot be worse off altogether.

We then provided two conditions to make minorities better off with the implementation of minority reserves. The first one is obtained by considering common preferences between agents: if all school priorities and student preferences are the same, then having minority reserves cannot make any minority student worse off. The second one is obtained by considering endogenous reserves. For any stable matching μ under no affirmative action, if the minority reserves are set such that all schools have reserves at least equal to the number of minority students matched with them in μ, then there exists a stable matching under minority reserves that Pareto dominates μ for the minority students.

Next, we considered the welfare effects of affirmative action policies under the top trading cycles algorithm. We showed that having minority reserves cannot make all minorities worse off. On the other hand, it is no longer true that allocations with minority reserves Pareto dominate allocations with majority quotas.

Negative results like the ones by Kojima (2010) and the ones presented in the current paper are generally obtained by finding counter examples to the conjectures. In the absence of a general positive result, one can still ask, on average, which policy is better. Therefore, we compared the performances of three policies using simulations. We performed simulations that take random preferences and priority orders and give matches that are obtained via the deferred acceptance (DA) and top trading cycles algorithms under different policies. We evaluated, on average, how much better/worse off both minorities and majorities are in comparison to other policies. In our simulations, we allowed for correlations between student preferences over schools and/or correlations between school priorities over students. The simulations show that DA with minority reserves not only Pareto dominates DA with majority quotas, but also significantly improves the welfare of both minorities and majorities. Also, we observed that majority quota based mechanisms are very sensitive to quota size,

especially for majorities, whereas minority reserves moderate the adverse effects of affirmative action policies on majorities.

Finally, we took the natural next step and generalized our model with minority reserves to a more general model of controlled school choice with floors and ceilings as in Ehlers (2010) and Abdulkadiroğlu (2010*a*). However, instead of considering these bounds as feasibility constraints, we viewed them as soft regulatory boundaries of preferential treatment. Our main results continue to hold in the generalization.

In conclusion, it is important to mention that our work is a normative study proposing how affirmative action policies should be, rather than an analysis or a characterization of affirmative action policies with hard bounds that is used in practice. Affirmative action with soft bounds has clear benefits over its hard-bound counterpart for the two celebrated algorithms that are more frequently used in practice. For school districts with diversity concerns such as the San Francisco school district, our work provides an alternative approach for implementation.

8. **Appendix: Proofs**

Here, we include the proofs.

Proof of Proposition 1. First, we show that the deferred acceptance algorithm with minority reserves produces a student-optimal stable matching.

Definition 1. *A school c's preference satisfies* ***substitutability*** *if for any group of students $\hat{S}$ that contains students s and s', $s \in Ch_c(\hat{S})$ implies $s \in Ch_c(\hat{S} \setminus \{s'\})$.*

Claim 1. *Every school's preference is substitutable.*

Proof. If $s \in S^M$, then $s \succ_c s''$ for every $s'' \in \hat{S} \setminus Ch_c(\hat{S})$. Therefore, $s \in Ch_c(\hat{S} \setminus \{s'\})$. Otherwise, $s \in S^m$. This implies that either (i) $|Ch_c(\hat{S}) \cap S^m| > r_c^m$ and $s \succ_c s''$ for every $s'' \in \hat{S} \setminus Ch_c(\hat{S})$ or (ii) $|Ch_c(\hat{S}) \cap S^m| \leq r_c^m$ and $s \succ_c s''$ for every $s'' \in (\hat{S} \setminus Ch_c(\hat{S})) \cap S^m$. In both cases, $s \in Ch_c(\hat{S} \setminus \{s'\})$. □

Therefore, each school's preference with strict priority and minority reserves can also be viewed as a substitutable preference profile. And similarly, by Theorem 6.8 of Roth and Sotomayor (1990), the student-proposing deferred acceptance algorithm with minority reserves produces the student-optimal stable matching.

To verify group strategy-proofness, we introduce a new school choice problem, where the student-proposing deferred acceptance algorithm produces the same matching with the student-proposing deferred acceptance algorithm with minority reserves.[22]

[22]An alternative proof can be done by an application of the main result in Martínez et al. (2004), or Hatfield and Kojima (2009).

Split each school c that has a quota of q_c and minority reserve r_c^m with preference $\succ_c$ into two schools c^1 (original) and c^2 (minority favoring): c^1 has a capacity of $q_c - r_c^m$ and preferences $\succ_c$, c^2 has a capacity of r_c^m and preferences $\succ'_c$:

$$s \succ'_c s' \iff \begin{cases} s \in S^m & \text{and } s' \in S^M \\ s, s' \in S^m & \text{and } s \succ_c s' \\ s, s' \in S^M & \text{and } s \succ_c s' \end{cases} .$$

For each student s we replace school c with its copies in the same order to get the new preference $\succ'_s$. For example, if $c_1 \succ_s c_2$ then $c_1^2 \succ'_s c_1^1 \succ'_s c_2^2 \succ'_s c_2^1$. In words, each student keeps the relative rankings of schools the same and prefers the minority favoring schools over the originals.

Let us call the original problem M^1 and the new one M^2. Any matching in M^2 can be transformed to a matching in M^1 in a straightforward manner: All students who are matched to c^1 and c^2 in M^2 will be matched to c in M^1. Now take a matching μ in M^1. We can transform this into a matching in M^2 as follows. If $|\mu(c) \cap S^m| \geq r_c^m$ then c^1 is matched to the highest-ranked minority students in $\mu(c)$ with respect to $\succ_c$, and the rest of the students in $\mu(c)$ are matched to c^2. Otherwise, if $|\mu(c) \cap S^m| < r_c^m$, then all the minority students in $\mu(c)$ and the best majority students from $\mu(c)$ with respect to $\succ_c$ are matched to c^1 until the quota of c^2 is reached or students are exhausted, and the rest of the students in $\mu(c)$ are matched to c^2. Let μ be a matching in M^1 and μ^2 be a matching in M^2 that correspond to each other by the preceding transformation. By construction, μ in M^1 is stable if and only if μ^2 is stable in M^2.

Therefore, the student-proposing deferred acceptance algorithm with minority reserves produces the same outcome as the student-proposing deferred acceptance algorithm in M^2. Suppose for contradiction that there exists a problem M^1 for which a set of students $\hat{S}$ can deviate from truth-telling in the student-proposing deferred acceptance algorithm to get better outcomes. If we look at the corresponding problem M^2, then $\hat{S}$ can also deviate from truth-telling to get better outcomes. This is a contradiction since the student-proposing deferred acceptance algorithm is group strategy-proof, which is the main result of Dubins and Freedman (1981). □

Proof of Theorem 1. If μ is stable under minority reserves with r^m, then we are done. Suppose, otherwise, that μ is not stable under minority reserves. Then there exists a blocking pair (c, s). Since (c, s) does not form a blocking pair under majority quotas; then s has to be a majority and $|\mu(c) \cap S_M| = q_c^M$ and $|\mu(c)| < q_c$. Therefore, c has an empty seat in μ and there exists a student who prefers c to its current match.

Whenever such a school exists we execute the improvement algorithm described after Theorem 1. Let μ' be the matching produced after applying the algorithm.

Note that all the students, both minorities and majorities, are weakly better off in μ' compared to μ. Moreover, at least one student is better off. To complete the proof we have to show that μ' is stable under minority reserves.

Assume otherwise. Since μ is an individually rational matching, so is μ'. Therefore, there exists a blocking pair (c', s') to violate stability under minority reserves. First note that $|\mu'(c')| = q_{c'}$. Let us separate the analysis into two cases depending on whether s' is a minority or majority.

Case 1 [**Minority**]: Suppose that s' is a minority student. If $\mu'(c') = \mu(c')$, then (c', s') would form a blocking pair for μ under majority quotas since $\mu'(s') \succeq_{s'} \mu(s')$. Therefore, $\mu'(c') \neq \mu(c')$. This means that c' filled some of its seats in the improvement procedure. At every step of the procedure when c' filled its seats, s' must have preferred c' to its match at that point since s' weakly improves its match at any step of the procedure. Therefore, for any student $s \in \mu'(c') \setminus \mu(c')$, $s \succ_{c'} s'$. For any student $s \in \mu'(c') \cap \mu(c')$, $s \succ_{c'} s'$ since (c', s') is not a blocking pair in μ under majority quotas. This contradicts the fact that (c', s') is a blocking pair under minority reserves.

Case 2 [**Majority**]: Suppose that s' is a majority student. If $\mu'(c') = \mu(c')$, then (c', s') would form a blocking pair for μ under majority quotas. Therefore, $\mu'(c') \neq \mu(c')$, which implies that school c filled some of its seats in the improvement procedure. At every step of the procedure when c' filled its seats, s' must have preferred c' to its match at that point since s' weakly improves its match at any step of the procedure. Therefore, for any student $s \in (\mu'(c') \setminus \mu(c')) \cap S^M$, $s \succ_{c'} s'$. Moreover, since (c', s') is not a blocking pair in μ under majority quotas, $s \in (\mu'(c') \cap \mu(c')) \cap S^M$, $s \succ_{c'} s'$. If we combine the last two statements, we get that $s \in \mu'(c') \cap S^M$, $s \succ_{c'} s'$. If $|\mu'(c') \cap S^m| \leq r^m_{c'}$ then c' cannot block since it has to keep the minority students and it prefers all the majority students to s'. Therefore, $|\mu'(c') \cap S^m| > r^m_{c'}$. Let s_m be the minority student who is minimal according to $\succ_{c'}$. Then $s_m \notin \mu(c')$, because otherwise either (i) $(\mu'(c') \setminus \mu(c')) \cap S^m \neq \emptyset$ and one of $(\mu'(c') \setminus \mu(c')) \cap S^m$ and c would form a blocking pair for μ under majority quotas, or (ii)$(\mu'(c') \setminus \mu(c')) \cap S^m = \emptyset$ and (c', s') would form a blocking pair for μ under majority quotas. Therefore, $s_m \notin \mu(c')$. This implies that s_m must have been matched to c' in the improvement procedure. Moreover, she must have been the last minority student to be matched to c'. At that step of the algorithm s' prefers her match to c', so s' should have been matched to c' rather than s_m according to the procedure. We get a contradiction. □

Proof of Theorem 2. Suppose for contradiction that for all $s \in S^m$, $\mu(s) \succ_s \mu^r(s)$.

When minority students submit their preferences truthfully, the resulting matching is μ^r with minority reserves. Now, we claim that if they jointly modify their preferences such that each minority student s lists $\mu(s)$ as the only acceptable choice, μ would be a stable matching under minority reserves. Let $\succ'_s$ be this preference ordering of $s \in S^m$.

We claim that if the preference profile is $((\succ'_s)_{s\in S^m}, (\succ_s)_{s\in S^M})$, then μ is a stable matching under minority reserves. First note that each minority student s is getting her top choice in μ according to $\succ'_s$. Therefore, none of the minorities will be in a blocking pair. Moreover, if (c, s) is a blocking pair where $s \in S^M$ for μ under minority reserves, then the same pair would also form a blocking pair for μ under no affirmative action. Therefore, there cannot be any blocking pairs and μ is stable under minority reserves for $((\succ'_s)_{s\in S^m}, (\succ_s)_{s\in S^M})$.

Therefore, the student-proposing deferred acceptance algorithm with minority reserves when students submit $((\succ'_s)_{s\in S^m}, (\succ_s)_{s\in S^M})$ must assign each minority student s her top choice, which is $\mu(s)$. Hence, all minority students get a strictly better outcome by jointly changing their preferences, which contradicts the fact that the student-proposing deferred acceptance algorithm under minority reserves is group strategy-proof (Proposition 1). □

Proof of Proposition 2. Without loss of generality, relabel schools such that for any $i, j \in \{1, \ldots, |C|\}$, all students prefer c_i to c_j if and only if $i < j$. Similarly, relabel students such that for any $i, j \in \{1, \ldots, |S|\}$, all schools prefer s_i to s_j if and only if $i < j$.

It is clear that under each affirmative action policy there is a unique stable matching because students' preferences and schools' priorities are all the same. Therefore, we start by characterizing the stable matchings under the policies.

No Affirmative Action: In the unique stable matching, c_1 is matched to top q_{c_1} students, $s_1, \ldots, s_{q_1}$; c_2 is matched to the next q_{c_2} students, $s_{q_1+1}, \ldots, s_{q_1+q_2}$; and so on. The unique stable matching in this case can be obtained by a serial dictatorship of schools in which c_k takes the k^{th} turn to choose its students.

Majority Quotas: In the unique stable matching, c_1 is matched to top $r^m_{c_1}$ minority students first and then top $q^M_{c_1} - r^m_{c_1}$ students among the remaining students. Next, c_2 is matched to top r^m_2 minority students among the remaining minority students, and top $q^M_{c_2} - r^m_{c_2}$ students among the remaining students, and so on. Even if there are not enough minority students to take $r^m_{c_k}$ seats at step k, school c_k cannot be matched to more than $q^m_{c_k} - r^m_{c_k}$ majority students. The unique stable matching in this case can be obtained by a serial dictatorship of schools in which c_k takes the k^{th} turn: First r^m_k minority students are admitted if there are enough minority students left, then $q^M_{c_k} - r^m_{c_k}$ students are admitted if there are enough students left.

Minority Reserves: In the unique stable matching, c_1 is matched to top $r^m_{c_1}$ minority students first, and then top students among the remaining ones to fill its capacity q_{c_1}; among the remaining students c_2 is matched to top $r^m_{c_2}$ minority students, and then top students among the remaining students to fill its quota q_{c_2}; and so on. The unique stable matching in this case can be obtained by a serial dictatorship of schools in which c_k moves in k^{th} turn: First $r^m_{c_k}$ minority students are admitted if

there are enough minority students left, then any type of students are admitted to fill its capacity q_{c_k}.

Next, note that minority students are matched with the same schools under majority quotas and minority reserves. The serial dictatorship mechanisms in both cases step by step give the same outcome as long as the minority reserves can be filled. If the minority reserves of c_k cannot be filled, then the minority students are exhausted. For c_k and remaining schools more seats are available to remaining majority students in minority reserves compared to majority quotas. Therefore, $\mu^r(s) = \mu^q(s)$ for any $s \in S^m$ and $\mu^r(s) \succeq_s \mu^q(s)$ for any $s \in S^M$.

Finally, to show that $\mu^r(s) \succeq_s \mu(s)$ for all $s \in S^m$ we prove the following. Let M_t^r and M_t be the set of majority students available to c_t during the serial dictatorship under minority reserves and no affirmative action, respectively. Similarly define m_t^r and m_t to be the set of minority students available at step t. Then the claim is $m_t^r \subseteq m_t$ and $M_t^r \supseteq M_t$.

The proof of this claim is by mathematical induction on t. When $t = 1$, $m_t^r = m_t = S^m$ and $M_t^r = M_t = S^M$, so the claim holds. Suppose that the claim holds for $t = k$. Since all schools have the same priorities over students, c_{k+1} prefers any student in $m_t \setminus m_t^r$ to any student in m_t^r, and similarly any student in $M_t^r \setminus M_t$ is preferred to any student in M_t. Note that either all students are chosen by c_{t+1} in both serial dictatorships if there is enough capacity, or the same number of students are chosen. In the first case, $m_{t+1}^r = m_{t+1} = \emptyset$ and $M_{t+1}^r = M_{t+1} = \emptyset$ and the claim holds. Now, consider the latter case. Suppose that a minorities and b majorities are chosen by c_{t+1} under no affirmative action. If $a \leq |m_t \setminus m_t^r|$, then $m_{t+1}^r \subseteq m_{t+1}$ (since only minority students from $m_t \setminus m_t^r$ are chosen under no affirmative action). Even if c_{t+1} chooses all majorities under minority reserves, we get that $M_{t+1}^r \supseteq M_{t+1}$ (since $a \leq |m_t \setminus m_t^r| = |M_t^r \setminus M_t|$ and at most a more majorities are chosen under minority reserves compared to no affirmative action). However, if $a > |m_t \setminus m_t^r|$, then c_{t+1} chooses $a - |m_t \setminus m_t^r|$ minorities among m_t^r when M_t is available. Therefore, even if all of $M_t^r \setminus M_t$ are chosen under minority reserves, which has the same cardinality as $|m_t \setminus m_t^r|$, at least $a - |m_t \setminus m_t^r|$ minorities are chosen. This implies $m_{t+1}^r \subseteq m_{t+1}$. Similarly, c_{t+1} has chosen b majorities among $M_t \cup m_t$ under no affirmative action, so it cannot choose more than $b + |M_t^r \setminus M_t|$ among $m_t^r \cup M_t^r$. This implies $M_{t+1}^r \supseteq M_{t+1}$.

□

Proof of Proposition 3. Assume for contradiction that μ is not a stable matching under minority reserves r^m. Since μ is a stable matching under no affirmative action, it is an individually stable matching. Therefore, there exists a blocking pair (c, s) when minority reserves are r^m. Since μ is a stable matching under no affirmative action, $|\mu(c) \cap S| = q_c$, i.e., there are no empty seats in c.

First suppose that s is a minority student. Since $|\mu(c) \cap S^m| \geq r_m^c$, there exists $s' \in \mu(c)$ such that $s \succ_c s'$. In this case, (c, s) also forms a blocking pair when there is no affirmative action policy, which is a contradiction.

Suppose now that s is a majority student. Then either (a) $|\mu(c) \cap S^m| \geq r_c^m + 1$ and there exists $s' \in \mu(c)$ such that $s \succ_c s'$, or (b) $|\mu(c) \cap S^m| = r_c^m$ and there exists $s' \in \mu(c) \cap S^M$ such that $s \succ_c s'$. In both cases (c, s) would form a blocking pair for μ with no affirmative action, which is a contradiction. □

Proof of Proposition 4. If μ is a stable matching under minority reserves with r^m, then the proof is complete. Suppose that μ is not stable under minority reserves. Then there exists a blocking pair (μ, s) such that $s \in S^m$, $|\mu(c) \cap S^m| < r_c^m$, and $c \succ_s \mu(s)$. In this case, we improve the matchings of minorities according to the following procedure.

Step 1: Take a school c that has not filled its minority reserve for which there exists a minority student who likes c to her current assignment. Consider all minority students who would like to switch to c. Assign the best minority students among these with respect to $\succ_c$ until the minority reserves are filled or there does not exist any more minority students who would like to switch to c. If k new minority students are assigned then release k majority students who are least preferred with respect to $\succ_c$.

Step t: If there exists no school c that has not filled its minority reserve for which there exists a minority student who prefers c to her current assignment, then stop. Otherwise, consider all minority students who would like to switch to c. Assign the best minority students with respect to $\succ_c$ until the minority reserves are filled or there does not exist any minority student who would like to switch to c. If k new minority students are assigned then release k majority students who are least preferred with respect to $\succ_c$.

The procedure ends in a finite number of steps since at any step of the algorithm a minority student improves her match. It ends when there exists no minority student who would like to switch to a school that has not filled its minority reserves. At the end of this procedure there are some unmatched majority students and schools with empty seats. Let μ_1 be the matching that we get after the procedure. We revise the matching using the improvement algorithm described after Theorem 1 to get a new matching. Let μ' be the matching at the end of this algorithm. For every minority student s, $\mu'(s) \succeq_s \mu(s)$. We prove that μ' is stable under minority reserves. Since the original matching is individually rational and we never assign a student to an unacceptable school in the above algorithms, μ' is also individually rational.

Assume for contradiction that there exists a blocking pair (c, s). Then $|\mu'(c)| = q_c$, i.e., there are no empty seats in c. We split the rest of the analysis into two cases depending on whether the student is a minority or majority.

Case 1 [**Minority**]: Let s be a minority student. Then there exists a student s' who is assigned to c during one of the improvement stages such that $s' \succ_c s$ since μ is a stable matching. However, in any of the improvement stages s' always prefers c to her assignment. Therefore, s' should have been selected instead of s at the step when s is assigned to c. This is a contradiction.

Case 2 [**Majority**]: Let s be a majority student. Then there exists either (i) $s' \in \mu(c) \cap S^M$ such that $s' \succ_c s$ or (ii) $s' \in \mu(c) \cap S^m$ such that $s' \succ_c s$ and $|\mu'(c) \cap S^m| > r_c^m$. In the former case, it cannot be that s' was matched to c in improvement stage II, because s would have been selected instead of s' since s would have liked to switch to c during any step of the algorithm. Since s' is a majority student then $s' \in \mu(c)$. This is also impossible since (c, s) would form a blocking pair in μ under no affirmative action. In the latter case, take the minority student who is the worst minority student according to $\succ_c$, say s''. This minority student must be assigned to c during the improvement stage II since the minority reserves of c are exceeded. But $s \succ_c s' \succ_c s''$, so s should have been selected instead of s'' at that step of improvement stage II. This is a contradiction.

□

Proof of Corollary 1. If $r_c^m \leq |\mu(c) \cap S^m|$ for all c, then μ is also stable under minority reserves with r^m by Proposition 3. Therefore, μ_r Pareto dominates μ for all students. However, if $r_c^m \geq |\mu(c) \cap S^m|$ for all c then there exists a stable matching μ' under minority reserves with r^m that Pareto dominates μ for all minority students by Proposition 4. Since μ_r is the student optimal stable matching under minority reserves with r^m, it Pareto dominates μ' for all students, which in turn Pareto dominates μ for all minority students. The conclusion follows. □

Proof of Proposition 5. For Pareto efficiency note that each student who is matched at the first step of the algorithm is getting her first choice, so she cannot be made better off. Each student getting matched at the next step cannot get a more preferred school without harming some of the students who are matched in step 1. By induction, the students who are matched at step k of the algorithm cannot get a better match without harming some of the students who are matched before step k.

For group strategy-proofness, suppose for contradiction that there exists a group of students $\hat{S}$ who can deviate from truth-telling to be matched with better schools (say c_s for all $s \in \hat{S}$). They can also achieve the same outcome by always pointing to these schools, i.e. by ranking these schools as their only acceptable choice. Denote this preference relation by $\succ'_s$ for $s \in \hat{S}$. Assume that $\hat{S}$ is using $\succ'_{\hat{S}}$.

The top trading cycles algorithm with $(\succ'_{\hat{S}}, \succ_{S\setminus\hat{S}})$ executes new cycles to match students in $\hat{S}$ with better schools. Suppose that the first new cycle executed is o at step k', and say s_1 is the student in o from $\hat{S}$ who is matched earliest under truthful reporting at step $k(\geq k')$. Let $o = \langle s_1, c_1, \ldots, s_n, c_n\rangle$. At step k' under the deviation strategy, s_i is pointing to c_i and c_i is pointing to s_{i+1} for $i \in \{1, \ldots, n\}$.[23] Note that an agent keeps pointing to the same agent until the former is removed. In addition, the algorithm executes the same cycles with truth-telling and the deviation before step k'. Therefore, c_n is still unmatched at step k of the algorithm under truth-telling. An inductive argument shows that all agents in the cycle are still in the problem at step k of the algorithm with truth-telling. This contradicts the fact that c_1 fills its quota before step k with truth-telling since s_1 is pointing to a less desirable school at step k. □

Proof of Theorem 3. The proof is by induction on the number of students.

Base Case: If there is only one student in the problem, then the claim is trivially true.

General Case: Consider the set of students and schools who are matched in the first step of the top trading cycles with minority reserves, say S_1 and C_1 respectively. If there exists a minority student among S_1, then we are done, since this student is matched to her top choice. Otherwise, $S_1 \subseteq S^M$. Moreover, each school in C_1 cannot be pointing to a minority student at the first step, since all students who are pointed to by schools in C_1 are matched at this step. Therefore, these schools have 0 minority reserves. Since each agent in $C_1 \cup S_1$ is pointing to her best choice, they must also be matched to each other in the first step of the top trading cycles without minority reserves. To implement the top trading cycles algorithm with or without minority reserves for the rest of the agents, we can consider a new problem with the set of students $S \setminus S_1$, and the capacities of C_1 reduced by one. By induction there exists at least one minority student s for which the outcome with the minority reserves is as good as the outcome without minority reserves. This completes the proof.

□

Proof of Proposition 7. In the next example, taken from Kojima (2010), we show that the top trading cycles can Pareto dominate, for the minority students, the top trading cycles with minority reserves.

Example 3. Consider the following problem: $C = \{c_1, c_2, c_3\}$, $S^M = \{s_1, s_2\}$, and $S^m = \{s_3, s_4\}$. All schools have a quota of one, $q = (1, 1, 1)$. Students' preferences are:

$$\succ_{s_1}\colon c_1, \quad \succ_{s_2}\colon c_2, \quad \succ_{s_3}\colon c_2, \quad \succ_{s_4}\colon c_3.$$

And colleges' priorities are:

$$\succ_{c_1}\colon s_4 \ \succ \ s_2, \quad \succ_{c_2}\colon s_1 \ \succ \ s_3, \quad \succ_{c_3}\colon s_1 \ \succ \ s_4 \ \succ \ s_2 \ \succ \ s_3.$$

[23]The indices are taken modulo n, so $s_{n+1} = s_1$.

Minority reserves are given by $r^m = (0, 0, 1)$. If we apply the top trading cycles mechanism then we obtain the matching μ:

$$\mu(c_1) = s_1,\ \mu(c_2) = s_3,\ \mu(c_3) = s_4,\ \mu(s_2) = s_2.$$

If we apply the top trading cycles with minority reserves then we obtain the matching μ':[24]

$$\mu'(c_1) = s_1,\ \mu'(c_2) = s_2,\ \mu'(c_3) = s_4,\ \mu'(s_3) = s_3.$$

In this problem s_3 prefers the top trading cycles algorithm, whereas s_4 is indifferent.

In the next example we show that the top trading cycles with minority reserves can Pareto dominate, for minority students, the top trading cycles.

Example 4. Consider the following problem: $C = \{c_1, c_2\}$, $S^M = \{s_2\}$, and $S^m = \{s_1\}$. All schools have a quota of one, $q = (1, 1)$. Students' preferences are:

$$\succ_{s_1}: c_1 \ \succ\ c_2,\ \ \succ_{s_2}: c_1 \ \succ\ c_2.$$

And colleges' priorities are:

$$\succ_{c_1}: s_2 \ \succ\ s_1,\ \ \succ_{c_2}: \dots.$$

Minority reserves are given by $r^m = (1, 0)$. If we apply the top trading cycles mechanism then we obtain the matching μ:

$$\mu(c_1) = s_2,\ \mu(c_2) = s_1.$$

If we apply the top trading cycles with minority reserves then we obtain the matching μ':

$$\mu'(c_1) = s_1,\ \mu'(c_2) = s_2.$$

In this problem s_1 prefers the top trading cycles with minority reserves.

In the next example we show that the top trading cycles mechanism with majority quotas can Pareto dominate, for minority students, the top trading cycles with minority reserves.

Example 5. Consider the following problem: $C = \{c_1, c_2, c_3\}$, $S^M = \{s_2\}$, and $S^m = \{s_1, s_3\}$. All schools have a quota of one, $q = (1, 1, 1)$. Students' preferences are:

$$\succ_{s_1}: c_2,\ \ \succ_{s_2}: c_1 \ \succ\ c_3,\ \ \succ_{s_3}: c_1 \ \succ\ c_3.$$

And colleges' priorities are:

$$\succ_{c_1}: s_1,\ \ \succ_{c_2}: s_2,\ \ \succ_{c_3}: s_3.$$

[24]This is also the outcome of the top trading cycles with majority quotas as shown by Kojima (2010).

Majority quotas are given by $q^M = (0, 1, 1)$, and corresponding minority reserves are $r^m = (1, 0, 0)$. If we apply the top trading cycles mechanism with majority quotas then we obtain the matching μ:

$$\mu(c_1) = s_3,\ \mu(c_2) = s_1,\ \mu(c_3) = s_2.$$

If we apply the top trading cycles with minority reserves then we obtain the matching μ':

$$\mu'(c_1) = s_2,\ \mu'(c_2) = s_1,\ \mu'(c_3) = s_3.$$

In this problem s_1 is indifferent between two algorithms, whereas s_3 prefers the top trading cycles with majority quotas.

In the last example, we show that the top trading cycles mechanism with minority reserves can Pareto dominate, for minority students, the top trading cycles mechanism with majority quotas.

Example 6. Consider the following problem: $C = \{c_1, c_2\}$, $S^M = \{s_2\}$, and $S^m = \{s_1\}$. All schools have a quota of one, $q = (1, 1)$. Students' preferences are:

$$\succ_{s_1}\colon c_2 \ \succ\ c_1,\ \ \succ_{s_2}\colon c_1 \ \succ\ c_2.$$

And colleges' preferences are:

$$\succ_{c_1}\colon s_1 \ \succ\ s_2,\ \ \succ_{c_2}\colon s_2 \ \succ\ s_1.$$

Majority quotas are given by $q^M = (0, 1)$, and the corresponding minority reserves are $r^m = (1, 0)$. If we apply the top trading cycles mechanism with majority quotas then we obtain the matching μ:

$$\mu(c_1) = s_1,\ \mu(c_2) = s_2.$$

If we apply the top trading cycles with minority reserves then we obtain the matching μ':

$$\mu'(c_1) = s_2,\ \mu'(c_2) = s_1.$$

In this problem s_1 prefers the top trading cycles with minority reserves.

□

References

Abdulkadiroğlu, Atila. 2005. "College admissions with affirmative action." *International Journal of Game Theory*, 33(4): 535–549.

Abdulkadiroğlu, Atila. 2010*a*. "Controlled school choice." mimeo: Duke University.

Abdulkadiroğlu, Atila. 2010*b*. "Generalized matching for school choice." mimeo: Duke University.

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

Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth. 2005. "The New York City high school match." *American Economic Review*, 95(2): pp. 364–367.

Abdulkadiroğlu, Atila, Parag Pathak, Alvin Roth, and Tayfun Sönmez. 2005. "The Boston public school match." *American Economic Review*, 95(2): 368–371.

Arcidiacono, Peter. 2005. "Affirmative action in higher education: How do admission and financial aid rules affect future earnings?" *Econometrica*, 73(5): 1477–1524.

Bertrand, Marianne, Rema Hanna, and Sendhil Mullainathan. 2010. "Affirmative action in education: Evidence from engineering college admissions in India." *Journal of Public Economics*, 94(1-2): 16–29.

Bird, Charles G. 1984. "Group incentive compatibility in a market with indivisible goods." *Economics Letters*, 14(4): 309–313.

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

Deshpande, Ashwini. 2006. "World Development Report 2006: Equity and Development." Chapter Affirmative Action in India and the United States. Washington D.C.: The World Bank.

Dubins, Lester E., and David A. Freedman. 1981. "Machiavelli and the Gale-Shapley algorithm." *American Mathematical Monthly*, 88(7): 485–494.

Ehlers, Lars. 2010. "School choice with control." mimeo: Université de Montréal.

Erdil, Aytek, and Haluk Ergin. 2008. "What's the matter with tie-breaking? Improving efficiency in school choice." *American Economic Review*, 98(3): 669–689.

Ergin, Haluk I. 2002. "Efficient resource allocation on the basis of priorities." *Econometrica*, 70(6): 2489–2497.

Fryer, Roland G. 2009. "Implicit quotas." *Journal of Legal Studies*, 38(1): 1–20.

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

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

Hatfield, John William, and Fuhito Kojima. 2009. "Group incentive compatibility for matching with contracts." *Games and Economic Behavior*, 67(2): 745–749.

Holzer, Harry, and David Neumark. 2000. "Assessing affirmative action." *Journal of Economic Literature*, 38(3): 483–568.

Jencks, Christopher. 1992. *Rethinking Social Policy: Race, Poverty, and the Underclass.*

Cambridge, MA: Harvard University Press.

Kesten, Onur. 2006. "On two competing mechanisms for priority-based allocation problems." *Journal of Economic Theory*, 127(1): 1297–1348.

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

Kesten, Onur, and Utku Ünver. 2009. "A theory of school choice lotteries." mimeo, Boston College and Carnegie Mellon University.

Kojima, Fuhito. 2010. "School choice: Impossibilities for affirmative action." Working Paper.

Loury, Linda Datcher, and David Garman. 1993. "Affirmative action in higher education." *American Economic Review*, 83(2): 99–103.

Martínez, Ruth, Jordi Massó, Alejandro Neme, and Jorge Oviedo. 2004. "On group strategy-proof mechanisms for a many-to-one matching model." *International Journal of Game Theory*, 33(1): 11–128.

McVitie, D. G., and L. B. Wilson. 1970. "Stable marriage assignment for unequal sets." *BIT*, 10(3): 295–309.

Pathak, Parag A. 2011. "The mechanism design approach to student assignment." Vol. 3.

Roth, Alvin, and Marilda Sotomayor. 1990. *Two-sided Matching: A Study in Game-Theoretic Modelling and Analysis.* Vol. 18 of *Econometric Society Monographs*, Cambridge University Press, Cambridge England.

Roth, Alvin E. 2008. "Deferred acceptance algorithms: History, theory, practice and open questions." *International Journal of Game Theory*, 36(3): 537–569.

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

Sönmez, Tayfun, and Utku Ünver. 2009. "Handbook of Social Economics." , ed. Alberto Bisin Jess Benhabib and Matthew Jackson, Chapter Matching, Allocation, and Exchange of Discrete Resources. Elsevier.

Sowell, Thomas. 2004. *Affirmative Action Around the World: An Empirical Study.* New Haven: Yale University Press.