Theoretical Economics 8 (2013), 325–363 1555-7561/20130325

Effective affirmative action in school choice

Isa E. Hafalir
Tepper School of Business, Carnegie Mellon University

M. Bumin Yenmez
Tepper School of Business, Carnegie Mellon University

Muhammed A. Yildirim
Center for International Development, Harvard University

The prevalent affirmative action policy in school choice limits the number of admitted majority students to give minority students higher chances to attend their desired schools. There have been numerous efforts to reconcile affirmative action policies with celebrated matching mechanisms such as the deferred acceptance and top trading cycles algorithms. Nevertheless, it is theoretically shown that under these algorithms, the policy based on *majority quotas* may be detrimental to minorities. Using simulations, we find that this is a more common phenomenon rather than a peculiarity. To circumvent the inefficiency caused by majority quotas, we offer a different interpretation of the affirmative action policies based on *minority reserves*. With minority reserves, schools give higher priority to minority students up to the point that the minorities fill the reserves. We compare the welfare effects of these policies. The deferred acceptance algorithm with minority reserves Pareto dominates the one with majority quotas. Our simulations, which allow for correlations between student preferences and school priorities, indicate that minorities are, on average, better off with minority reserves while adverse effects on majorities are mitigated.

Keywords. School choice, affirmative action, deferred-acceptance algorithm, top trading cycles algorithm.

JEL classification. C78, D61, D78, I20.

1. Introduction

Affirmative action is a popular, albeit controversial, scheme that is implemented to close socioeconomic gaps that exist between groups as a result of historic discrimination. To this end, it involves policies designed to increase the representation of some

Isa E. Hafalir: isaemin@cmu.edu
M. Bumin Yenmez: byenmez@andrew.cmu.edu
Muhammed A. Yildirim: muhammed_yildirim@hks.harvard.edu
We thank the co-editor, Gadi Barlevy, and two anonymous referees, as well as Onur Kesten, Fuhito Kojima, Dimitar Simeonov, Tayfun Sönmez, and Alistair Wilson. We also thank seminar participants at Bilkent University, California Institute of Technology, Tepper School of Business, University of Maryland, and University of Montreal.

Copyright © 2013 Isa E. Hafalir, M. Bumin Yenmez, and Muhammed A. Yildirim. Licensed under the Creative Commons Attribution-NonCommercial License 3.0. Available at http://econtheory.org.
DOI: 10.3982/TE1135

groups in public areas such as employment, education, and business contracting. This paper studies affirmative action in school choice, the so-called *controlled choice problem* (Abdulkadiroğlu and Sönmez 2003), where the goal of affirmative action is to maintain diversity at schools by giving underrepresented groups (usually minorities) higher chances to attend better schools.

Many members 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 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 quotas (e.g., historically in Seattle (WA), Jefferson County (KY), Louisville (KY), Minneapolis (MN), and White Plains (NY)). Alternatively, 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)).[1]

In a seminal paper, Abdulkadiroğlu and Sönmez (2003) approach the school choice problem from a mechanism-design perspective. They illustrate that the mechanisms used in practice had shortcomings, and propose as alternatives two celebrated algorithms, *the student-proposing deferred acceptance algorithm* (DA) and *the top trading cycles algorithm* (TTC). Abdulkadiroğlu and Sönmez (2003) extend their analysis to accommodate a simple affirmative action policy with type-specific quotas. In a recent paper, Kojima (2012) investigates the consequences of these proposed affirmative action policies on students' welfare in a setup where there are two student types (minority and majority) and quotas for majority students only. Surprisingly, he shows that these policies may hurt minority students, the purported beneficiaries. To be more explicit, he finds examples in which all minority students are made worse off under these mechanisms, and he concludes that caution should be exercised when implementing such policies.

Although Kojima (2012) gives some specific scenarios to show that minority students could be worse off under affirmative action policies with majority quotas, in our simulations, we find that this might be a more common phenomenon rather than a peculiarity (see Section 5 for more detail). In some instances, up to 25% of minority students are worse off along with 55% of majority students under such policies with rigid quotas.

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 who can be assigned to c even if there are empty seats.[2] This, in turn, increases the competition for other schools and thus can even make the minority students worse off.

[1]Historically, the affirmative action policies in public school admissions took the form of racial quotas. In 2007, the Supreme Court banned the use of race-based admissions policies (Parents Involved in Community Schools v. Seattle School District No. 1 and Meredith v. Jefferson County Board of Education). This decision shifted the framing of affirmative action policies to promote other measures of diversity, which are not solely based on race or ethnicity.

[2]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).

The problem, however, is not just about setting the appropriate quotas for majorities. The number of minority students who prefer one school to another is not known a priori by the policymakers. Even most intelligent guesses of quota levels are prone to small deviations in minority students' realized desire to attend a particular school, which might cascade inefficiencies throughout the system. Indeed, in our simulations, when we set the majority quotas to the expected levels of majority students, small variations translate into adverse welfare effects in the simulated society. Moreover, these quotas are usually set by third parties such as courts or school districts, which means that they cannot be readjusted easily if schools have empty seats. Therefore, we are in dire need of revisiting the issue of affirmative action for the school choice problem.

In this paper, we circumvent these inefficiencies 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 still be admitted to fill up that school's reserved seats. Therefore, minority reserve mechanisms also avoid wasting the capacity in schools on top of resolving inefficiencies. 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. 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 desirable properties.

1.1 *Main results*

First, 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).[3] Next, the student-proposing deferred acceptance algorithm (DA) with minority reserves is never strictly Pareto dominated by DA with *no affirmative action* for minority students (Theorem 2).[4] When all schools and all students have the same priorities/preferences, then the stable matchings under minority reserves and majority quotas Pareto dominate the stable matching under no affirmative action for minority students (Proposition 2). Furthermore, 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 for minorities (Proposition 4).

[3]Stability, which is a fairness notion, 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. The second property is also called *no justified envy* in the school choice context.

[4]This is in contrast to the result of Kojima (2012) that all minorities can be hurt by an affirmative action policy with majority quotas.

We then analyze the performance of these three policies in the top trading cycles algorithm (TTC). We first show that there is no mechanism that is weakly preferred by all students to TTC with majority quotas and satisfies the desirable properties of TTC (Theorem 3). Next, we introduce TTC with minority reserves that keeps the properties of TTC while giving minorities an edge. Similar to our result for the deferred acceptance algorithm, TTC with minority reserves is never strictly Pareto dominated by TTC with no affirmative action for minority students (Theorem 4). However, 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).

To complement our theoretical results, we devise computer simulations that 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.

1.2 *Related literature*

To study controlled choice, we build on the work of Abdulkadiroğlu and Sönmez (2003), who were the first to approach the school choice problem from a mechanism-design perspective.[5] They propose two celebrated algorithms, the student-proposing deferred acceptance algorithm (DA) and the top trading cycles algorithm (TTC) as alternatives to some popular mechanisms. 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. Moreover, the student-proposing deferred acceptance algorithm is *weakly group strategy-proof*, i.e., there exists no group of students who can jointly manipulate their preferences such that all of them are strictly better off (Dubins and Freedman 1981, Roth 1982a). The TTC was first studied by Shapley and Scarf (1974), who attribute it to David Gale. The TTC is *Pareto efficient*, hence one cannot make any student better off without hurting others. Moreover, it is also *strongly group strategy-proof*, so there exists no group of students who can jointly manipulate their preferences such that all of them are weakly better off and at least one of them

[5]See also Balinski and Sönmez (1999) for a preliminary study. In general, 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 to three excellent reviews for recent applications (Roth 2008, Pathak 2011, Sönmez and Ünver 2011).

is strictly better off (Roth 1982b, Bird 1984, Pycia and Ünver 2011). The main 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.[6,7]

Abdulkadiroğlu and Sönmez (2003) also model a simple affirmative action policy with quotas and show that modified versions of the two aforementioned mechanisms maintain their desirable properties. Subsequently, Abdulkadiroğlu (2005) considers *college admissions* with affirmative action policies where colleges have preferences rather than given priorities. He shows that two assumptions on school preferences are sufficient to recover the desirable properties of the deferred acceptance algorithm.

In an independent work, Westkamp (forthcoming) studies the German university admissions system in which there are transferable quotas on different subpopulations. In this *matching with complex constraints* problem, affirmative action with minority reserves can be accommodated as a special case. However, Westkamp (forthcoming) does not study the welfare effects of affirmative action policies, which is the main question of our work. In another recent paper, Kamada and Kojima (2011) study the Japanese Residency Matching Program, where there are quotas (regional caps) on the number of residents that each region can admit. In the current mechanism, the government sets target capacities for hospitals to implement regional quotas. Instead, Kamada and Kojima propose a new algorithm based on deferred acceptance in which hospitals can admit more than their target capacities as long as regional caps are not violated. They demonstrate that imposing target capacities to satisfy regional quotas may result in avoidable efficiency losses that can be corrected by violating these target capacities. Although the idea of their paper is similar to ours, the setups are completely different (for instance, there are no doctor types in their model) as are the suggested solutions.

In a subsequent paper, Ehlers et al. (2011) consider a controlled school choice model with multiple student types. In their model, each type has floors and ceilings as enrollment targets. They consider these targets both as hard bounds (i.e., feasibility constraints), and as soft bounds that regulate school priorities.[8] With hard bounds, the existence of stable (fair) matchings is not guaranteed. Therefore, they introduce a weaker stability notion and provide an algorithm that finds such matchings. Alternatively, they adapt the deferred acceptance algorithm to soft bounds to get the student-optimal stable matching. However, they do not offer detailed welfare comparison results or simulations as we have done in this paper.

[6] 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 is 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 or inefficient matchings.

[7] 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 (2013) approach the problem from an ex ante perspective instead of randomly breaking ties.

[8] Abdulkadiroğlu (2010) considers only hard bounds in the same model.

From a general perspective, affirmative action has been a source of debate in philosophy, law, and economics since its introduction. It is of great importance that we understand the social and economic effects of affirmative action policies, yet the consequences of these policies receive surprisingly little attention (Sowell 2004). Although there is no consensus on whether affirmative action policies result in overall efficiency gains or losses, affirmative action seems to offer significant redistribution of welfare toward women and minorities with relatively small efficiency consequences (Holzer and Neumark 2000). In the economics of education literature, it has been shown that minority students give importance to the presence of affirmative action policies while deciding on their higher education (Loury and Garman 1993, Arcidiacono 2005). More recently, Bertrand et al. (2010) and Bagde et al. (2011) examine affirmative action programs for lower-caste groups in Indian engineering colleges. They show empirically that affirmative action benefits targeted students.[9]

The main objective of affirmative action policies is to increase diversity of the schools by setting up targets for minority representation. Minority reserve-based affirmative action policies resemble the "soft" quota-based ones where the soft quotas are targets that institutions try to reach but inevitably may fail (Jencks 1992). 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 (preferences of both students and schools can be accessed by an auditor): all admissions are simultaneously done by a central authority and the system is open to legal actions. Hence, the implementation of minority reserves would not lead to hard quotas in the school choice problem because the school districts can openly justify the admission process. The system might fail if some schools discourage minority students applications by other means. This is beyond the scope of our paper, but we believe that the provisions in the legal system prevent such discriminatory practices.[10]

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 concludes. All proofs are given in Appendix A and all supplementary figures are provided in Appendix B.

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

[9] In India, affirmative action policies have been used since the 1930's 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 (2005).

[10] Please see footnotes 1 and 2 for examples.

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

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 (2012), 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 $c \in C$
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 matching depends on the affirmative action policy. Below, we define three different affirmative action policies; 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 not only make the majority students worse off, but also the minority students (Kojima 2012). 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) $s \in S^m$ and $s' \succ_c s$ for all $s' \in \mu(c)$

(ii) $s \in S^M$, $|\mu(c) \cap S^m| > r_c^m$, and $s' \succ_c s$ for all $s' \in \mu(c)$

(iii) $s \in S^M$, $|\mu(c) \cap S^m| \leq r_c^m$, and $s' \succ_c s$ for all $s' \in \mu(c) \cap S^M$.

Condition (i) describes a situation where (c, s) does not form a blocking pair because s is a minority student and c prefers all students in c to s. In condition (ii), whereas blocking does not happen because s is a majority student, the number of minority students in c exceeds minority reserves and c prefers all students in c to s. Finally, in condition (iii), (c, s) does not form a blocking pair 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 last 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 have formed 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 quotas 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 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 represent the outcome of the mechanism by $\phi(\succ_S)$.

A matching mechanism ϕ is *strategy-proof* if for each student s and 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 *weakly group strategy-proof* if for any group of students $\hat{S} \subseteq S$ and 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 weakly 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. In addition, ϕ is *strongly group strategy-proof* if for any group of students $\hat{S} \subseteq S$ and for any $\succ_S$, there exists no $\succ'_{\hat{S}}$ such that $\phi_s(\succ'_{\hat{S}}, \succ_{S \setminus \hat{S}}) \succeq_s \phi_s(\succ_S)$ for all $s \in \hat{S}$ and $\phi_s(\succ'_{\hat{S}}, \succ_{S \setminus \hat{S}}) \succ_s \phi_s(\succ_S)$ for some $s \in \hat{S}$. If a mechanism is strongly group strategy-proof, then there exists no group of students who can jointly change their preference profiles to make each student in the group weakly better off and at least one of them strictly better off. 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 $\psi(\succ_S)$.

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

3. Deferred acceptance algorithm with minority reserves

We first 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 at step $k-1$. 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.[13]

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 weakly group strategy-proof for students.

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 weakly group strategy-proof.*

In the proof, we show that an equivalent way to implement the deferred acceptance algorithm with minority reserves is first to create a new matching problem with no affirmative action and then to apply the original deferred acceptance algorithm to this market.[14] 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

$$s \succ_c' s' \quad \Longleftrightarrow \quad \begin{cases} s \in S^m \quad \text{and} \quad s' \in S^M \\ s, s' \in S^m \quad \text{and} \quad s \succ_c s' \\ s, s' \in S^M \quad \text{and} \quad s \succ_c s'. \end{cases}$$

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,

[13]Note that this algorithm is not equal to the standard deferred acceptance algorithm where for each school c, we modify $\succ_c$ as follows: If minority student s is one of the top r_c^m ranked minority students with respect to $\succ_c$, then she has higher priority than all majority students.

[14]This result also follows from Theorem 2 of Westkamp (forthcoming).

each student keeps the relative rankings of schools the same and prefers the minority-favoring schools over the originals.[15] 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 that 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 that 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 school 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 $\succ_{c_k}$ 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. Moreover, it produces a stable matching under minority reserves (see Appendix A 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.[16]

In our simulations, we found that the positive welfare effects on minority students are substantial, with improvements for up to 30% of minority students (see Section 5 on our simulations). But even more drastic welfare benefits are achieved for majority students, with up to 50% better off under the deferred acceptance algorithm with minority reserves compared to the one with majority quotas.

[15]The relative ranking of the two copies of the same school is not important. All our results hold with the alternative choice.

[16]When each school has a quota of 1, the algorithm corresponds to the decentralized process of offers and acceptances studied in Blum et al. (1997).

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, the Pareto domination relationship in Theorem 1 holds if and only if it holds for the student-optimal stable matchings under the corresponding policies.

Kojima (2012) 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, to the contrary, 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 under 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 weakly group strategy-proof (Proposition 1).

Even though Theorem 2 guarantees only one minority to be weakly better off under the deferred acceptance algorithm with minority reserves compared to that with no affirmative action, in our simulations we found that the number of minority students who are better off is, on average, around 50 times more than those who are worse off under minority reserves. Alternatively, on very peculiar cases, such as the example below, imposing minority reserves can make some minorities worse off while leaving the rest indifferent.

Example 1. Consider the 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 1: $q = (1, 1, 1)$. Students' preferences and schools' priorities are given by the table

$\succ_{s_1}$	$\succ_{s_2}$	$\succ_{s_3}$	$\succ_{c_1} = \succ_{c_2} = \succ_{c_3}$
c_1	c_3	c_1	s_1
c_3	c_1	c_2	s_2
c_2	c_2	c_3	s_3

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

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

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

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

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 the next two subsections, we provide two positive results that guarantee that minorities are better off with minority reserves policies. The first one is obtained by considering common preferences of students together with common priorities of schools, whereas the second one is obtained by considering smart reserves.

3.1 *Common preferences and priorities*

In some countries, such as India (Bertrand et al. 2010), China (Chen and Kesten 2011), and Turkey (Balinski and Sönmez 1999), and some schools in the United States (such as EdOpt schools in New York (Abdulkadiroğlu et al. 2005a)), students take a centralized exam that determine common school priorities over students. Similarly, students may have the same preferences over schools as evidenced by Abdulkadiroğlu et al. (2011). In the next proposition, we consider this case and show that the student-proposing deferred acceptance algorithm with minority reserves Pareto dominates those with no affirmative action and 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 over-demanded, 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 differ only

for majority students. This happens when minority students are exhausted at some step of the serial dictatorship. After this step, more majority students can be admitted with minority reserves than can be with majority quotas, and this makes majority students better off.[17]

3.2 *Smart reserves*

In the absence of assumptions about agents' preferences and priorities, we can guarantee only 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 arbitrarily. However, we now argue that if the reserves are chosen by calculating the number of admitted minority students in a stable matching with no affirmative action, 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 for minorities.

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. Alternatively, 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.

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.*

In Appendix A, we show that whenever minority reserves exceed the number of minority students in μ, then the outcome of the deferred acceptance algorithm with minority reserves is at least as good as the outcome of μ for all minority students.

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

[17]The result that all minority students are weakly better off with minority reserves instead of no affirmative action cannot be made stronger by assuming only common preferences of students or common priorities of schools. Indeed, one can come up with examples showing the contrary.

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.[18]

We have the following corollary to Propositions 3 and 4.

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 either $\mu^r = \mu$ or μ^r Pareto dominates μ for minorities.*

Therefore, if minority reserves are set by calculating the number of admitted minority students in a stable matching with no affirmative action, DA with minority reserves can guarantee better results for minorities (as compared to no affirmative action).

Remark 2. If we set minority reserves to be the capacities for all schools ($r^m = q$), then Proposition 4 implies that 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 guarantees that all minorities are better off.

4. Top trading cycles algorithm with minority reserves

In the previous section, we introduced the deferred acceptance algorithm with minority reserves that improves on the deferred acceptance algorithm with majority quotas (Theorem 1) and keeps the desirable properties of the deferred acceptance algorithm (Proposition 1). Unfortunately, the corresponding result for the top trading cycles algorithm does not hold.

Theorem 3. *There exists no Pareto efficient and strongly group strategy-proof mechanism that is weakly preferred by all students to the top trading cycles algorithm with majority quotas.*

In the proof, provided in Appendix A, we give an example in which either students can jointly manipulate their preferences to get better outcomes or the mechanism assigns an inefficient matching.

In light of Theorem 3, we must give up at least one of the stated properties to get a positive result. Therefore, we keep the desirable properties of the top trading cycles algorithm, namely, strongly group strategy-proofness and Pareto efficiency, while using the minority reserves to give minorities an edge over majorities. We provide the following adaptation of the top trading cycles to minority reserves. Even though the top trading cycles algorithm with minority reserves does not Pareto dominate its majority quotas counterpart, students, on average, are better off (see Section 5).

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

[18]We do not propose a scheme in which DA without affirmative action is run first and then minority reserves are assigned. This scheme may be manipulable. Hence, it is important to use past data.

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. Alternatively, 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 strongly group strategy-proof.*

For Pareto efficiency, note that at each step of the algorithm, students point to the school with empty seats they like the most. 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). For strongly group strategy-proofness, we use an invariance property that the outcome of the algorithm remains the same if the top choice of a student is changed in a certain way; see Appendix A for the detailed proof.

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

Theorem 4. *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 a minority exists among the set of students who are 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 1. Both ψ^r and ψ produce the same matching in the smaller problem that they produce in the larger one. The conclusion follows from this induction argument.

Theorem 4 tells us only that we cannot make all minority students worse off by having minority reserves.[19] However, in our simulations, we found that, on average, up to 80% of minorities are better off compared to less than 1% who are worse off (see Section 5). In addition, we establish that if each school sets a positive minority reserve size then we obtain a stronger result and guarantee that at least some minority students are 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 Appendix A for which one mechanism outcome Pareto dominates the outcome of the other mechanism. A brief discussion about the different results of Proposition 7 and Theorem 1 is in order. Roughly, Theorem 1 obtains by noting that minority reserves does not waste capacity, thus it Pareto improves on majority quotas. The same intuition does not hold in TTC. While applying TTC, although having minority reserves may help minorities by facilitating cycles that are otherwise impossible (since a school with minority reserves points to minorities and not to majorities), some cycles formed earlier in the procedure may involve majority students. That is, some majority students can be assigned to a school in which he/she has a very low priority. This in turn can make some minority students, who are not in earlier cycles, worse off. Alternatively, TTC with majority quotas prevents majority students from pointing to a school that has no majority quotas, assuring that some majority students are worse off, and might make minority students better off. For a specific example, see Example 5 in Appendix A.

[19]The corresponding result does not hold for majority quotas. Consider the example $C = \{c_1, c_2\}$, $S^M = \{s_2\}$, and $S^m = \{s_1\}$. All schools have a capacity of 1, $q = (1, 1)$. Preferences and priorities are given as $c_1 \succ_{s_1} c_2$, $c_2 \succ_{s_2} c_1$, $s_2 \succ_{c_1} s_1$, and $s_1 \succ_{c_2} s_2$. With no affirmative action, both students get their top choices in the 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.

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 to our result for the student-proposing deferred acceptance algorithm (Remark 2).

Example 2. Consider the 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 1, $q = (1, 1, 1)$. Students' preferences and schools' priorities are given by the table[20]

$\succ_{s_1}$	$\succ_{s_2}$	$\succ_{s_3}$	$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$
c_2	c_2	c_3	s_2	s_3	s_1
c_1	c_3		s_1	s_2	s_3
c_3	c_1			s_1	

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

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

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

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

Therefore, in this example, one of the minorities (s_1) 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 with 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 and how many drop in their matches should also be taken into account while determining which policy to use.

To this end, we devise 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 with other policies.[21] We defined utility

[20] In all of the examples, unlisted schools/students are unacceptable to the corresponding agents.

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

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 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. Students are denoted by $s_1, \dots, s_n$ and schools are denoted by $c_1, \dots, c_m$. 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 independent and identically distributed normally distributed random variables with zero mean and variance 1. 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 $(n, m, p, M, r, \alpha, \theta)$ and randomly generate the utility functions. We define the preference order for each student s_i for all pairs of schools $(c_j, c_{j'})$ by using the relation $c_j \succ_{s_i} c_{j'} \iff U_{s_i}(c_j) > U_{s_i}(c_{j'})\ \forall j, j' \in 1, \dots, m$. Similarly, a school's priority order can be determined by comparing utility levels for each student pair $(s_i, s_{i'})$: $s_i \succ_{c_j} s_{i'} \iff U_{c_j}(s_i) > U_{c_j}(s_{i'})\ \forall i, i' \in 1, \dots, n$. For each set of parameters, we perform 100 simulations to capture representative behavior of different matching models. We implement all six matching algorithms in PERL and ran more than five million simulations in total to sample throughout the parameter space.[22]

In our first set of simulations, we set the number of students to $n = 1{,}000$, the number of schools to $m = 20$, each school size to $M = 50$, and the proportion of minority students to $p = 20\%$,[23] and varied minority reserve ratio r,[24] α, and θ. Note that the expected ratio of minority students assigned by $\mathrm{DA_{NAA}}$ and $\mathrm{TTC_{NAA}}$ is equal to p (20%) in each school.

For each simulation result, we show the median of 100 simulations with 25% and 75% quartiles. Initially, we set $r = 20\%$, and change, α and θ from 0 to 1 in steps of 0.1.

[22]Simulation code is available on request.

[23]Our simulation results are robust for various instances of these parameters. We systematically tried different values for these variables; for instance, we changed the number of students to $n = 5{,}000$ and 1,200, the number of schools to $m = 50$, and the proportion of minority students to $p = 15\%$ or used variable reserve sizes for each school, and our conclusions were not affected. Moreover, we also ran simulations when the number of seats was greater than the number of students ($m \times M > n$) and vice versa ($n > m \times M$), but we did not observe any qualitative changes.

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

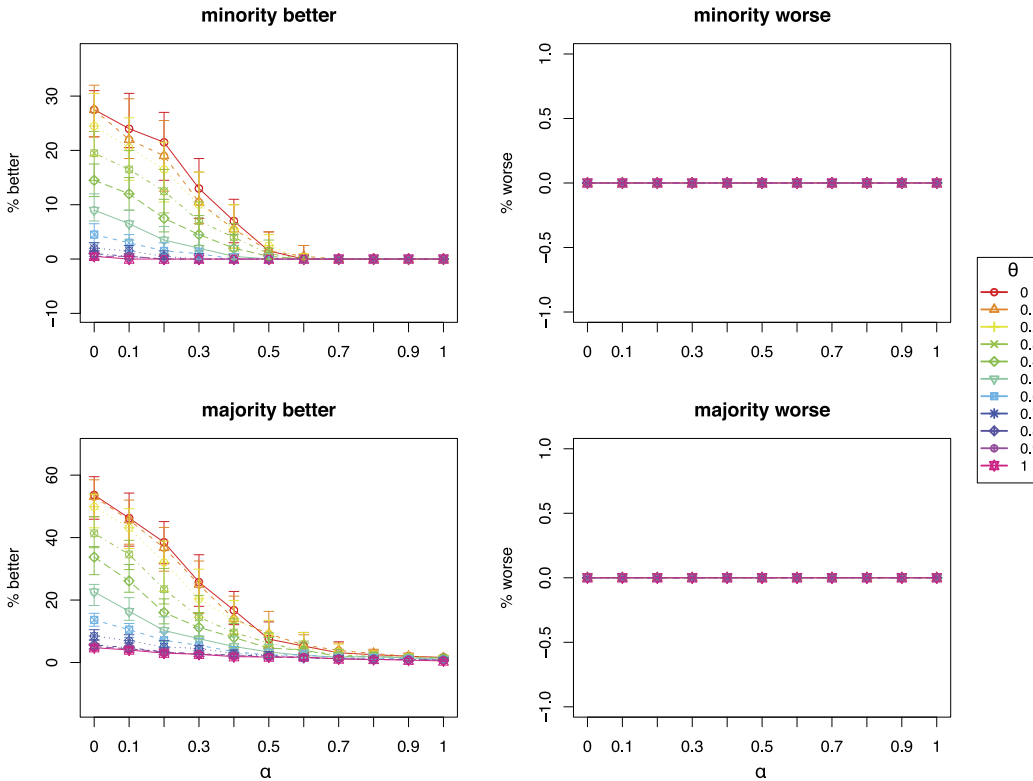

FIGURE 1. Median percentage of minorities and majorities who are better off under $\mathrm{DA}_{\mathrm{MiR}}$ than under $\mathrm{DA}_{\mathrm{MaQ}}$ after 100 simulations. The error bars indicate inter-quartile range. $\mathrm{DA}_{\mathrm{MiR}}$ Pareto dominates $\mathrm{DA}_{\mathrm{MaQ}}$.

We first compare $\mathrm{DA}_{\mathrm{MiR}}$ to $\mathrm{DA}_{\mathrm{MaQ}}$. As a sanity check, our simulations confirm the Pareto dominance of $\mathrm{DA}_{\mathrm{MiR}}$ over $\mathrm{DA}_{\mathrm{MaQ}}$ (Figure 1). 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 $\mathrm{DA}_{\mathrm{MiR}}$ 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* $\mathrm{DA}_{\mathrm{MiR}}$ *in median.* When either school priorities or student preferences are perfectly correlated, both methods give rise to the same assignments for minorities.

Under the same settings, $\mathrm{DA}_{\mathrm{MiR}}$ increases the match quality of 5–40% of minorities in median, but there are some peculiar cases where few minorities are worse off, although the number of minorities who are worse off is not statistically different from 0 (Figure 2). When $\alpha = \theta = 1$, we corroborate the results of Proposition 2, with 40% of minorities being better off in median under the minority reserves policy. Alternatively, $\mathrm{DA}_{\mathrm{MaQ}}$ makes 5–40% of minorities better off, on average, while decreasing the match quality of 5–60% of majorities in median compared to $\mathrm{DA}_{\mathrm{NAA}}$ (Figure 3). *Most surprisingly, for low levels of α and θ, ∼20% of minorities are worse off in median under* $\mathrm{DA}_{\mathrm{MaQ}}$ *than under* $\mathrm{DA}_{\mathrm{NAA}}$*, corroborating that the observations of Kojima (2012) are not peculiarities.*

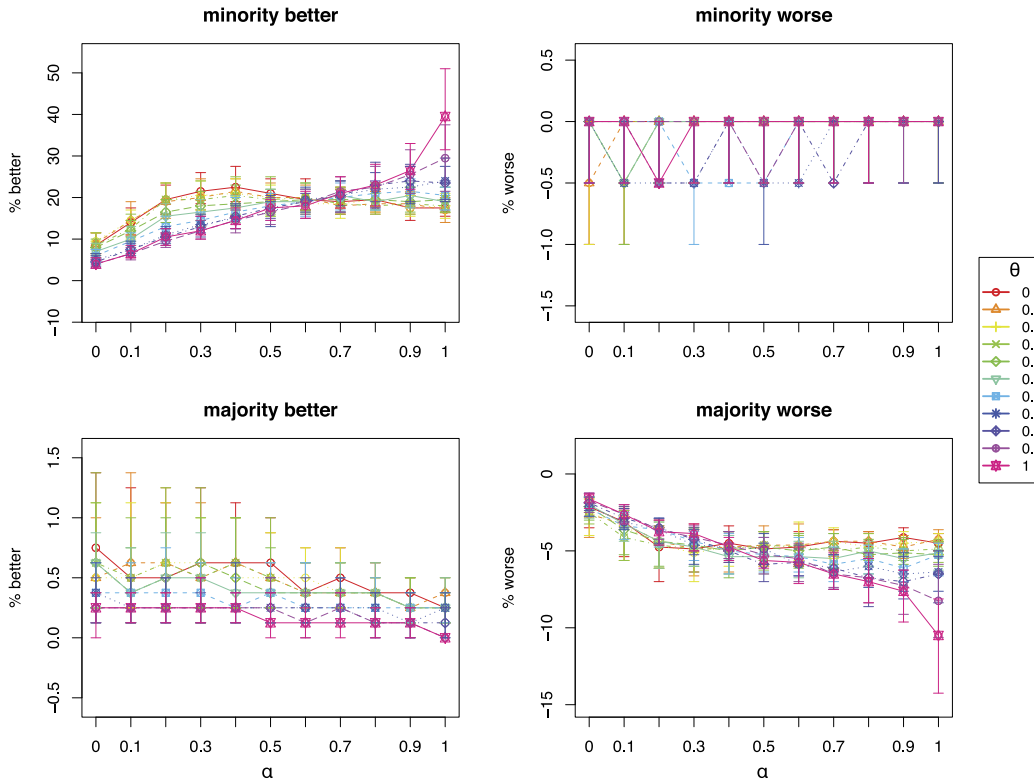

FIGURE 2. Median percentage of minorities and majorities who are better/worse off under DA_{MiR} than under DA_{NAA} after 100 simulations. The error bars indicate interquartile range. Here, there are some cases where minorities can be worse off or majorities can be better off, but neither cases is significant.

The differences between matches of minorities under different TTC algorithms are almost exclusively α dependent and θ independent, showing the power bestowed to students by TTC algorithms (Figure S.1).[25] When $\alpha < 1$ TTC_{MiR} increases the match qualities of minority students compared to both TTC_{MaQ} and TTC_{NAA} more significantly. 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.

Next, we want to assess the effects of setting various reserve sizes. For this purpose, we change the minority reserve ratio, r, from 0% to 20% in steps of 4%, set the α and θ parameters equal, and vary them simultaneously from 0 to 1 in steps of 0.1. When the reserve size is much smaller than the minorities present in the environment (e.g., $0\% < r < 12\%$), we do not see much effect of affirmative action policies, DA_{MiR} and DA_{MaQ}, compared to no affirmative action based policy, DA_{NAA} (Figure S.2). But when

[25]In the main text, we report only the simulations concerning DA mechanisms. We also highlight results for the TTC mechanism and alternative specifications, although summaries of the simulations are delegated to Appendix B as supplementary figures.

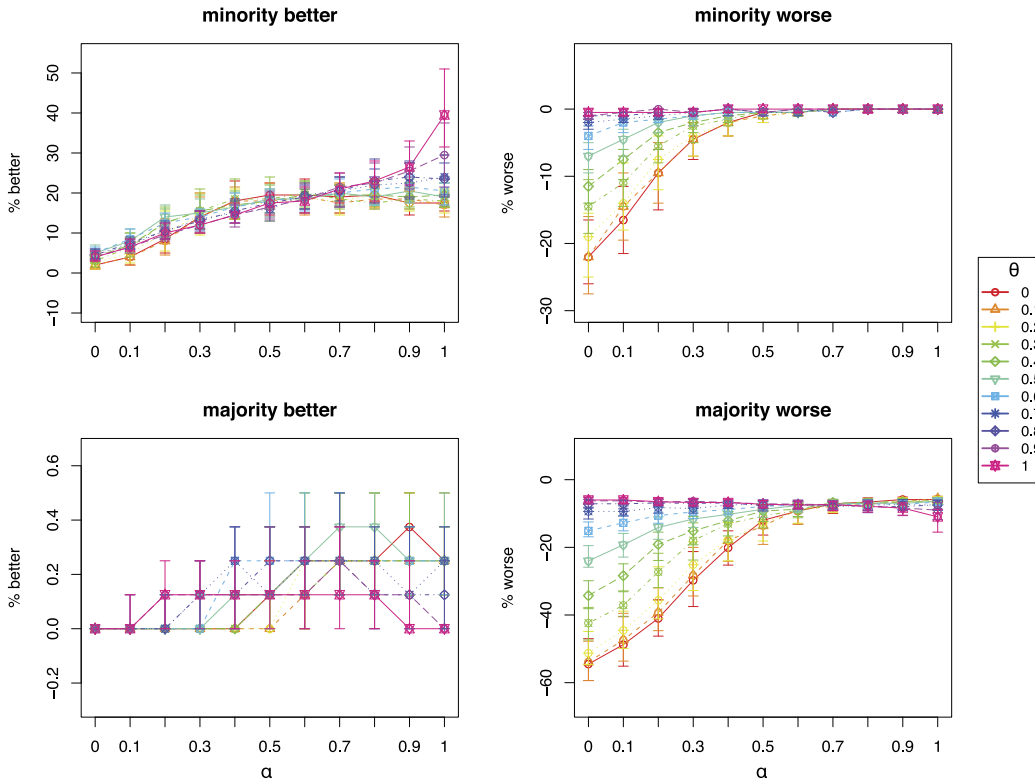

FIGURE 3. Median percentage of minorities and majorities who are better/worse off under DA_{MaQ} than under DA_{NAA} after 100 simulations. The error bars indicate interquartile range. The number of majorities who are better of is not statistically significant, but for small α levels, some minorities can be worse off.

we set $r = 16\%$ and 20%, we start to observe the adverse effects of DA_{MaQ} for both majorities and minorities (Figure S.2). With TTC_{MiR}, however, minority reserve sizes start to show positive redistribution effects for minority students even for very small reserve sizes (Figure S.2).

In real-life situations, affirmative action policies are directed toward groups who tend to be left behind for a variety of reasons. For instance, there might be observed differences in 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 introduce a new variable, $\Delta \leq 0$, the average overall shared preference toward minority students. We can define an updated school utility function for minority students as

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

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

With this new utility function in hand, we first check the effects of Δ and its interactions with α and minority reserve, r, parameters. Initially, we vary Δ from 0 to -2 in

steps of 0.2,[26] vary 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 increase match qualities of minorities up to 80% in median for smaller Δ values compared with their no affirmative action counterparts (Figure S.3). Moreover, as Δ decreases from 0 to -2, the amount of improvements under $\mathrm{DA}_{\mathrm{MiR}}$ compared to $\mathrm{DA}_{\mathrm{MaQ}}$ decreases for both minorities and majorities (Figure S.3).

Next, we set $\alpha = \theta = 50\%$ and analyze the interaction between Δ and minority reserve size r. With decreasing values of Δ, affirmative action policies make minorities better off more dramatically (Figure S.4). For lower values of Δ, positive effects of affirmative action 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.

Last, we compare the student-proposing deferred acceptance algorithms with the top trading cycles algorithms. Overall, the ratio of students who are better off to worse off under $\mathrm{TTC}_{\mathrm{NAA}}$ compared to $\mathrm{DA}_{\mathrm{NAA}}$ is around 4, validating the notion that TTC based algorithms improve the overall social welfare of students (Figure S.1). 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 (Figure S.1).

6. Conclusion

In recent years, public school admissions have been improved by implementing market-design-rooted mechanisms (Abdulkadiroğlu et al. 2005a, 2005b). 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 so as to increase minority welfare. These quotas, when taken as hard feasibility constraints, may lead to negative consequences such as perverse comparative statistics.

Instead of considering these bounds as hard feasibility constraints, we view them as soft regulatory boundaries that regulate school priorities dynamically. With this view, the deferred acceptance algorithm outcome Pareto dominates any stable matching under majority quotas. In addition, simulations show that the new affirmative action policy mitigates the perverse comparative statistics caused by the old one.

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. Moreover, the

[26]A value of $\Delta = -2$ corresponds to the case where the average utility of minorities is 2 standard deviations lower than the average utility of majorities.

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. Therefore, an important contribution of our paper is the analysis of the welfare effects of different affirmative action policies.

Another contribution of this paper is to provide a simulation method to analyze "on average" effects of different affirmative action policies. Simulations allow us to look at some policy-oriented questions that we cannot study by theoretical analysis while enriching our model with realistic features. In future work, we plan to run simulations so as to determine the minority reserves that benefit minorities while minimizing adverse effects on majorities. One can also run simulations using data or more realistic models (such as more student types, floors, and ceilings) to determine the effects of affirmative action policies in different subpopulations.

In conclusion, it is important to mention that our work is a normative study that proposes how affirmative action policies should operate in centralized mechanisms, rather than an analysis or a characterization of affirmative action policies with hard bounds that are used in practice. The proposed affirmative action with minority reserves has clear benefits over majority quotas. For school districts with diversity concerns such as San Francisco or Jefferson County, our work provides an alternative approach for implementation.

Appendix A: Proofs

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

The choice function of a school c, $\mathrm{Ch}_c : 2^S \to 2^S$, is defined as follows. For a given subset of S, say $\hat{S}$, $\mathrm{Ch}_c(\hat{S})$ consists of r_c^m minority students from $\hat{S}$ with the highest priorities if there are enough minority students and of students with the highest priorities from the remaining students in $\hat{S}$ without exceeding q_c.

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

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

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

Therefore, each school's preference with strict priority and minority reserves can also be viewed as a substitutable preference profile. Thus, 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 weakly 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.[27]

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 $\succ_c'$. Hence,

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

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 call 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 are 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^2 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^1. 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^2 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^1. Let μ be a matching in M^1 and let 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, to the contrary, 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 weakly 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, $|\mu(c) \cap S_M| = q_c^M$, 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 strictly better off. To complete the proof, we have to show that μ' is stable under minority reserves.

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

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') forms 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') forms 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) $(\mu'(c') \setminus \mu(c')) \cap S^m \neq \varnothing$, and one of $(\mu'(c') \setminus \mu(c')) \cap S^m$ and c forms a blocking pair for μ under majority quotas or (ii)$(\mu'(c') \setminus \mu(c')) \cap S^m = \varnothing$ and (c', s') forms 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, to the contrary, 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 $\succ'_s$. Thus, none of the minorities is 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 $((\succ'_s)_{s \in S^m}, (\succ_s)_{s \in S^M})$.

Consequently, 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 with minority reserves is weakly group strategy-proof (Proposition 1). □

Proof of Proposition 2. Without loss of generality, relabel schools such that for any $i, j \in \{1, \dots, |C|\}$, all students prefer c_i to c_j if and only if $i < j$. Similarly, relabel students such that for any $i, j \in \{1, \dots, |S|\}$, 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 the top q_{c_1} students, $s_1, \dots, s_{q_1}$, c_2 is matched to the next q_{c_2} students, $s_{q_1+1}, \dots, s_{q_1+q_2}$, 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 kth turn to choose its students.

Majority quotas. In the unique stable matching, c_1 is matched to the top $r^m_{c_1}$ minority students first and then to the top $q^M_{c_1} - r^m_{c_1}$ students among those remaining. Next, c_2 is matched to the top r^m_2 minority students among the remaining minority students and to the 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 kth 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 the top $r^m_{c_1}$ minority students first and then to the top students among those remaining to fill its capacity q_{c_1}. Among the remaining students, c_2 is matched to the top $r^m_{c_2}$ minority students and then to the top students among those remaining 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 the kth 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 give the same outcome step by step 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 with 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 claim. Let M^r_t 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^r_t 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; 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, $m_{t+1}^r = m_{t+1} = \varnothing$ and $M_{t+1}^r = M_{t+1} = \varnothing$, 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 $m_t^r \cup M_t^r$. Consequently, $M_{t+1}^r \supseteq M_{t+1}$.

Since $m_t^r \subseteq m_t$ for all t, each minority student is chosen under minority reserves no later than she is chosen under no affirmative action. Therefore, $\mu^r(s) \succeq_s \mu(s)$ for all $s \in S^m$. □

Proof of Proposition 3. Assume, to the contrary, 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 (otherwise (c, s) is a blocking pair).

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) forms a blocking pair for μ with no affirmative action, which is a contradiction. □

Proof of Proposition 4. We show that the outcome of the student-proposing deferred acceptance algorithm with minority reserves r^m ($\mathrm{DA_{MiR}}$) is at least as good as the outcome of μ for all minorities. Let ν be the outcome of $\mathrm{DA_{MiR}}$, and let ν^k be the tentative matching at step k of $\mathrm{DA_{MiR}}$.

Suppose, to the contrary, that there exists a nonempty set $T \subseteq S^m$ such that for all $s \in T$, $\mu(s) \succ_s \nu(s)$. For each $s \in T$, we have $\mu(s) \in C$. This is because ν is an individually rational matching, i.e., $\nu(s) \succeq_s s$ for all s. Therefore, each $s \in T$ has been rejected by school $\mu(s)$ at some step of the $\mathrm{DA_{MiR}}$. Consider one student in T who has been rejected

at the earliest step, say student s at step k (if there are multiple students rejected at this step, choose one randomly). Denote $\mu(s)$ by c and $\nu(s)$ by c'.

Then we claim that there has to be a new minority student in $\nu^k(c)$ who was not matched to school c in μ. That is, there exists $s' \in S^m$ such that $\nu^k(s') = c$ and $\mu(s') \equiv c'' \neq c$. Otherwise, if the set of students who are matched to c in ν^k is a subset of $\mu(c)$ (i.e., $(\nu^k(c) \cap S^m) \subseteq (\mu(c) \cap S^m)$), then there can be a maximum of $|\mu(c) \cap S^m| - 1$ minority students assigned to c in ν^k since $s \in \mu(c) \setminus \nu^k(c)$. Therefore, school c has not filled its minority reserves in ν^k, which contradicts with the rejection of student s at step k. Consequently, there exists a minority student $s' \in \nu^k(c) \setminus \mu(c)$. But since s is rejected but s' is tentatively accepted to school c in ν^k, we get $s' \succ_c s$. Moreover, we know that the original matching μ is stable. Therefore, for s' not to form a blocking pair with c in μ, s' should be matched to a school that is preferred by s'. Hence, $\mu(s') = c'' \succ_{s'} c$. In $\mathrm{DA}_{\mathrm{MiR}}$, s' applies to schools according to her preference list and since s' is assigned to c by ν^k, it means that s' was rejected by c'' in an earlier stage than k. By construction, s is among the students who were rejected by their original matching in μ at the earliest step. Hence, we get a contradiction. Consequently, we prove that $\nu(s)$ is at least as good as $\mu(s)$ for all $s \in S^m$. □

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 Theorem 3. Suppose, to the contrary, that there exists such a mechanism μ. The proof is by means of an example: $C = \{c_1, c_2, c_3\}$, $S^m = \{s_1, s_2\}$, and $S^M = \{s_3\}$. All schools have a quota of 1: $q = (1, 1, 1)$. Students' preferences and schools' priorities are given by the table

$\succ_{s_1}$	$\succ_{s_2}$	$\succ_{s_3}$	$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$
c_3	c_3	c_1	s_2	s_1	s_3
c_1	c_2	c_2	s_1	s_2	s_1
c_2	c_1	c_3	s_3	s_3	s_2

Majority quotas are given by $q^M = (0, 0, 1)$. If we apply the top trading cycles algorithm with majority quotas, then we obtain the matching ν:

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

There are only two Pareto efficient matchings, say ν^1 and ν^2, that Pareto dominate ν:

$$\nu^1(c_1) = s_3, \qquad \nu^1(c_2) = s_2, \qquad \nu^1(c_3) = s_1$$
$$\nu^2(c_1) = s_1, \qquad \nu^2(c_2) = s_3, \qquad \nu^2(c_3) = s_2.$$

Therefore, with these preferences, either $\mu(\succ_{s_1}, \succ_{s_2}, \succ_{s_3}) = \nu^1$ or $\mu(\succ_{s_1}, \succ_{s_2}, \succ_{s_3}) = \nu^2$. We show that both are impossible.

Case 1 ($\mu = \nu^1$). In this case, s_2 and s_3 can jointly submit the preferences $\succ'_{s_2}$: $c_3 \succ'_{s_2} c_1 \succ'_{s_2} c_2$ and $\succ'_{s_3}$: $c_1 \succ'_{s_2} c_3 \succ'_{s_2} c_2$. The outcome of the top trading cycles algorithm with majority quotas is $\{(c_1, s_2), (c_2, s_1), (c_3, s_3)\}$.[28] There is a unique Pareto efficient matching that improves on this: $\{(c_1, s_3), (c_2, s_1), (c_3, s_2)\}$. Therefore, $\mu(\succ_{s_1}, \succ'_{s_2}, \succ'_{s_3}) = \{(c_1, s_3), (c_2, s_1), (c_3, s_2)\}$. But s_2 strictly prefers $\mu(\succ_{s_1}, \succ'_{s_2}, \succ'_{s_3})$ to $\mu(\succ_{s_1}, \succ_{s_2}, \succ_{s_3})$, while s_3 is indifferent. This is a contradiction since μ is strongly group strategy-proof.

Case 2 ($\mu = \nu^2$). In this case, s_3 can submit the preference $\succ'_{s_3}$: $c_1 \succ'_{s_2} c_3 \succ'_{s_2} c_2$. The outcome of the top trading cycles algorithm with majority quotas is $\{(c_1, s_1), (c_2, s_2), (c_3, s_3)\}$. There is a unique Pareto efficient matching that improves on this: $\{(c_1, s_3), (c_2, s_2), (c_3, s_1)\}$. Therefore, $\mu(\succ_{s_1}, \succ_{s_2}, \succ'_{s_3}) = \{(c_1, s_3), (c_2, s_2), (c_3, s_1)\}$. But s_3 strictly prefers $\mu(\succ_{s_1}, \succ_{s_2}, \succ'_{s_3})$ to $\mu(\succ_{s_1}, \succ_{s_2}, \succ_{s_3})$. This is a contradiction since μ is strongly group strategy-proof (which implies strategy-proofness). □

Proof of Proposition 5. For Pareto efficiency, note that all students who are matched at the first step of the algorithm get their first choice schools, so they cannot be made better off. Similarly, all students who get matched at the next step cannot get into more preferred schools without harming some of the students who are matched in step 1. By induction, students who are matched at step k of the algorithm cannot get into more preferred schools without harming some of the students who are matched before step k, which proves Pareto efficiency.

We prove the group strategy-proofness of the top trading cycles algorithm with minority reserves in three steps. First, we prove individual strategy-proofness. Suppose that μ is the outcome of TTC with minority reserves. Let $\succ'_s$ be a preference relation for student s that assigns the best outcome to student s with respect to true preferences $\succ_s$ (i.e., for all $\widehat{\succ}_s$, $\mu(\succ'_s, \succ_{S\setminus\{s\}}) \succeq_s \mu(\widehat{\succ}_s, \succ_{S\setminus\{s\}})$). Note that by the nature of top trading cycles algorithms, s can get the same outcome as $\mu_s(\succ'_s, \succ_{S\setminus\{s\}})$ by stating $\mu_s(\succ'_s, \succ_{S\setminus\{s\}})$ as the only acceptable choice. Similarly, listing choices that are worse than $\mu_s(\succ'_s, \succ_{S\setminus\{s\}})$ truthfully does not change the outcome. Finally, listing choices that are better than $\mu_s(\succ'_s, \succ_{S\setminus\{s\}})$ truthfully cannot harm s. By construction, listing choices that are better than $\mu_s(\succ'_s, \succ_{S\setminus\{s\}})$ truthfully also cannot improve the outcome of s, so the outcome is the same regardless of how s submits her preferences. Hence, we have $\mu_s(\succ'_s, \succ_{S\setminus\{s\}}) = \mu_s(\succ_s, \succ_{S\setminus\{s\}})$, which proves the strategy-proofness.

Second, we prove the *invariance* property of μ. If a student modifies her submitted preference list by changing only her top school to a better school than assigned by μ or to the school assigned by μ, while keeping the ranking of other schools the same, then the outcome of μ does not change. Formally, for any $a \in C \cup \{s\}$ such that $a \succeq_s \mu_s(\succ) \equiv \hat{c}$,[29] the invariance property requires that $\mu(\succ) = \mu(\succ_s^{*(a)}, \succ_{S\setminus\{s\}})$, where $\succ_s^{*(a)}$ is defined by $a \succeq_s^{*(a)} a'$ for all $a' \in C \cup \{s\}$ and for all $a', a'' \in C \cup \{s\} \setminus \{a\}$, we have $a \succ_s^{*(a)} a'$ if and only if $a \succ_s a'$. This follows from the following observations. If $a = \hat{c}$, by the argument we use

[28] Here, we use an alternative notation for one-to-one matchings.

[29] Here $\hat{c}$ does not have to be a school; it can also be the outside option s.

above, s is assigned to $\hat{c}$ and the choices after $\hat{c}$ do not matter; hence μ executes the same cycles under $\succ$ and $(\succ_s^{*(a)}, \succ_{S\setminus\{s\}})$. If $a \neq \hat{c}$, then by strategy-proofness, we know that s cannot be assigned to any school better than $\hat{c}$ under $(\succ_s^{*(a)}, \succ_{S\setminus\{s\}})$. This means that until all schools better than $\hat{c}$ (under $\succ_s$) exhaust their quotas, the cycles under $\succ$ and $(\succ_s^{*(a)}, \succ_{S\setminus\{s\}})$ are the same. Furthermore, once all of these schools exhaust their quotas, the same cycles are executed under both preference profiles from that point on, since at that point the network graphs are the same, proving the invariance property.

In the third and the last step, we argue that the invariance property implies strongly group strategy-proofness. Suppose that there exist $\{s_1, \dots, s_{|T|}\} \equiv T \subseteq S$, $\succ'_T$, and $\succ$ such that $c'_s \equiv \mu_s(\succ'_T, \succ_{S\setminus T}) \succeq_s \mu_s(\succ) \equiv c_s$ for all $s \in T$.[30] We claim that $c'_s = c_s$ for all $s \in T$. For all $s \in T$, let $\succ_s^{*(c'_s)}$ be the preference profile defined as above and let $\succ_T^* \equiv (\succ_s^{*(c'_s)})_{s\in T}$. Since $c'_{s_1} \succeq c_{s_1}$, by the invariance property, $\mu(\succ_{s_1}^{*(c'_{s_1})}, \succ_{S\setminus\{s_1\}}) = \mu(\succ)$. Then, again by the invariance property, $\mu(\succ_{s_1}^{*(c'_{s_1})}, \succ_{s_2}^{*(c'_{s_2})}, \succ_{S\setminus\{s_1,s_2\}}) = \mu(\succ_{s_1}^{*(c'_{s_1})}, \succ_{S\setminus\{s_1\}})$. By mathematical induction, we conclude that

$$\mu(\succ_T^*, \succ_{S\setminus T}) = \mu(\succ). \tag{1}$$

Similarly, we have $\mu(\succ_{s_1}^{*(c'_{s_1})}, \succ'_{T\setminus\{s_1\}}, \succ_{S\setminus T}) = \mu(\succ'_T, \succ_{S\setminus T})$, $\mu(\succ_{s_1}^{*(c'_{s_1})}, \succ_{s_2}^{*(c'_{s_2})}, \succ'_{T\setminus\{s_1,s_2\}}, \succ_{S\setminus T}) = \mu(\succ_{s_1}^{*(c'_{s_1})}, \succ'_{T\setminus\{s_1\}}, \succ_{S\setminus T})$, and so on. Finally, by mathematical induction,

$$\mu(\succ_T^*, \succ_{S\setminus T}) = \mu(\succ'_T, \succ_{S\setminus T}). \tag{2}$$

Combining (1) and (2), we get $\mu(\succ'_T, \succ_{S\setminus T}) = \mu(\succ)$ and, in particular, $c'_s = c_s$ for all $s \in T$, which shows that μ is strictly group strategy-proof. □

Proof of Theorem 4. 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 that 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 zero minority reserves. Since each agent in $C_1 \cup S_1$ is pointing to her best choice, those agents 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 1. 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 (2012), we show that the top trading cycles algorithm can Pareto dominate, for the minority students, the top-trading cycles algorithm with minority reserves.

[30] As before, c_s and c'_s are not necessarily schools; they can also represent the outside options.

Example 3. Consider the 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 1: $q = (1, 1, 1)$. Students' preferences and schools' priorities are given by the table

$\succ_{s_1}$	$\succ_{s_2}$	$\succ_{s_3}$	$\succ_{s_4}$	$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$
c_1	c_2	c_2	c_3	s_4	s_1	s_1
				s_2	s_3	s_4
						s_2
						s_3

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

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

If we apply the top trading cycles algorithm with minority reserves, then we obtain the matching μ'[31]:

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

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 algorithm with minority reserves can Pareto dominate, for minority students, the top trading cycles algorithm.

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

$\succ_{s_1} = \succ_{s_2}$	$\succ_{c_1} = \succ_{c_2}$
c_1	s_2
c_2	s_1

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

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

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

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

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

[31] This is also the outcome of the top trading cycles with majority quotas as shown by Kojima (2012).

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

Example 5. Consider the 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 1: $q = (1, 1, 1)$. Students' preferences and schools' priorities are given by the table

$\succ_{s_1}$	$\succ_{s_2} = \succ_{s_3}$	$\succ_{c_1}$	$\succ_{c_2}$	$\succ_{c_3}$
c_2	c_1	s_1	s_2	s_3
	c_3			

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 algorithm with majority quotas, then we obtain the matching μ:

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

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

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

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

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

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

$\succ_{s_1}$	$\succ_{s_2}$	$\succ_{c_1}$	$\succ_{c_2}$
c_2	c_1	s_1	s_2
c_1	c_2	s_2	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 algorithm with majority quotas, then we obtain the matching μ:

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

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

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

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

Appendix B: Supplementary figures

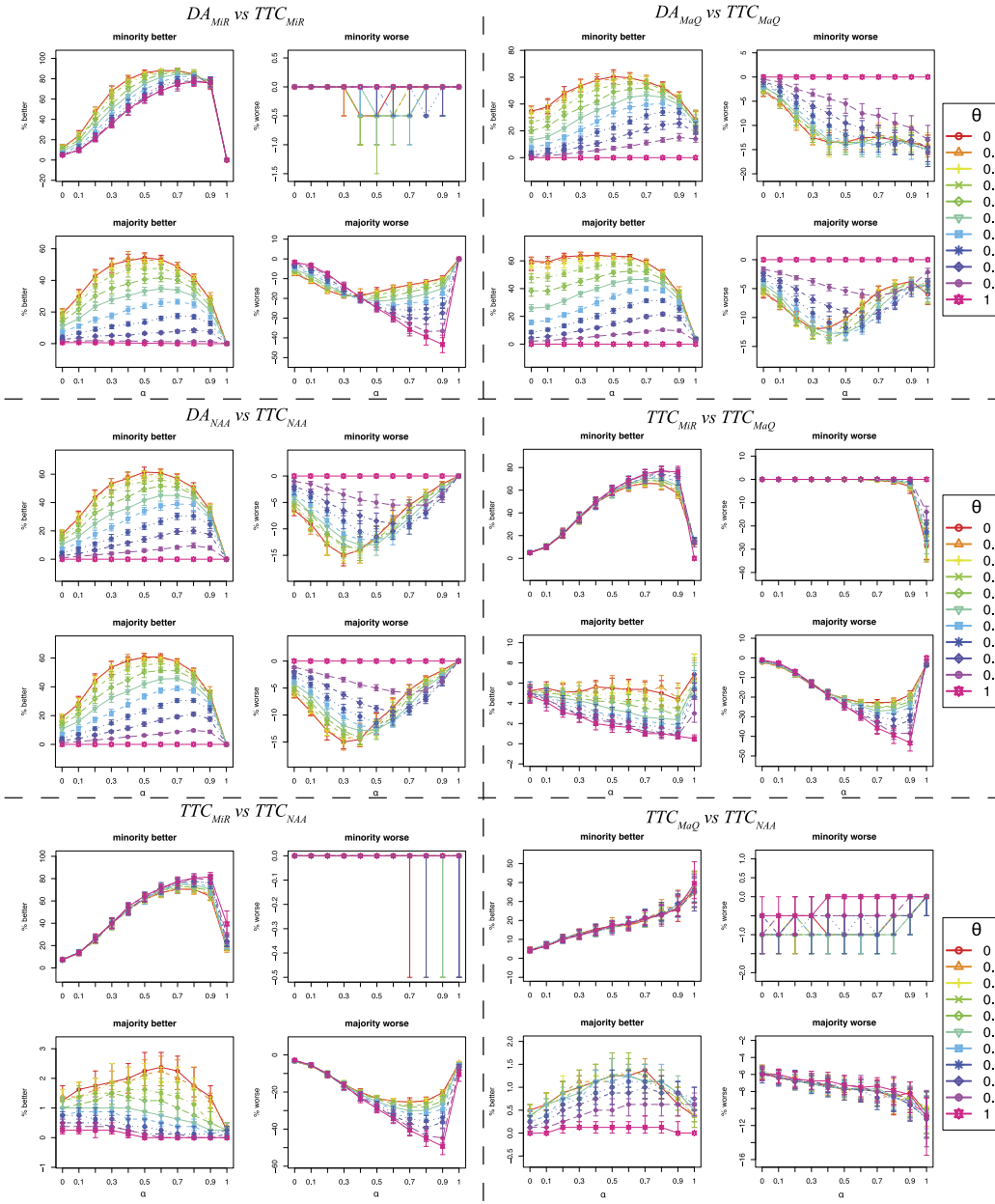

Figure S.1. Median percentage of minorities and majorities who are better/worse off under different mechanisms after 100 simulations. The error bars indicate interquartile range. We set the number of students to $n = 1{,}000$, the number of schools to $m = 20$, each school size to $M = 50$, the proportion of minority students to $r = 20\%$, and the minority reserve ratio to $q = 20\%$, and vary α and θ.

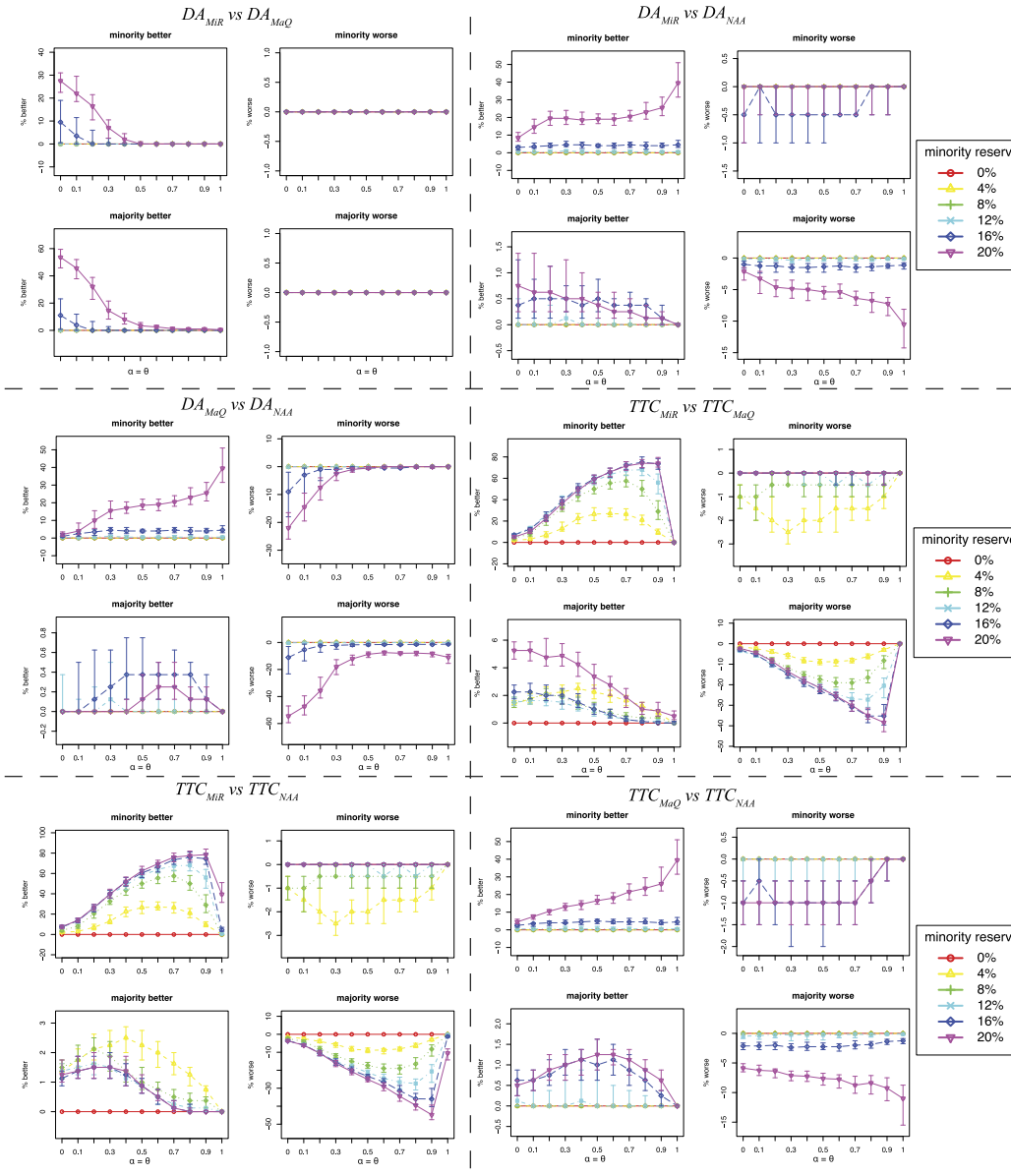

FIGURE S.2. Median percentage of minorities and majorities who are better/worse off under different mechanisms after 100 simulations. The error bars indicate interquartile range. We set the number of students to $n = 1{,}000$, the number of schools to $m = 20$, each school size to $M = 50$, and the proportion of minority students to $r = 20\%$. We set $\alpha = \theta$ and vary them along with the minority reserve ratio.

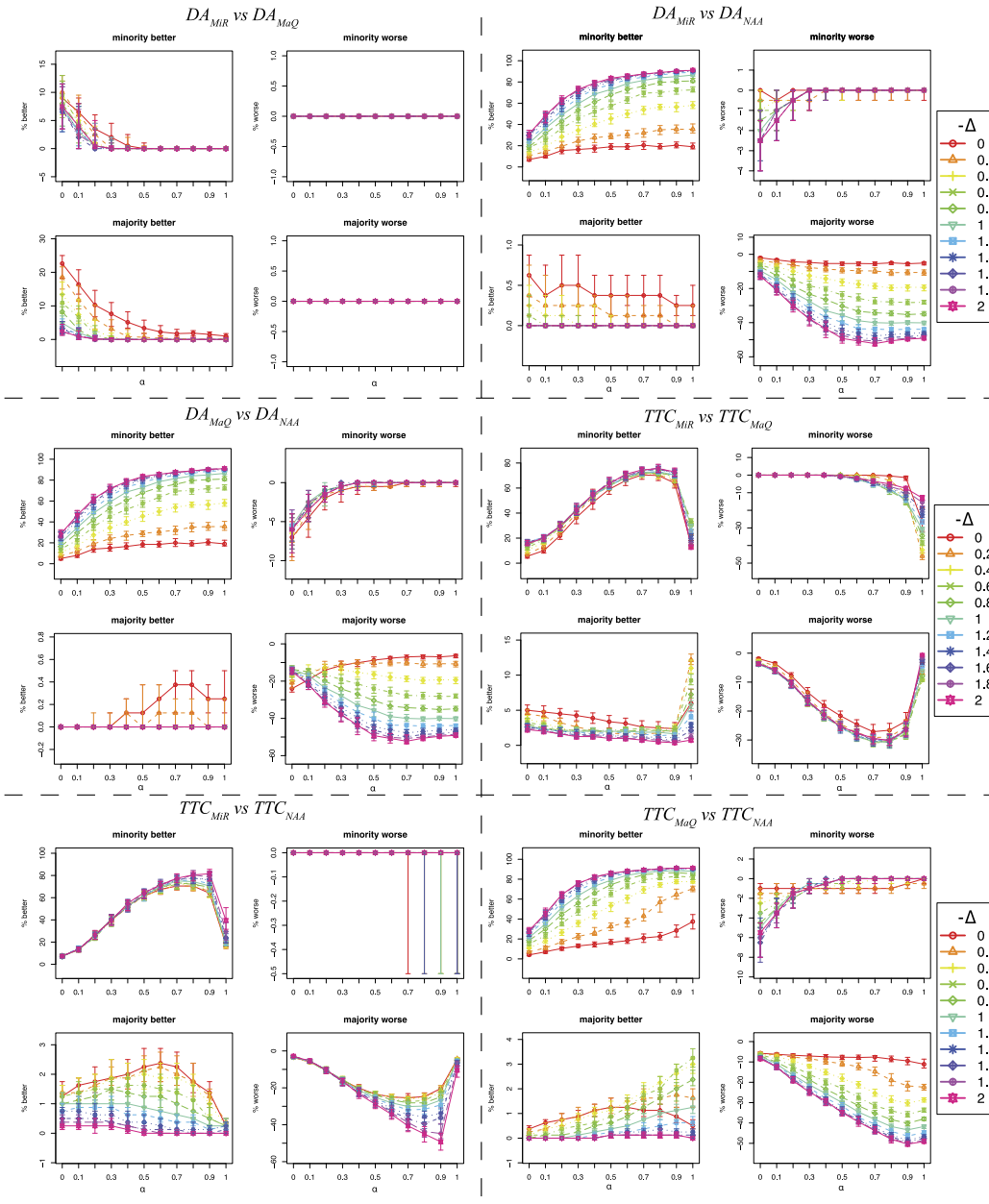

FIGURE S.3. Median percentage of minorities and majorities who are better/worse off under different mechanisms after 100 simulations. The error bars indicate interquartile range. We set the number of students to $n = 1{,}000$, the number of schools to $m = 20$, each school size to $M = 50$, and the proportion of minority students to $r = 20\%$. We introduce a new variable, Δ, which is the average preference of schools toward minority students. We set $\theta = 0.5$ and minority reserve $q = 20\%$, and vary α and Δ.

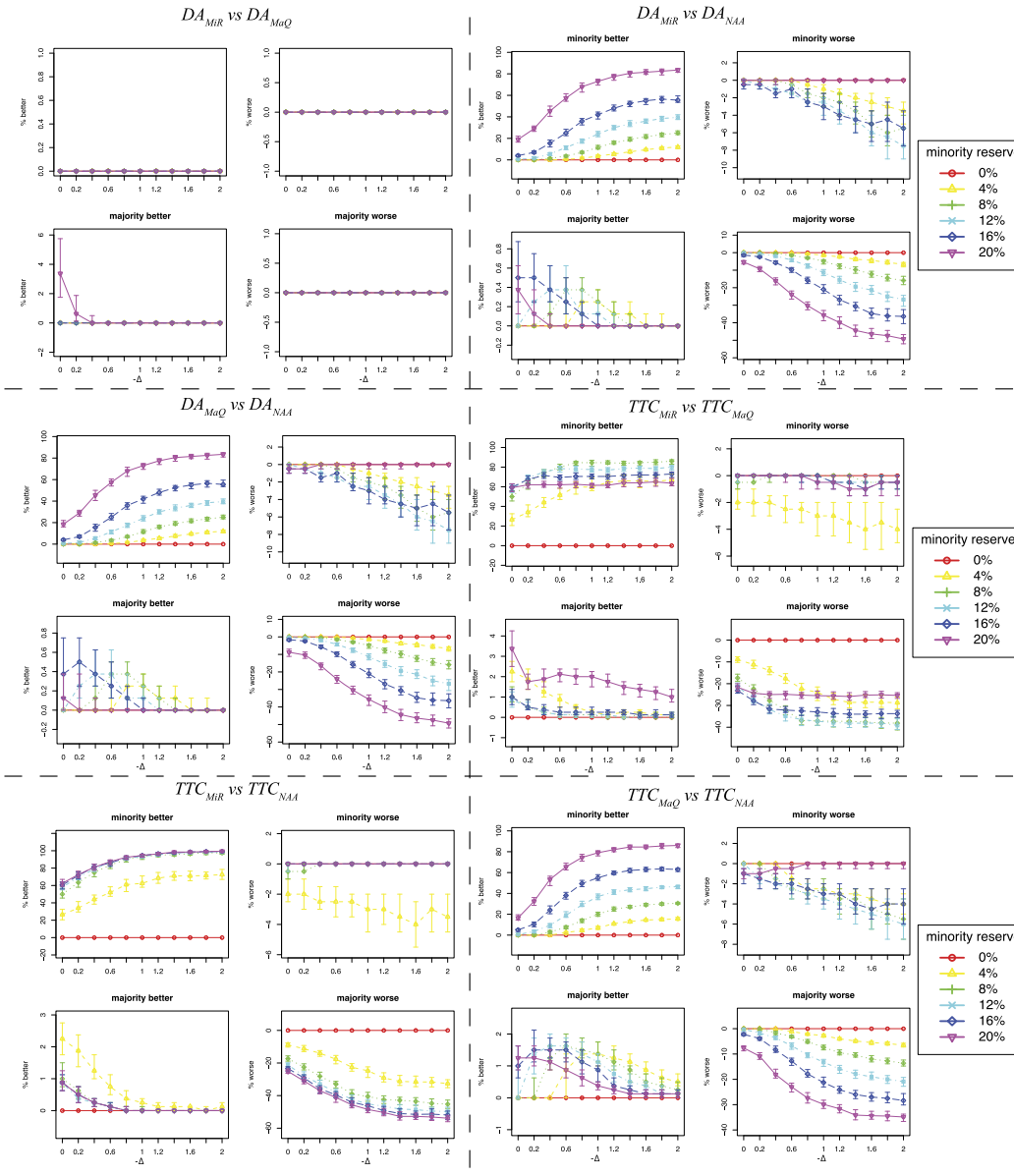

FIGURE S.4. Median percentage of minorities and majorities who are better/worse off under different mechanisms after 100 simulations. The error bars indicate interquartile range. We set the number of students to $n = 1{,}000$, the number of schools to $m = 20$, each school size to $M = 50$, and the proportion of minority students to $r = 20\%$. We introduce a new variable, Δ, which is the average preference of schools toward minority students. We set $\alpha = \theta = 0.5$ and vary Δ along with the minority reserve ratio.

References

Abdulkadiroğlu, Atila (2005), "College admissions with affirmative action." *International Journal of Game Theory*, 33, 535–549. [329]

Abdulkadiroğlu, Atila (2010), "Controlled school choice." Unpublished paper. [329]

Abdulkadiroğlu, Atila, Yeon-Koo Che, and Yosuke Yasuda (2011), "Resolving conflicting preferences in school choice: The 'Boston mechanism' reconsidered." *American Economic Review*, 101, 399–410. [336]

Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth (2005a), "The New York City high school match." *American Economic Review: Papers and Proceedings*, 95, 364–367. [336, 346]

Abdulkadiroğlu, Atila, Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez (2005b), "The Boston Public School match." *American Economic Review: Papers and Proceedings*, 95, 368–371. [346]

Abdulkadiroğlu, Atila and Tayfun Sönmez (2003), "School choice: A mechanism design approach." *American Economic Review*, 93, 729–747. [326, 328, 329, 339]

Arcidiacono, Peter (2005), "Affirmative action in higher education: How do admission and financial aid rules affect future earnings?" *Econometrica*, 73, 1477–1524. [330]

Bagde, Surendrakumar, Dennis Epple, and Lowell J. Taylor (2011), "Dismantling the legacy of caste: Affirmative action in Indian higher education." Unpublished paper. [330]

Balinski, Michel and Tayfun Sönmez (1999), "A tale of two mechanisms: Student placement." *Journal of Economic Theory*, 84, 73–94. [328, 336]

Bertrand, Marianne, Rema Hanna, and Sendhil Mullainathan (2010), "Affirmative action in education: Evidence from engineering college admissions in India." *Journal of Public Economics*, 94, 16–29. [330, 336]

Bird, Charles G. (1984), "Group incentive compatibility in a market with indivisible goods." *Economics Letters*, 14, 309–313. [329]

Blum, Yosef, Alvin E. Roth, and Uriel G. Rothblum (1997), "Vacancy chains and equilibration in senior-level labor markets." *Journal of Economic Theory*, 76, 362–411. [334]

Chen, Yan and Onur Kesten (2011), "From Boston to Shanghai to deferred acceptance: Theory and experiments on a family of school choice mechanisms." Unpublished paper. [336]

Chen, Yan and Tayfun Sönmez (2006), "School choice: An experimental study." *Journal of Economic Theory*, 127, 202–231. [341]

Deshpande, Ashwini (2005), "Affirmative action in India and the United States." World Development Report Background Paper, World Bank, Washington, DC. [330]

Dubins, Lester E. and David A. Freedman (1981), "Machiavelli and the Gale–Shapley algorithm." *American Mathematical Monthly*, 88, 485–494. [328, 348]

Ehlers, Lars, Isa E. Hafalir, M. Bumin Yenmez, and Muhammed A. Yildirim (2011), "School choice with controlled choice constraints: Hard bounds versus soft bounds." Unpublished paper. [329]

Erdil, Aytek and Haluk Ergin (2008), "What's the matter with tie-breaking? Improving efficiency in school choice." *American Economic Review*, 98, 669–689. [329, 341]

Fryer, Roland G. (2009), "Implicit quotas." *Journal of Legal Studies*, 38, 1–20. [330]

Gale, David and Lloyd S. Shapley (1962), "College admissions and the stability of marriage." *American Mathematical Monthly*, 69, 9–15. [328]

Haeringer, Guillaume and Flip Klijn (2009), "Constrained school choice." *Journal of Economic Theory*, 144, 1921–1947. [329]

Hatfield, John W. and Fuhito Kojima (2009), "Group incentive compatibility for matching with contracts." *Games and Economic Behavior*, 67, 745–749. [348]

Holzer, Harry and David Neumark (2000), "Assessing affirmative action." *Journal of Economic Literature*, 38, 483–568. [330, 346]

Jencks, Christopher (1992), *Rethinking Social Policy: Race, Poverty, and the Underclass*. Harvard University Press, Cambridge, Massachusetts. [330]

Kamada, Yuichiro and Fuhito Kojima (2011), "Improving efficiency in matching markets with regional caps: The case of the Japan residency matching program." Unpublished paper. [329]

Kesten, Onur (2006), "On two competing mechanisms for priority-based allocation problems." *Journal of Economic Theory*, 127, 155–171. [329]

Kesten, Onur (2010), "School choice with consent." *Quarterly Journal of Economics*, 125, 1297–1348. [329]

Kesten, Onur and M. Utku Ünver (2013), "A theory of school–choice lotteries." Unpublished paper. [329]

Kojima, Fuhito (2012), "School choice: Impossibilities for affirmative action." *Games and Economic Behavior*, 75, 685–693. [326, 327, 331, 335, 343, 354, 355]

Loury, Linda Datcher and David Garman (1993), "Affirmative action in higher education." *American Economic Review: Papers and Proceedings*, 83, 99–103. [330]

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, 115–128. [348]

McVitie, D. G. and L. B. Wilson (1970), "Stable marriage assignment for unequal sets." *BIT*, 10, 295–309. [334]

Pathak, Parag A. (2011), "The mechanism design approach to student assignment." *Annual Review of Economics*, 3, 513–536. [328]

Pycia, Marek and M. Utku Ünver (2011), "Trading cycles for school choice." Unpublished paper. [329]

Roth, Alvin E. (1982a), "The economics of matching: Stability and incentives." *Mathematics of Operations Research*, 7, 617–628. [328]

Roth, Alvin E. (1982b), "Incentive compatibility in a market with indivisible goods." *Economics Letters*, 9, 127–132. [329]

Roth, Alvin E. (2008), "Deferred acceptance algorithms: History, theory, practice, and open questions." *International Journal of Game Theory*, 36, 537–569. [328]

Roth, Alvin E. and Marilda A. Oliveira Sotomayor (1990), *Two-Sided Matching: A Study in Game-Theoretic Modelling and Analysis*. Cambridge University Press, Cambridge. [328, 347]

Shapley, Lloyd S. and Herbert E. Scarf (1974), "On cores and indivisibility." *Journal of Mathematical Economics*, 1, 23–37. [328]

Sönmez, Tayfun and M. Utku Ünver (2011), "Matching, allocation, and exchange of discrete resources." In *Handbook of Social Economics*, volume 1B (Jess Benhabib, Matthew O. Jackson, and Alberto Bisin, eds.), 781–852, North-Holland, San Diego. [328]

Sowell, Thomas (2004), *Affirmative Action Around the World*. Yale University Press, New Haven. [330]

Westkamp, Alexander (forthcoming), "An analysis of the German university admissions system." *Economic Theory*. [329, 333]

Submitted 2011-12-18. Final version accepted 2012-4-26. Available online 2012-4-26.