College Admissions with Affirmative Action*

Atila Abdulkadiroğlu
Department of Economics
Columbia University
New York, NY 10025

February 2002, This version: September 2003

Abstract

This paper first shows that, when colleges' preferences are substitutable, there does not exist any stable matching mechanism that makes truthful revelation of preferences a dominant strategy for every student. It introduces student types and captures colleges' preferences towards affirmative action via type specific quotas: A college always prefers a set of students that respects its type specific quotas to another set that violates them. Then it shows that a stable mechanism that makes truthful revelation of preferences a dominant strategy for every student exists if each college's preferences satisfy responsiveness over acceptable sets of students that respect its type specific quotas. These results have direct policy implications in several entry-level labor markets (Roth 1991). Furthermore, the algorithm, the related incentive theory and a fairness notion developed here is applied to controlled choice in the context of public school choice by Abdulkadiroğlu and Sönmez (2003).

*I am grateful to Al Roth for his valuable feedback. I would also like to thank Ron Jones, Bahar Leventoglu, Paul Milgrom, Tayfun Sönmez and William Thomson for their helpful comments. I would like to thank Alfred P. Sloan Foundation for its research fellowship.

1 Introduction

In several real-life applications of college admissions problems, colleges' preferences over sets of students are determined by gender, racial and ethnic composition of sets. Roth (1991) makes the following observation in an entry-level labor market in United Kingdom: A doctor in UK can become eligible for full registration with the General Medical Council only if that doctor completes 12 months in a preregistration position, typically six months in a medical position and six months in a surgical position under consultants. The entry-level labor markets for unregistered doctors are cleared via centralized matching mechanisms. In the Edinburgh case, some consultants may specify that they will not employ more than one female doctor in any six-month period. In some other applications, preferences are determined by the composition of professional specialities of students. Roth and Peranson (1999) point out the following in the American resident matching market that is cleared via a centralized resident matching procedure: For particular residency programs, if some positions remain unfilled, then these positions may revert to other programs (see also Roth 2002, Milgrom 2003).

A similar case arises in public school choice in the United States (Abdulkadiroğlu and Sönmez 2003): Public school choice gives parents the opportunity to choose the public school their child will attend. However, in some states, choice is limited by court-ordered desegregation guidelines. In Missouri, for example, St. Louis and Kansas City must observe strict racial guidelines for the placement of students in city schools. Donald Hirch (1994, page 120) points out similar constraints in UK: City Technology Colleges are required to admit a group of students from across the ability range and their student body should be representative of the community in the catchment area.

The fact that colleges have preferences over different compositions of students is acknowledged by Lee C. Bollinger, the former president of the University of Michigan and the current president of Columbia University, as well. Professor Bollinger argues that "[a]dmissions is not and should not be a linear process of lining up applicants to their grades and test scores and then drawing a line through the list. It shows the importance of seeing racial and ethnic diversity in a broader context of diversity, which is geographic and international and socio-economic and athletic and all various forms of differences, complementary differences, that we draw on to compose classes year after year" (Alkan and Gale 2001).

In fact, we find the following in a standard application form for Columbia University:[1] "Columbia attempts to draw students from diverse ethnic and racial backgrounds. We ask you to assist us in this effort by describing yourself as a member of one of the following groups. Please add more specific information where relevant (such as tribal affiliation or country of origin)." Students can select one of the following in the form: African American/Black, Asian/Asian American/Pacific Islander, Biracial/Multiracial, Caucasian, Chicano/Mexican American, Dominican, Hawaiian Native/Alaskan Native, Hispanic/Latino, Native American/American Indian, Puerto Rican, South Asian, Southeast Asian, Other.

In each of these examples, preferences of hospitals/schools/colleges are quite different than a simple ordering of individual students. In this paper, we give a class of preferences for colleges that, we believe, capture preferences towards gender or racial and ethnic compositions, which we refer as preferences towards affirmative action.

A college admissions problem is a many-to-one, two-sided matching problem, in which there is a finite set of students and a finite set of colleges. Each college has a finite capacity to enroll students. Preference relation of each student over colleges is a linear order of colleges, where as preference relation of each college over *sets* of students is a linear order of *sets* of students. A matching matches each student with a college or with herself, and each college with a set of students that are no more than the capacity of that college. A student blocks a matching if she prefers to remain unmatched rather than being matched to a college under that matching. A college blocks a matching if it rather prefers a strict subset of the students that it is matched. A student-college pair blocks a matching if the student prefers that college to her match and the college rather prefers her with a subset of its match. A matching is stable if no student, no college and no student-college pair block the matching.

Faced with a set S of students, a college c can determine which subset of S it most prefers. We refer this subset as c's choice among S. A college c has substitutable preferences when its preferences over sets of students satisfy the following condition for every set S of students: When a student s is in c's choice among S, s is in c's choice among $S - s'$ for any other student $s' \neq s$. That is, the college continues to prefer to admit a student even if some of the other students in its choice set become unavailable. It regards students

[1] The form can be downloded from the official web site of Columbia University.

as substitutes not as complements. When colleges have substitutable preferences, the set of stable matchings is non-empty (Kelso and Crawford 1982, Roth 1991, Alkan and Gale 2001, Milgrom 2003).[2] Furthermore, there exists an algorithm, namely a deferred acceptance algorithm with students proposing, that selects a stable matching for every preference profile, such that every student likes this matching as well as any other stable matching (Theorem 6.8, Roth and Sotomayor 1990). Similarly, there exists an algorithm, namely a deferred acceptance algorithm with colleges proposing, that selects a stable matching for every preference profile, such that every college likes this matching as well as any other stable matching (Theorem 6.7, Roth and Sotomayor 1990). For later reference, let us give the deferred acceptance algorithm with students proposing:

Step 1: Each student proposes to her most preferred college. Each college rejects all but those in its choice among its proposers.[3]

In general, at

Step k: Each student who was rejected in the previous step proposes to her next preferred college. Each college considers the students it has been holding together with its new proposers. It rejects all but those in its choice among these students.

The algorithm terminates when no student proposal is rejected. Then each student is matched with the college that she proposes last and not rejected by. We refer this algorithm as the Gale-Shapley student optimal algorithm.

A (direct) stable mechanism is a preference revelation game in which students and colleges report their preferences, and a matching that is stable with respected to the stated preferences is produced. Roth (1982) shows that there is no stable mechanism that makes truthful revelation of preferences a dominant strategy for *all agent*.

A college c's preferences over sets of students induce an ordering of individual students, which we refer as c's preferences over individual students. A student s is ranked higher than s' in that ordering if, when c is faced with options of "enrolling s only" and "enrolling s' only", c prefers to enroll s

[2]In fact, Proposition 3 of Roth (1990) proves existence of stable matchings in a more general many-to-two matching environment with substitutable preferences.

[3]Roth (1991) gives an algorithm for the case where colleges may have indeferences among different sets of students. This simplified version of Roth's algorithm works in our case since we assume linear preferences. Note that we follow Milgrom (2003)'s terminology in defining the algorithm.

only. s is ranked below c if c prefers to keep all positions unfilled rather than enroll s only. We say that c's preferences over *sets* of students are responsive to its preferences over *individual* students if the following holds: For any two sets S and S' of students such that $S' = (S - s) \cup x$ for some $s \in S$ and $x \in (S - s) \cup c$: The college c prefers S to S' if and only if s is ranked higher than x in c's preferences over individual students. When c's preferences are responsive and it has q_c position to fill, the choice of c among a set S is the q_c highest ranked students among those that are ranked above c in c's preference relation over individual students. Hence, responsive preferences are also substitutable but not vice versa.

When colleges' preferences are *responsive*, a stable mechanism that is coupled with the Gale-Shapley student optimal algorithm makes truthful revelation of preferences a dominant strategy for *all students* (Dubins and Freedman 1981, Roth 1982). However, Milgrom (2003) shows that this mechanism fails to make truthful revelation of preferences a dominant strategy for all students, when colleges have *substitutable* preferences. Our first theorem strengthens this negative result as follows: When colleges' preferences are substitutable, there does not exist any stable mechanism that makes truthful revelation of preferences a dominant strategy for all students (Theorem 1).

This negative result brings up the following question: Is there a non-trivial subclass of substitutable preferences that capture preferences towards affirmative action, and at the same time allows for existence of a stable mechanism that is incentive compatible for every student? Our answer to this question positive. We capture preferences towards affirmative action via type specific quotas as follows: There exists a finite type space, such as {African American/Black, Asian/Asian American/Pacific Islander, ..., Other} in case of Columbia University. Each student is of one of these types. In addition to its capacity, each college has a type specific quota. We assume for each college that, facing two sets of students, one respecting its type specific quotas, the other violating them, the college prefers the former to the latter. In this case, we say that the former set respects affirmative action constraints at that college. We refer to this assumption as AA (resembling "**A**ffirmative **A**ction").

Furthermore, we impose responsiveness on a college's preferences only over sets of students that satisfy AA and that the college would prefer to enroll: For any two sets S and S' of students such that both S and S' respect affirmative action constraints at c, c prefers both S and S' to leaving

all positions unfilled, $S' = (S - s) \cup x$ for some $s \in S$ and $x \in (S - s) \cup c$: The college c prefers S to S' if and only if s is ranked higher than x in c's preferences over individual students. We refer this assumption as RR (resembling "**R**estricted **R**esponsiveness"). Responsiveness implies RR, but a preference profile that satisfies RR may fail to be responsive.

Roth (1991) makes similar assumptions in a many-to-two matching market when the type space contains two types, namely {female, male}, in his Proposition 6. Our assumptions AA and RR are generalizations of Roth's assumption in a many-to-one matching market with possibly more than two types. We refer to these problems as college admissions with affirmative action problems.

When colleges' preferences satisfy AA and RR, they are substitutable (Lemma 1), so that the set of stable matchings is non-empty (Theorem 2). Furthermore, the Gale-Shapley student optimal algorithm given above produces a stable matching that every student likes as well as any other stable matching (Theorem 3). Once we have substitutability, Theorem 2 and Theorem 3 are straightforward implications of the existing results in the literature.

When colleges' preferences satisfy AA and RR, the direct mechanism that is coupled with the Gale-Shapley student optimal algorithm makes truthful revelation of preferences a dominant strategy for all students (Theorem 4). We will refer this mechanism as the Gale-Shapley student optimal mechanism. Theorem 4 provides a positive dominant strategy result as opposed to Milgrom's negative result and our even stronger result in Theorem 1. Also, the positive dominant strategy result with responsive preferences that is due to Dubins and Freedman 1981 and Roth 1982 becomes a corollary of our theorem with a singleton type space. The proof of Theorem 1 reveals that once we impose AA, RR becomes almost a necessary condition for the existence of a stable mechanism that makes truthful revelation of preferences a dominant strategy for all students.

Next we apply our theory to the National Resident Matching Program (NRMP) and public school choice in the US.

For particular residency programs in the NRMP, if some positions remain unfilled, then these positions may revert to other programs. But this is equivalent to saying that a hospital prefers candidates of a certain program over the candidates of another program. So (i) setting two type specific quotas, one for each program, and each quota being equal to the capacity constraint, (ii) obtaining a one single ranking of all candidates, and (iii) applying the Gale-Shapley student optimal mechanism with these quotas

would solve this allocation problem.

Abdulkadiroğlu and Sönmez (2003) introduces a controlled choice problem in the context of public school choice in the US. A controlled choice problem is essentially a college admissions with affirmative action problem with one distinction: In college admissions, colleges themselves are agents which have preferences over students, whereas in school choice, schools are objects to be consumed by the students and the ranking of students at each school does not represent preferences of that school over individual students but a priority ordering of students imposed by state or local laws. While stability is key in a college admissions problem, it is *fairness* that plays a crucial role in controlled choice: An assignment of schools to students is fair if the following is satisfied: If there is an unmatched student-school pair (s, c) where student s prefers school c to her assignment and she has higher priority than some other student s' who is assigned a seat at school c then (i) students s and s' are of different types, and (ii) the quota for the type of student s is met at school c. It is easy to show that an assignment in controlled choice problem is fair if and only if it is stable in the corresponding college admissions with affirmative problem (Theorem 5).[4] Then our Theorems 3 and 4 imply that the Gale-Shapley student optimal mechanism produces a fair assignment that every student likes as well as any other fair assignment. Furthermore, it makes truthful revelation of preferences a dominant strategy for all students (Theorem 6). Since schools are not strategic agents, this immediately implies that this mechanism is strategy-proof.

We introduce our model and give our results in Section 2. We devote Section 3 to the proofs of our results. In Section 4, we apply our theory to the NRMP and controlled public school choice in the US. We discuss further theoretical and practical issues in Section 5.

2 The Model and the Results

2.1 College Admissions

A **college admissions problem** consists of:

1. a finite set of students $S = \{s_1, \ldots, s_n\}$;

[4]The original observation that connects stability in two sided matching problem to fairness in a one sided matching problem is made by Balinski and Sönmez (1999) in the context of Turkish college admissions.

2. a finite set of colleges $C = \{c_1, \ldots, c_m\}$,

3. a capacity vector $q = (q_{c_1}, \ldots, q_{c_m})$ where q_c is the capacity of college $c \in C$,

4. a list of strict preference profile $P^S = (P_{s_1}, \ldots, P_{s_n})$, where P_s is the strict preference relation of student $s \in S$ over $C \cup \{s\}$,

5. a strict preference profile $P^C = (P_{c_1}, \ldots, P_{c_m})$, where P_c is the strict preference relation of college $c \in C$ over subsets of S.

Each preference relation is assumed to be complete and transitive. The first four items are common in every paper on college admissions. We will comment on the fifth item after giving the building blocks. From now on, small letters will represent individual agents and singleton sets of individuals, whereas capital letter will represent non-singleton sets.

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

i. $|\mu(s)| = 1$ for every student s, and $\mu(s) = s$ if $\mu(s) \notin C$;

ii. $\mu(c) \subset S$ and $|\mu(c)| \leq q_c$ for every college c;

iii. $\mu(s) = c$ if and only if $s \in \mu(c)$.

$\mu(s)$ denotes the college that student s is enrolled in; $\mu(c)$ denotes the set of students college c enrolls.

Let $Ch_c(S')$ denote the most preferred subset of $S' \subset S$ for college c, i.e. $Ch_c(S') \subset S'$ and for any other $\hat{S} \subset S'$, $Ch_c(S') P_c \hat{S}$. We will refer $Ch_c(S')$ as **c's choice among** S'.

A matching μ is **blocked by a student** s if s prefers to stay unmatched rather than be matched to $\mu(s)$, i.e. $sP_s\mu(s)$. It is **blocked by a college** c if c prefers a strict subset of $\mu(c)$ to $\mu(c)$, i.e. $\mu(c) \neq Ch_c(\mu(c))$. It is **blocked by a student-college pair** (s, c) if s and c are not matched by μ but would both prefer if s was enrolled in c, i.e. $\mu(s) \neq c$, $cP_s\mu(s)$ and $s \in Ch_c(\mu(c) \cup s)$.

A matching μ is **stable** if it is not blocked by any individual agent or any student-college pair.

College c has **substitutable preferences** if for any $S' \subset S$, $s' \in S'$, $s'' \in S - s'$, when s' is in $Ch_c(S')$, s' is in $Ch_c(S' - s'')$ as well. The set of stable matchings is nonempty when every college has substitutable preferences (Kelso and Crawford 1982, Roth 1991, Alkan and Gale 2001, Milgrom 2003).

Furthermore, consider the following deferred acceptance algorithm with students proposing:

Step 1: Each student proposes to her most preferred college. Each college rejects all but those in its choice among its proposers.

In general, at
Step k: Each student who was rejected in the previous step proposes to her next preferred college. Each college considers the students it has been holding together with its new proposers. It rejects all but those in its choice among these students.

The algorithm terminates when no student proposal is rejected. Then each student is matched with the college that she proposes last and not rejected by.

When all the colleges have substitutable preferences, this algorithm produces a stable matching that every student likes as well as any other stable matching (Theorem 6.8, Roth and Sotomayor 1990, p.176). Henceforth we refer this algorithm as the **Gale-Shapley student optimal stable algorithm** (GS^S).

A (direct) **mechanism** requires agents to reveal their preferences and selects a matching based on these submitted preferences according to a predetermined algorithm. A mechanism is (dominant strategy) **incentive compatible for an agent** if reporting her true preferences is a dominant strategy for that agent in the preference revelation game induced by that mechanism. A **stable mechanism** is a direct mechanism that selects a matching that is stable with respect to the stated preference profile.

It is well known that there does not exist any stable mechanism that is incentive compatible for every agent (Theorem 3, Roth 1982). However, there exist restrictions on preferences that is sufficient for the existence of a stable mechanism that is incentive compatible for every student. Before discussing these restrictions, we need to introduce some definitions.

Each P_c induces a preference relation for c over $S \cup c$, i.e. *singletons of students and c*. In that representation, c is interpreted as leaving all positions at c unfilled. Then, when the alternatives c faces are only singletons and leaving all the positions unfilled, $sP_cs'P_ccP_cs''$ is read as follows: College c prefers to enroll s rather than s'; c prefers to enroll s' rather than leave positions unfilled; and c prefers to leave positions unfilled rather than enroll s''. We refer to this preference relation as c's **preferences over individual students**.

Now consider a **standard college admissions problem**, in which the fifth item of a college admissions problem is replaced by a list of *strict* preference relations for colleges over individual students. Then a simple assumption of *responsiveness* is used to connect colleges' preferences over groups of students to their preferences over individual students: c's preference relation over groups of students is **responsive** (to its preferences over individual students) if, for any $S', S'' \subset S$, $s' \in S'$, $s'' \in S - S'$, (i) whenever $S'' = S' - s'$, c prefers S' to S'' if and only if c prefers to enroll s' rather than leave positions unfilled with respect to its preferences over *individual students*; (ii) whenever $S'' = (S' - s') \cup s''$, c prefers S' to S'' if and only if c prefers to enroll s' rather than s'' with respect to its preferences over individual students. In this case, we say that c has **responsive preferences**. If c has responsive preferences, then $Ch_c(S')$ is the smaller of the following two sets: (i) the first q_c highest ranked students in c's preference ordering of individual students; (ii) all the students c prefers to enroll rather than leave positions unfilled.

A responsive preference relation is substitutable. Therefore, the above claims about GS^S are valid in a standard college admission problem as well[5]. Furthermore, the direct mechanism coupled with GS^S, which we refer as GS^S **mechanism**, is incentive compatible for every student when all colleges have responsive preferences (Theorem 5.16, Roth and Sotomayor 1990). However, Milgrom (2003) shows that this result does not generalize to substitutable preferences. We strengthen this negative results as follows:

Theorem 1: When colleges can admit any substitutable preferences, there does not exist any stable mechanism that is incentive compatible for every student.

Proof: The proof is via a counterexample. There are three students $S = \{s_1, s_2, s_3\}$ and two colleges $C = \{c_1, c_2\}$ with capacities $q_{c_1} = 2$ and $q_{c_2} = 1$.

[5]When colleges' preferences are responsive to their preferences over individual students, one does not need to assume strict preferences over sets of students. Strict preferences over individual students suffices. For further discussion, see Roth and Sotomayor (1990) page 129.

Consider the following preference profile P:

P_{s_1}	P_{s_2}	P_{s_3}
c_1	c_2	c_2
c_2	c_1	c_1
s_1	s_2	s_3

P_{c_1}	P_{c_2}
s_3	s_1
$\{s_1, s_2\}$	s_2
s_1	s_3
s_2	$\varnothing$
$\varnothing$	

Note that both P_{c_1} and P_{c_2} are substitutable. There is a unique stable matching μ for P: $\mu(c_1) = s_3$, $\mu(c_2) = s_1$ and $\mu(s_2) = s_2$, i.e. s_2 remains unmatched.

Now consider the preference profile $P' = (P_{-s_2}, P'_{s_2})$ where P'_{s_2} reverses the ranking of colleges, i.e. $c_1 P'_{s_2} c_2 P'_{s_2} s_2$. There are two stable matchings $\mu'_1 = \mu$ and μ'_2 for P' : $\mu'_2(c_1) = \{s_1, s_2\}$ and $\mu'_2(c_2) = s_3$.

Next consider the preference profile $P'' = (P'_{-s_3}, P''_{s_3})$ where P''_{s_3} ranks c_1 as "unacceptable," i.e. $c_2 P''_{s_3} s_3 P''_{s_3} c_1$. There is a unique stable matching $\mu'' = \mu'_2$.

Suppose in contrary that there is a stable matching mechanism m that is incentive compatible for every student. If $m(P') = \mu'_1$, then when the true preference profile is P', s_3 would be better off by misrepresenting her preferences as P''_{s_3}, since then m must pick the unique stable matching μ'' under P'', i.e. $m(P'') = \mu''$, and $\mu''(s_3) = c_2 P'_{s_3} c_1 = \mu'_1(s_3)$. This is a contradiction. If $m(P') = \mu'_2$, then when the true preference profile is P, s_2 would be better off by misrepresenting her preferences as P'_{s_2}, since m must pick the unique stable matching under P, i.e. $m(P) = \mu$, and $\mu'_2(s_2) = c_1 P_{s_2} s_2 = \mu(s_2)$. This is a contradiction.

To complete the proof, let us show why we need at least three students and at least two colleges one with capacity of at least two in that counterexample:
(i) If there is one single college, there is no room for beneficial misrepresentation of preferences by students. Because, stability implies that the college will be matched with its choice among those who prefer that college.
(ii) If there is one single students, there is no room for beneficial misrepresentation of preferences by that student. Because, stability implies that the student has to be matched with her most preferred one among those that prefer to enroll her.
(iii) When there are two students $\{s, s'\}$ and at least two colleges, we will show that there is no room for beneficial misrepresentation of preferences

by any student. Consider a stable mechanism. Fix colleges' preferences at P_C. Suppose that this mechanism produces the matching μ when preferences are given by $P = (P_s, P_{s'}, P_C)$. For notational simplicity, let $\mu(s) = c$ and $\mu(s') = c'$. Also suppose in contrary that s can benefit by misrepresenting his preferences as $\hat{P}_s$. Let $\hat{\mu}$ be the matching that the stable mechanism produces under $\hat{P} = (\hat{P}_s, P_{s'}, P_C)$. Then $\hat{\mu}(s) P_s c$ so that $\hat{\mu}(s) \neq c$. If $\hat{\mu}(s) = c'' \neq c'$, then (s, c'') blocks μ under P. So $\hat{\mu}(s) = c'$. Also, $c' P_{s'} \hat{\mu}(s')$. Otherwise: (a) If $\hat{\mu}(s') = \mu(s') = c'$, stability of $\hat{\mu}$ under $\hat{P}$ implies that $\{s, s'\} P_{c'} s'$. Then (s, c') blocks μ under P. (b) If $\hat{\mu}(s') P_{s'} c'$ and $\hat{\mu}(s') \neq c$, then $(s', \hat{\mu}(s'))$ blocks μ under P. (c) If $\hat{\mu}(s') P_{s'} \mu(s')$ and $\hat{\mu}(s') = c$, then stability of μ under P implies that $s P_c \{s, s'\}$. Since $c \hat{P}_s c'$, (s, c) blocks $\hat{\mu}$ under $\hat{P}$. This proves that $c' P_{s'} \hat{\mu}(s')$, which in turn implies that $\hat{\mu}(s') \neq c'$. Stability of μ under P implies that $s' P_{c'} s$. But then, (s', c') blocks $\hat{\mu}$ under $\hat{P}$. So, s cannot benefit from misrepresentation. This proves our claim that there is no room for beneficial misrepresentation of preferences by any student when there are two students.
(iv) When every college has a capacity of one, colleges' preferences are necessarily responsive, so GS^S is incentive compatible.

This completes the proof. ■

Under P', the students prefer μ'_2 to μ'_1. Milgrom (2003) uses P and P' to show that GS^S is not incentive compatible when colleges can admit any substitutable preferences.

Note that even if we assume responsive preferences, there does not exist a stable mechanism that is incentive compatible for every college (Proposition 2, Roth 1985).

Then, our negative result brings up the following question: Is there a non-trivial subclass of substitutable preferences that capture preferences towards affirmative action, and at the same time allows for existence of a stable mechanism that is incentive compatible for every student? In the next section we characterize such a subclass of preferences.

2.2 College Admissions with Affirmative Action

In addition to the five items in a college admissions problem, a **college admissions with affirmative action problem** consists of

6. a type space $T = \{\tau_1, ..., \tau_k\}$

7. a type function $\tau : S \to T$; $\tau(s)$ is the type of student s

8. for each college c, a vector of type specific quotas $q_c^T = (q_c^{\tau_1}, ..., q_c^{\tau_k})$ such that $q_c^\tau \leq q_c$ for each c, each τ; and $\sum_{\tau \in T} q_c^\tau \geq q_c$. Sometimes, we will refer to these quotas as affirmative action constraints.

Each student has a type. We use type specific quotas to capture colleges' preferences towards affirmative action. We interpret q_c^τ as the maximum number of slots that college c would like to allocate to type τ students. To be precise, let us give our assumptions on the class of preferences more formally.

A list of students $S' \subset S$ **respects affirmative action constraints at college** c if

i. S' respects the capacity limit at c, that is $|S'| \leq q_c$;

ii. S' respects type specific quotas at c, that is $|\{s \in S' : \tau(s) = \tau\}| \leq q_c^\tau$ for each $\tau \in T$.

The following assumption will give us the restriction on preferences imposed by affirmative action constraints.

Assumption AA: (Affirmative Action) For any $S', S'' \subset S$ such that S' respects affirmative action constraints and S'' does not respect affirmative action constraints, $S' P_c S''$.

Next we impose responsiveness only on "acceptable" sets of students that respect affirmative action constraints.

Assumption RR: (Restricted Responsiveness) For any $S', S'' \subset S$ such that $S' P_c \varnothing$, $S'' P_c \varnothing$, and both S' and S'' respect affirmative action constraints,

(RP1) if $S'' = S' - s'$ for some $s' \in S'$, then for each $c \in C$

$$S' P_c S'' \text{ if and only if } s' P_c c;$$

(RP2) if $S'' = (S' - s') \cup s''$ for some $s' \in S', s'' \in S - S'$. Then, for each $c \in C$

$$S' P_c S'' \text{ if and only if } s' P_c s''.$$

RR reduces to Martinez et al. (2000)'s q_F-responsiveness property when the type space is a singleton. Note that this assumption does *not* imply the responsiveness property. To see this, consider the following example.

Example: There are two female students $\{f_1, f_2\}$ and two male students $\{m_1, m_2\}$. A college c has two seats, $q_c = 2$; and it prefers to enroll at most one student of each gender, i.e. $q_c^f = 1$ and $q_c^m = 1$. Otherwise, its preference relation over groups of students is responsive to the following ranking: $f_1 P_c f_2 P_c m_1 P_c m_2 P_c c$. Then $\{f_1, m_2\} P_c \{f_1, f_2\}$ but $f_2 P_c m_2$. So, c's preferences are not responsive although it satisfies RR.

However, assumptions AA and RR imply substitutability:
Lemma 1: If c's preference relation P_c satisfies AA and RR, then P_c is substitutable.

Proof: Suppose that P_c satisfies AA and RR. Take any $S' \subset S$, $s' \in S'$, $s'' \in S - s'$. Suppose that s' is in $Ch_c(S')$. Since $\varnothing$ satisfies AA and it is feasible to choose, AA implies that $Ch_c(S')$ respects affirmative action constraints, so does $Ch_c(S') - s''$. Furthermore, since s' is in $Ch_c(S')$, RR implies that $s' P_c \hat{s}$ for any other $\hat{s} \in S' - Ch_c(S')$ such that $(Ch_c(S') - s') \cup \hat{s}$ respects affirmative action constraints. Then RR implies that s' should be in $Ch_c(S' - s'') \subset (Ch_c(S') - s'')$. Hence, P_c is substitutable. ■

Roth (1991) obtains a similar result in a many-to-two matching model, when students can be one of two types {male, female} and colleges may specify that they will employ no more than one female students. Then college preferences satisfying this constraint but otherwise responsive to a simple rank-ordering are substitutable (Proposition 6, Roth 1991). Lemma 1 generalizes this observation in a many-to-one matching framework when student types may be more than two. Then we can deduce the following result:

Theorem 2: Set of stable matchings is nonempty when colleges' preferences satisfy AA and RR.

Substitutability is sufficient for the existence of a stable matching (Roth 1991, Alkan and Gale 2001, Milgrom 2003)[6]. So the proof of Theorem 1 follows directly from Lemma 1.

The following result follows from Theorem 6.8 of Roth and Sotomayor (1990).

[6] Also see Roth and Sotomayor (1990), Proposition 5.22.

Theorem 3: When colleges' preferences satisfy AA and RR, GS^S produces a stable matching that every student likes as well as any other stable matching.

Later, we will construct the GS^S algorithm for our environment. Then, we will give direct proofs of these results. The proofs will enhance our understanding of the algorithm that will be helpful in the proof of the following result.

Theorem 4: When colleges' preferences satisfy AA and RR, the GS^S mechanism is incentive compatible for every student.

Theorem 4 provides a positive dominant strategy result as opposed to Milgrom's negative result and our even stronger result in Theorem 1. Also, the positive dominant strategy result with responsive preferences that is due to Dubins and Freedman 1981 and Roth 1982 becomes a corollary of our theorem with a singleton type space. Note also that once we impose AA, RR becomes almost necessary for the existence of a stable mechanism that makes truthful revelation of preferences a dominant strategy for all students if and only if colleges' preferences satisfy RR.[7]

We devote the next section to the proofs of Theorems 2, 3 and 4.

3 Proofs

First, let us give the following definitions:

If $cP_s s$, we say that c is **acceptable** to s, i.e. s prefers to be matched to c rather than be unmatched. If $sP_s c$, then c is **unacceptable** to s, i.e. s prefers to be unmatched rather than be matched to c. s is acceptable to s.

[7] Assume that, in every problem we study, there exists a type such that there are at least three students of that type, two colleges, one with a quota of at least two, the other with a positive quota for that type. Note that each of these colleges may have positive quotas for other types as well. Then, the following strong result holds: Assume that colleges' preferences satisfy AA. For a given set of students, colleges, capacities and affirmative action constraints, there exists a stable mechanism that is incentive compatible for every student if and only if colleges' preferences satisfy RP. The proof follows from the proof of Theorem 1. This assumption is likely to be satisfied in all relevant real-life applications of college admissions, since the number of colleges, the number of positions and the number of students are large in such applications.See Roth and Peranson (1999) for data on the American National Resident Match Program, see Rees (2000) for data on school choice in the US.

Similarly, if sP_cc, s is **acceptable** to c. If cP_cs, s is **unacceptable** to c. c is acceptable to c. We can easily generalize this definition for sets of students.

A matching μ **respects affirmative action constraints** if for each $c \in C$, $\mu(c)$ respects affirmative action constraints at c. **The quota for type τ at college c is met under μ** if $|\{s \in \mu(c) \text{ and } \tau(s) = \tau\}| = q_c^\tau$. A matching μ is **individually rational** if (i) for each $s \in S$, $\mu(s)$ is acceptable to s; (ii) for each $c \in C$, $\mu(c)$ respects affirmative action constraints, and $\mu(c) = Ch_c(\mu(c))$.

By Assumptions AA and RR, we can give the following equivalent stability definition:

Definition: A matching μ is **stable** if

i. it is individually rational;

ii. there do not exist $s, s' \in S$, $c \in C$ such that

(a) $\mu(s') = c$

(b) $cP_s\mu(s)$,

(c) $(\mu(c) - s') \cup s$ respects affirmative action constraints at c,

(d) sP_cs'.

Now consider the following equivalent variant of GS^S :

Step 1: Each student proposes to her most preferred college among acceptable ones. Each college c orders the individual students that propose to c with respect to P_c. Then, c tentatively admits one acceptable student at a time in this order such that c does not exceed its capacity and c respects type specific quotas. If the type specific quota for type τ is met at c, all the remaining type τ students are rejected by c. If the capacity is met at c, all the remaining students are rejected by c.[8] If no student that proposes to some college is rejected, then GS^S terminates, and the matches are finalized. Otherwise, GS^S proceeds to step 2.

In general at
Step $k > 1$: Each student, who has been rejected at step $k-1$, proposes to her most preferred college among acceptable ones by which she has not been

[8] This is equivalent to finding the choice set among proposers under the assumptions AA and RP.

rejected yet. Each college c orders, with respect to P_c, the students that c has tentatively admitted at step $k-1$ and the students that propose to c at step k. Then, c tentatively admits one acceptable student at a time in this order such that c does not exceed its capacity and c respects type specific quotas. If the type specific quota for type τ is met at c, all the remaining type τ students are rejected by c. If the capacity is met at c, all the remaining students are rejected by c. If no student that proposes to some college is rejected, then GS^S terminates, and the matches are finalized. Otherwise, GS^S proceeds to step $k+1$.

Let $P = (P^S, P^C)$ be given, and $GS^S(P)$ denote both the procedure and the matching that GS^S produces under P. We will drop the argument when it will not cause any confusion.

3.1 Existence

We will give the proofs of Theorem 2 and Theorem 3 below for the sake of completeness, although, as mentioned in the previous section, they are straightforward implications of the previous results in the literature.

The following observations about GS^S will be helpful later: Take two students s and s', and a college c.

Observation 1. Suppose that s and s' are of the same type. Let c prefer s to s'. If s is rejected by c at some step k, then s' is not tentatively matched to c at step k or later.

Observation 2. If s is rejected by c at step k, then one of the following is true:

(a) The type specific quota for s's type at c is not met at the end of this step, i.e. the number of students who are of the same type as s and tentatively matched to c at this step is less than the quota for s's type. Then, c prefers s less than each student who is tentatively matched to c at step k or later.

(b) Or, the type specific quota for s's type at c is met, i.e. the number of such students is equal to the quota for s's type. Then c prefers s less than any student of type $\tau(s)$ that is tentatively admitted by c at step k or later. .

Proposition 1: $\mu = GS^S(P)$ is stable with respect to P.

Proof: First, if a student is rejected at some step, she proposes to a less preferred college or does not propose to any college at the next step. By finiteness of agents, this guarantees that GS^S terminates at finite steps. Next, each student proposes only to acceptable colleges and each college admits only its choice among proposers. So, $GS^S(P)$ is individually rational. Finally, suppose in contrary that $GS^S(P)$ is not stable. Then there exists $s, s' \in S$, $s \neq s'$ such that

1. $c = \mu(s')$
2. $cP_s\mu(s)$,
3. $(\mu(c) - s') \cup s$ respects affirmative action constraints at c,
4. sP_cs'.

Once s is rejected by c at step k of $GS^S(P)$, by Observation 1, all other students of type $\tau(s)$ that are assigned a slot at c at step k or later are preferred to s by c. So, (4) implies that $\tau(s') \neq \tau(s)$. Then, the quota for type $\tau(s)$ at c is not met under μ, by (3) above. Let step k' of $GS^S(P)$ be the first step after step $k-1$ at which the quota for type $\tau(s)$ at c is not met. Any type $\tau(s)$ student who is rejected at this step, and s as well, will be preferred by c less than any other student who is tentatively matched to c at this step or later, by Observation 2a. So s' can be matched to c only if $s'P_cs$. This contradicts with (3) above. So, the proof is completed. ■

Theorem 1 is an immediate implication of this result.

3.2 Optimality

Define the available set of colleges for s as follows: $A(s) = \{c \in C \cup s : \exists \mu,\ \mu$ is stable, $\mu(s) = c\}$. Let $c(s)$ denote the best alternative in $A(s)$ with respect to s's preference ordering.

Proposition 2: GS^S matches each s to $c(s)$, i.e. GS^S produces a stable matching that every student likes as well as any other stable matching.

Proof: Let μ^k be the tentative matching produced by $GS^S(P)$ at the end of step k. We will show the following: If a student s is rejected by a college

at some step of GS^S, this college is not in s's available set $A(s)$. In contrary, suppose that there exists some student who is rejected by some college in his available set. Let k be the first step of GS^S such that some student s is rejected by some c at step k, and there exists a stable matching ν such that $\nu(s) = c$. (Note that a student is never rejected by himself, if he proposes to himself.)

Our choice of k implies that $\mu^k(s')R_{s'}\nu(s')$ for each $s' \neq s$. Because, if $\nu(s')P_{s'}\mu^k(s')$ for some s', then s' should have been rejected by $\nu(s')$ at a step before step k, then this would contradict with the choice of k. In particular, for each s' with $\mu(s') = c$, either $\nu(s') = c$ or $\mu^k(s') = cP_{s'}\nu(s')$. There are two possible cases:

1) The quota for type $\tau(s)$ at c is not met under μ^k. Then, (i) the capacity limit at c, q_c, is met under μ^k; and (ii) c prefers s less than each student who is tentatively matched to c at step k, i.e. for every $s' \in \mu^k(c)$, we have $s'P_cs$. But then, for each s' with $\mu^k(s') = cP_{s'}\nu(s')$, the quota for type $\tau(s')$ at c should be met under ν. Otherwise, (s', c) would block ν since (i) $(\nu(c)-s)\cup s'$ respects affirmative action constraints at c since the quota for $\tau(s')$ at c is not met under ν, (ii) $cP_{s'}\nu(s')$, and (iii) $s'P_cs$. This contradicts with stability of ν. Then, this implies that if $\mu^k(s') = cP_{s'}\nu(s')$ then

$$\left|\left\{s'' : \tau(s'') = \tau(s') \text{ and } \mu^k(s'') = c\right\}\right| \leq \left|\{s'' : \tau(s'') = \tau(s') \text{ and } \nu(s'') = c\}\right| = q_c^{\tau(s')}$$

for each such student. That is for any s' such that $\mu^k(s') = c$ but $\nu(s') \neq c$, removing s' from $\mu^k(c)$ while switching from μ^k to ν does not free up a slot at c. For other students $s' \in \mu^k(c)$, we have that $\nu(s') = c$. Moreover, remember that the capacity limit at c, q_c, is met under μ^k in this case. These imply that the capacity limit at c, q_c, is met under ν. Since $\mu^k(s) \neq c$ and $\nu(s) = c$ and the capacity limit at c is met under both μ^k and ν, there exists some s' such that $\mu^k(s') = c$ and $\nu(s') \neq c$. If $\tau(s') = \tau(s)$, then (i) $(\nu(c) - s) \cup s'$ respects affirmative action constraints at c, (ii) $cP_{s'}\nu(s')$, and (iii) sP_cs'. This contradicts with stability of ν. If $\tau(s') \neq \tau(s)$, removing any such s' from $\mu^k(c)$ while switching from μ^k to ν does not free up a slot at c by the above observation. So $\nu(s) = c$ implies that the number of students in $\nu(c)$ would exceed q_c. Contradiction.

2) The quota for type $\tau(s)$ at c is met under μ^k. Then for each $s' \in \mu^k(c)$ with $\tau(s') = \tau(s)$, we have that $s'P_cs$. Among such students, there should exist some s' such that $\nu(s') \neq c$. Otherwise, $\nu(s) \neq c$. Because (i) the quota for type $\tau(s)$ at c is met under μ^k, and (ii) if for each $s' \in \mu^k(c)$ with

$\tau(s') = \tau(s)$ we have $\nu(s') = c$, then $\nu(s) = c$ would imply that the quota for type $\tau(s)$ at c is exceeded under ν, contradiction. But then, we should have $\nu(s')P_{s'}c$ for such s'. Otherwise, (s', c) would block ν since (i) $cP_{s'}\nu(s')$, (ii) $(\nu(c) - s) \cup s'$ respects affirmative action constraints at c, and (iii) $sP_c s'$. In this case, $\nu(s')P_{s'}c = \mu^k(s')$ implies that s' has been rejected by $\nu(s')$ at a step $k' < k$. This contradicts with the choice of k.

So, if a student s is rejected by a college at some step of GS^S, this college is not in s's available set $A(s)$. Since each student proposes to successively less desirable alternative, GS^S matches each s to $c(s)$, i.e. GS^S always selects the student optimal stable matching. This completes the proof. ■

Theorem 3 follows immediately.

3.3 Incentive Compatibility

Our proofs resemble the original proofs in the Roth (1982) with one distinction: In the literature, the main results are obtained first in marriage models (one-to-one matching problems). Then by responsiveness assumption, these results easily generalize to college admissions (many-to-one matchings). In our setting, we cannot reduce our problem to a standard marriage problem. Because, the possibility of $\sum_{\tau \in T} q_c^\tau > q_c$ precludes a fixed and well defined preference relation in a corresponding marriage market. To see this, consider the following example.

Example: c has two seats. There are two types: black and white. c prefers to allocate at most one seat for each type. Consider two female students f_1 and f_2 and two male students m_1 and m_2. Suppose that c ranks female students above male students. Now, let us try to construct a corresponding marriage market by dividing c into two separate colleges, c_1 and c_2, each with one slot. In a standard college admissions problem, the preference relation of each c_i coincides with the preference relation of c. In our problem, c_2 will prefer a female student to a male if c_1 admits a male student. Otherwise, the quota for female students is met, so c_2 prefers the other male student to females.

Hence, we have to derive all the results in our model directly from our model.

Let P_{-i} denote the preference relations of all agents except agent $i \in S \cup C$. Let μ' be the matching produced by $GS^S(P'_s, P_{-s})$. Let Q_s be such that

$\mu'(s)Q_s c$ for all $c \neq \mu'(s)$. We will refer to Q_s as **a simple misrepresentation.**

Lemma 2: If Q_s is a simple misrepresentation, then $GS^S(Q_s, P_{-s}) = GS^S(P'_s, P_{-s})$.

Proof: μ is stable under (P'_s, P_{-s}), so μ is stable under (Q_s, P_{-s}). Moreover, $c(s') = \mu(s')$ for all $s' \in S$ under (Q_s, P_{-s}). Therefore, $GS^S(Q_s, P_{-s})$ produces μ, as well, since GS^S always selects the student optimal stable matching. ■

So, for any misrepresentation, there is a simple misrepresentation that works as well. Let $\mu = GS^S(P_s, P_{-s})$ when P_s is the true preference relation of s. Then,

Lemma 3: If a simple misrepresentation by s leaves s at least as well off as μ, then no student will suffer.

Proof: Let Q_s be a simple misrepresentation and $\nu = GS^S(Q_s, P_{-s})$. Also assume that either $\nu(s)P_s\mu(s)$ or $\nu(s) = \mu(s)$. Suppose, on contrary, that for some s', $\mu(s')P_{s'}\nu(s')$. Since $s' \neq s$, s' states the same preferences, so that s' should be rejected by $\mu(s')$ at some step of $GS^S(Q_s, P_{-s})$. Let k be the first step of $GS^S(Q_s, P_{-s})$ at which some student, say s', is rejected by $\mu(s')$. Since Q_s is a simple misrepresentation, Q_s ranks $\nu(s)$ the first. So, s is tentatively assigned to $\nu(s)$ at step 1 of $GS^S(Q_s, P_{-s})$, and remains at $\nu(s)$ thereafter. Define $S' = \{s'' \in S : s''$ did not propose to $\mu(s')$ in $GS^S(P_s, P_{-s})\}$. Then, for any $s'' \in S'$, $\mu(s'')P_{s''}\mu(s')$. If any $s'' \in S'$ proposes to $\mu(s')$ at step k of $GS^S(Q_s, P_{-s})$, then s'' should have been removed from $\mu(s'')$ at some earlier step $k' < k$. This contradicts with the choice of k. So, no $s'' \in S'$ points to $\mu(s')$ at step k of $GS^S(Q_s, P_{-s})$. Then, s' is rejected by $\mu(s')$ at step k of $GS^S(Q_s, P_{-s})$ in favor of s. Therefore, $\mu(s') = \nu(s)$ and $\mu(s')$ prefers s to s', i.e. $sP_{\mu(s')}s'$.

There are two possibilities:
(1) $\tau(s') = \tau(s)$. In this case, s blocks μ if $\mu(s') = \nu(s)P_s\mu(s)$, since $sP_{\mu(s')}s'$. So, if $\tau(s') = \tau(s)$ then $\nu(s) = \mu(s)$. Remember that s is tentatively assigned to $\nu(s)$ at step 1 of $GS^S(Q_s, P_{-s})$, and remains at $\nu(s)$ thereafter. So, in this case, under $GS^S(Q_s, P_{-s})$, each school receives either the same proposals as under $GS^S(P_s, P_{-s})$, or a subset of these proposals. Therefore, each student is at least better off under $GS^S(Q_s, P_{-s})$. This contradicts with the supposition that $\mu(s')P_{s'}\nu(s')$.
(2) $\tau(s') \neq \tau(s)$. Then $\mu(s') = \nu(s) \neq \mu(s)$ implies that the quota for type $\tau(s)$ at $\mu(s') = \nu(s)$ should be met under μ, since $sP_{\mu(s')}s'$ and $\mu(s') = \nu(s)P_s\mu(s)$. Furthermore, for every s'' such that $\mu(s'') = \mu(s')$ and $\tau(s'') =$

$\tau(s)$, we have $s'' P_{\mu(s')} s$. Then some of these students should not propose to $\mu(s') = \nu(s)$ at step k of $GS^S(Q_s, P_{-s})$. Otherwise, s would be rejected by $\mu(s') = \nu(s)$ at step k of $GS^S(Q_s, P_{-s})$, a contradiction. Then, there should exist some $s'' \in S'$ who proposes to $\mu(s')$ at step k of $GS^S(Q_s, P_{-s})$, otherwise s' would not be rejected by $\mu(s')$ even if the quota for type $\tau(s)$ at $\mu(s') = \nu(s)$ were met at step k of $GS^S(Q_s, P_{-s})$. Then s'' should have been rejected by $\mu(s'')$ at a step $k' < k$. This contradicts with the choice of k. So, $\nu(s) = \mu(s)$, then we obtain the same contradiction as above.

This completes the proof. ■

Proposition 3: Let Q_s be a simple misrepresentation that leaves s at least as well off as μ. Let $\nu = GS^S(Q_s, P_{-s})$. Then, for each $c \in C : |\nu(c)| = |\mu(c)|$.

Proof: By the Proposition 2, if a student does not propose to a college c in $GS^S(P_s, P_{-s})$, then she does not propose to c in $GS^S(Q_s, P_{-s})$ either. We will refer to this result as **Argument***. Moreover, the number of students that are tentatively assigned to a college never decreases from one step to the next one in GS^S. So, $|\nu(c)| \leq |\mu(c)|$ for each $c \in C$. Again, by the proposition above, $\sum_{c \in C} |\nu(c)| \geq \sum_{c \in C} |\mu(c)|$, since the number of unmatched students under ν will be less than or equal to the number of unmatched students under μ. So, for each $c \in C$ $|\nu(c)| = |\mu(c)|$. ■

Proposition 4: No student can successfully misrepresent his preferences, i.e. reporting the true preferences is a dominant strategy for each student under GS^S.

Proof: We do not need to check unsuccessful misrepresentations. Let Q_s be a simple misrepresentation. Suppose that $\mu = GS^S(P_s, P_{-s})$, $\nu = GS^S(Q_s, P_{-s})$, and either $\nu(s) P_s \mu(s)$ or $\nu(s) = \mu(s)$. We will show that $\nu(s) P_s \mu(s)$ is not possible.

For any s', we say that s' makes a match at step k of $GS^S(P_s, P_{-s})$ if s' proposes to $\mu(s')$ at step k.

Let t be the final step of $GS^S(P_s, P_{-s})$. Consider a student s' who makes a match at step t. Then, if a student is rejected by $\mu(s')$ at some step in $GS^S(P_s, P_{-s})$, this is due to the quota limit for the type of this student, since there will always be some empty slots in $\mu(s')$ before step t. Otherwise, in order to match s' to $\mu(s')$, some other student would be rejected at step t, then t would not be the final step of $GS^S(P_s, P_{-s})$, contradiction. Moreover, no type $\tau(s')$ student would be rejected by $\mu(s')$ in $GS^S(P_s, P_{-s})$. Because, if a type $\tau(s')$ student is rejected by $\mu(s')$ in $GS^S(P_s, P_{-s})$, this should be due

to the quota limit for the type $\tau(s')$, because of the same reason. However, then some type $\tau(s')$ student would be rejected in favor of s' at step t. Then step t would not be the final step of $GS^S(P_s, P_{-s})$. Contradiction.

Let us summarize the scenario at step t : There is some empty slot available for s' at the beginning of step t. No other type $\tau(s')$ student is rejected by $\mu(s')$ in $GS^S(P_s, P_{-s})$. If some student is rejected by $\mu(s')$ at some step in $GS^S(P_s, P_{-s})$, this is due to the quota limit for the type of this student.

Now, we will show that $\mu(s') = \nu(s')$. Suppose that $\mu(s') \neq \nu(s')$. Then $|\nu(\mu(s'))| < |\mu(\mu(s'))|$. Because:
• By Argument*, no type $\tau(s')$ student who did not propose to $\mu(s')$ in $GS^S(P_s, P_{-s})$ proposes to $\mu(s')$ in $GS^S(Q_s, P_{-s})$.
• All type $\tau(s')$ students who propose to $\mu(s')$ in $GS^S(P_s, P_{-s})$ are assigned to $\mu(s')$.
• So, they will be assigned to $\mu(s')$ in $GS^S(Q_s, P_{-s})$ if they propose.
• All other type τ students, who are rejected by $\mu(s')$ in $GS^S(P_s, P_{-s})$, are rejected because of the quota limit for type τ.
• So, no other type student who is rejected by $\mu(s')$ in $GS^S(P_s, P_{-s})$ will fill the slot that s' empties at $\mu(s')$ in $GS^S(Q_s, P_{-s})$.
• Therefore, the slot emptied by s' at $\mu(s')$ in $GS^S(Q_s, P_{-s})$ will not be filled in $GS^S(Q_s, P_{-s})$ (i.e. under ν).

So, $|\nu(\mu(s'))| < |\mu(\mu(s'))|$. But, this contradicts with Proposition 4. Therefore, $\mu(s') = \nu(s')$ for any s' that makes his match at the final step of $GS^S(P_s, P_{-s})$.

The same conclusion holds for a group of students S' who are matched to c by $GS^S(P_s, P_{-s})$ such that students in S' are the only ones who propose to c in $GS^S(P_s, P_{-s})$. If s makes a match at t or s is included in such a S', then $\nu(s) = \mu(s)$.

Suppose that s makes a match at step $k < t$ of $GS^S(P_s, P_{-s})$. Let r be such that $k \leq r < t$. Assume that $\mu(s') = \nu(s')$ for any s' who makes his match at step $r+1$ or at a later step of $GS^S(P_s, P_{-s})$. We have just showed that this is true for $r = t-1$. Next we will show that $\mu(s') = \nu(s')$ for any s' who makes his match at step r, as well. In turn, this will prove by induction that $\mu(s') = \nu(s')$ for any s' who makes his match at step k or later.

Consider a student s' who makes his match at step r. Suppose $\mu(s') \neq \nu(s')$. Then $\nu(s') P_{s'} \mu(s')$ by Argument*. Consider the slot at $\mu(s')$ that s' empties under ν. By Proposition 4, this slot will be filled under ν by a student s'' such that $\nu(s'') = \mu(s') \neq \mu(s'')$. Then by Argument*, $\mu(s') P_{s''} \mu(s'')$ so that s'' is rejected by $\mu(s')$ in $GS^S(P_s, P_{-s})$. Let s'' is the student that is

most preferred by $\mu(s')$ among all students who are rejected by $\mu(s')$ in $GS^S(P_s, P_{-s})$. There are two possibilities:
(1) $(\mu(\mu(s')) - s') \cup s''$ respects affirmative action constraints.
• Then s'' will be matched to $\mu(s')$ prior to step r in $GS^S(P_s, P_{-s})$.
• In turn, s'' will be rejected by $\mu(s')$ at step r in $GS^S(P_s, P_{-s})$ in favor of s', and she will make her match at step $r+1$ or later.
• So, $\mu(s'') = \nu(s'')$ by the induction hypothesis.
• Since s'' will be rejected by $\mu(s')$, we also have $\mu(s'') \neq \mu(s')$.
• On the other hand, consider $GS^S(Q_s, P_{-s})$.
• Note that (i) s'' is the student that is most preferred by $\mu(s')$ among all students who are rejected by $\mu(s')$ in $GS^S(P_s, P_{-s})$, (ii) s' does not propose to $\mu(s')$ in $GS^S(Q_s, P_{-s})$, (iii) $(\mu(\mu(s')) - s') \cup s''$ respects affirmative action constraints, s'' will not be rejected by $\mu(s')$ in $GS^S(Q_s, P_{-s})$, and (iv) by Argument*, if a student does not propose to a college $\mu(s')$ in $GS^S(P_s, P_{-s})$, then she does not propose to $\mu(s')$ in $GS^S(Q_s, P_{-s})$ either.
• Then s'' will not be rejected by $\mu(s')$ in $GS^S(Q_s, P_{-s})$, so that $\nu(s'') = \mu(s')$, which contradicts with $\nu(s'') = \mu(s'') \neq \mu(s')$.
(2) So, $(\mu(\mu(s')) - s') \cup s''$ does not respect affirmative action constraints. Then repeat the arguments in (1) with the student that is most preferred by $\mu(s')$ among all students who were rejected by $\mu(s')$ in $GS^S(P_s, P_{-s})$ except s''. Obtain the same contradiction. Whenever such a contradiction is arrived, repeat the same arguments with a similar student that is preferred next. By finiteness of students, we arrive a final contradiction.

So no s'' such that $\nu(s'') = \mu(s') \neq \mu(s'')$ exists. But then $|\nu(c)| < |\mu(c)|$, which contradicts with the result of Proposition 4.

Thus, $\mu(s') = \nu(s')$ for any s' who makes his match at step r. Then, the induction proves that $\mu(s') = \nu(s')$ for any s' who makes his match at step k or later, in particular for s. Thus, s cannot successfully manipulate GS by misrepresenting his preferences. Therefore, reporting the true preferences is a dominant strategy for each student under GS^S. This completes the proof. ■

Theorem 4 follows immediately.

4 Applications

4.1 Entry-level labor markets

As mentioned in the introduction, in the American resident matching market,[9] for particular residency programs, if some positions remain unfilled, then these positions may revert to other programs (Roth and Peranson 1999, see also Roth 2002, Milgrom 2003). This allocation problem is easily solved by the following modification of the Gale-Shaley algorithm: For any such program, specify type specific quotas. For example, consider a urology program that offers six positions. Suppose that any unfilled urology position will revert to internal medicine. Then, set the capacity at six. Set the type specific quotas as follows: Six positions for urology, six positions for internal medicine. Now ask for the rankings of the candidates in each program. Construct a new preference ordering by listing urology candidates above internal medicine candidates by respecting the original preference orderings within each category. Then the Gale-Shapley algorithm tries to fill these six positions by urology candidates first. If this is not possible, then the algorithm tries to fill the remaining slots by internal medicine candidates. Note that the algorithm does not go back or restart at any point.[10]

As mentioned before, a doctor in UK can become eligible for full registration with the General Medical Council only if that doctor completes 12 months in a preregistration position, typically six months in a medical position and six months in a surgical position under consultants. In the Edinburgh case, some consultants may specify that they will not employ more than one female doctor in any six-month period. In this case, the modified Gale-Shapley algorithm asks for a single preference ordering of all candidates, and type specific quotas. The algorithm works as described before.

4.2 Controlled Choice in Public Schools

Abdulkadiroğlu and Sönmez (2003) introduce a new class of problems, namely controlled choice problems, in the context of public school choice. A **controlled choice problem** consists of the following:

1. a finite set of students $S = \{s_1, \ldots, s_n\}$;

[9] For recent advances, see Ehlers (2002), Klaus and Klijn (2003)).

[10] Note that for some programs, a single position may revert to more than one program. Theorem 1 provides an impossibility result for incentive compatibility in such cases.

2. a finite set of schools $C = \{c_1, \ldots, c_m\}$,

3. a capacity vector $q = (q_{c_1}, \ldots, q_{c_m})$ where q_c is the capacity of school $c \in C$,

4. a list of strict preference profile $P^S = (P_{s_1}, \ldots, P_{s_n})$, where P_s is the strict preference relation of student $s \in S$ over $C \cup \{s\}$,

5. a strict priority profile $P^C = (P_{c_1}, \ldots, P_{c_m})$, where P_c is the strict priority ordering of students at $c \in C$. sP_cs' means that s has a higher priority at c than s'.

6. a type space $T = \{\tau_1, ..., \tau_k\}$

7. a type function $\tau : S \to T$; $\tau(s)$ is the type of student s

8. for each school c, a vector of type specific quotas $q_c^T = (q_c^{\tau_1}, ..., q_c^{\tau_k})$ such that $\sum_{\tau \in T} q_c^\tau \geq q_c$. We refer these constraints as **controlled choice constraints.**

Here, priorities do not represent school preferences. They are imposed by state or local laws. For example, in the Boston Public School Choice Program, for each school a priority ordering is determined according to the following hierarchy: Students who have siblings already attending a school and living within the walk zone of that school constitute the first priority group at that school. Students who only have siblings already attending that school constitute the second priority group. Students who only live within the walk zone of that school constitute the third priority group. All the other students constitute the forth priority group. Students in the same priority group are ordered based on a previously announced lottery.

Controlled choice attempts to provide choice to students while maintaining the racial and ethnic balance at schools. In some states, choice is limited by court-ordered desegregation guidelines, while these guidelines are adopted voluntarily in some schools districts. Controlled choice constraints reflect the restrictions imposed by these desegregation guidelines. For example, in Minneapolis, the district is allowed to go above or below the district-wide average enrollment rates by up to 15 percent points in determining the racial quotas. So consider a school district in Minneapolis, where the average enrollment rates of majority students versus minority students are 60%, 40% respectively, and consider a school with 100 seats. Racial quotas for this school would be 75 for majority students, and 55 for minority students.

In a school choice program, each student should be assigned a seat at one of the schools. An **assignment** is a matching of students and schools. The final assignment should respect capacity constraints and controlled choice constraints. Furthermore, we introduce the following fairness requirement:

Definition: An assignment μ is **fair** in a controlled choice problem if

i. the list of students at each school respects controlled choice constraints under μ;

ii. there do not exist students $s, s' \in S$, and a school $c \in C$ such that

   (a) s' is assigned to c, i.e. $\mu(s') = c$

   (b) s prefers c to his assignment, i.e. $cP_s\mu(s)$,

   (c) If s' is replaced by s at c, the resulting list of students, $(\mu(c) - s') \cup s$, respects controlled choice constraints at c,

   (d) s has a higher priority at c than s', i.e. sP_cs'.

In college admissions, colleges themselves are agents which have preferences over students, whereas in school choice, schools are objects to be consumed by the students. Despite this important difference between the two models, school preferences and school priorities are similar mathematical objects: They both rank students. Hence there is a close connection between stability in college admissions and fairness in school choice: For any given controlled choice problem, write a corresponding college admissions with affirmative action problem as follows: The same set of students; the same set of schools renamed as colleges; the same preference profile for students; each college's preferences over individual students are given by the priority ordering of students at that college in the original controlled choice problem; each college's preferences satisfy AA and RR. Then,

Theorem 5: An assignment in a controlled choice problem is fair if and only if it is stable in the corresponding college admissions with affirmative action problem.

The proof of this result follows from a simple comparison of the definitions of stability and fairness. The original observation that connects fairness in a one sided matching problem to stability in a corresponding two sided matching problem is made by Balinski and Sönmez (1999) in the context of

Turkish college admissions, where they study a college admissions problem with responsive preferences.

Consider the following version of GS^S in controlled choice:

Step 1: Each student proposes to her first choice. Each school tentatively assigns its seats to its proposers one at a time following their priority order. If the quota of a type fills, the remaining proposers of that type are rejected and the tentative assignment proceeds with the students of the other types. Any remaining proposers are rejected.

In general, at
Step k: Each student who was rejected in the previous step proposes to her next choice. Each school considers the students it has been holding together with its new proposers and tentatively assigns its seats to these students one at a time following their priority order. If the quota of a type fills, the remaining proposers of that type are rejected and the tentative assignment proceeds with the students of the other types. Any remaining proposers are rejected.[11]

It is easy to see that this mechanism satisfies the following equivalent version of the fairness requirement (Abdulkadiroğlu and Sönmez, 2003) : If there is an unmatched student-school pair (s, c) where student s prefers school c to her assignment and she has higher priority than some other student s' who is assigned a seat at school c then (i) students s and s' are of different types, and (ii) the quota for the type of student s is met at school c.

Furthermore, our previous results with Theorem 4 imply the following:

[11] One can give equivalent versions of this mechanism, among which we have the following: 1. For each school set a capacity counter at its capacity, and set a type counter at its type specific quota limit for each type. 2. Take an arbitrary student s and assign s a tentative slot at her most preferred school among the ones that have a slot available for her type. Reduce the capacity counter and corresponding type counter of this school by one. 3. In general, given a tentative matching for some students, take an arbitrary student s': (i) Assign s' a tentative slot at her most preferred school c if c has a slot available for her type. Reduce the capacity counter and corresponding type counter of this school by one. (ii) Otherwise, if there is another student s'' that is of the same type as s' and ranked lower than s' with respect to c's preferences over individual students, find such s'' with the lowest rank, then remove s'' and assign s' a tentative slot at c. (iii) If s' cannot be assigned a tentative slot, repeat i and ii with the next preferred school of s'. (iv) If s' cannot be assigned a seat, keep her unmatched. 4. Stop when no more students can be assigned a tentative slot. Then each student is assigned her final tentative slot.

Theorem 6: In the class of controlled choice problems, GS^S produces a fair assignment that every student likes as well as any other fair assignment. Furthermore, it is dominant strategy incentive compatible.

The incentive compatibility result is rewritten in Abdulkadiroğlu and Sönmez (2003).[12]

5 Further Discussion

We have assumed that the type of each student is a one dimensional variable. However, a college might have preferences towards affirmative action along various dimensions. For example, a college might prefer a class that is racially balanced as well as balanced in terms of gender. In case of multi-dimensional type space, assumptions AA and RR are no longer sufficient to guarantee substitutability. To see this, consider the following example.

Example: There are four students, $S = \{bm, bf, wm, wf\}$. A student xy is of the race "x" and the gender "y". College c has a capacity of two and prefers to enroll at most one student of each race and at most one student of each gender. Its preferences are given by

P_c
$\{bm, wf\}$
bm
$\{bf, wm\}$
wf
bf
wm
$\varnothing$

Note that c's preference relation over sets of students is responsive to the following preference relation over individual students: $bmP_cwfP_cbfP_cwmP_cc$. Also check that $Ch_c(S' = \{bf, wm, wf\}) = \{bf, wm\}$ whereas $bf \notin Ch_c(S' - wm) = wf$. So, although c's preferences satisfy AA and RR, they are not substitutable.

[12] The meaning of priorities of students is open to interpretation. The above fairness notion eliminates all envy that can be justifiable among students via priorities. Abdulkadiroğlu and Sönmez (2003) provides an alternative interpretation for priority orderings and also offers a variant of top trading cycles algorithm, which is Pareto efficient and strategy proof, for that alternative interpretation.

So, existence of a stable matching in case of multidimensional type space does not follow from previous results, since we lose substitutability. However, the following example demonstrates that a stable matching may exist even when substitutability is violated.

Example: There are four students, $S = \{bm, bf, wm, wf\}$, and two colleges, $C = \{c, c'\}$. Each college has a capacity of two and prefers to enroll at most one student of each race and at most one student of each gender. The preference profile is given as follows:

P_{bf}	P_{bm}	P_{wf}	P_{wm}
s	s	s'	s'
s'	s'	s	s

P_c	$P_{c'}$
$\{bm, wf\}$	$\{bm, wf\}$
bm	wf
$\{bf, wm\}$	$\{bf, wm\}$
wf	wm
bf	bf
wm	bm
$\varnothing$	$\varnothing$

The unique stable matching matches c with bm, s' with wf, leaving bf and wm unmatched, which can be obtained via GS^S.

Multidimensional type space and control along each dimension brings a different type of challenge in the context of controlled choice in public schools. In a controlled choice problem, *every* student has to be assigned a seat in one of the schools. Consider the following controlled choice problem.

Example: There are four students, $S = \{bm, bf, wm, wf\}$, and two schools, $C = \{c, c'\}$. Each school has a capacity of two and can enroll at most one student of each race and at most one student of each gender. The preference profile of students and the priorities of students at schools are given as follows:

P_{bf}	P_{bm}	P_{wf}	P_{wm}
s	s	s'	s'
s'	s'	s	s

P_c	$P_{c'}$
bm	wf
bf	wm
wm	bf
wf	bm

There are two feasible assignments μ_1 and $\mu_2 : \mu_1(c) = \mu_2(c') = \{bm, wf\}$ and $\mu_1(c') = \mu_2(c) = \{bf, wm\}$. And both of these assignments are fair. However the standard Gale-Shapley algorithms fail to produce these assignments.

Finally, one has to note that imposing type specific quotas alone does not guarantee desegregation unless these quotas are chosen appropriately at

each school by the district authority. Consider the example in the previous section: A school in Minneapolis with racial quotas of 75 for majority students, and 55 for minority students. Enrolling 75 majority students and no minority students would not violate the fairness notion above. Furthermore, such unacceptable allocations may be produced by the Gale-Shapley student optimal mechanism. Imposing minimum quotas solves this problem. However it comes with additional theoretical challenges. Abdulkadiroğlu (2003) studies controlled choice in public schools with minimum quotas, as well as maximum quotas.

References

[1] Abdulkadiroğlu, Atila. "Controlled Choice in Public Schools," mimeo, 2003.

[2] Abdulkadiroğlu, Atila and Sönmez, Tayfun. "School Choice: A Mechanism Design Approach," *American Economic Review*, June 2003, 93(3), 729-747

[3] Alkan, Ahmet and Gale, David. "Stable Schedule Matching Under Revealed Preferences," *Journal of Economic Theory*, forthcoming.

[4] Balinski, Michel and Sönmez, Tayfun. "A Tale of Two Mechanisms: Student Placement," *Journal of Economic Theory*, January 1999, 84(1), pp. 73-94.

[5] Dubins, L.E. and D.A. Freedman. "Machiavelli and the Gale-Shapley Algorithm." *American Mathematical Monthly*, 1981, 88, 485-494.

[6] Ehlers, Lars. "In Search of Advice for Physicians in Entry-Level Medical Markets," June 2002, mimeo.

[7] Kelso, Alexander S.,Jr. and Vincent P. Crawford. "Job Matching, Coalition Formation, and Gross Substitutes", *Econometrica*, 1982, 50, 1483-1504.

[8] Klaus, Bettina and Flip Klijn. "Stable Matchings and Preferences of Couples: Some things couples always wanted to know about stable matchings (but were afraid to ask)," July 2003, mimeo

[9] Martínez, Ruth; Massó, Jordi; Neme, Alejandro and Oviedo, Jorge. "Single Agents and the Set of Many-to-One Stable Matchings," *Journal of Economic Theory*, 2000, 91, pp. 91-105.

[10] Milgrom, Paul. "Matching with Contracts," mimeo, March 2, 2003.

[11] Roth, Alvin E. "The Economics of Matching: Stability and Incentives," *Mathematics of Operations Research*, 1982, 7, pp. 617-628.

[12] Roth, Alvin E. "The College Admissions Problem is not Equivalent to the Marriage Problem," *Journal of Economic Theory*, 1985, 36, 277-288.

[13] Roth, Alvin E. "A Natural Experiment in the Organization of Entry-Level Labor Markets: Regional Markets for New Physicians and Surgeons in the United Kingdom," *American Economic Review*, June 1991, 81(3), pp. 414-440.

[14] Roth, Alvin E. "The Economist as an Engineer: Game Theory, Experimentation, and Computation as Tools for Design," *Econometrica*, July 2002, 70(4), pp. 1341-1378.

[15] Roth, Alvin E. and Peranson, Elliot. "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design," *American Economic Review*, September 1999, 89(4), pp. 748-780.

[16] Roth, Alvin E. and Sotomayor, Marilda A.O. *Two-Sided Matching: A Study in Game Theoretic Modeling and Analysis*. New York: Cambridge University Press, 1990.

[17] Rees, Nina Shokraii. "School Choice 2000 Annual Report," Backgrounder #1354, March 30, 2000. Heritage Foundation.