

A Tale of Two Mechanisms: Student Placement*

Michel Balinski

C.N.R.S., and Ecole Polytechnique, Paris, France

and

Tayfun Sönmez

University of Michigan, Ann Arbor, Michigan

Received November 26, 1997; revised June 29, 1998

A new class of matching problems that models centralized college admissions via standardized tests is presented. The allocation mechanism that is used in real-life applications of this problem in Turkey is analyzed. It is shown that this mechanism, multi-category serial dictatorship, has a number of serious deficiencies, most notably inefficiency, vulnerability to manipulation, and the potential of penalizing students for improved test scores. Exploiting the relation between this class of problems and the celebrated college admissions model (Gale & Shapley [4]), an alternative mechanism is proposed that overcomes these deficiencies. This mechanism—the Gale–Shapley student optimal mechanism—is characterized as “best” in this context. *Journal of Economic Literature* Classification Numbers: C71, C78, D71, D78. © 1999 Academic Press

1. INTRODUCTION

In this paper a new class of matching problems is introduced that models centralized college admissions via standardized tests. The model precisely mimics the current college admissions practices in Turkey: there is a central

* We are grateful to Sergiu Hart, Andreu Mas-Colell, and Erol Sezer for discussions that motivated this paper. We thank Ted Bergstrom, John Conley, Ali Dođramacı, Roger Gordon, Faruk Gül, Matthew Jackson, Hervé Moulin, Ünal Oktay, Alvin Roth, Marilda Sotomayor, Eres Söylemez, Ennio Stacchetti, Ayşe Sibel Uzer, Lin Zhou, seminar participants at Alabama, Duke, Koç, Michigan, Penn State, Pittsburgh, Rochester, the 1997 Spring Midwest Economic Theory Meetings at Northwestern, the third International Conference of the Society for the Advancement of Economic Theory at Antalya, Turkey, and especially an anonymous associate editor, an anonymous referee, and William Thomson for their helpful comments. Sönmez gratefully acknowledges the research support of National Science Foundation via Grant SBR-9709138. All errors are regrettably our own responsibility.

authority that places students in colleges. This authority annually offers a number of tests and ranks all students who wish to enroll in a college in a number of skill categories. Colleges are public goods and they have no say in the admissions process. However, for each faculty there is an associated skill category and a student with a better score in this category has a higher priority for this faculty. This requirement is called *fairness*.

This class of problems is closely related to the celebrated *college admissions* model (Gale & Shapley [4]). There is one key difference between them. In the student placement problem the only agents are the students: their examination scores and their preferences for colleges determine assignments. The colleges are merely objects to be consumed. On the other hand, in the college admissions problem both students and colleges are agents, students express their preferences for colleges and colleges for students (perhaps using examination scores).

The student placement mechanism currently in use in Turkey, namely *multi-category serial dictatorship*, is fair. However, it has a number of serious deficiencies. It is not *Pareto efficient* and not even a *second-best mechanism* among *fair* mechanisms. It is not *strategy-proof*: it can be *manipulated* by students. It does not necessarily *respect improvements* in test scores: a student may increase her score in one or more skill categories and yet be punished with a worse assignment when everything else stays the same. Exploiting the relation between this model and the college admissions model, we propose an alternative placement mechanism, namely the *Gale-Shapley student optimal mechanism*, and show that it is the only second-best mechanism, essentially the only strategy-proof mechanism,¹ and essentially the only mechanism that respects improvements. These results not only suggest a potential improvement over the current mechanism in use in Turkey, but also show that there is a best mechanism to use in this context.

Some might argue that among all fair assignments society should prefer one in which students are assigned according to their comparative advantages (as determined by test scores), rather than by their personal preferences. This argument is refuted. We believe that students should be given the greatest freedom of choice consistent with their aptitudes, since motivation is more important in ultimate success than mere scores on standardized exams. We *prove* that the only reasonable mechanism to be used assigns students according to their preferences rather than by their comparative advantages.

¹ This result is an immediate corollary of results in Dubins & Freedman [3], Roth [10] and Alcalde & Barberà [2].

2. STUDENT PLACEMENT PROBLEMS

In Turkey college admissions is centralized. It is the responsibility of a student placement office to assign students to colleges, in fact to the particular faculties (e.g., engineering, medical, dental, business) of colleges, with no student assigned to more than one college-faculty or *slot*. Every year this office offers a standardized test that every student who wishes to attend a college is required to take. This examination consists of several component tests, including a mathematics test, a science test, a verbal aptitude test, etc. Different faculties prize different skills; accordingly, they use different combinations of tests to arrive at rankings of the students. For example, the engineering faculties combine scores from the mathematics and science tests, the medical and dental faculties use the scores of the science test, and the business faculties combine scores from the mathematics and verbal aptitude tests. Every faculty-type ranks the applying students identically, strictly according to the relevant scores.

Colleges are considered to be public services: they have no say in the admissions process. Each faculty of each college has a capacity for the number of students it can admit fixed in advance. Prior to the examination, each student submits to the placement office his preferences over the faculties in colleges he is willing to attend (e.g., first choice engineering at Bilkent University, second choice engineering at Boğaziçi University, third choice business at Koç University, etc.). In order to lighten the terminology we will throughout take the term *college* to mean a college-faculty, so that a *college* is associated with a particular well-defined skill category.

These are the raw data on which the decision to place students in college-faculty slots depend: the preferences of the students, their relevant examination scores, and the capacities of the colleges.

We now formalize this model. A (*student*) *placement problem* consists of

1. a set of students $S = \{s_1, \dots, s_n\}$,
2. a set of colleges $C = \{c_1, \dots, c_m\}$,
3. a capacity vector $q = (q_{c_1}, \dots, q_{c_m})$ where q_{c_i} is the capacity of college c_i ,
4. a list of student preferences $P_S = (P_{s_1}, \dots, P_{s_n})$ where P_{s_i} is the preference relation of student s_i over colleges including the *no-college* option,
5. a set of skill categories $T = \{t_1, \dots, t_k\}$,
6. a list of student test scores $f = (f^{s_1}, \dots, f^{s_n})$ where $f^{s_i} = (f_{t_1}^{s_i}, \dots, f_{t_k}^{s_i})$ is a vector which gives the test score of student s_i in each category, and
7. a function $t: C \rightarrow T$ where $t(c)$ is the skill category required by college c .

Each student has a strict preference on $C \cup \{c_0\}$, where c_0 denotes the no-college option and $q_{c_0} = |S|$. Let R_s denote the at-least-as-good-as relation associated with the preference relation P_s for all $s \in S$. That is, for all $c, c' \in C \cup \{c_0\}$ we have cR_sc' if and only if cP_sc' or $c = c'$. For all $\tilde{s} \in S$, let $P_{-\tilde{s}} = (P_s)_{s \in S \setminus \{\tilde{s}\}}$. It is assumed that there are no ties in the test scores of any skill category, that is $f_t^{s_i} \neq f_t^{s_j}$ for all $t \in T$ and $s_i, s_j \in S$ with $s_i \neq s_j$.² This assumption implies that the test scores induce a strict ranking of the students in each skill category. Throughout the paper S, C, T, t are fixed and hence each triple list of preferences, test scores, and capacities defines a placement problem (P_S, f, g) .

A *matching* is an allocation of college slots to students such that no student occupies more than one college slot. Formally it is a function $\mu: S \rightarrow C \cup \{c_0\}$ such that $|\mu^{-1}(c)| \leq q_c$, for all $c \in C$. If $\mu(s) = c_0$ student s is not assigned any college slot. Given a preference relation P_s of a student s , initially defined over $C \cup \{c_0\}$, it is extended to the set of matchings in the following natural way: student s prefers the matching μ to the matching μ' if and only if she prefers $\mu(s)$ to $\mu'(s)$. By an abuse of notation we also use P_s to denote this extension.

A matching μ is *individually rational* if no student is assigned to a college that is worse than the no-college option. Formally a matching μ is *individually rational* if $\mu(s)R_sc_0$ for all $s \in S$.

A matching μ is *non-wasteful* if whenever a student prefers a college c to his assignment, the college c has all its slots filled. Formally a matching μ is *non-wasteful* if $cP_s\mu(s)$ implies $|\mu^{-1}(c)| = q_c$ for all $s \in S$ and for all $c \in C$.

A matching η *Pareto dominates* a matching μ if no student prefers μ to η and there is at least one student who prefers η to μ . Formally, a matching η *Pareto dominates* a matching μ if $\eta(s_i)R_{s_i}\mu(s_i)$ for all $s_i \in S$ and $\eta(s_j)P_{s_j}\mu(s_j)$ for some $s_j \in S$. A matching μ is *Pareto efficient* if it is not Pareto dominated by any other matching. Note that Pareto efficiency implies individual rationality and non-wastefulness in this context.

A *placement mechanism* is a systematic procedure that selects a matching for each placement problem. A placement mechanism is *individually rational* if it always selects an individually rational matching, it is *non-wasteful* if it always selects a non-wasteful matching, and it is *Pareto efficient* if it always selects a Pareto efficient matching. A placement mechanism φ *Pareto dominates* another placement mechanism ψ if no student ever prefers the matching selected by ψ to the matching selected by φ and there is at least one placement problem where at least one student prefers the matching selected by φ to the matching selected by ψ .

² In practice ties are rare but possible. Whenever there are ties they are broken by the Turkish placement office on the basis of additional criteria such as student age.

3. FAIRNESS

The following *fairness* criterion is essential for a mechanism to be acceptable: students with better test scores should be assigned their better choices. Formally a matching μ is *fair* if for all students $s, \tilde{s} \in S$ with $\mu(\tilde{s}) = \tilde{c}$,

$$\tilde{c}P_s\mu(s) \quad \text{implies} \quad f_{t(\tilde{c})}^{\tilde{s}} > f_{t(\tilde{c})}^s.$$

For example, suppose a student is assigned her third choice and her first choice c is in the mathematics category whereas her second choice \tilde{c} is in the science category. Then no student with a lower mathematics score should be assigned a slot at college c and no student with a lower science score should be assigned a slot at college \tilde{c} .³

The following lemma characterizes the set of fair matchings in a very intuitive way: a base score is announced for each college and each student selects the best college among those with lower base scores than his or her scores. If feasible, the induced matching is fair. Conversely all fair matchings may be obtained in this way.⁴

LEMMA 1. *A matching μ is fair if and only if there exists a list of base scores $f_* = (f_{c_1}, \dots, f_{c_m})$ that satisfies the following conditions for all $s \in S$ and $c \in C$:*

- (i) $\mu(s) = c$ implies $f_{t(c)}^s \geq f_c$,
- (ii) $f_{t(c)}^s \geq f_c$ implies $\mu(s) R_s c$.

Proof. Let μ be a fair matching. For each college c define $f_c = \min_{s \in \mu^{-1}(c)} f_{t(c)}^s$ and let $f_* = (f_{c_1}, \dots, f_{c_m})$. We claim that μ and f_* satisfy the conditions. The first holds by the definition of the f_c . Suppose the second does not, namely, there is a student s and a college c with $f_{t(c)}^s \geq f_c$ and yet $cP_s\mu(s)$. Then $\mu(s) \neq c$, and the construction of f_c together with the fact that there are no ties in scores imply that $f_{t(c)}^s > f_c$. So there must be a student $\tilde{s} \in \mu^{-1}(c)$ with $f_{t(c)}^{\tilde{s}} > f_{t(c)}^s$: that is, student s prefers c , and c ranks s higher than student \tilde{s} , contradicting the fairness of μ . This establishes necessity.

To prove sufficiency, suppose μ is a matching and $f_* = (f_{c_1}, \dots, f_{c_m})$ is a list of scores that satisfy the conditions. Let the students s, \tilde{s} be such that

³ This *fairness* notion reduces to the *weak fairness* notion of Svensson [17] whenever there is one skill category.

⁴ We are grateful to John Conley and Matthew Jackson whose comments led to this characterization.

$\mu(\tilde{s}) = \tilde{c}$ and $\tilde{c} P_s \mu(s)$. By the first condition, $f_{u(\tilde{c})}^{\tilde{s}} > f_{\tilde{c}}$, whereas by second condition $f_{\tilde{c}} > f_{u(\tilde{c})}^s$, so $f_{u(\tilde{c})}^{\tilde{s}} > f_{u(\tilde{c})}^s$, showing that μ is fair. Q.E.D.

In fact, the student placement office in Turkey announces the base scores for each college together with the matching it selected. In this way the fairness of the selected matching can be easily verified by each student.

A placement mechanism is *fair* if it always selects a fair matching. If there is only one category (and hence only one test score for each student) then it is easy to see that there is only one placement mechanism that is fair and Pareto efficient:⁵ *serial dictatorship*, where the student with the highest test score is assigned her top choice, the student with the next highest score is assigned his top choice among the remaining slots, and so on. Indeed, admissions to a set of prestigious Turkish public high schools is done in exactly this manner. Every year a test is offered which ranks students (in only one category) and the serial dictatorship matching induced by this ranking is used to assign students to these high schools.

4. ASSOCIATED COLLEGE ADMISSIONS PROBLEM

A *college admissions problem* (Gale & Shapley [4]) consists of a set of students $S = \{s_1, \dots, s_n\}$, a set of colleges $C = \{c_1, \dots, c_m\}$, a capacity vector $q = (q_{c_1}, \dots, q_{c_m})$ where q_{c_i} is the capacity of college c_i , a list of student preferences $P_S = (P_{s_1}, \dots, P_{s_n})$ where P_{s_i} is the preference of student s_i over colleges that includes the no-college option, and a list of college preferences $P_C = (P_{c_1}, \dots, P_{c_m})$ where P_{c_i} is the preference of college c_i over students that includes the *no-student* option. For all $\tilde{c} \in C$, let $P_{-\tilde{c}} = (P_c)_{c \in C \setminus \{\tilde{c}\}} \cdot S$, C are fixed throughout, so each triple (P_S, P_C, q) of student preferences, college preferences, and capacities define a college admissions problem. As before the notation R_s and R_c is used for the at least as good relation associated with the preferences P_s and P_c .

It is tempting to treat the class of placement problems as a special case (or a sub-class) of college admissions problems. But this is emphatically not valid: in college admissions problems, colleges are agents, they have their preferences for students and they may choose their strategies in expressing their preferences. In student placement problems, colleges are merely objects to be consumed: the students are the only agents, the only strategic players.⁶ Notions such as the *core* and *stability* that are central for college admissions problems do not have any direct meaning in placement problems. Moreover, the very definition of Pareto efficient depends on

⁵ See the discussion following Lemma 2 for a simple proof of this observation.

⁶ Another class of problems that is closely related is *office allocation problems* (Hylland & Zeckhauser [5]). See also Abdulkadiroğlu & Sönmez [1] and Zhou [20].

whether or not colleges are agents. This does not mean that the findings for college admissions problems are irrelevant in the present context. Quite the contrary, they are crucial for student placement problems as well.

To each placement problem (P_S, f, q) define the *associated college admissions problem* (P_S, P_C, q) by constructing a preference relation P_c for each college c based on the test scores in its category $t(c)$. That is, for all $c \in C$ the preference P_c is such that

$$sP_c\tilde{s} \quad \text{if and only if} \quad f_{t(c)}^s > f_{t(c)}^{\tilde{s}}, \quad \text{for all } s, \tilde{s} \in S,$$

and

$$sP_cs_0, \quad \text{for all } s \in S,$$

where s_0 denotes the no-student option.

The definitions of *matching* and *individual rationality* carry over to college admissions problems, and an *admissions mechanism* is simply a procedure that selects a matching for each college admissions problem.

A student-college pair $(s, c) \in S \times C$ *blocks* a matching μ if

$$cP_s\mu(s) \quad \text{and} \quad |\mu^{-1}(c)| < q_c,$$

or

$$cP_s\mu(s) \quad \text{and} \quad sP_c\tilde{s} \quad \text{for some } \tilde{s} \in \mu^{-1}(c).$$

A matching is *stable* if it is individually rational and is blocked by no student-college pair. Let $\mathcal{S}(P_S, P_C, q)$ denote the *set of stable matchings* for the college admissions problem (P_S, P_C, q) . The following lemma follows from the definitions.

LEMMA 2. *A matching is individually rational, fair, and non-wasteful for a placement problem if and only if it is stable for its associated college admissions problem.*

There is a unique placement mechanism that is fair and Pareto efficient on the subclass of placement problems with one skill category, as mentioned earlier. This observation immediately follows from Lemma 2 together with the following two facts:

1. If all colleges have the same preferences in a college admissions problem then there is a unique stable matching.

2. Pareto efficiency implies individual rationality and non-wastefulness in the context of placement problems.

Stability not only plays a central role in the theoretical literature concerning two-sided matching problems, but is also essential for real-life applications of these problems (see Roth [11] and Roth & Xing [14]). It is by now well-known that for each college admissions problem there is a stable matching that is preferred to any other stable matching by all the students. This stable matching is also the worst stable matching for all the colleges. We refer to this matching as the *student optimal stable matching*. There is an analogous stable matching that is preferred by the colleges, and we refer to it as the *college optimal stable matching*. Let $\mu^S(P_S, P_C, q)$ denote the *student optimal stable matching* for the college admissions problem (P_S, P_C, q) and $\mu^C(P_S, P_C, q)$ denote the *college optimal stable matching*. These key concepts in college admissions problems permit the definition of the following placement mechanisms. The *Gale–Shapley student optimal mechanism* selects the student optimal stable matching of the associated college admissions problem for each placement problem. Similarly, the *Gale–Shapley college optimal mechanism* selects the college optimal stable matching of the associated college admissions problem for each placement problem.

The student optimal stable matching can be obtained using the *student proposing deferred acceptance algorithm* (Gale & Shapley [4]):

Step 1. Each student proposes to his top choice among those colleges for which he is acceptable. Each college c rejects all but the best q_c students among those students who proposed to it. Those that remain are “tentatively” assigned one slot at college c .

In general,

Step k. Each student who is rejected in the last step proposes to his top choice among those colleges that has not as yet rejected him and for which he is acceptable. (If there is no such college the student stops proposing.) Each college c rejects all but the best q_c students among those students who have just proposed and those that were tentatively assigned to it at the last step. Those that remain are “tentatively” assigned one slot at college c . The algorithm terminates when no student proposal is rejected. Each student is assigned to her or his final tentative assignment. If a student is rejected by all colleges to which she or he has applied, the student is assigned to no college.

Similarly, the college optimal stable matching can be obtained using the *college proposing deferred acceptance algorithm* where the roles of colleges and students are interchanged.

5. MULTI-CATEGORY SERIAL DICTATORSHIP

Since *fairness* is the essential criterion in Turkish college admission practices, it is natural to presume that the mechanism used there is fair. It is. We call this mechanism *multi-category serial dictatorship*, for it is a natural extension of serial dictatorship modified to deal with several rankings in different categories. The following notation is useful to define it. Let a *tentative placement* be a correspondence $v: S \rightarrow C \cup \{c_0\}$ such that no college is assigned to more students than its capacity. Note that unlike a matching, a tentative placement allows a student to be assigned more than one college slot. So a tentative placement is not necessarily a matching (though of course every matching is a tentative placement). Each category t has a total capacity $q^t = \sum_c \{q_c: t(c) = t\}$.

Multi-category serial dictatorship is defined by means of the following recursive algorithm applied to any student placement problem (P_S, f, q) . For step 1, $(P_S^1, f, q) = (P_S, f, q)$.

Step k. Given (P_S^k, f, q) ,

(a) consider a category t and the ranking induced by the test scores in this category. Assign (only) the colleges in category t (i.e., $\{c \in C: t(c) = t\}$) to (at most q^t) students using the serial dictatorship applied to this ranking. That is, student with the highest score in category t is assigned his top choice among those colleges in category t , the student with the next highest score is assigned her top choice among the remaining slots in this category, and so on. Do the same for all categories. Assign c_0 to all students who are not assigned a college. In general, this leads to a tentative placement since a student may be assigned slots in two or more colleges, which are all in different categories. Name the tentative placement v_k . Note the following property that characterizes v_k : $c \in v_k(s)$ if and only if any student s^* ranked higher than student s in category $t = t(c)$, but not tentatively assigned to c , must prefer to c either another college c^* in the same category $t = t(c^*)$ or the no college option.

(b) For each student s construct the preference relation P_s^{k+1} from P_s^k as follows. If s is assigned no slot, then $P_s^{k+1} = P_s^k$. If s is assigned one or more slots, then define P_s^{k+1} by moving the no-college option c_0 directly after the best of these assignments in P_s^k (so that only one of the assignments is acceptable to the student in this new preference relation). The rankings of the colleges are not changed. Let $P_S^{k+1} = (P_{s_1}^{k+1}, \dots, P_{s_n}^{k+1})$ be the list of new preferences.

The algorithm terminates when no student is assigned more than one college slot. The multi-category serial dictatorship mechanism selects this

final tentative placement, which is a matching. Denote *multi-category serial dictatorship* by φ^D .

The following example demonstrates the algorithm.

EXAMPLE 1. Let $S = \{s_1, s_2, s_3, s_4, s_5\}$, $C = \{c_1, c_2, c_3\}$, $q = (q_{c_1}, q_{c_2}, q_{c_3}) = (2, 1, 1)$, $T = \{t_1, t_2\}$, $t(c_1) = t_1$, $t(c_2) = t(c_3) = t_2$. Let preferences $P_S = (P_{s_1}, P_{s_2}, P_{s_3}, P_{s_4}, P_{s_5})$ and test scores $f = (f^{s_1}, f^{s_2}, f^{s_3}, f^{s_4}, f^{s_5})$ be as follows:

$$\begin{array}{ll} c_2 P_{s_1} c_1 P_{s_1} c_0 P_{s_1} c_3 & f^{s_1} = (f_{t_1}^{s_1}, f_{t_2}^{s_1}) = (90, 90) \\ c_1 P_{s_2} c_2 P_{s_2} c_3 P_{s_2} c_0 & f^{s_2} = (f_{t_1}^{s_2}, f_{t_2}^{s_2}) = (80, 60) \\ c_1 P_{s_3} c_3 P_{s_3} c_2 P_{s_3} c_0 & f^{s_3} = (f_{t_1}^{s_3}, f_{t_2}^{s_3}) = (70, 70) \\ c_1 P_{s_4} c_2 P_{s_4} c_0 P_{s_4} c_3 & f^{s_4} = (f_{t_1}^{s_4}, f_{t_2}^{s_4}) = (60, 80) \\ c_2 P_{s_5} c_3 P_{s_5} c_1 P_{s_5} c_0 & f^{s_5} = (f_{t_1}^{s_5}, f_{t_2}^{s_5}) = (50, 50) \end{array}$$

These scores induce the following rankings in categories t_1 and t_2 :

$$\begin{array}{l} t_1: s_1 \quad s_2 \quad s_3 \quad s_4 \quad s_5 \\ t_2: s_1 \quad s_4 \quad s_3 \quad s_2 \quad s_5 \end{array}$$

Step 1.

$$\begin{array}{l} t_1: \begin{array}{cc} s_1 & s_2 \\ c_1 & c_1 \end{array} \quad t_2: \begin{array}{ccc} s_1 & s_4 & s_3 \\ c_2 & \text{---} & c_3 \end{array} \end{array}$$

This diagram is obtained as follows. Consider serial dictatorship applied to skill category t_1 . The only college in this category is college c_1 and it has 2 slots. The first student in this category is s_1 , and since c_1 is acceptable to her under P_{s_1} she is tentatively assigned one of its slots. The next student is s_2 for whom c_1 is also acceptable, and she is tentatively assigned the second slot of college c_1 . There are no other slots in this category. Consider serial dictatorship associated with category t_2 . There are two colleges in this category, college c_2 and college c_3 , both with 1 slot. The first student in this category is s_1 and the best acceptable college for her under P_{s_1} is college c_2 and she is tentatively assigned this slot. The only remaining slot in this category is one slot at college c_3 . The next student in this category is s_4 for whom college c_3 is not acceptable under P_{s_4} . The next student is s_3 for whom college c_3 is acceptable under P_{s_3} and he is tentatively assigned the only slot at college c_3 . Therefore Step 1 yields the following tentative placement:

$$v_1 = \begin{pmatrix} s_1 & s_2 & s_3 & s_4 & s_5 \\ c_1, c_2 & c_1 & c_3 & c_0 & c_0 \end{pmatrix}.$$

Since each of students s_1, s_2 , and s_3 are assigned at least one slot, their preferences are changed to:

$$\begin{aligned} c_2 P_{s_1}^2 c_0 P_{s_1}^2 c_1 P_{s_1}^2 c_3, \\ c_1 P_{s_2}^2 c_0 P_{s_2}^2 c_2 P_{s_2}^2 c_3, \\ c_1 P_{s_3}^2 c_3 P_{s_3}^2 c_0 P_{s_3}^2 c_2. \end{aligned}$$

For the other two students, $P_{s_4}^2 = P_{s_4}^1$, and $P_{s_5}^2 = P_{s_5}^1$.

Step 2. Find the serial dictatorship outcomes for the preference profile P_S^2 .

$$t_1: \begin{array}{ccccc} s_1 & s_2 & s_3 & & \\ \text{---} & c_1 & c_1 & & \end{array} \quad t_2: \begin{array}{ccccc} s_1 & s_4 & s_3 & & \\ & c_2 & \text{---} & & c_3 \end{array}$$

This yields the following tentative placement:

$$v_2 = \begin{pmatrix} s_1 & s_2 & s_3 & s_4 & s_5 \\ c_2 & c_1 & c_1, c_3 & c_0 & c_0 \end{pmatrix}$$

and the following list of preferences:

$$\begin{aligned} c_2 P_{s_1}^3 c_0 P_{s_1}^3 c_1 P_{s_1}^3 c_3, \\ c_1 P_{s_2}^3 c_0 P_{s_2}^3 c_2 P_{s_2}^3 c_3, \\ c_1 P_{s_3}^3 c_0 P_{s_3}^3 c_3 P_{s_3}^3 c_2, \\ c_1 P_{s_4}^3 c_2 P_{s_4}^3 c_0 P_{s_4}^3 c_3, \\ c_2 P_{s_5}^3 c_3 P_{s_5}^3 c_0 P_{s_5}^3 c_1. \end{aligned}$$

Step 3. Find the serial dictatorship outcomes for the preference profile P_S^3 .

$$t_1: \begin{array}{ccccc} s_1 & s_2 & s_3 & & \\ \text{---} & c_1 & c_1 & & \end{array} \quad t_2: \begin{array}{ccccc} s_1 & s_4 & s_3 & s_2 & s_5 \\ c_2 & \text{---} & \text{---} & \text{---} & c_3 \end{array}$$

This yields the following tentative placement (which is a matching):

$$v_3 = \begin{pmatrix} s_1 & s_2 & s_3 & s_4 & s_5 \\ c_2 & c_1 & c_1 & c_0 & c_3 \end{pmatrix}.$$

Since no student is assigned more than one college the algorithm terminates and $\varphi^D(P_S, f, q) = v_3$.

Note that, in the special case where every college is in a different category, this algorithm is equivalent to the college proposing deferred acceptance algorithm. While the two algorithms are different in general, they always yield the same outcome.

THEOREM 1. *The multi-category serial dictatorship and the Gale–Shapley college optimal mechanisms are equivalent.*

Proof. Consider a placement problem (P_S, f, q) and the sequence of placement problems (P_S^l, f, q) , $l = 1, \dots, \bar{l}$, generated by the algorithm, where $P_S^1 = P_S$. Associate with each the corresponding college admissions problem (P_S^l, P_C, q) .

The sets of stable matchings of any two of these problems are identical, $\mathcal{S}(P_S^l, P_C, q) = \mathcal{S}(P_S^{l+1}, P_C, q)$.

To see this suppose, first, that $\mu \in \mathcal{S}(P_S^l, P_C, q)$. Then μ must be individually rational in (P_S^{l+1}, P_C, q) . For otherwise, $\mu(s^1) = c$ for some student s^1 and some college c with $c_0 P_{s^1}^{l+1} c$, whereas $c P_{s^1}^l c_0$ since μ is individually rational in (P_S^l, P_C, q) . By the algorithm this means that student s^1 is tentatively assigned a better college c^1 in the placement problem (P_S^l, f, q) . That is, $c^1 \in v_l(s^1)$ with $c^1 P_{s^1}^l c P_{s^1}^l c_0$. Let $t(c^1) = t^*$.

Either $|\mu^{-1}(c^1)| < q_{c^1}$ or $|\mu^{-1}(c^1)| = q_{c^1}$. In the first case it is immediately clear that the pair (s^1, c^1) blocks μ in (P_S^l, P_C, q) , a contradiction. In the second case, $s^1 \notin \mu^{-1}(c^1)$, $s^1 \in v_l^{-1}(c^1)$, and $|v_l^{-1}(c^1)| \leq q_{c^1}$ imply that the set $\mu^{-1}(c^1)$ must include some student $s^2 \notin v_l^{-1}(c^1)$. If s^2 ranks below s^1 in category $t^* = t(c^1)$, then (s^1, c^1) again blocks μ . Otherwise, s^2 ranks above s^1 in category t^* and since $s^2 \notin v_l^{-1}(c^1)$ and $c^1 P_{s^1}^l c_0$, he must be tentatively assigned a better college c^2 in the same category $t(c^2) = t^*$ at v_l . The situation concerning (s^1, c^1) and c is now repeated for (s^2, c^2) and c^1 , where s^2 ranks higher than s^1 in category t^* : $\mu(s^2) = c^1$ and $c^2 \in v_l(s^2)$ in the placement problem (P_S^l, f, q) with $c^2 P_{s^2}^l c^1$. So either (s^2, c^2) blocks μ or the situation repeats again for some (s^3, c^3) and c^2 , where s^3 ranks higher than s^2 in category t^* , $t(c^3) = t^*$, $c^3 P_{s^3}^l c^2$ and $c^3 \in v_l(s^3)$. This situation cannot be repeated forever; the ranks of the students in category t^* keep increasing, so at some stage (s^k, c^k) must block μ , a contradiction.

So μ is individually rational in (P_S^{l+1}, P_C, q) , as claimed. The absence of any blocking pair in (P_S^l, P_C, q) implies the same for (P_S^{l+1}, P_C, q) , since the restriction of the student preferences to colleges are identical. Thus, $\mu \in \mathcal{S}(P_S^{l+1}, P_C, q)$.

To show the reverse inclusion suppose that $\mu \notin \mathcal{S}(P_S^l, P_C, q)$. Then either μ is not individually rational in (P_S^l, P_C, q) , which immediately implies that μ is not individually rational in (P_S^{l+1}, P_C, q) by construction;

or there is a pair (s, c) that blocks μ in (P_S^l, P_C, q) , which implies that it also blocks μ in (P_S^{l+1}, P_C, q) . So $\mu \notin \mathcal{S}(P_S^{l+1}, P_C, q)$, and therefore $\mathcal{S}(P_S^l, P_C, q) = \mathcal{S}(P_S^{l+1}, P_C, q) = \mathcal{S}(P_S, P_C, q)$.

Observe that at termination, when the tentative assignment $v_{\bar{l}}$ is a matching, the algorithm stops, but the problem $(P_S^{\bar{l}+1}, P_C, q)$ may be constructed and it's stable set is, as has been proven, equivalent to its predecessors'. But the matching $v_{\bar{l}}$ assigns to *every* college $c \in C$ its q_c highest ranked students under P_c (or all students if there are fewer) from among those who do not prefer the no-college option c_0 to c . This is the college optimal stable matching $\mu^C(P_S^{\bar{l}+1}, P_C, q)$. But since it is the college optimal stable matching for this problem it is also the college optimal stable matching for the original problem, $\mu^C(P_S, P_C, q)$. Q.E.D.

A variant of the Gale–Shapley college optimal mechanism has been used to match medical interns and hospitals in the United States since 1951.⁷ The algorithm used for this purpose is different than both the college proposing deferred acceptance algorithm and the algorithm used for Turkish student placement. (See Roth [9] for a description of this algorithm.) In May 1997, the Board of Directors of the National Resident Matching Program decided to switch to a variant of the Gale–Shapley student optimal mechanism starting with the 1998 match (see Roth & Peranson [12]). Both of these mechanisms have their strengths and weaknesses in the context of college admissions problems. But in the following sections it is shown that the Gale–Shapley student optimal mechanism is the clear choice in the context of student placement problems.

6. PARETO EFFICIENCY

Multi-category serial dictatorship has a serious drawback. It is not Pareto efficient. The following simple example makes the point.

EXAMPLE 2. Let $S = \{s_1, s_2\}$, $C = \{c_1, c_2\}$, $q = (q_{c_1}, q_{c_2}) = (1, 1)$, $T = \{t_1, t_2\}$, $t(c_1) = t_1$, $t(c_2) = t_2$. Let preferences $P_S = (P_{s_1}, P_{s_2})$ and test scores $f = (f^{s_1}, f^{s_2})$ be as follows:

$$c_1 P_{s_1} c_2 P_{s_1} c_0 \quad f^{s_1} = (f_{t_1}^{s_1}, f_{t_2}^{s_1}) = (80, 90)$$

$$c_2 P_{s_2} c_1 P_{s_2} c_0 \quad f^{s_2} = (f_{t_1}^{s_2}, f_{t_2}^{s_2}) = (90, 80)$$

⁷Note that this is an application in college admissions problems, not in placement problems since hospitals are agents and they have preferences over medical interns.

Step 1 of the multi-category serial dictatorship algorithm assigns student s_1 one slot at college c_2 , and student s_2 one slot at college c_1 . Since both students are assigned only one slot the algorithm terminates and assigns both students their second choices:

$$\varphi^D(P_S, f, q) = \begin{pmatrix} s_1 & s_2 \\ c_2 & c_1 \end{pmatrix}.$$

Clearly the matching that assigns both students their top choices—which is fair—Pareto dominates this matching.

The Pareto inefficiency of multi-category serial dictatorship is apparent, given its equivalence to the Gale–Shapley college optimal mechanism. What about other mechanisms? Does there exist a mechanism that is both fair and Pareto efficient? The answer is unfortunately negative. An example in Roth & Sotomayor [13] can easily be adopted to make this point.

EXAMPLE 3. Let $S = \{s_1, s_2, s_3\}$, $C = \{c_1, c_2\}$, $q = (q_{c_1}, q_{c_2}) = (1, 1)$, $T = \{t_1, t_2\}$, $t(c_1) = t_1$, $t(c_2) = t_2$. Let preferences $P_S = (P_{s_1}, P_{s_2}, P_{s_3})$ and test scores $f = (f^{s_1}, f^{s_2}, f^{s_3})$ be as follows:

$$c_2 P_{s_1} c_1 P_{s_1} c_0 \quad f^{s_1} = (f_{t_1}^{s_1}, f_{t_2}^{s_1}) = (90, 80)$$

$$c_1 P_{s_2} c_2 P_{s_2} c_0 \quad f^{s_2} = (f_{t_1}^{s_2}, f_{t_2}^{s_2}) = (80, 70)$$

$$c_1 P_{s_3} c_2 P_{s_3} c_0 \quad f^{s_3} = (f_{t_1}^{s_3}, f_{t_2}^{s_3}) = (70, 90)$$

First recall that any matching that is Pareto efficient should be both individually rational and non-wasteful. There are six such matchings:

$$\mu_1 = \begin{pmatrix} s_1 & s_2 & s_3 \\ c_1 & c_2 & c_0 \end{pmatrix}, \quad \mu_2 = \begin{pmatrix} s_1 & s_2 & s_3 \\ c_1 & c_0 & c_2 \end{pmatrix}, \quad \mu_3 = \begin{pmatrix} s_1 & s_2 & s_3 \\ c_2 & c_1 & c_0 \end{pmatrix},$$

$$\mu_4 = \begin{pmatrix} s_1 & s_2 & s_3 \\ c_2 & c_0 & c_1 \end{pmatrix}, \quad \mu_5 = \begin{pmatrix} s_1 & s_2 & s_3 \\ c_0 & c_1 & c_2 \end{pmatrix}, \quad \mu_6 = \begin{pmatrix} s_1 & s_2 & s_3 \\ c_0 & c_2 & c_1 \end{pmatrix}.$$

Among these matchings only μ_2 is fair.⁸ But it is Pareto dominated by μ_4 . An immediate implication is the following impossibility result.

LEMMA 3. *There is no placement mechanism that is both fair and Pareto efficient.*

⁸ Indeed, μ_1 is not fair since $c_2 P_{s_1} c_1$ and $f_{t(c_2)}^{s_1} > f_{t(c_2)}^{s_2}$; μ_3 is not fair since $c_2 P_{s_3} c_0$ and $f_{t(c_2)}^{s_3} > f_{t(c_2)}^{s_1}$; μ_4 is not fair since $c_1 P_{s_2} c_0$ and $f_{t(c_1)}^{s_2} > f_{t(c_1)}^{s_3}$; μ_5 is not fair since $c_1 P_{s_1} c_0$ and $f_{t(c_1)}^{s_1} > f_{t(c_1)}^{s_2}$; finally μ_6 is not fair since $c_1 P_{s_1} c_0$ and $f_{t(c_1)}^{s_1} > f_{t(c_1)}^{s_3}$.

Can this impossibility result be an excuse for using multi-category serial dictatorship? The answer is negative. In Example 2 the matching that assigns both students their top choices is fair and Pareto dominates the matching selected by multi-category serial dictatorship. Therefore the multi-category serial dictatorship is not even a second-best mechanism. What then is the class of *second-best* placement mechanisms? It turns out that there is only one such mechanism.

THEOREM 2. *The Gale–Shapley student optimal mechanism Pareto dominates any other fair mechanism.*

Proof. It must be shown that, for any placement problem (P_S, f, q) the student-optimal stable matching $\mu^S(P_S, P_C, q)$ of the associated college admissions problem (P_S, P_C, q) Pareto dominates any other fair matching. Let μ be a fair matching for the placement problem (P_S, f, q) . There are three cases to consider.

(1) The matching μ is individually rational and non-wasteful: Lemma 2 implies that $\mu \in \mathcal{S}(P_S, P_C, q)$ and the conclusion follows from the definition of the student-optimal stable matching.

(2) The matching μ is individually rational but wasteful: In this case there are colleges with empty slots under μ . Consider the placement problem (P_S, f, \bar{q}) with $\bar{q}_c = |\mu^{-1}(c)|$ for all $c \in C$, where all empty slots are eliminated. The matching μ is fair, individually rational, and non-wasteful for (P_S, f, \bar{q}) , and therefore $\mu \in \mathcal{S}(P_S, P_C, \bar{q})$. Hence $\mu^S(P_S, P_C, \bar{q})$ either Pareto dominates, or is equal to μ . But increasing the capacities can never hurt a student under the student-optimal stable matching, and therefore $\mu^S(P_S, P_C, q)$ either Pareto dominates, or is equal to $\mu^S(P_S, P_C, \bar{q})$. Since μ is wasteful, $\mu \neq \mu^S(P_S, P_C, q)$ and therefore the transitivity of the Pareto dominance relation implies that $\mu^S(P_S, P_C, q)$ Pareto dominates μ .

(3) The matching μ is not individually rational: There are students who prefer the no-college option to their assignments. Let ν be the matching that assigns the no-college option to all such students, and $\mu(s)$ to all other students. The matching ν Pareto dominates μ . Moreover it is fair and individually rational. Therefore by cases 1 and 2, ν is either equal to or Pareto dominated by $\mu^S(P_S, P_C, q)$. The transitivity of the Pareto dominance relation implies that $\mu^S(P_S, P_C, q)$ Pareto dominates μ . Q.E.D.

It is well-known that the student optimal stable matching Pareto dominates any other stable matching. In addition, as is shown by Theorem 2, the student optimal stable matching of the associated college admissions problem Pareto dominates any other fair matching as well. This proves that there is a unique second-best mechanism.

7. STRATEGY-PROOFNESS

What about the strategic properties of multi-category serial dictatorship? Is it immune to manipulation by the students? Consider the following example.

EXAMPLE 4. (The same as Example 2). $S = \{s_1, s_2\}$, $C = \{c_1, c_2\}$, $q = (q_{c_1}, q_{c_2}) = (1, 1)$, $T = \{t_1, t_2\}$, $t(c_1) = t_1$, $t(c_2) = t_2$ and

$$c_1 P_{s_1} c_2 P_{s_1} c_0 \quad f^{s_1} = (f_{t_1}^{s_1}, f_{t_2}^{s_1}) = (80, 90)$$

$$c_2 P_{s_2} c_1 P_{s_2} c_0 \quad f^{s_2} = (f_{t_1}^{s_2}, f_{t_2}^{s_2}) = (90, 80)$$

Recall from Example 2 that

$$\varphi^D(P_S, f, q) = \begin{pmatrix} s_1 & s_2 \\ c_2 & c_1 \end{pmatrix}.$$

Now suppose student s_1 announces the false preference relation $c_1 \tilde{P}_{s_1} c_0 \tilde{P}_{s_1} c_2$, where only college c_1 is acceptable. In this case

$$\varphi^D((\tilde{P}_{s_1}, P_{s_2}), f, q) = \begin{pmatrix} s_1 & s_2 \\ c_1 & c_2 \end{pmatrix},$$

and student s_1 successfully manipulates the multi-category serial dictatorship mechanism.

A placement mechanism is *strategy-proof* if no student can ever benefit by unilaterally misrepresenting his or her preferences. Strategy-proofness is a very desirable and yet demanding property. If achieved, it makes truthtelling a dominant strategy. Is there any fair placement mechanism that is strategy-proof? The answer is well-known.

THEOREM 3. (Dubins & Freedman [3], Roth [8]). *The Gale–Shapley student optimal mechanism is strategy-proof.*⁹

Moreover, it is “essentially” the only fair mechanism that is strategy-proof.

⁹ Strictly speaking, what is shown is that truthtelling is a dominant strategy for all students under the Gale–Shapley student optimal mechanism in the context of college admissions problems. Since colleges are merely public goods in the context of placement problems, Theorem 3 immediately follows.

THEOREM 4. (Alcalde & Barberà [2]). *The Gale–Shapley student optimal mechanism is the only student placement mechanism that is individually rational, non-wasteful, fair, and strategy-proof.*

8. RESPECTING IMPROVEMENTS

Multi-category serial dictatorship is subject to yet another major flaw.

EXAMPLE 5 (Again, Example 2). $S = \{s_1, s_2\}$, $C = \{c_1, c_2\}$, $q = (q_{c_1}, q_{c_2}) = (1, 1)$, $T = \{t_1, t_2\}$, $t(c_1) = t_1$, $t(c_2) = t_2$ and

$$c_1 P_{s_1} c_2 P_{s_1} c_0 \quad f^{s_1} = (f_{t_1}^{s_1}, f_{t_2}^{s_1}) = (80, 90)$$

$$c_2 P_{s_2} c_1 P_{s_2} c_0 \quad f^{s_2} = (f_{t_1}^{s_2}, f_{t_2}^{s_2}) = (90, 80)$$

Recall from Example 2 that

$$\varphi^D(P_S, f, q) = \begin{pmatrix} s_1 & s_2 \\ c_2 & c_1 \end{pmatrix}.$$

Now suppose student s_1 scores worse in the tests and his new test scores are $\tilde{f}^{s_1} = (\tilde{f}_{t_1}^{s_1}, \tilde{f}_{t_2}^{s_1}) = (70, 70)$. In this case

$$\varphi^D(P_S, (\tilde{f}^{s_1}, f^{s_2}), q) = \begin{pmatrix} s_1 & s_2 \\ c_1 & c_2 \end{pmatrix},$$

and student s_1 is rewarded by getting his top choice as a result of a worse performance! Alternatively, going in the other direction, student s_1 is punished by getting his second choice for doing better in the tests.

A placement problem (P_S, \bar{f}, q) is an *improvement for student s* over (P_S, f, q) if $\bar{f}^{s_i} = f^{s_i}$ for all $s_i \neq s$ and $\bar{f}^s \geq f^s$. The two problems are the same except that the scores of one student s may have improved.

In terms of college admissions problems (P_S, \bar{P}_C, q) is an *improvement for student s* over (P_S, P_C, q) if for all $c \in C$

$$s P_c \bar{s} \quad \text{implies} \quad s \bar{P}_c \bar{s} \quad \text{for all } \bar{s} \in S \cup \{s_0\}, \quad \text{and}$$

$$s^* P_c \bar{s} \quad \text{if and only if} \quad s^* \bar{P}_c \bar{s} \quad \text{for all } s^*, \bar{s} \in S \cup \{s_0\}, \quad s^* \neq s, \quad \bar{s} \neq s.$$

Again the two problems are the same except that some student may have moved up in the preferences of one or more colleges. But these definitions of “improvement” harbor a subtle difference. There is more latitude for changes in the preferences of colleges in admissions problems than in the changes of the rankings of colleges in placement problems. In placement a

change of examination score may force a simultaneous improvement across many colleges, which is not the case in admissions.

A placement mechanism φ *respects improvements*¹⁰ if for any $s \in S$, (P_s, \bar{f}, q) an improvement for s over (P_s, f, q) , implies

$$\varphi(P_s, \bar{f}, q) R_s \varphi(P_s, f, q).$$

Similarly an admissions mechanism φ *respects improvements* if for any $s \in S$, (P_s, \bar{P}_C, q) an improvement for s over (P_s, P_C, q) , implies

$$\varphi(P_s, \bar{P}_C, q) R_s \varphi(P_s, P_C, q).$$

The Gale–Shapley student optimal mechanism is essentially the only mechanism that respects improvements. This is established in two steps due to the differences in the meanings of “improvement” as applied to admissions and placement problems. We begin with college admissions mechanisms.

THEOREM 5. *The unique stable admissions mechanism that respects improvements is the Gale–Shapley student optimal mechanism.*

Proof. We first show that the Gale–Shapley student optimal mechanism respects improvements. Let (P_s, \bar{P}_C, q) be an improvement for student s over (P_s, P_C, q) , and $\bar{\mu}^S$ and μ^S be their respective student optimal stable matchings. It must be shown that $\bar{\mu}^S(s) R_s \mu^S(s)$. Suppose on the contrary $\mu^S(s) P_s \bar{\mu}^S(s)$. To begin, note that $\mu^S \notin \mathcal{S}(P_s, \bar{P}_C, q)$ by the definition of student optimal stable matching.

Let $c = \mu^S(s)$. Suppose that student s announces the fake preferences P_s^* where

$$c P_s^* c_0 P_s^* c^* \quad \text{for all } c^* \in C, c^* \neq c.$$

That is, under the fake preference relation she prefers only college c to the no-college option. Consider the college admissions problem $(P_s^*, P_{-s}, \bar{P}_C, q)$.

We claim that $\mu^S \in \mathcal{S}(P_s^*, P_{-s}, \bar{P}_C, q)$. Since μ^S is individually rational in (P_s, P_C, q) and student s is assigned her top choice at μ^S it is also individually rational in $(P_s^*, P_{-s}, \bar{P}_C, q)$. Suppose that μ^S is blocked by some pair in the problem $(P_s^*, P_{-s}, \bar{P}_C, q)$. It cannot be blocked by a pair involving a student other than s , for then this same pair would necessarily block μ^S in (P_s, P_C, q) , a contradiction. Clearly student s cannot be

¹⁰ This property can alternatively be interpreted as *non-manipulability through destroying endowments* where the test scores are “endowments.” See Postlewaite [7], Thomson [18] for applications of such manipulation in exchange economies, Sertel [15], Thomson [19] for applications in public goods economies and Sönmez [16] for applications in two-sided matching markets.

partner to a blocking pair either, so $\mu^S \in \mathcal{S}(P_s^*, P_{-s}, \bar{P}_C, q)$ as claimed. Moreover since $\mu^S(P_s^*, P_{-s}, \bar{P}_C, q)$ Pareto dominates any other stable matching of $(P_s^*, P_{-s}, \bar{P}_C, q)$, we must have $\mu^S(P_s^*, P_{-s}, \bar{P}_C, q)(s) = c$. But then

$$\underbrace{\mu^S(P_s^*, P_{-s}, \bar{P}_C, q)(s)}_{=c} \quad P_s \underbrace{\mu^S(P_s, \bar{P}_C, q)(s)}_{=\mu^S(s)}$$

which means that student s can manipulate the Gale–Shapley student optimal mechanism. This contradicts Theorem 3 and completes the proof of the first part.

To establish uniqueness, let φ be an admissions mechanism that is stable and that respects improvements. Suppose φ is not the Gale–Shapley student optimal mechanism. Then there exists an admissions problem (P_S, P_C, q) such that

$$\varphi(P_S, P_C, q) \neq \mu^S(P_S, P_C, q).$$

For simplicity of notation let ϕ be the matching $\varphi(P_S, P_C, q)$, and μ^S be the matching $\mu^S(P_S, P_C, q)$. In particular, this means there exists a student $s \in S$ with $\phi(s) \neq \mu^S(s) = c$.

Consider the admissions problem $(P_S, P_c, \hat{P}_{-c}, q)$, where for all $\hat{c} \in C$, $\hat{c} \neq c$ the preferences $\hat{P}_{\hat{c}}$ are the same as $P_{\hat{c}}$ except that student s is the least acceptable student. That is, for all $\hat{c} \in C$, $\hat{c} \neq c$,

$$\begin{aligned} \tilde{s} \hat{P}_{\hat{c}} s^* &\text{ if and only if } \tilde{s} P_{\hat{c}} s^* && \text{for all } \tilde{s}, s^* \in S, \quad \tilde{s} \neq s, \quad s^* \neq s, \\ \tilde{s} \hat{P}_{\hat{c}} s \hat{P}_{\hat{c}} s_0 &&& \text{for all } \tilde{s} \in S, \quad \tilde{s} \neq s. \end{aligned}$$

The key to the proof is the following *Claim*:

$$\mu \in \mathcal{S}(P_S, P_c, \hat{P}_{-c}, q) \text{ implies } \mu(s) R_s c.$$

To prove this Claim observe first that $\mu^S \in \mathcal{S}(P_S, P_c, \hat{P}_{-c}, q)$, since otherwise a pair that blocks μ^S in $(P_S, P_c, \hat{P}_{-c}, q)$ would block it in (P_S, P_C, q) as well.

Now suppose, to the contrary, that $c P_s \mu(s)$, and let $\hat{\mu}^C = \mu^C(P_S, P_c, \hat{P}_{-c}, q)$. Then since $\hat{\mu}^C$ is the student-worst stable matching for $(P_S, P_C, \hat{P}_{-c}, q)$ this means that $c P_s \hat{\mu}^C(s)$, which in turn implies $c^* = \hat{\mu}^C(s) \neq c$. Moreover $c^* \neq c$ implies that college c^* ranks student s worse than any other student under \hat{P}_{c^*} .

The situation is then:

$$\mu^S, \hat{\mu}^C \in \mathcal{S}(P_S, P_c, \hat{P}_{-c}, q),$$

and

$$\mu^S(s) = c, \hat{\mu}^C(s) = c^*, c \neq c^*.$$

McVitie & Wilson [6] and Roth [10] show that any college is always assigned the same number of students by any pair of stable matchings, so there must exist a student $\hat{s} \neq s$ with $\mu^S(\hat{s}) = c^*$ and $\hat{\mu}^C(\hat{s}) \neq c^*$. We have $\hat{s} \hat{P}_{c^*} s$, so the stability of $\hat{\mu}^C$ implies $\hat{\mu}^C(\hat{s}) P_{\hat{s}} c^*$, or equivalently $\hat{\mu}^C(\hat{s}) P_{\hat{s}} \mu^S(\hat{s})$. But this contradicts the fact that $\hat{\mu}^C$ is the student-worst stable matching for the problem $(P_S, P_c, \hat{P}_{-c}, q)$, and so establishes the Claim.

The proof is now easily finished. Since φ is stable, $\hat{\phi} = \varphi(P_S, P_c, \hat{P}_{-c}, q) \in \mathcal{S}(P_S, P_c, \hat{P}_{-c}, q)$ and so by the Claim, $\hat{\phi}(s) R_s \mu^S(s)$. Moreover, $\varphi(P_S, P_c, q) = \phi \in \mathcal{S}(P_S, P_c, q)$ and $\phi(s) \neq \mu^S(s)$ imply $\mu^S(s) P_s \phi(s)$, and therefore $\hat{\phi}(s) P_s \phi(s)$. This contradicts the fact that φ respects improvements and completes the proof. Q.E.D.

The theorem does not carry over to placement mechanisms because in the context of placement problems it may be impossible to single out one college c and construct a placement problem whose associated admissions problem has the properties of $(P_S, P_c, \hat{P}_{-c}, q)$. One natural way of handling this difficulty is asking a placement mechanism be *college admissions invariant*: namely, that it selects the same matching for any two placement problems that have the same associated college admissions problem. In fact this condition is obviously satisfied by both the Gale–Shapley student optimal and multi-category serial dictatorship mechanisms. Invoking invariance, the uniqueness result carries over to placement problems.

THEOREM 6. *The unique fair, individually rational, and non-wasteful placement mechanism that is college admissions invariant and respects improvements is the Gale–Shapley student optimal mechanism.*

9. CONCLUDING REMARKS

The argument in support of the Gale–Shapley student optimal mechanism for placement problems is overwhelming: it is the only *second-best* placement mechanism, it is essentially the only *fair* mechanism that is *strategy-proof*, and it is also essentially the only *fair* mechanism that *respects improvements*. The results are summarized in the table below, where φ is any fair mechanism except the Gale–Shapley student optimal mechanism, φ^D is the multi-category serial dictatorship, φ^C is the Gale–Shapley college optimal mechanism, and φ^S is the Gale–Shapley student optimal mechanism.

	Second- best	Strategy- proof	Respects improvements
$\varphi^D = \varphi^C$ (Theorem 1)	No (Example 2)	No (Example 4)	No (Example 5)
φ^S	Yes (Theorem 2)	Yes (Theorem 3)	Yes (Theorem 6)
φ	No (Theorem 2)	No (Theorem 4)	No (Theorem 6)

These are not merely theoretical arguments, they are practical reasons. To be acceptable, a mechanism cannot penalize a student for improving her examination scores. Moreover, to offer all students effective equal treatment it is essential that expressing their true preferences be in their best interest. Today's mechanism—the college optimal mechanism—induces misrepresentation and so assignments that do not even correctly incorporate the students' preferences. Society is not well served by assigning students according to their comparative advantages.

REFERENCES

1. A. Abdulkadiroğlu, and T. Sönmez, Random serial dictatorship and the core from random endowments in house allocation problems, *Econometrica* **66** (1998), 689–701.
2. J. Alcalde and S. Barberà, Top dominance and the possibility of strategy-proof stable solutions to matching problems, *Econ. Theory* **4** (1994), 417–435.
3. L. E. Dubins and D. A. Freedman, Machiavelli and the Gale–Shapley algorithm, *Amer. Math. Monthly* **88** (1981), 485–494.
4. D. Gale and L. Shapley, College admissions and the stability of marriage, *Amer. Math. Monthly* **69** (1962), 9–15.
5. A. Hylland and R. Zeckhauser, The efficient allocation of individuals to positions, *J. Polit. Econ.* **87** (1979), 293–314.
6. D. G. McVitie and L. B. Wilson, Stable marriage assignments for unequal sets, *BIT* **10** (1970), 259–309.
7. A. Postlewaite, Manipulation via endowments, *Rev. Econ. Stud.* **46** (1979), 255–262.
8. A. E. Roth, The economics of matching: Stability and incentives, *Math Oper. Res.* **7** (1982), 617–628.
9. A. E. Roth, The evolution of the labor market for medical interns and residents: A case study in game theory, *J. Polit. Econ.* **92** (1984), 991–1016.
10. A. E. Roth, The college admissions problems is not equivalent to the marriage problem, *J. Econ. Theory* **36** (1985), 277–288.
11. A. E. Roth, A natural experiment in the organization of entry level labor markets: Regional markets for new physicians and surgeons in the U.K., *Amer. Econ. Rev.* **81** (1991), 415–440.

12. A. E. Roth and E. Peranson, The effects of a change in the NRMP matching algorithm, *J. Amer. Medical Assoc.* **278** (1997), 729–732.
13. A. E. Roth and M. Sotomayor, “Two-Sided Matching: A Study in Game Theoretic Modeling and Analysis,” Cambridge Univ. Press, London/New York, 1990.
14. A. E. Roth, and X. Xing, Jumping the Gun: Imperfections and institutions related to the timing of market transactions, *Amer. Econ. Rev.* **84** (1994), 992–1044.
15. M. Sertel, Manipulating Lindahl equilibrium via endowments, *Econ. Lett.* **51** (1994), 167–171.
16. T. Sönmez, Manipulation via capacities in two-sided matching markets, *J. Econ. Theory* **77** (1997), 197–204.
17. L.-G. Svensson, Queue allocation of indivisible goods, *Soc. Choice Welfare* **11** (1994), 323–330.
18. W. Thomson, Monotonic allocation mechanisms, mimeo, University of Rochester, 1987.
19. W. Thomson, Monotonic allocation mechanisms in economies with public goods, Mimeo, University of Rochester, 1987.
20. L. Zhou, On a conjecture by Gale about one-sided matching problems, *J. Econ. Theory* **52** (1990), 123–135.