Efficient Matching in the School Choice Problem*

Philip J. Reny

Department of Economics
University of Chicago

Abstract

We consider the problem of efficient matching in school choice by defining a matching μ to be *priority-neutral* iff no matching ν can make any student whose priority is violated by μ better off unless ν violates the priority of some student who is made worse off. Clearly, every stable matching is priority-neutral. A matching is *priority-efficient* iff it is both priority-neutral and Pareto efficient. We show that there is a unique priority-efficient matching. Moreover every student weakly prefers the unique priority-efficient matching to every priority-neutral matching and so to every stable matching. We also consider the mechanism that selects the unique priority-efficient matching when students report their preferences to the mechanism. We provide practical advice for students participating in this mechanism and we provide conditions that are both novel and natural under which truth-telling is an equilibirum.

Keywords: school choice, stable matchings, fair matchings, Pareto efficient matchings, priority-efficiency, priority-neutrality.

1 Introduction

Many U.S. cities (including New York City, Boston, Seattle, Cambridge, Charlotte, Denver, Minneapolis, and Columbus) allow some form of school choice wherein families can choose a school for their children that is outside the district in which they live. But because there may not be enough seats at any given school to accommodate all students for whom that school is their first choice, school districts must set priority rules in order to resolve the conflicts that inevitably arise.

For example, in Boston, children in a school's predefined walk zone who have a sibling at that school have priority over children who only have a sibling at the school, and the latter children have priority over children who are only in the school's walk zone. All remaining

*Current Version: January 2, 2021. I thank Atila Abdulkadiroğlu, Ben Brooks, Aram Grigoryan, Thayer Morrill, Parag Pathak, Doron Ravid, Joe Root, Al Roth, Tayfun Sönmez, Leeat Yariv, Utku Unver and Tina Danting-Zhang for helpful discussions. Financial support from the National Science Foundation (SES-1724747) is gratefully acknowledged. Author contact: preny@uchicago.edu

students have lowest priority. Any conflicts between students in the same priority group are resolved according to the outcome of a random lottery. So after receiving their lottery numbers, all students are strictly ranked in terms of their priority at each Boston school.

A matching of students to schools creates a priority violation when some student i prefers another school s to their own school and, either, school s has a vacant seat, or, student i has higher priority at s than some student assigned to s. Such a matching is said to violate i's priority (at school s). We will assume here that, like Boston public schools, each school strictly orders the students from highest to lowest priority. Consequently, any conflicts between students over a given school can be resolved by that school's priority order. Even so, it is not at all clear whether the priority orders across all of the schools are mutually compatible, i.e., whether there is a matching of students to schools that does not violate any student's priority at any school. When such a matching does exist, it is called *stable*.

Remarkably, stable matchings always do exist, regardless of the schools' priority orders and regardless of the students' preferences over schools. Even more remarkable perhaps is that, among all of the stable matchings, there is one (and only one) that all of the students agree is best. Both of these results are due to Gale and Shapley (1962), who call the stable matching that is best for all the students, *student-optimal*.

Unfortunately, as is well-known, the student-optimal stable matching need not be Pareto efficient.[1] In fact, the extent of the inefficiency can be very large. For example, Kesten (2010) shows that for any set of schools and seat quotas, there are school priorities, students, and student preferences over schools such that the student-optimal stable matching assigns each student to his or her worst or second-worst school. While this theoretical possibility is indeed an extremely poor outcome, one might wonder whether any significant inefficiencies actually occur in practice.

According to Abdulkadiroğlu, Pathak, and Roth (2009), in a New York City school district in 2006-2007, over 4,000 grade 8 students could have been made better off by reassigning them to a school different than their match in the student-optimal stable matching, without hurting any other students. Thus the extent of the inefficiencies that can arise in the student-optimal stable matching is a matter of real practical importance.

So it is no surprise that a good deal of attention has been paid to the problem of selecting a Pareto efficient matching for the school choice problem, and this will be our main goal here as well. At the same time however, we will pay due attention to two important practical matters that must be addressed if a theoretical solution to the problem is to have any hope of finding success in the field.

The first important practical matter is that any theoretical solution must be described

[1] We follow the school-choice literature convention, starting with Balinski and Sönmez (1999), that schools are objects to be allocated to students. So Pareto efficiency is always with respect to students only.

in straightforward language that is meaningful to school board officials, city councillors, and state legislators. For if a solution cannot be so explained, then it is difficult to imagine how a school board would be willing to move away from its current system, potentially risking lawsuits brought by upset parents whose child's priority is violated as a result of a new system that no one seems to understand. The second important practical matter is that it must be possible to give simple and effective advice to students as to how they should behave in any mechanism that is intended to implement the solution, so that the outcome of the mechanism is more likely than not to produce the desired matching.[2] We address each of these matters in turn.

To explain our solution in language that will resonate with officials in charge of the system, we must give clear guidance as to the "rights" that are embodied in student priorities. Because our primary goal is to select a Pareto efficient matching, violating the priority of one or more students will be unavoidable whenever the student-optimal stable matching is not Pareto efficient. Consequently, we cannot give students whose priority is violated at a school the absolute right to take a seat at that school, since that would mean that only stable matchings could ever be chosen. Instead, we will be guided by two natural principles.

The Right to Relief. Every student has the right to seek relief from a priority violation by replacing the offending matching with any other matching so long as the Equal Priority Rights principle is respected.

Equal Priority Rights. No student may gain relief from a priority violation by replacing the offending matching with one that violates the priority of a student that it makes worse off.

Under these principles, a student's priority at any school protects that student in two ways. First, it gives the student the right to seek relief when their priority is violated. Second, it protects the student from harm when any other student seeks relief from their own priority violation. Notice that this second layer of protection is absent when each student is given the absolute right to take a seat at any school that violates their priority, and it is for this reason that students will turn out to be better off under the two guiding principles here.

Say that a matching is *priority-neutral* if and only if it is not possible to make any student whose priority is violated better off without violating the priority of some student who is made worse off.

Evidently, priority-neutral matchings are precisely those matchings that conform to the two principles above. Notice that every stable matching is priority-neutral because there are no priority violations at all.

[2] See Roth and Rothblum (1999) for advice that can be given to participants in two-sided matching market mechanisms that, in contrast to our goal here, select a stable matching.

Of course, our goal is to select a Pareto efficient matching.[3] Consequently, we seek matchings that are both Pareto efficient and priority-neutral, and so let us call any such matching *priority-efficient*. It is not at all clear whether there are many priority-efficient matchings or whether there are none at all.

Our first main result is that there always exists precisely one priority-efficient matching. Moreover, every student weakly prefers this matching to every priority-neutral matching and so to every stable matching as well. In particular, this means that students are weakly better off when our two guiding principles are used to select a Pareto efficient matching than when every student is given the absolute right to a seat at any school that violates their priority.

To implement a priority-efficient matching in practice, we will ask students to submit their preferences over schools and then we will choose the unique priority-efficient matching for the submitted preferences (school priorities and quotas are assumed known). Let us call this mechanism the *priority-efficient (PE) mechanism*.

Because the PE mechanism always selects a Pareto efficient matching that dominates the student optimal stable matching, it is not strategy-proof.[4] Nevertheless, without specific knowledge of other students' reported preferences, it is typically far from obvious how a student can gain an advantage by submitting an untruthful report. This brings us to the second important practical matter, namely to provide accurate advice to students that is likely to lead many of them to submit their true preferences.

That students might try to "game the system" by reporting false preferences is a difficulty that has occurred in practice with other mechanisms. Indeed, as one New York City Department of Education official, Peter Kerr, wrote when commenting about the change from an old system for assigning students to New York City high schools to a new system that selects a stable matching (New York Times, November 2, 2003): "The new process is a vast improvement.. . . students will be able to rank schools without the risk that naming a competitive school as their first choice will adversely affect their ability to get into the school they rank lower."

If students submit untruthful preferences, then the matching that is ultimately selected by the PE mechanism will be priority-efficient only for the submitted preferences, but perhaps not for the true preferences, a situation that we wish to avoid. Our strategic analysis of the PE mechanism permits us to give the following advice to students. Notice that the first piece of advice addresses the gaming of the old New York City system described by Peter Kerr.

[3] Pareto efficiency is a concept that is straightforward to communicate to non experts.

[4] See Abdulkadiroğlu, Pathak, and Roth (2009, Theorem 1), or Kesten (2010, Theorem 4).

Advice for students participating in the PE mechanism:

1. Always list your top choice first. Lowering your top choice can change your placement only when leaving it first would have given you your top choice.

2. Changing your reported preferences by switching the positions of two schools so that the one of them that you prefer is now truthfully ranked above the other makes it more likely that you will be placed at the one you prefer and less likely that you will be placed at the other.

3. For any k, changing your reported preferences by truthfully ranking your top k schools makes it more likely that you will be placed at a school among your top k.

4. Untruthful reporting is risky because, no matter what priorities and quotas the schools have set, for any untruthful preference that you report, either there are preferences that other students could report that would make your placement worse than had you reported truthfully, or, no matter what preferences other students report, your placement will be the same as had you reported truthfully.

In addition to the above advice that is always valid, we provide natural conditions under which truthful reporting by all students is an equilibrium of the PE mechanism. Taken altogether, these results suggest that many, even most, students would find it in their best interests to report their preferences truthfully in the PE mechanism.[5]

The remainder of the paper is organized as follows. Section 2 provides an example of a priority-efficient matching. Section 3 contains our model and notation, as well as our results on existence and uniqueness of priority-efficient matchings. Section 3 also introduces feasible sequences of matchings which play an important role in all of our analysis, especially our strategic anlaysis. Additionally, Section 3 shows that an algorithm developed by Kesten (2010) and modified by Tang and Yu (2014) can be used to compute the unique priority-efficient matching in any school choice problem. Section 4 contains our results for the strategic analysis of the PE mechanism. Section 5 reviews the literature, and Section 6 contains the proofs of all of the results stated in the main text.

2 An Example

To gain some familiarity with priority-efficient matchings, consider the situation illustrated in Figure 1 involving five students, $i_1, ..., i_5$, and five schools, $s_1, ..., s_5$ each with a quota of

[5] Empirical analysis may be helpful in establishing the extent to which false reporting might be advantageous in practice.

one seat. Student preferences are given by the table on the left and school priorities are given by the table on the right. For example, the table on the left indicates that student i_1 ranks school s_2 highest, s_1 second-highest, etc., while the table on the right indicates that school s_2 gives highest priority to student i_3, second-highest priority to student i_5, etc. Dots indicate that the remaining rankings do not matter for the purposes of this example.

Figure 2 displays several matchings for the school choice problem given in Figure 1. In each of the panels (a), (b), and (c) we have reproduced a copy of the table on the left in Figure 1 that describes the students' preferences. The shaded squares in Figure 2 indicate the student-optimal stable matching, while the three other matchings in Figure 2, $\tilde{\mu}$ in panel (a), $\hat{\mu}$ in panel (b), and μ^* in panel (c), are Pareto efficient, and two of them, $\hat{\mu}$ and μ^*, Pareto dominate the student-optimal stable matching.

Student Preferences

i_1	i_2	i_3	i_4	i_5
s_2	s_3	s_3	s_1	s_4
s_1	s_1	s_4	s_2	s_1
s_3	s_5	s_2	s_4	s_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	s_2
				s_5

School Priorities

s_1	s_2	s_3	s_4	s_5
i_2	i_3	i_1	i_4	$\vdots$
i_1	i_5	i_5	i_3	
i_5	i_4	i_2	i_5	
i_4	i_1	i_3	$\vdots$	
i_3	i_2	i_4		

(each school has one seat)

Figure 1

To see how priority-efficiency works to select a unique Pareto efficient matching, let us return to the three Pareto efficient matchings, $\tilde{\mu}$, $\hat{\mu}$, and μ^* shown in Figure 2. We will be content to show that only the μ^* matching in panel (c) is priority-efficient. Because all three matchings are Pareto efficient, to establish whether or not they are priority-efficient, we need only check whether they are priority-neutral.

Consider the matching $\tilde{\mu}$ in panel (a). This matching violates student i_2's priority at school s_3 (and also at school s_1), because student i_2 prefers school s_3 to school s_5 where she is assigned, and student i_2 has priority over student i_3 at school s_3 where i_3 is assigned. Consider now the student-optimal stable matching indicated by the green-shaded cells in Figure 2. Let us call this stable matching $\bar{\nu}$. Student i_2 prefers $\bar{\nu}$ to $\tilde{\mu}$ because $\bar{\nu}$ assigns i_2 to school s_1, which he prefers to school s_5 to which he is assigned under $\tilde{\mu}$. Moreover, because $\bar{\nu}$

is stable, it violates no student's priority. Consequently, $\tilde{\mu}$ is not priority-neutral because the matching $\bar{\nu}$ makes a student, i_2, whose priority is violated by $\tilde{\mu}$ better off without violating the priority of any student at all. So $\tilde{\mu}$ is not priority-efficient.

Next, consider the matching $\hat{\mu}$ in panel (b). This matching too violates student i_2's priority at school s_3 for exactly the same reasons as in the previous paragraph. Let us compare the matching $\hat{\mu}$ to the matching μ^* given in panel (c). Student i_2 prefers μ^* to $\hat{\mu}$, and the only student who finds μ^* worse than $\hat{\mu}$ is student i_3, and student i_3's priority is not violated by μ^*. But this means that $\hat{\mu}$ is not priority-neutral because switching to the matching μ^* makes student i_2, whose priority is violated by $\hat{\mu}$, better off without violating the priority of any student who is made worse off. So $\hat{\mu}$ too is not priority-efficient.

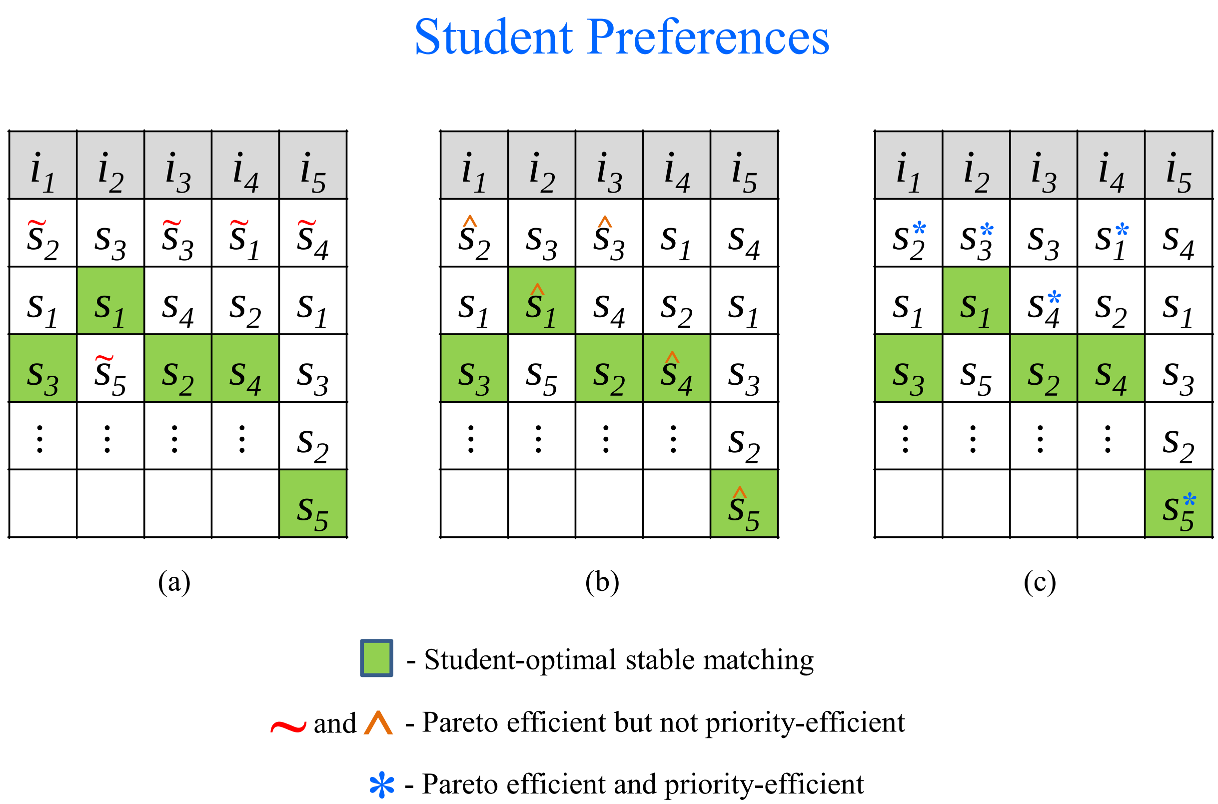

Figure 2

Finally, consider the matching μ^*. We wish to show that μ^* is priority-neutral. To see this, notice first that the only student whose priority is violated by μ^* is student i_5. But in order to make student i_5 better off, one of the other students would have to be assigned to school s_5, which would make that student worse off. Moreover, it is easy to check that, no matter which of the students $i_1, ..., i_4$ is assigned to school s_5, that student's priority will be violated. Consequently, it is not possible to change the matching from μ^* so as to make the only student whose priority is violated by μ^*, student i_5, better off without violating the priority of a student who is made worse off. Hence, μ^* is priority-neutral and therefore, being Pareto efficient, it is priority-efficient. Notice also that μ^* Pareto dominates the student-optimal stable matching.

3 Formal Model and First Results

Let I denote the nonempty finite set of students and let S denote the nonempty finite set of schools. Each school $s \in S$ has a finite number of available seats, or quota, $q_s \in \{1, 2, ...\}$, and has a strict priority ordering P_s over the set of students. Each student $i \in I$ has a strict preference ordering P_i over the set of schools, and we write sR_it to mean sP_it or $s = t$. We assume that $\#I \leq \sum_{s \in S} q_s$, so that the total number of students is not greater than the total number of available seats. All of these elements are fixed throughout the analysis, unless stated otherwise.

A *matching* is any mapping $\mu : I \to S$ such that for every $s \in S$, $\#\mu^{-1}(s) \leq q_s$. Implicit in this definition is the assumption that each student is assigned to some school.[6]

For any two matchings μ and ν, we reduce notation by writing $\mu P_i \nu$ instead of $\mu(i)P_i\nu(i)$ and by writing $\mu R_i \nu$ instead of $\mu(i)R_i\nu(i)$.

Say that a matching μ *violates student i's priority* iff there is $s \in S$ such that $sP_i\mu(i)$ and, either, $\#\mu^{-1}(s) < q_s$,[7] or, iP_sj for some $j \in \mu^{-1}(s)$. We then also say that μ violates i's priority at school s.

A matching μ is (pairwise) *stable* iff μ does not violate any student's priority (at any school).[8]

Say that a matching μ *dominates* a matching ν iff $\mu R_i \nu$ for every $i \in I$. So by strict preferences, a matching μ Pareto dominates a matching ν if and only if μ dominates ν and $\mu \neq \nu$.

A matching is a *student-optimal stable matching* iff it is stable and it dominates every other stable matching.

Gale and Shapley (1962) show that a student-optimal stable matching always exists and that it is unique.

[6] One can accommodate the possibility of home-schooling by including in S, for each student i, a school, h_i, (the "home-school" for student i) with quota one that gives highest priority to student i, where each student ranks every other student's home school below every non-home school and below his own home school.

[7] By convention, every student has priority over every empty seat.

[8] As is well known, nothing changes if stability were to also require the absence of priority violations against subsets of students if it is assumed that schools have responsive preferences since, in that case, a school can violate the priority of a subset of students only if it violates the priority of some student in the subset. So a matching would be stable is this more restrictive sense if and only if it is pairwise stable as defined here.

Say that a matching μ is *priority-neutral* iff no matching ν can make any student whose priority is violated by μ better off unless ν violates the priority of some student who is made worse off.

A matching is a *student-optimal priority-neutral matching* iff it is priority-neutral and it dominates every other priority-neutral matching.

Say that a matching μ is *priority-efficient* iff it is priority-neutral and Pareto efficient.

We can now state our first result. All proofs are in Section 6.

Theorem 3.1 *There is a unique priority-efficient matching. This priority-efficient matching is also the unique student-optimal priority-neutral matching. Every stable matching is priority-neutral and therefore the unique priority-efficient matching dominates the student-optimal stable matching.*

Remark 3.2 *It can be shown that the set of priority-neutral matchings is a lattice with respect to the coordinatewise partial order defined by the students' preferences. We do not know whether this lattice is distributive nor do we know whether it is a sublattice of the set of all matchings, both of which hold for the lattice of stable matchings. While the join of any two priority-neutral matchings (i.e., the smallest priority-neutral matching that is larger than both) is always their coordinatewise maximum, we do not know whether the meet is always their coordinatewise minimum. Nevertheless, for any two priority-neutral matchings there is always a largest priority-neutral matching that is smaller than both. So their meet is always well-defined within the set of priority-neutral matchings.*

We next give a convenient characterization of the unique priority-efficient matching.

Theorem 3.3 *A matching μ is priority-efficient if and only if no matching ν can make any student better off unless ν violates the priority of some student who is made worse off.*

Notice that in the statement of Theorem 3.3, the matching ν can make any student better off. In particular, ν does not need to make some student whose priority is violated by μ better off, as is required in the definition of priority-neutrality. So checking that a matching is priority-efficient is very much like checking that it is Pareto efficient, except that a "dominating" matching can ignore the preferences of students whose priorities it does not violate.

We next show that the unique priority-efficient matching happens to be the output of a simple, elegant, and well-studied modification of the deferred acceptance algorithm.

3.1 The Kesten-Tang-Yu Algorithm

Kesten (2010), in his study of Pareto-efficient matching in the school choice problem, introduced an important modification of Gale and Shapley's (1962) student-proposing deferred acceptance algorithm (henceforth simply *the DA algorithm*). Kesten's (2010) algorithm, called EADA (efficiency adjusted deferred acceptance) has since been modified by Tang and Yu (2014). The matching that is produced by Tang and Yu's (2014) modification is identical to that produced by Kesten's (2010) algorithm (Tang and Yu 2014, Theorem 3), but the Tang and Yu algorithm is rather more efficient and it turns out to be simpler to work with for our purposes. Before describing Tang and Yu's modification of Kesten's algorithm, we first briefly review Gale and Shapley's DA algorithm, which proceeds in steps and works as follows.

In the first step, all students apply to their favorite school. Then each school places each applicant, in order of highest priority, on their waitlist until their quota is reached, and rejects all remaining applicants. In each subsequent step, all students rejected in the previous step apply to their favorite school among those who have not yet rejected them. Then, each school places each of its applicants (both new and waitlisted), in order of highest priority, on their new waitlist until their quota is reached, and rejects all remaining applicants. The algorithm stops after any step in which no school rejects any students. Each student is then assigned to the school at which he is currently waitlisted.

Tang and Yu's (2014) algorithm makes use of so-called "underdemanded schools." Say that a school s is *underdemanded* at a matching μ iff no student prefers s to the school to which they are assigned by μ.

Gale and Shapley (1962, Theorem 2) showed that the DA algorithm always stops in finitely many steps and always produces the student-optimal stable matching, $\hat{\mu}$ say. Notice that, because students apply to schools in order starting with their most preferred, they prefer a school s to the school to which they are assigned by $\hat{\mu}$ if and only if s rejected them at some point during the DA algorithm. Consequently, a school is underdemanded at the student-optimal stable matching if and only if that school did not reject any students during the DA algorithm.

We can now describe Tang and Yu's modification of Kesten's algorithm. The Tang and Yu (2014) algorithm, called sEADA* (simplified EADA) proceeds in rounds and works as follows.[9]

In the first round, run the DA algorithm with the entire set of schools and with the entire

[9] The asterisk in sEADA* indicates that we are considering here only the special case of the Tang and Yu algorithm in which all students "consent" to allowing their priorities to be violated. See also Kesten (2010) where the idea of consent originated.

set of students, yielding the student-optimal stable matching, μ_1 say. Each school s that is underdemanded at μ_1 (i.e., each school that did not reject any students during the execution of the DA algorithm),[10] is permanently assigned its students $\mu_1^{-1}(s)$ for the remainder of the algorithm and school s and its students are called *settled*. Remove all settled schools and their students and proceed to round two.[11] The second round proceeds exactly as the first but where the DA algorithm is applied only to the "submarket" of unsettled schools and students. This submarket's underdemanded schools at its DA matching are permanently assigned their students, and these students and schools become settled and are removed, altogether resulting a second-round matching μ_2 (which includes the permanently assigned students and their schools from the first round). These rounds repeat, with each round n producing a matching μ_n, and with smaller sets of unsettled schools and students with each successive round. The algorithm ends after the round, N say, in which all remaining schools and students become settled, thereby defining the matching, μ_N, that is the output of the algorithm. We will then call $\mu_1, ..., \mu_N$ the sEADA* *output sequence*.

3.2 Feasible Sequences of Matchings

Following Tang and Yu (2014), for any matching μ and for any student i, say that student i is *(Pareto) μ-improvable* if and only if there is a matching that dominates μ and that makes student i strictly better off. Say that student i is *μ-unimprovable* if and only if student i is not μ-improvable.

We next define a class of finite sequences of matchings that will turn out to be very important for our analysis of priority-efficient matchings.

Say that a finite sequence of matchings $\mu_1, \mu_2, ..., \mu_N$ is *feasible* if and only if μ_1 is stable, μ_N is Pareto efficient, and, for each $n > 1$, μ_n dominates μ_{n-1} and μ_n does not violate the priority of any μ_{n-1}-improvable student.

An immediate question is whether any feasible sequences of matchings exist. Tang and Yu (2014) establish that the sEADA* algorithm is well-defined (i.e., that it always produces a matching in finitely many rounds). They also establish a number of properties of the sEADA* output sequence that we use to establish the following result.

Theorem 3.4 *The sEADA* output sequence is well-defined and feasible.*

Theorem 3.4 establishes the existence of at least one feasible sequence of matchings. But there can be others. In fact we can typically generate many feasible sequences by

[10] It is well-known that at least one such school always exists. See, e.g. Gale and Sotomayor (1985) for the case of one to one matching.

[11] In contrast to Tang and Yu (2014), we find it more convenient to remove underdemanded schools at the end of each round rather than at the beginning of each round.

adjusting the sEADA* algorithm in various ways. For example, if in any number of rounds we remove only some, but not necessarily all underdemanded schools and their students, but otherwise leave the algorithm unchanged, the sequence of matchings that is produced will change, but will nevertheless remain feasible. Also in any round, in addition to removing all underdemanded schools and their students one can remove additional schools and their students so long as those students are unimprovable.[12] In addition, one can change the algorithm by choosing, in any round's submarket, any stable matching instead of the student-optimal stable matching, so long as there is at least one underdemanded school in the submarket at the chosen stable matching. Once again the sequence of matchings will typically change, but will remain feasible. These facts will be especially useful to us in our strategic analysis.[13]

The following result sheds some light on why feasible sequences of matchings are so relevant for our analysis.

Theorem 3.5 *If $\mu_1, ..., \mu_N$ is feasible, then μ_N is the unique priority-efficient matching.*

Remark 3.6 *Theorems 3.4 and 3.5 have two immediate implications. First, because the sEADA* algorithm produces the same matching as Kesten's (2010) algorithm, Kesten's algorithm (in his case when all students consent) always yields the unique priority-efficient matching. Second, the unique priority-efficient matching can be efficiently computed by running the sEADA* algorithm, which requires at most $\#S - 1$ executions of the DA algorithm (since at least one school is settled after each round of the sEADA* algorithm and if there is ever just one unsettled school remaining, it is not actually necessary to execute the DA algorithm in that last trivial round).*

We next turn our attention to strategic matters.

4 Strategic Considerations

We consider here the direct mechanism in which students submit their preference lists and the mechanism (which is assumed to have access to the true profiles of school priorities and quotas) chooses the unique priority-efficient matching given the submitted preferences. Let us call this mechanism the *priority-efficient (PE) mechanism.*

Abdulkadiroğlu, Pathak, and Roth (2009) show that no mechanism that Pareto dominates the mechanism that chooses the student-optimal stable matching is strategy-proof. In

[12] For example, one can remove any school and its students if that school is the top choice of all its students.
[13] See the proofs of Theorems 4.1 and 4.2 in Part B of Section 6.

particular therefore, the PE mechanism is not strategy-proof.[14] Nevertheless, building on Roth and Rothblum (1999), Ehlers (2008), and Kesten (2010), we can identify plausible circumstances under which it is an equilibrium of the PE mechanism for all students to report their preferences truthfully.

So consider a situation in which the set of schools S and the set of students I are known, but there is incomplete information about school priorities, quotas, and student preferences. A *state* specifies each school's priority list over students, each school's quota, and each student's preferences over schools. Incomplete information about the state is captured by a prior distribution over all possible states, where these are restricted so that in every state the total number of available seats across all schools is at least as large as the total number of students.

Let Σ denote the set of all possible states, with typical element σ, where each $\sigma \in \Sigma$ takes the form $\sigma = (P, \Psi, q)$, where $P = (P_i)_{i\in I}$ is any profile of (strict) student preferences and $\Psi = (\Psi_s)_{s\in S}$ is any profile of (strict) school priority lists, and $q = (q_s)_{s\in S}$ is any specification of nonnegative school quotas satisfying $\sum_{s\in S} q_s \geq \#I$. So for any student i, if $\sigma = (P, \Psi, q)$, then $\sigma_i = P_i$ specifies i's preferences and σ_{-i} specifies all other students' preferences and all school priorities and quotas.

For any state $\sigma \in \Sigma$, let $f(\sigma) \in M$ denote the priority-efficient matching when the state is σ, and let $f_i(\sigma)$ denote the school to which student i is assigned under the matching $f(\sigma)$.

The PE mechanism works as follows. After a state is chosen according to the prior, each student observes only his own preferences and the school district observes only the realized priorities and quotas of each school. Students are asked to submit their preferences to the school district, and the school district is committed to selecting the priority-efficient matching given the submitted preferences and the realized school priorities and quotas.[15]

The following result justifies the advice for students given in Section 1.

Theorem 4.1 *Consider any student i who is participating in the PE mechanism. Fix the preferences submitted by students other than i, and fix the realized priorities and quotas of the schools. Let P_i and P_i' be two possible preferences for student i, neither of which need be i's true preference. Then we have the following.*

1. Suppose that P_i' is obtained from P_i by lowering only the top-ranked school s according to P_i. If i's assigned school changes when his submitted preference changes from P_i to P_i', then he is assigned to s when he submits P_i.

[14] Example 7 in Kesten (2010) shows that Kesten's mechanism is not strategy-proof. So by Remark 3.6, or by direct computation, the same example shows that the PE mechanism is not strategy-proof.

[15] So the students are the only strategic agents in this situation.

2. Suppose that P_i' is obtained from P_i by switching the positions of two schools, s and t, where sP_it. If i's assigned school is s when he submits the preference order P_i', then it is also s when he submits the preference order P_i. In fact, every other student's assigned school remains unchanged as well.[16]

3. Let K be any set of k schools. Suppose that P_i is obtained from P_i' by raising only the schools in K so that they become the top k schools (their relative order need not be unchanged). If i's assigned school is in K when he submits the preference order P_i', then it is also in K when he submits the preference order P_i.

4. Suppose that i's true preference is P_i and that i prefers his assigned school when he submits P_i' to his assigned school when he submits P_i. Then, keeping the realized school priorities and quotas unchanged, there exist preferences that other students could submit that would make i prefer his assigned school when he submits P_i to his assigned school when he submits P_i'.

Statements 1-3 in Theorem 4.1 each provide a sense in which truthful reporting can be advantageous,[17] while statement 4 provides a sense in which untruthful reporting can be disadvantageous. All four statements are valid without any conditions on the prior distribution over states, the preferences that others might submit or the priorities and quotas submitted by schools. We next develop conditions under which truthful reporting is an equilibrium.

Let $Q \in \Delta(\Sigma)$ be the prior distribution over states, and let $\tilde{\sigma}$ denote the lottery (random variable) over states that is generated by Q. Not all states $\sigma \in \Sigma$ need receive positive probability. If the state σ has positive probability and student i observes that his true preferences are σ_i, then we let $\tilde{\sigma}_{-i}|_{\tilde{\sigma}_i=\sigma_i}$ denote the lottery over the remaining components of the state conditional on the observation $\tilde{\sigma}_i = \sigma_i$.

Under the PE mechanism, if the true state is σ and student i alone deviates from truthful reporting and reports the preference P_i instead of σ_i, then the outcome of the mechanism would be the unique priority-efficient matching $f(P_i, \sigma_{-i})$ and $f_i(P_i, \sigma_{-i})$ would be student i's assigned school under this matching. Consequently, after observing that his preferences are σ_i, if student i reports P_i, then conditional on his information and conditional on all other students reporting truthfully, i's assigned school is random and is given by the lottery $f_i(P_i, \tilde{\sigma}_{-i}|_{\tilde{\sigma}_i=\sigma_i})$.

In order to define a concept of equilibrium here, we must specify how each student i ranks lotteries over schools. But in fact, we will not need to rank all pairs of lotteries over

[16] Thus, in particular, the PE mechanism satisfies the condition that Elhers (2008) calls "positive association," which only requires that student i's assignment remains unchanged.

[17] Let P_i^* be i's true preference. For statement 1 suppose that s is the top choice according to P_i^*. For statement 2, suppose that sP_i^*t. For statement 3, suppose that the schools in K are the k most-preferred schools according to P_i^*.

schools. It will be enough to rank only pairs of lotteries that are comparable in a first-order stochastic dominance sense.

For any student i, for any preference P_i for i, and for any two lotteries $\tilde{s}$ and $\tilde{t}$ over schools, following Ehlers (2008) say that $\tilde{s}$ *stochastically P_i-dominates* $\tilde{t}$ iff for every school $s \in S$,

$$\Pr(\tilde{s} R_i s) \geq \Pr(\tilde{t} R_i s).$$

It is not difficult to show that one lottery stochastically P_i-dominates another if and only if the one can be obtained from the other by successively shifting probability weight from a school that is ranked lower by P_i to a school that is ranked higher. We can now use Elhers' (2008) concept of an "ordinal" equilibrium.

Say that truth-telling is an *ordinal equilibrium* of the PE mechanism iff for every positive probability state $\sigma \in \Sigma$, for every student i, and for every possible preference P_i for student i,

$$f_i(\sigma_i, \tilde{\sigma}_{-i}|_{\tilde{\sigma}_i=\sigma_i}) \text{ stochastically } \sigma_i\text{-dominates } f_i(P_i, \tilde{\sigma}_{-i}|_{\tilde{\sigma}_i=\sigma_i}).$$

We next provide conditions on the prior Q under which truth-telling is an ordinal equilibrium of the PE mechanism.

4.1 Similarity Partitions and Student-Oriented Preferences and Priorities

Let $\mathcal{S}$ be a finite partition of the set of schools S. Let P be any profile of student preferences. Say that $\mathcal{S}$ is a *similarity partition* for P iff for each pair of sets in $\mathcal{S}$, and for each student $i \in I$, P_i ranks every school in one of the sets above every school in the other. (Note that different students can rank the sets in $\mathcal{S}$ differently.) Each of the sets in S is called a *similarity set* (for P).

For any profile P of student preferences and for any profile Ψ of school priorities, say that (P, Ψ) is *student-oriented* with respect to a similarity partition $\mathcal{S}$ for P iff for any two schools s and t in different similarity sets in $\mathcal{S}$, and for any two students i and j, if sP_it and tP_js, then $i\Psi_sj$ and $j\Psi_ti$. In words, (P, Ψ) is student-oriented with respect to a similarity partition of the schools if and only if for any two schools in distinct similarity sets, if any two students rank the two schools differently, then each of the two students has higher priority at the school he prefers than the other student.[18]

The existence of a similarity partition means that each student can order the similarity sets in the partition from best to worst, but it allows different students to order the sets

[18] Reny (2020) shows that if preferences and priorities are student-oriented with respect to the school-partition whose elements are all singletons, then there is a unique stable matching and it is Pareto efficient.

differently. If preferences and priorities are also student-oriented, then in case two students do order two similarity sets differently, each of those students has higher priority at each school in the set that he prefers than the other student.

Notice that if all students order the similarity sets within a similarity partition the same way, as Kesten (2010) assumes, then the profile of preferences and priorities is trivially student-oriented. But preferences and priorities can be student-orientated even when students rank similarity sets differently, which can be plausible in practice. For example, athletes might prefer schools with strong football programs and musicians might prefer schools with strong music programs. If "schools for the arts" form one similarity set and "schools for athletics" form another (disjoint set), then athletes and musicians are likely to rank these similarity sets differently but may also be likely to receive higher priority at the schools in the similarity set they prefer. Similarly student-orientated preferences and priorities are also plausible when public school priority rules are designed to align with student (parent) preferences, as would be the case, for example, when priority is given to students with siblings already enrolled in a school precisely because enrolling all of their children at the same school is preferred by parents.

Following Kesten (2010) (see also Roth and Rothblum 1999, Elhers 2008, and Elhers and Morril 2020), say that the prior Q *induces symmetric information* on a partition $\mathcal{S}$ of the set of schools S iff for any two schools s and t in the same element of the partition and for any student i, each state σ that is given positive probability by Q is as likely as the state that is identical to σ in every respect except that school s has the priority order and quota of school t and vice versa, and schools s and t are switched in the preference list of every student except student i.[19]

Symmetric information on a similarity partition $\mathcal{S}$ means that any student, after observing his own preferences, views any two schools in the same similarity set symmetrically insofar as how other students rank the two schools and what priorities and quotas they may have. Notice that correlation between preferences and quotas is permitted, which may be plausible if, for example, a student knows that some students prefer larger schools while others prefer smaller schools.[20] Note also that Q induces symmetric information trivially

[19] Our definition here is identical to Kesten's (2010) because Kesten implicitly conditions on student i's observation of his own preferences. Ehlers (2008) and Ehlers and Morrill (2020) make the stronger assumption that symmetric information holds when there is a single similarity set consisting of the entire set of schools.

[20] But symmetric information on a similarity partition does imply independence for various marginal distributions. For example, for any even number of students in the same similarity set, the marginal distribution over profiles of their preferences must be uniform over all possible profiles of their preferences. Consequently, if, as in Ehlers (2008) and Ehlers and Morrill (2020), there is symmetric information on the trivial partition of S into the single similarity set S itself, then the marginal distribution over the preferences of all students is uniform and has full support if $\#I$ is even, while if $\#I$ is odd, then the marginal distribution over the preferences of any $\#I - 1$ students is uniform and has full support.

when similarity sets are singletons.

Theorem 4.2 *Suppose that $\mathcal{S}$ is a partition of S and that the prior Q induces symmetric information on $\mathcal{S}$. Furthermore, suppose that for any state $\sigma = (P, \Psi, q)$ that is given positive probability by Q, $\mathcal{S}$ is a similarity set for P and (P, Ψ) is student-oriented with respect to $\mathcal{S}$. Then, truth-telling is an ordinal equilibrium of the PE mechanism.*

Remark 4.3 *(a) If the set of schools does not include the possibility of home-schooling, then the above strategic analysis does not permit truncation strategies. But these are easily accommodated by including in S a home-school for each student as described in footnote 6, and letting each home-school be a singleton similarity set. Theorem 4.2 remains valid and establishes that no deviation, including a deviation to a truncation strategy, can improve upon truthful reporting when others report truthfully. (b) Theorem 4.2 is stated as an equilibrium result. But playing equilibrium strategies, even truth-telling strategies, can sometimes entail more coordination than may be plausible for the circumstances under study. But in fact, the proof of Theorem 4.2 shows that student i can do no better than to report truthfully even if he has a personal prior Q_i that describes the joint distribution over his own preferences, the reported preferences of other students, and the school priorities and quotas that will be used to compute the priority-efficient matching, so long as Q_i satisfies the conditions of Theorem 4.2. (c) When the conditions of Theorem 4.2 hold and all similarity sets in $\mathcal{S}$ are singletons, then (see Reny 2020) in every positive probability state, the student-optimal stable matching will be unique and Pareto efficient. Hence, the student-optimal stable matching will coincide with the priority-efficient matching and so by Dubins and Freedman (1981), truth-telling is an ex-post Nash equilibrium of the PE mechanism.*

5 Related Literature

Balinski and Sönmez (1999) initiated the application of matching theory to the problem of assigning students to schools by recognizing that the rules that Turkish universities use to map multiple test scores for each student into a ranking of the students defines, for each university, a surrogate preference relation over students. With this observation, Balinski and Sönmez could apply results from Gale and Shapley (1962) to this school choice problem even though schools themselves are not active participants, but are merely objects to be allocated among students.[21] Balinski and Sönmez (1999) also recognized the problem of

[21] We do not include in this literature review papers in which schools are active participants with preferences over students, as for example, Ehlers (2007) and Mauleon, Vannetelbosch, and Vergote (2011), both of which introduce interesting set-valued solutions that are based upon the concept of von Neumann-Morgenstern stable sets (von Neumann and Morgenstern, 1944).

Pareto inefficiency, noting that no matching mechanism can always produce a matching that is stable and Pareto efficient for students (see their Lemma 3).

Abdulkadiroğlu and Sönmez (2003) focus more directly on the conflict between stability and Pareto efficiency in school choice, and, building upon Balinski and Sönmez (1999), they even more explicitly open up the line of research that applies matching theory to the problem of school choice. They compare two mechanisms, one of which (the student-optimal stable matching (SOSM) mechanism) always produces a stable matching and the other of which (a novel Gale-inspired top trading cycles (TTC) mechanism) always produces a Pareto efficient matching. Both mechanisms are strategy-proof and so both mechanisms can sometimes produce matchings that fail to be priority-efficient. Abdulkadiroğlu and Sönmez (2003) describe how both the SOSM and TTC mechanisms can be adapted so as to satisfy additional constraints that arise in actual school-choice problems in various U.S. cities, and they establish several properties of these adapted mechanisms.

Pápai (2000) characterizes the class of matching mechanisms that are Pareto efficient, group strategy-proof, and reallocation-proof. This class contains, for example, Gales top trading cycles algorithm but contains other related mechanisms as well. Since the PE mechanism is not strategy-proof (let alone group strategy-proof), every mechanism in the class characterized by Pápai will, for some profile of submitted preferences, fail to select the unique priority-efficient matching.

Roth and Rothblum (1999) were the first to investigate whether, in a mechanism that always selects the firm-optimal stable matching, workers have an incentive to misrepresent their preferences when they have little information about the preferences of other workers and of firms. They show that workers can restrict their untruthful preference reports to truncated lists that truthfully rank firms in order from best to worst, but perhaps stopping before the first unacceptable firm. Ehlers (2008) generalizes these results to a large class of mechanisms.

Erdil and Ergin (2008) are concerned with the constrained efficiency problem that seeks a stable matching that is undominated among all stable matchings in the presence of priority rules with ties, and they develop a novel algorithm for finding such matchings. Abdulkadiroğlu, Pathak, and Roth (2009) are also concerned with this constrained efficiency problem. They establish the theoretical result (see their Theorem 1) that, for any priority tie-breaking rule, there is no strategy-proof mechanism that always produces a matching that dominates the student-optimal stable matching for that tie-breaking rule. Most relevant for our purposes is their empirical analysis of the unconstrained efficiency losses that arise from choosing a stable matching rather than a Pareto efficient matching. As we have already remarked in Section 1, in a New York City school district in 2006-2007, Abdulkadiroğlu, Pathak, and

Roth (2009) find that over 4,000 grade 8 students could have been made better off by being reassigned to a school different than their match in the student-optimal stable matching, without hurting any other students. More precisely, given the preferences submitted by the roughly 80,000 grade 8 students participating in the match that year and given the priorities and quotas of the NYC high-schools, on average 4,000 students could be made better off over 250 random and uniform draws for breaking ties in priority rules. So a significant number of students are likely to benefit if instead a priority-efficient matching were chosen, assuming that students and schools would submit similar preferences, priorities and quotas. Our strategic analysis, in which only students are considered strategic actors, suggests that many students would find it in their best interest to submit their true preferences and so we might indeed expect a significant improvement in the efficiency of the match.

Kesten (2010) modifies the DA algorithm so as to produce a Pareto efficient matching in the school choice problem when all students consent in advance to allow their priorities to be violated if necessary. Kesten's modification eliminates the effect of students who are eventually rejected by a school to which they apply and who preclude some other student from attending that school as a result of their initial application. Kesten's (2010) algorithm exhibits important incentive properties when viewed as a direct mechanism that maps submitted preferences into a matching and when students have enough imperfect information about other students' preferences and about school priorities and quotas. In particular, no student's assignment depends on whether he consents or not, and so consenting cannot ever be harmful. As we have seen in Section 3.1, Tang and Yu's (2014) modification of Kesten's algorithm yields the same matching as Kesten's algorithm and the matching that is produced by both of these algorithms is always the unique priority-efficient matching when all students consent.

Alcalde and Romero-Medina (2017) show that any Pareto efficient matching that dominates the student optimal stable matching is characterized by the property that it is immune to what they call "admissible objections." Consequently, by Theorem 3.1, the unique priority-efficient matching is immune to their admissible objections.

Dur et. al. (2019) analyze a generalized school choice problem in which school priorities can be partitioned into two groups, "hard priorities" that can never be violated and "soft priorities" that can be violated. When the collection of soft priority violations satisfies a natural "connectedness" condition, Dur et. al. provide a family of algorithms that always produce a matching that is Pareto efficient among the set of matchings that do not violate any hard priorities. Algorithms of this nature are especially helpful when the set of soft priorities are known in advance. See Dur et. al. (2019) for examples. In contrast, whether a priority-efficient matching will violate a student's priority at some school generally depends upon the preferences of other students. So it is not possible, in general, to know in advance which priorities would need to be designated as soft when implementing the mechanism that selects the unique priority-efficient matching. Kesten (2010) makes the same point in the context of his EADA mechanism, which is why, in his mechanism, students are asked to give consent to have their priorities violated in advance, i.e., at the same time at which they submit their preferences.

Troyan et. al. (2020) define a matching to be "essentially stable" if reassigning any student i to a school s at which his priority is violated always causes a reassignment chain (i.e., a sequence of students starting with i who are successively reassigned to their favorite school at which there is a student – the next student in the sequence – with lower priority) that eventually displaces student i from school s. Clearly, every stable matching is essentially stable. Troyan et. al. show that there is always at least one, and sometimes more than one, essentially stable matching that is Pareto efficient. So essential stability is not always able to select a unique matching among those that are Pareto efficient. They also show that the outcome of Kesten's (2010) EADA algorithm is essentially stable and so, by Theorems 3.4 and 3.5, the unique priority-efficient matching is essentially stable.

Ehlers and Morrill (2020) define a set of matchings to be *legal* if and only if that set satisfies certain "internal" and "external" consistency properties. They show that there is exactly one legal set of matchings and that among this set there is a matching that dominates all the others. Thus, there is a unique student-optimal legal matching. Ehlers and Morrill (2020) do not assume that school preferences are responsive, but instead assume that schools have substitute preferences as in Blair (1988) and so their setting is more general than the setting we consider here. However, when their schools are assumed to have responsive preferences, it can be shown that their student-optimal legal matching is the unique priority-efficient matching.

6 Proofs.

This section is broken into two parts, A and B. In part A, student preferences, P, and school priorities and quotas, (Ψ, q), are fixed, but can be any that satisfy the conditions in Section 3. That is, (P, Ψ, q) is any state in Σ, where Σ is as defined in Section 4. Since (P, Ψ, q) is fixed throughout Part A, it is omitted from our notation there.

In contrast, the proofs in Part B sometimes require the consideration of multiple states $\sigma, \sigma', \sigma'' \in \Sigma$ and so in Part B we will be explicit about the state when necessary.

6.1 Part A.

Throughout this section, we let M denote the set of all matchings. For any pair of matchings $\mu, \nu \in M$, define the functions $\mu \vee \nu$ and $\mu \wedge \nu$ (the "join" and "meet" of μ and ν, respectively), each mapping I into S, as follows. For each $i \in I$, $[\mu \vee \nu](i)$ is the one school, $\mu(i)$ or $\nu(i)$, that student i weakly prefers to the other, and $[\mu \wedge \nu](i)$ is the one school, $\mu(i)$ or $\nu(i)$, that student i prefers weakly less than the other. Note that these functions are well-defined because student preferences are strict and so indifference between $\mu(i)$ or $\nu(i)$ can occur only if $\mu(i) = \nu(i)$.

Our first lemma provides a modest but useful generalization of the standard lattice property of stable matchings that is basic to much of matching theory. Ehlers and Morrill (2020, Lemmas 4 and 5) prove a similar result, but restrict attention to pairs of matchings that are individually rational. In contrast, we do not impose this, albeit typically harmless, restriction here.

Lemma 6.1 *Let μ and ν be any two matchings and suppose that for every student i, μ does not violate i's priority at $\nu(i)$ and ν does not violate i's priority at $\mu(i)$. Then both $\mu \vee \nu$ and $\mu \wedge \nu$ are matchings. Moreover, for each school, the number of students assigned to that school is the same under μ, ν, $\mu \vee \nu$, and $\mu \wedge \nu$.*

Proof. Let $B := \{i \in I : \nu P_i \mu\}$ be the set of students in I who strictly prefer the matching ν to the matching μ. Suppose that the matching switches from μ to ν. Then every student in B changes schools. Let i be any student in B. Student i's new school $\nu(i)$ must be at its quota under μ and every student assigned by μ to $\nu(i)$ must have higher priority at school $\nu(i)$ than student i, since otherwise i's priority would be violated by μ at school $\nu(i)$ (because $\nu P_i \mu$) violating one of our assumptions. Consequently, if we imagine that all seats are numbered, then in the move from μ to ν, each student $i \in B$ can point to a unique other student, $f(i)$ say, whose seat i takes (i.e., $\nu(i) = \mu(f(i))$), who switches schools (since i's new school $\nu(i)$ was at its quota under μ), and who has higher priority than i at i's new school $\nu(i)$. The mapping $f : B \to I$ so defined satisfies $\nu(i) = \mu(f(i))$ for every $i \in B$, and is one to one since no student is pointed to by more than one student. We next show that f is a bijection on B. Since we already know that f is one to one, we need only show that the image of f is contained in B.

Let i be any student in B and let $j = f(i)$ be the student that i points to. We must show that j is in B. To see this, note first that $\nu(i) = \mu(j) \neq \nu(j)$, because i takes j's seat and j changes schools. Since $\mu(j) \neq \nu(j)$ and preferences are strict, either $\nu P_j \mu$ or $\mu P_j \nu$. If we suppose that the latter holds, then ν violates j's priority at school $\mu(j) = \nu(i)$ because

j has priority over i at school $\nu(i)$, violating one of our assumptions. Hence, we must have $\nu P_j \mu$ and so $j \in B$ as desired. Hence f is a bijection on B.

It remains only to show that $\mu \vee \nu$ and $\mu \wedge \nu$ are matchings. For every $i \in B$, $[\mu \vee \nu](i) = \nu(i)$. Therefore, since $\nu(i) = \mu(f(i))$ for $i \in B$, we have that $[\mu \vee \nu](i) = \mu(f(i))$ for $i \in B$ and $[\mu \vee \nu](i) = \mu(i)$ for $i \in I \backslash B$. This says that $\mu \vee \nu$ can be obtained from μ by having each student $i \in B$ take the seat of student $f(i) \in B$, leaving the school assignments of students outside B unchanged. Since $f : B \to B$ is a bijection, this seating change is feasible, and so changing the matching from μ to $\mu \vee \nu$ does not change the number of students attending any school. Similarly, $[\mu \wedge \nu](i) = \nu(f^{-1}(i))$ for $i \in B$ and $[\mu \wedge \nu](i) = \nu(i)$ for $i \in I \backslash B$, and so $\mu \wedge \nu$ is a matching because it can be obtained from ν by having each student $i \in B$ take the seat of student $f^{-1}(i) \in B$, leaving the school assignments of students outside B unchanged. So changing the matching from ν to $\mu \wedge \nu$ does not change the number of students attending any school.

Switching the roles of μ and ν throughout the argument establishes that changing the matching from ν to $\mu \vee \nu$ or from μ to $\mu \wedge \nu$ does not change the number of students attending any school. Hence, for any school, the number of students attending that school is the same under μ, ν, $\mu \vee \nu$, and $\mu \wedge \nu$. ■

We can now give an alternative characterization of priority-efficient matchings. The next lemma is essentially the same as Theorem 3.3, but without the uniqueness claim. The proof of Theorem 3.3 follows once uniqueness is established (see below).

Lemma 6.2 *A matching μ is priority-efficient if and only if no matching ν can make any student better off unless ν violates the priority of some student that it makes worse off.*

Proof. We begin with the "if" direction. So suppose that no matching ν can make any student better off unless ν violates the priority of some student that it makes worse off. Then, clearly, no matching ν can make any student better off without making some student worse off, and so μ is Pareto efficient. It remains to show that μ is priority-neutral. But this follows immediately because if ν makes some student whose priority is violated μ better off, then by hypothesis ν violates the priority of some student that it makes worse off.

For the "only if" direction, suppose that μ is priority-efficient and that some student prefers ν to μ. We must show that ν violates the priority of some student who prefers μ to ν.

Let $B = \{i \in I : \nu P_i \mu\}$ and let $W = \{i \in I : \mu P_i \nu\}$. Since B is the set of students who prefer ν to μ, we know that B is nonempty. Since W is the set of students who prefer μ to ν, we must show that ν violates the priority of some student in W.

If there is a student $i \in B$ whose priority is violated by μ, then because μ is priority-neutral, ν must violate the priority of some student who prefers μ to ν and we are done. So we may henceforth suppose that there is no student in B whose priority is violated by μ. In particular, we may suppose that for every student $i \in I$, μ does not violate i's priority at $\nu(i)$.

We must show that ν violates the priority of some student in W. Suppose, by way of contradiction, that there is no student in W whose priority is violated by ν. In particular then, for every student $i \in I$, ν does not violate i's priority at $\mu(i)$.

So for every student $i \in I$, μ does not violate i's priority at $\nu(i)$ and ν does not violate i's priority at $\mu(i)$. Hence, by Lemma 6.1, $\mu \vee \nu$ is a matching. Since μ is priority-efficient, it is Pareto efficient. Therefore, because $\mu \vee \nu$ dominates μ, we must have $\mu \vee \nu = \mu$ and so $\mu R_i \nu$ for every student i, contradicting the nonemptiness of B. ■

We remind the reader of our convention to settle students and schools at the end of each round of the sEADA* algorithm, not at the beginning of each round as in Tang and Yu (2014). See Section 3.1.

Lemma 6.3 *Suppose that $\mu_1, ..., \mu_N$ is the output sequence of the sEADA* algorithm and that student i is settled in round n. Then student i is μ_n-unimprovable.*

Proof. Throughout the proof, by "round one," "round two," etc., we will always mean the corresponding round of the sEADA* algorithm.

Let i, n and $\mu_1, ..., \mu_N$ be as given in the statement of the lemma and let ν dominate μ_n. We must show that $\nu(i) = \mu_n(i)$.

Let J_n and T_n be the sets of students and schools, respectively, that are unsettled at the start of round n (so $J_1 = I$ and $T_1 = S$). Then $i \in J_n$.

Let j be any student in J_n and let s be any school outside T_n. Hence, there is $k < n$ such that s is settled in round k. Therefore, school s is underdemanded at the student-optimal stable matching for round k's submarket of unsettled students and schools. Hence, $\mu_k(j) P_j s$ since student j is included in that submarket (student $j \in J_n$ is not settled until round $n > k$ or later) and since $\mu_k(j)$ is j's school in that submarket's student-optimal stable matching. Hence, $\mu_n(j) P_j s$ since $j \in J_n$ implies that $\mu_n R_j \mu_k$ (by Lemma 2 of Tang and Yu, 2014). Consequently, the only schools that student j prefers to $\mu_n(j)$ are schools that are unsettled at the start of round n, i.e., schools in T_n.

Since ν dominates μ_n, we have that for any $j \in J_n$, either, $\nu(j) P_j \mu_n(j)$ in which case $\nu(j) \in T_n$ by the conclusion of the previous paragraph, or, $\nu(j) = \mu_n(j)$ in which case $\nu(j) \in T_n$ because, by definition, the round n matching μ_n assigns every student in J_n to a school in T_n. Hence, ν assigns every student in J_n to a school in T_n, which means that the restriction of ν to J_n is a matching for the submarket (J_n, T_n) consisting of students in J_n and schools in T_n. By definition, the restriction of μ_n to J_n is the student-optimal stable matching for the submarket (J_n, T_n). Since, by hypothesis, student i is settled in round n, i's assigned school $\mu_n(i)$ is underdemanded at this student-optimal stable matching in this submarket. Also, since ν dominates μ_n, the restriction of ν to J_n dominates the restriction of μ_n to J_n. Hence, we may apply Lemma 1 of Tang and Yu (2014) to the submarket (J_n, T_n) to conclude that $\nu(i) = \mu_n(i)$ as desired. ■

Proof of Theorem 3.4. By Tang and Yu (2014) Proposition 1, the sEADA* algorithm is well defined and ends in finitely many rounds, N say. Therefore it produces a finite sequence of matchings, $\mu_1, \mu_2, ..., \mu_N$, where μ_n is the matching produced in the n-th round. We must show that this sequence is feasible. Henceforth, by "first round," "second round," etc., we mean the corresponding round of the sEADA* algorithm.

Observe first that the matching μ_1 that is produced in the first round is the student-optimal stable matching. In particular, μ_1 is stable. Second, μ_N is Pareto efficient by Tang and Yu (2014) Theorem 1. Third, for each $n > 1$, $\mu_{n-1}(j) = \mu_n(j)$ for any student j that is settled before round n, and $\mu_n R_j \mu_{n-1}$ for any student j who is unsettled at the start of round n by Tang and Yu (2014) Lemma 2. Hence, μ_n dominates μ_{n-1}. So it remains only to show that for each $n > 1$, μ_n does not violate the priority of any μ_{n-1}-improvable student.

Suppose that student i is μ_{n-1}-improvable. Then for every $k \leq n-1$, student i is μ_k-improvable because μ_{n-1} dominates μ_k. Hence, by Lemma 6.3, student i is not settled before round n. Therefore student i is included in the submarket consisting of students and schools that are unsettled at the start of round n. Since, by definition, the restriction of μ_n to students in that submarket is the student-optimal stable matching for that submarket, student i's priority is not violated by μ_n at any school that is unsettled at the start of round n. For any other school, i.e., any school s that is settled in some round $k < n$, school s is underdemanded at μ_k restricted to round k's submarket of unsettled students and schools. Hence, $\mu_k(i) P_i s$ since student i is included in that submarket, and so $\mu_n(i) P_i s$ since μ_n dominates μ_k. Therefore, μ_n does not violate i's priority at any school, whether that school is unsettled at the start of round n or not. Since i was arbitrary, we may conclude that μ_n does not violate the priority of any μ_{n-1}-improvable student. Q.E.D.

Lemma 6.4 *Let Π^* be the set of priority-neutral matchings and let $\mu_1, ..., \mu_N$ be any feasible sequence. Then,*

$$\Pi^* \subseteq \{\mu \in M : \textit{ if } \mu \textit{ violates } i\textit{'s priority, then } \mu R_i \mu_N\}. \tag{6.1}$$

Proof. Let $\bar{\mu}$ be a priority-neutral matching. We must show that $\bar{\mu} R_i \mu_N$ for every student

i whose priority is violated by $\bar{\mu}$.

We begin by showing that for every $n \in \{1, ..., N\}$,

(1) if μ_n violates any student i's priority, then $\mu_n R_i \bar{\mu}$, and

(2) if $\bar{\mu}$ violates any student i's priority, then $\bar{\mu} R_i \mu_n$.

We proceed by induction starting with $n = 1$.

So suppose that $n = 1$. Then (1) holds trivially since μ_1 is stable. To see (2), suppose that $\bar{\mu}$ violates student i's priority. Then $\bar{\mu} R_i \mu_1$ since otherwise, μ_1 would make i better off without violating the priority of any other student (μ_1 is stable), which would contradict the priority-neutrality of $\bar{\mu}$. Hence, (2) holds.

Assume as an induction hypothesis that (1) and (2) hold for n. We must show that (1) and (2) hold for $n + 1$.

To see that (1) holds, suppose that μ_{n+1} violates i's priority. We must show that $\mu_{n+1} R_i \bar{\mu}$. By the induction hypothesis, μ_n and $\bar{\mu}$ satisfy the hypotheses of Lemma 6.1. Consequently, $\bar{\mu} \vee \mu_n$ is a matching that dominates μ_n. Also, by feasibility μ_{n+1} dominates μ_n. Since μ_{n+1} violates i's priority, feasibility implies that student i is μ_n-unimprovable. Hence, $[\bar{\mu} \vee \mu_n](i) = \mu_n(i)$ and $\mu_{n+1}(i) = \mu_n(i)$ (since $\bar{\mu} \vee \mu_n$ and μ_{n+1} dominate μ_n), from which we obtain $\mu_{n+1}(i) = \mu_n(i) R_i \bar{\mu}(i)$ and so (1) holds for $n + 1$. It remains to show that (2) also holds for $n + 1$.

Suppose, by way of contradiction, that (2) fails for $n+1$. That is, suppose that $\bar{\mu}$ violates i's priority and that $\mu_{n+1} P_i \bar{\mu}$. Since $\mu_{n+1}(i)$ makes i better off than $\bar{\mu}$, and since $\bar{\mu}$ is priority-neutral, there must be a student j whose priority is violated by μ_{n+1} such that $\bar{\mu} P_j \mu_{n+1}$. But then (1) would fail for $n+1$, which is a contradiction and completes the induction argument.

Since (2) holds for each $n = 1, ..., N$, we have $\bar{\mu} R_i \mu_N$ for every i such that $\bar{\mu}$ violates i's priority, as desired. ■

Lemma 6.5 *Let $\mu_1, ..., \mu_N$ be any feasible sequence. Then*

$$\{\mu \in M : \text{ if } \mu \text{ violates } i\text{'s priority, then } \mu R_i \mu_N\} \subseteq \{\mu \in M : \mu_N R_i \mu \text{ for every } i \in I\} \tag{6.2}$$

Proof. Let A^* denote the set on the left-hand side of (6.2). We first show that,

$$\mu \vee \mu_n \text{ is a matching for every } n \in \{1, ..., N\}, \text{ and for every } \mu \in A^*. \tag{6.3}$$

We will establish (6.3) by induction on n starting with $n = 1$.

Let μ be any matching in A^*. We must show that $\mu \vee \mu_1$ is a matching. It suffices to show that μ and μ_1 satisfy the hypotheses of Lemma 6.1. For any student i, μ cannot violate i's priority at $\mu_1(i)$. Otherwise, $\mu_1 P_i \mu R_i \mu_N$, where the weak preference follows because $\mu \in A^*$.

But then $\mu_1 P_i \mu_N$ contradicting the fact that, by feasibility, μ_N dominates μ_1. Hence, μ does not violate the priority of any student i at $\mu_1(i)$. Therefore, since μ_1 is stable, μ and μ_1 satisfy the hypotheses of Lemma 6.1 and so $\mu \vee \mu_1$ is a matching.

Next, assume as an induction hypothesis that (6.3) holds for n. We must show that (6.3) holds for $n + 1$.

So suppose that $\mu \in A^*$. We must show that $\mu \vee \mu_{n+1}$ is a matching, and so it suffices to show that μ and μ_{n+1} satisfy the hypotheses of Lemma 6.1.

For any student i, μ cannot violate i's priority at $\mu_{n+1}(i)$. Otherwise, $\mu_{n+1} P_i \mu R_i \mu_N$, where the weak preference follows because $\mu \in A^*$. But then $\mu_{n+1} P_i \mu_N$ contradicting the fact that, by feasibility, μ_N dominates μ_{n+1}. Hence, μ does not violate the priority of any student i at $\mu_{n+1}(i)$.

Suppose next that μ_{n+1} violates i's priority. Then, by feasibility, i is μ_n-unimprovable. Since $\mu \vee \mu_n$ is a matching (induction hypothesis), $\mu \vee \mu_n$ dominates μ_n. Also, by feasibility, μ_{n+1} dominates μ_n. Therefore, since i is μ_n-unimprovable, $[\mu \vee \mu_n](i) = \mu_n(i) = \mu_{n+1}(i)$ and so $\mu_{n+1}(i) = [\mu \vee \mu_n](i) R_i \mu(i)$. So μ_{n+1} does not violate i's priority at $\mu(i)$. Hence, μ and μ_{n+1} satisfy the hypotheses of Lemma 6.1 and so we may conclude that $\mu \vee \mu_{n+1}$ is a matching. This completes the induction and establishes (6.3).

Setting $n = N$ in (6.3), we may conclude that $\mu \vee \mu_N$ is a matching for every $\mu \in A^*$. Consequently, because, by definition, $\mu \vee \mu_N$ dominates μ_N, and because, by feasibility, μ_N is Pareto efficient, we must have $\mu \vee \mu_N = \mu_N$ for every $\mu \in A^*$. But this means that $\mu_N R_i \mu$, for every $i \in I$, and for every $\mu \in A^*$ as desired. ■

We can now prove Theorems 3.1 and 3.5 by combining them into a single theorem.

Theorem 6.6 *There is a unique priority-efficient matching and it dominates every priority-neutral matching and every stable matching. If $\mu_1, ..., \mu_N$ is any feasible sequence then μ_N is the unique priority-efficient matching.*

Proof. Let $\mu_1, ..., \mu_N$ be any feasible sequence. Such a sequence exists by Theorem 3.4. We begin by showing that μ_N is priority-efficient. Since feasibility implies that μ_N is Pareto efficient, we need only show that μ_N is priority-neutral. So suppose that student i's priority is violated by μ_N and that there is a matching ν such that $\nu P_i \mu_N$. We must show that there is a student j whose priority is violated by ν such that $\mu_N P_j \nu$. Observe that ν is not in the set on the right-hand side of (6.2) because $\nu P_i \mu_N$. Consequently, by Lemma 6.5, ν is not in the set on left-hand side of (6.2). But this means that there is a student j whose priority is violated by ν such that $\mu_N P_j \nu$, as desired. Hence, μ_N is priority-efficient. In particular, a priority-efficient matching always exists.

Next, observe that Lemmas 6.4 and 6.5 together imply that μ_N dominates every priority-neutral matching. Since every stable matching is priority-neutral, μ_N dominates every stable matching. It remains only to show that no matching other than μ_N is priority-efficient.

Let ν be any priority-efficient matching. Then, in particular, ν is priority-neutral and so μ_N dominates ν. Since ν is Pareto efficient (because ν is priority-efficient), we must then have $\nu = \mu_N$, as desired. ■

Proof of Theorem 3.3. The proof of Theorem 3.3 follows from Lemma 6.2 and the uniqueness part of Theorem 3.1. Q.E.D.

Lemma 6.7 *Let μ^* be the unique priority-efficient matching and let μ be any matching such that $\mu R_j \mu^*$ for every student j whose priority is violated by μ. Then μ^* dominates μ.*

Proof. Let $\mu_1, ..., \mu_N$ be any feasible sequence. Such a sequence exists by Theorem 3.4, and $\mu_N = \mu^*$ by Theorem 3.5. The result now follows from Lemma 6.5. ■

6.2 Part B.

The notation below is as in Section 4.

For any state $\sigma = (P, \Psi, q) \in \Sigma$, say that a matching μ is σ-*stable* iff μ is stable when the state is (P, Ψ, q), and say that μ σ-*dominates* ν iff μ dominates ν when the profile of student preferences is P. Recall that $f(\sigma)$ is the unique priority-efficient matching in state σ.

Lemma 6.8 *Let $\sigma = (P, \Psi, q) \in \Sigma$ be any state and let $\mu_1, ..., \mu_N$ be the sEADA* output sequence for state σ. For any student i, let P_i' be any preference order for i that is obtained without lowering any school $\mu_n(i)$ in the preference order P_i (i.e., $\mu_n(i) P_i s \Rightarrow \mu_n(i) P_i' s$, for every n and for every school s). Then the priority-efficient matching is the same in states σ and $\sigma' = (P_i', P_{-i}, \Psi, q)$.*

Proof. Because $\mu_n(i) P_i s \Rightarrow \mu_n(i) P_i' s$, for every n and for every school s, it is straightforward to check that for any n, (a) if μ_n does not violate a student's priority in state σ then μ_n does not violate that students priority in state σ', (b) if a matching ν dominates μ_n in state σ', then ν dominates μ_n in state σ, (c) if μ_n dominates a matching ν in state σ, then μ_n dominates ν in state σ', and (d) if a student is μ_n-improvable in state σ', then that student is μ_n-improvable in state σ. Using (a)-(d) it is straightforward to establish that because $\mu_1, ..., \mu_N$ is a feasible sequence in state σ (by Theorem 3.4), it also a feasible sequence in state σ'. Hence, by Theorem 3.5, μ_N is the priority-efficient matching in states σ and σ'. ■

Lemma 6.9 *Consider any student $i \in I$ and any two states $\sigma = (P, \Psi, q)$ and $\sigma' = (P'_i, P_{-i}, \Psi, q)$ that differ only in students i's preference order. Suppose that $f(\sigma')P_i f(\sigma)$ and that $\mu_1, ..., \mu_N$ is any feasible sequence of matchings in state σ. Then there exists $n \in \{1, ..., N\}$ such that μ_n violates i's priority in state σ' but μ_n does not violate i's priority in state σ.*

Proof. By Theorem 3.5, $f(\sigma) = \mu_N$. Hence, $f(\sigma')P_i f(\sigma) = \mu_N$. Therefore, since $\mu_N R_i \mu_n$ for every n (by the sequential dominance property of the state-σ feasible sequence $\mu_1, ..., \mu_N$), we may conclude by transitivity that $f(\sigma')P_i \mu_n$ for every n.

Since only student i's preferences change between states σ and σ', it will be convenient to define $P'_j = P_j$ for every student $j \neq i$, which allows us to denote any student j's preferences in state σ' by P'_j.

We claim that μ_N must violate the priority of some student j such that $f(\sigma')P'_j \mu_N$. For suppose to the contrary that $\mu_N R'_j f(\sigma')$ for every student j whose priority is violated by μ_N. Then, by Lemma 6.7, $f(\sigma')$ σ'-dominates μ_N. Since $P'_j = P_j$ for every $j \neq i$, we may conclude that $f(\sigma')R_j \mu_N$ for every $j \neq i$. Since $f(\sigma')P_i \mu_N$ (because $f(\sigma')P_i \mu_n$ for every n), $f(\sigma')$ Pareto dominates μ_N in state σ. But this contradicts the feasibility of $\mu_1, ..., \mu_N$ in state σ (which requires that μ_N is Pareto efficient in state σ) and establishes the claim.

By the claim just established, we may let $\bar{n}$ be the smallest $n \in \{1, ..., N\}$ such that, in state σ', μ_n violates the priority of some student who strictly prefers $f(\sigma')$ to μ_n. Hence, we may choose $j \in I$ so that $\mu_{\bar{n}}$ violates j's priority in state σ' and so that $f(\sigma')P'_j \mu_{\bar{n}}$.

We next claim that, if $\bar{n} > 1$, then students i and j (it is possible that $j = i$) are $\mu_{\bar{n}-1}$-improvable in state σ. To see why this claim is true, note first that, by the definition of $\bar{n}$, if $\mu_{\bar{n}-1}$ violates any student k's priority in state σ, then $\mu_{\bar{n}-1} R'_k f(\sigma')$. Therefore, by Lemma 6.7, $f(\sigma')$ σ'-dominates $\mu_{\bar{n}-1}$. Hence, since each student $k \neq i$ has the same preferences in states σ and σ', $f(\sigma')R_k \mu_{\bar{n}-1}$ holds for every $k \neq i$. Furthermore, we have already established that $f(\sigma')P_i \mu_n$ for every n, and so, in particular, $f(\sigma')P_i \mu_{\bar{n}-1}$ from which we may conclude that student i is $\mu_{\bar{n}-1}$-improvable in state σ. So if $j = i$ the claim is established. It therefore remains only to show that, if $j \neq i$, then $f(\sigma')P_j \mu_{\bar{n}-1}$. So suppose that $j \neq i$. We know that $f(\sigma')P'_j \mu_{\bar{n}}$ by our choice of $\bar{n}$ and j. Since $j \neq i$ we have $P'_j = P_j$, and so $f(\sigma')P_j \mu_{\bar{n}} R_j \mu_{\bar{n}-1}$, where the weak preference follows from the sequential dominance property of the state-σ feasible sequence $\mu_1, ..., \mu_N$. This establishes the claim.

We next claim that $j = i$. To establish this claim, suppose first that $\bar{n} = 1$. By the feasibility of $\mu_1, ..., \mu_N$ in state σ, μ_1 is stable in state σ and so, in particular, μ_1 does not violate the priority of any student $k \neq i$. Since $P'_k = P_k$ for every $k \neq i$, μ_1 does not violate the priority of any student $k \neq i$ in state σ'. Since $\mu_1 = \mu_{\bar{n}}$ violates the priority of student j in state σ', we may conclude that $j = i$. Suppose next that $\bar{n} > 1$, and suppose, contrary to the claim, that $j \neq i$. Then, because $\mu_{\bar{n}}$ violates j's priority in state σ' (by our choice of $\bar{n}$ and j), and because j's preferences do not change between states σ and σ' (because $j \neq i$), $\mu_{\bar{n}}$ violates j's priority also in state σ. Then, by the feasibility of $\mu_1, ..., \mu_N$ in state σ, student j is $\mu_{\bar{n}-1}$-unimprovable in state σ, which contradicts the conclusion in the previous paragraph and establishes the claim that $j = i$.

So we have so far established that $\mu_{\bar{n}}$ violates i's priority in state σ' (since $j = i$) and that if $\bar{n} > 1$ then student i is $\mu_{\bar{n}-1}$-unimprovable in state σ. Hence, whether $\bar{n} = 1$ or whether $\bar{n} > 1$, $\mu_{\bar{n}}$ does not violate i's priority in state σ because, by the feasibility of $\mu_1, ..., \mu_N$ in state σ, μ_1 is stable in state σ, and, if $\bar{n} > 1$ $\mu_{\bar{n}}$ does not violate the priority of any $\mu_{\bar{n}-1}$-improvable student in state σ and so in particular $\mu_{\bar{n}}$ does not violate i's priority in state σ.

Therefore, the matching $\mu_{\bar{n}}$ violates i's priority in state σ', but does not violate i's priority in state σ. ■

Proof of Theorem 4.1. Let P_i and P_i' be any two preferences that student i is considering submitting. For each student $j \neq i$, let P_j be any preference submitted by j, and for each school $s \in S$, let P_s be any priority list and let q_s be any nonnegative quota satisfying only the restriction that $\sum_{s \in S} q_s \geq \#I$. Let $\sigma = (P, \Psi, q)$ and let $\sigma' = (P_i', P_{-i}, \Psi, q)$. So σ and σ' are arbitrary states that differ only in student i's preference order.

Given the preferences P_j submitted by other students and given the school priorities and quotas (Ψ_s, q_s), $f(\sigma)$ is the priority-efficient matching that is chosen when student i submits the preference P_i, and in that case i's assigned school is $f_i(\sigma)$, and $f(\sigma')$ is the priority-efficient matching that is chosen when student i submits the preference P_i', and in that case i's assigned school is $f_i(\sigma')$.

To prove statement (1), let P_i' be obtained from P_i by lowering only the top-ranked school s in P_i. If $f_i(\sigma) = t \neq s$, then none of i's assigned schools in the sEADA* output sequence in state σ are s because the schools for i in this sequence successively (weakly) increase in his preference P_i until reaching t. Hence, P_i and P_i' satisfy the hypotheses of Lemma 6.8 and so $f_i(\sigma')$ is also t. This proves (1).

To prove statement (2), notice that we may suppose that $f_i(\sigma') = s$ and that P_i is obtained from P_i' by switching the positions of schools s and t, where $tP_i's$. We must show that $f(\sigma) = f(\sigma')$. So let μ be a matching such that for some student j, $\mu P_j f(\sigma')$ and $f(\sigma')$ violates j's priority in state σ. We must show that there is a student k whose priority is violated by μ in state σ such that $f(\sigma')P_k\mu$.

Suppose first that $j = i$. Then $\mu(i)P_i f_i(\sigma') = s$ and $f(\sigma')$ violates i's priority in state σ, and so $\mu(i)P_i' f_i(\sigma')$ and $f(\sigma')$ violates j's priority in state σ' (because $rP_is \Rightarrow rP_i's$ for every school r). Therefore, because $f(\sigma')$ is priority-efficient in state σ', there is a student

$k \neq i$ whose priority is violated by μ in state σ' and who prefers $f(\sigma')$ to μ in state σ'. But then μ violates k's priority in state σ, and k prefers $f(\sigma')$ to μ in state σ because only i's preferences change between states σ and σ'. So we have found the requisite student k when $j = i$.

Suppose next that $j \neq i$. Then, because only i's preferences change between states σ and σ', student j prefers μ to $f(\sigma')$ in state σ', and $f(\sigma')$ violates j's priority in state σ'. Therefore, because $f(\sigma')$ is priority-efficient in state σ', there is a student k whose priority is violated by μ in state σ' and who prefers $f(\sigma')$ to μ in state σ'. If $k \neq i$, then because only i's preferences change between states σ and σ', k's priority is violated by μ in state σ' and k prefers $f(\sigma')$ to μ in state σ'. So if $k \neq i$ we have found the requisite student k. If $k = i$, then $f_i(\sigma') = sP_i'\mu(i)$ because $k = i$ prefers $f(\sigma')$ to μ in state σ'. Therefore, $rP_i'\mu(i) \Rightarrow rP_i\mu(i)$ for every school r. Consequently, i's priority is violated by μ in state σ because i's priority is violated by μ in state σ', and $sP_i\mu(i)$ because $sP_i'\mu(i)$. Hence student $k = i$ is the requisite student, which completes the proof of statement (2).

For statement (3), notice that if $f_i(\sigma)$ is not in K, then each school in the state-σ sEADA* output sequence of schools for i, $\mu_1(i), ..., \mu_N(i)$ say, is ranked below every school in K according to P_i. Therefore P_i' is obtained from P_i without lowering the schools $\mu_n(i)$ in P_i. By Lemma 6.8 we may conclude that $f_i(\sigma')$ is also outside the set K, which proves (3). It remains only to prove statement (4).

Since student i is arbitrary, and $\sigma = (P, \Psi, q)$ and $\sigma' = (P_i', P_{-i}, \Psi, q)$ are arbitrary states that differ only in student i's preference, we can establish (4) by supposing that untruthfully reporting P_i' is better for i in state σ, i.e., by supposing that,

$$f_i(P_i', P_{-i}, \Psi, q)P_if_i(P_i, P_{-i}, \Psi, q), \tag{6.4}$$

and showing that this untruthful report can be worse for some preferences of the other students, i.e., and showing that there exist preferences P_j'' for every other student $j \neq i$, such that,

$$f_i(P_i, P_{-i}'', q)P_if_i(P_i', P_{-i}'', q). \tag{6.5}$$

By Theorem 3.1, for any given state, any stable matching is dominated by the unique priority-efficient matching, and any matching that is stable and Pareto efficient is the unique priority-efficient matching. So it suffices to find matchings μ and ν and preferences P_j'' for all $j \neq i$ such that μ is stable in state (P_i, P_{-i}'', Ψ, q), ν is Pareto efficient and stable in state $(P_i', P_{-i}'', \Psi, q)$, and,

$$\mu P_i \nu, \tag{6.6}$$

because, as just pointed out, this will imply that $f_i(P_i, P_{-i}'', \Psi, q)R_i\mu P_i\nu = f_i(P_i', P_{-i}'', \Psi, q)$, and so (6.4) will hold.

By Theorem 3.4, feasible sequences of matchings exist in any state, and so we may let $\mu_1, ..., \mu_N$ be a feasible sequence of matchings in state $\sigma = (P, \Psi, q)$. Since $\sigma' = (P_i', P_{-i}, \Psi, q)$ and since (6.4) says that $f(\sigma')P_i f(\sigma)$, by Lemma 6.9 we may chooses $n \in \{1, ..., N\}$ such that μ_n violates i's priority in state σ', but μ_n does not violate i's priority in state σ.

Let s be the P_i'-most-preferred school among those at which i's priority is violated by μ_n in state σ'. Hence, in particular, $sP_i'\mu_n(i)$. Moreover, because μ_n does not violate i's priority in state σ, and because school priorities are the same in states σ and σ', we must have $\mu_n(i)P_i s$.

There are two cases to consider. Either school s is at its quota under μ_n or it is not. We consider each case in turn starting with the first.

Suppose that school s is at its quota under μ_n. Since i's priority is violated at school s by μ_n in state σ', student i has higher priority at s than the student, j say, with lowest priority assigned to s under μ_n. Define P_j'' to be any preference order for student j that ranks school $\mu_n(j) = s$ first and school $\mu_n(i)$ second. For any student k distinct from i and j, define P_k'' to be any preference order for k that ranks school $\mu_n(k)$ first.

Let $\sigma'' = (P_i', P_{-i}'', \Psi, q)$. We claim that the matching, ν say, that assigns i to $s = \mu_n(j)$ and j to $\mu_n(i)$, and that assigns every other student k to $\mu_n(k)$ is stable and Pareto efficient in state σ''. Notice that ν is obtained from μ_n by having only students i and j switch schools and so school s remains at its quota under ν.

To see that ν is stable in state σ'', note first that for every student k distinct from i and j, k is assigned to his P_k''-most-preferred school and so ν does not violate k's priority. Consider next student j. The only school that student j prefers to $\nu(j) = \mu_n(i)$ according to P_j'' is school $s = \mu_n(j)$. By our choice of student j, student j has lower priority at s than student i ($\nu(i) = s$) and lower priority at s than every other student assigned to s by ν since these other students are also assigned to s by μ_n. Hence, j's priority is not violated by ν in state σ''.

Lastly, we must show that ν does not violate i's priority in state σ''. If t is any school that student i prefers to $\nu(i)$ in state σ'', then $tP_i'\nu(i) = s$. Hence, $tP_i'sP_i'\mu_n(i)$. Therefore, by our choice of s, i's priority is not violated at school t by μ_n in state σ'. Since t is distinct from $s = \mu_n(j)$ and $\mu_n(i)$, the students assigned to t under μ_n are precisely the students who are assigned to t under ν. Hence, i's priority is not violated at school t by ν in state σ'. Since i's preferences, P_i', as well all school priorities and quotas are unchanged between the states σ' and σ'', i's priority is not violated at school t by ν in state σ''. Since school t was arbitrary, ν does not violate i's priority in state σ'' and so ν is stable in state σ''.

To see that ν is Pareto efficient in state $\sigma'' = (P_i', P_{-i}'', \Psi, q)$, let $K = I\backslash\{i, j\}$. Since each student in K is assigned to his P_k''-most-preferred school under ν, no student in K can change schools in any Pareto improvement (over ν in state σ''). Any school that student i P_i'-prefers to $\nu(i)$ is at its quota under ν (because ν is σ''-stable) and is distinct from $\nu(j)$ because $\nu(i) = sP_i'\mu_n(i) = \nu(j)$. Hence, any school that student i P_i'-prefers to $\nu(i)$ is filled with students in K under ν and so must remain filled with those same students in any Pareto improvement. Therefore student i cannot change schools in any Pareto improvement. So only student j can possibly change schools in any Pareto improvement. However, since the only school that student j P_j''-prefers to $\nu(j) = \mu_n(i)$ is school $s = \nu(i)$, and since we have already observed that school s is at its quota under ν, student j cannot change schools in any Pareto improvement. Therefore, in state σ'', no student can change schools in any Pareto improvement over ν and so ν is Pareto efficient in state σ''.

Next, we claim that μ_n is stable in state (P_i, P_{-i}'', Ψ, q). Indeed, because i's priority is not violated by μ_n in state $\sigma = (P, \Psi, q)$, student i's priority is not violated by μ_n in state (P_i, P_{-i}'', Ψ, q). Also, every student $k \neq i$ is assigned his P_k''-most-preferred school by μ_n and so no such student's priority is violated by μ_n in state (P_i, P_{-i}'', Ψ, q), establishing the claim. Since $\mu_n(i)P_i s = \nu(i)$, defining $\mu = \mu_n$, establishes (6.6) and establishes (4) when s is at its quota under μ_n.

It remains to consider the (simpler) case in which s is not at its quota under μ_n. In this case, starting from the matching μ_n, transfer only student i from $\mu_n(i)$ to s and let ν denote the matching that results. For every $k \neq i$, let P_k'' be any matching such that $\mu_n(k)$ is student k's most preferred school. Then, the same arguments used above, but now ignoring those parts that refer to student j, establish that ν is stable and Pareto efficient in state σ'' and that μ_n is stable in state σ. Since $\mu_n(i)P_i s = \nu(i)$, defining $\mu = \mu_n$ establishes (6.6) and completes the proof of (4). Q.E.D.

Lemma 6.10 *Let $P = (P_w)_{w \in I \cup S}$ be any profile of student preferences and school priorities, let $\mathcal{S}$ be a similarity partition for P and suppose that P is student-oriented with respect to $\mathcal{S}$. Let i be any student, let P_i' be any preference over schools for student i, and let σ' be the state $\sigma' = (P_i', P_{-i}, q)$. Let μ' be any matching that, in state σ', does not violate the priority of any μ'-improvable student, and let ν be any matching that σ'-dominates μ'. Then, either $\mu' P_i \nu$ or the schools $\nu(i)$ and $\mu'(i)$ are in the same similarity set in $\mathcal{S}$.*

Proof. Suppose, by way of contradiction, that $\nu(i)$ and $\mu'(i)$ are in distinct similarity sets (hence $\nu(i) \neq \mu'(i)$) and that $\nu P_i \mu'$ (they cannot be indifferent because preferences are strict). We can draw two conclusions from the fact that ν dominates μ' in the state (P_i', P_{-i}, q). First, $\nu P_i' \mu'$, and second, ν dominates μ' in the state (P, q) because $\nu P_i \mu'$. We use both of these conclusions next.

We claim that there is a finite sequence of students $i_1, i_2, ..., i_N$, such that $i_N = i$, the schools $\mu'(i_1)$ and $\mu'(i_N)$ are in distinct similarity sets, and for each $n = 1, ..., N$,

$$\mu'(i_{n+1}) P_{i_n} \mu'(i_n) \;\; (i_{N+1} := i_1). \tag{6.7}$$

To see why this claim is true, consider switching from the matching μ' to the matching ν. Since $\mu'(i) \neq \nu(i)$ and preferences are strict, student i changes schools, from $\mu'(i)$ to $\nu(i)$, and is made strictly better off under both P_i and P_i'. In particular, because $\nu P_i' \mu'$, student i is μ'-improvable in state $\sigma' = (P_i', P_{-i}, q)$ and so μ' does not violate i's priority in state σ'. Hence, school $\nu(i)$ must be at its quota under μ'. Therefore, when the matching switches from μ' to ν, some student i_1 assigned to $\nu(i)$ under μ' (i.e., such that $\mu'(i_1) = \nu(i) P_i \mu'(i)$) must be displaced by i and so, because ν σ'-dominates μ', i_1 moves to a P_{i_1}-preferred school $\nu(i_1)$. Evidently then, i_1 is μ'-improvable in state σ'. Hence, by hypothesis, μ' does not violate i_1's priority in state σ' and so $\nu(i_1)$ must be at its quota under μ'. Student i_1 must therefore displace a student i_2 from school $\nu(i_1) = \mu'(i_2)$, who similarly displaces a student i_3, etc. So if we imagine that all seats are numbered, each student $i_n \neq i$ who is made P_{i_n}-better off when the matching switches from μ' to ν can point to the unique student i_{n+1} whose seat he takes (and so $\mu'(i_{n+1}) = \nu(i_n) P_{i_n} \mu'(i_n)$) and who moves to a new school. Since there are finitely many students, and no student is pointed to by more than one student (because each seat can be occupied by only one student), this improvement chain, of length N say, consists of distinct students and must eventually lead back to student i. So we have a sequence of distinct students $i_1, ..., i_N$, such that $i_N = i$, $\mu'(i_1) = \nu(i)$ and $\mu'(i_N) = \mu'(i)$ are in distinct similarity sets, and (6.7) holds for each n. This establishes the claim.

Without loss of generality, we will assume that for every $n > 1$, $\mu'(i_1)$ and $\mu'(i_n)$ are in distinct similarity sets. Otherwise, there is a largest $n > 1$ such that $\mu'(i_1)$ is in the same similarity set as $\mu'(i_n)$, in which case the shorter sequence $i_n, ..., i_N$ will do. (Note that $n < N$ here since $\mu'(i_1)$ and $\mu'(i_N)$ are in distinct similarity sets.) To see that the shorter sequence will do, one needs to verify only that $\mu'(i_n) P_{i_N} \mu'(i_N)$. Since $\mu'(i_1) P_{i_N} \mu'(i_N)$, and since $\mu'(i_1)$ and $\mu'(i_n)$ are in the same similarity set, S_k say, and $\mu'(i_N)$ is in another, $S_{k'}$ say, the desired conclusion, i.e., that $\mu'(i_n) P_{i_N} \mu'(i_N)$, follows because S_k and $S_{k'}$ are similarity sets for P.

We claim next that $\mu'(i_n) P_{i_n} \mu'(i_1)$ for every $n > 1$. Otherwise, choose the smallest $n > 1$ such that $\mu'(i_1) R_{i_n} \mu'(i_n)$. Then, because $\mu'(i_n) \neq \mu'(i_1)$ (they are in distinct similarity sets), strict preferences imply that $\mu'(i_1) P_{i_n} \mu'(i_n)$. If $n = 2$ then $\mu'(i_n) P_{i_{n-1}} \mu'(i_1)$ follows from (6.7), and if $n > 2$ then $\mu'(i_n) P_{i_{n-1}} \mu'(i_1)$ follows by transitivity since $\mu'(i_n) P_{i_{n-1}} \mu'(i_{n-1})$ by (6.7) and $\mu'(i_{n-1}) P_{i_{n-1}} \mu'(i_1)$ by the choice of n. Consequently, $\mu'(i_n) P_{i_{n-1}} \mu'(i_1)$ and so, given their preferences in P, students i_n and i_{n-1} disagree about the ranking of two colleges, $\mu'(i_1)$ and

$\mu'(i_n)$, that are in distinct similarity sets in $\mathcal{S}$. Therefore, because P is student-oriented with respect to $\mathcal{S}$, student i_{n-1} is ranked higher by school $\mu'(i_n)$ than student i_n. But then, since $i_{n-1} \neq i_N = i$, μ' violates student i_{n-1}'s priority at school $\mu'(i_n)$ in state $\sigma' = (P_i', P_{-i}, q)$, contradicting our hypotheses since, like all students in the cycle, student i_{n-1} is μ'-improvable in state σ'. This establishes the claim.

By the claim just established, $\mu'(i_N) P_{i_N} \mu'(i_1)$. But this contradicts (6.7) when $n = N$ and completes the proof. ■

Corollary 6.11 *Let $P = (P_w)_{w \in I \cup S}$ be any profile of student preferences and school priorities, let $\mathcal{S}$ be a similarity partition for P and suppose that P is student-oriented with respect to $\mathcal{S}$. Let i be any student, let P_i' be any preference over schools for student i, and let σ' be the state $\sigma' = (P_i', P_{-i}, q)$. If $\mu_1, ..., \mu_N$ is feasible in state σ', then for every $m < n$, either $\mu_m(i)$ and $\mu_n(i)$ are in the same similarity set, or $\mu_m P_i \mu_n$.*

Proof. Fix any $m < n$. We claim that, in state σ', μ_m does not violate the priority of any μ_m-improvable student. If $m = 1$ this follows because μ_1 is σ'-stable by feasibility. If $m > 1$, then in state σ', μ_m does not violate the priority of any μ_{m-1}-improvable student by feasibility. Since in state σ', μ_m dominates μ_{m-1} (by feasibility), every μ_m-improvable student is μ_{m-1}-improvable, from which the claim follows. Since μ_n σ'-dominates μ_m by feasibility, the result now follows from Lemma 6.10. ■

Proof of Theorem 4.2. For any student $i \in I$ and for any preferences P_i, P_i', and P_i'' for i, say that P_i'' *stochastically P_i-dominates* P_i' iff $f_i(P_i'', \tilde{\sigma}_{-i}|_{\tilde{\sigma}_i = \sigma_i})$ stochastically P_i-dominates $f_i(P_i', \tilde{\sigma}_{-i}|_{\tilde{\sigma}_i = \sigma_i})$ (see Section 4). That is, P_i'' *stochastically P_i-dominates* P_i' if and only if reporting the preference P_i'' to the PE mechanism yields a lottery over student i's assigned school that stochastically P_i-dominates the lottery that would result if student i instead reported the preference P_i', when all other students always report their preferences truthfully.

So to complete the proof we must show that for every student $i \in I$ and for pair of preferences P_i and P_i' for student i, if P_i occurs with positive probability under Q, then

$$P_i \text{ stochastically } P_i\text{-dominates } P_i'. \tag{6.8}$$

For any student $i \in I$, if P_i is student i's true preference order, then for any two schools s and t, and for any preference order P_i' for i, say that P_i' *truthfully ranks* s and t iff $[sP_i't \Leftrightarrow sP_it]$ holds, and say that P_i' *falsely ranks* s and t otherwise.

We claim that the PE mechanism satisfies two conditions introduced by Elhers' (2008), namely *anonymity* (permuting the names of schools permutes the outcome accordingly) and *positive association* (changing one's submitted preference list by interchanging in that list the school to you which you would otherwise have been matched with a school that is ranked higher in that list, does not change the school to which you are matched). That the PE mechanism satisfies anonymity is obvious. That the PE mechanism satisfies positive association is also straightforward to verify.[22] Consequently, because our Q-symmetry assumption implies that student i has symmetric information (in the sense of Ehlers 2008) about any pair of schools in the same similarity set in $\mathcal{S}$, we may apply Elhers (2008, Theorem A.1), to conclude that for any student i and for any true preference P_i for i that occurs with positive probability under Q, any preference order P_i' that falsely ranks any two schools in the same similarity set is stochastically P_i-dominated by the preference order that is derived from P_i' by switching the positions of (only) those two schools in the preference order P_i'. Consequently, since the stochastic P_i-domination relation is transitive, starting from P_i' and successively switching the positions of pairs of schools in the same similarity set that are falsely ranked, we will, after finitely many position switches, arrive at a preference order that truthfully ranks every pair of schools that are in the same similarity set and that stochastically P_i-dominates P_i'. Therefore, it suffices to show that P_i stochastically P_i-dominates every $\bar{P}_i$ that truthfully ranks every pair of schools that are in the same similarity set in $\mathcal{S}$.

Fix any student $i \in I$ and fix any state $\sigma = (P, q)$ that is given positive probability by Q. So in this state, student i's true preference order is P_i.

By hypothesis, $\mathcal{S}$ is a similarity set for P and P is student-oriented with respect to $\mathcal{S}$. Let $\bar{P}_i$ be any preference order for student i that truthfully ranks any two schools that are in the same similarity set in $\mathcal{S}$. Note that $\bar{P}_i$ need not occur with positive probability under Q. The proof will be complete if we can establish that i's assigned school when he submits P_i cannot be worse for him (according to his true preference order P_i) than his assigned school if he submits $\bar{P}_i$. So it suffices to show that,

$$f_i(P_i, \sigma_{-i}) R_i f_i(\bar{P}_i, \sigma_{-i}). \tag{6.9}$$

Let $\bar{\sigma} = (\bar{P}_i, P_{-i}, q)$, and let $\bar{\mu}_1, ..., \bar{\mu}_N$, be the output-sequence of the sEADA* algorithm when computed using the state $\bar{\sigma}$, i.e., in state $\bar{\sigma}$. Hence, $f(\bar{P}_i, \sigma_{-i}) = \bar{\mu}_N$ by Theorems 3.4 and 3.5 (notice that, $\bar{\sigma}_{-i} = \sigma_{-i}$ since $\sigma = (P, q)$).

Let $\hat{\mu}$ be the student-optimal stable matching for the state $\sigma = (P, q)$. By the definition of the sEADA* algorithm, $\bar{\mu}_1$ is the student-optimal stable matching for the state $\bar{\sigma} = (\bar{P}_i, P_{-i}, q)$. Consequently, because the mechanism that selects the student-optimal stable

[22] Indeed, the priority-efficient matching, μ say, before switching the two schools in student i's preference order remains priority-efficient after the the switch because, after the switch, no additional students' priorities are violated by μ and any matching ν that makes any of those students better off without violating the priority of any student that ν makes worse off, would have had exactly these properties before the switch, and so there can be no such ν because μ is priority-efficient before the switch.

matching is strategy-proof for students (Dubins and Freedman 1981), we may conclude that $\hat{\mu}R_i\bar{\mu}_1$. By Theorem 3.1, for any state, and in particular for the state $\sigma = (P, q)$, the priority-efficient matching σ-dominates the student-optimal stable matching, and so $f_i(P_i, \sigma_{-i})R_i\hat{\mu}$. Hence, by transitivity,

$$f_i(P_i, \sigma_{-i})R_i\bar{\mu}_1. \tag{6.10}$$

We now break the proof of (6.9) into two cases.

Case I. Not all of the schools $f_i(P_i, \sigma_{-i})$, $\bar{\mu}_1(i)$, $\bar{\mu}_2(i), ..., \bar{\mu}_N(i) = f_i(\bar{P}_i, \sigma_{-i})$ are in the same similarity set in $\mathcal{S}$.

Case II. All of the schools $f_i(P_i, \sigma_{-i})$, $\bar{\mu}_1(i)$, $\bar{\mu}_2(i), ...,$ $\bar{\mu}_N(i) = f_i(\bar{P}_i, \sigma_{-i})$ are in the same similarity set in $\mathcal{S}$.

Let us begin with Case I.

Proof of (6.9) in Case I. In this case, not all of the schools $f_i(P_i, \sigma_{-i})$, $\bar{\mu}_1(i)$, $\bar{\mu}_2(i), ..., \bar{\mu}_N(i)$ are in the same similarity set in $\mathcal{S}$.

Suppose first that not all of the schools $\bar{\mu}_1(i)$, $\bar{\mu}_2(i), ..., \bar{\mu}_N(i)$ are in the same similarity set in $\mathcal{S}$. We claim that $\bar{\mu}_1 P_i \bar{\mu}_N$. To see this, suppose to the contrary that $\bar{\mu}_N R_i \bar{\mu}_1$. Then, by Corollary 6.11, $\bar{\mu}_1(i)$ and $\bar{\mu}_N(i)$ are in the same similarity set in $\mathcal{S}$. Since not all of the schools $\bar{\mu}_1(i)$, $\bar{\mu}_2(i), ..., \bar{\mu}_N(i)$ are in the same similarity set, there must be $\bar{\mu}_n$ such that $\bar{\mu}_n(i)$ is not in the same similarity set as $\bar{\mu}_1(i)$ and $\bar{\mu}_N(i)$. Then, applying Corollary 6.11 first to $\bar{\mu}_1(i)$ and $\bar{\mu}_n(i)$ gives $\bar{\mu}_1 P_i \bar{\mu}_n$, and second to $\bar{\mu}_n(i)$ and $\bar{\mu}_N(i)$ gives $\bar{\mu}_n P_i \bar{\mu}_n$. But then transitivity gives $\bar{\mu}_1 P_i \bar{\mu}_N$, which is a contradiction and establishes the claim. Hence, $\bar{\mu}_1 P_i \bar{\mu}_N = f_i(\bar{P}_i, \sigma_{-i})$ and so transitivity and (6.10) yield (6.9).

Suppose next that all of the schools $\bar{\mu}_1(i)$, $\bar{\mu}_2(i), ..., \bar{\mu}_N(i)$ are in the same similarity set in $\mathcal{S}$, but that $f_i(P_i, \sigma_{-i})$ is not in that same similarity set. Then in particular, $f_i(P_i, \sigma_{-i}) \neq \bar{\mu}_1(i)$ and so strict preferences and (6.10) imply that $f_i(P_i, \sigma_{-i})P_i\bar{\mu}_1$. Since $\mathcal{S}$ is a similarity partition for P, student i P_i-prefers $f_i(P_i, \sigma_{-i})$ to every school in the same similarity set as $\bar{\mu}_i(i)$. In particular, $f_i(P_i, \sigma_{-i})P_i\mu_N(i) = f_i(\bar{P}_i, \sigma_{-i})$, and so (6.9) holds, which completes the proof for Case I.

Proof of (6.9) in Case II. In this case, all of the schools $f_i(P_i, \sigma_{-i})$, $\bar{\mu}_1(i)$, $\bar{\mu}_2(i), ...,$ $\bar{\mu}_N(i) = f_i(\bar{P}_i, \sigma_{-i})$ are in the same similarity set in $\mathcal{S}$.

Since (6.9) clearly holds when $\bar{P}_i = P_i$, we may assume that $\bar{P}_i \neq P_i$. Let s be the lowest-ranked school according to $\bar{P}_i$ such that P_i and $\bar{P}_i$ agree when restricted to the set of schools that $\bar{P}_i$ ranks weakly above s (i.e., such that if $s'\bar{R}_i s$ and $s''\bar{R}_i s$, then $[s'\bar{P}_i s'' \Leftrightarrow s' P_i s'']$). Let t be the school that $\bar{P}_i$ ranks just below s (i.e., such that $s\bar{P}_i t$ and there is no school r satisfying

$s\bar{P}_i r \bar{P}_i t$; such a school t must exist since $P_i \neq \bar{P}_i$). Call (s,t) the *$\bar{P}_i$-first P_i-disagreement pair* (of schools). Notice that, since $s\bar{P}_i t$, we must have $tP_i s$ by the definition of s, and so P_i disagrees with $\bar{P}_i$ about the ranking of schools s and t. Notice also that s and t are in distinct similarity sets in $\mathcal{S}$ because P_i and $\bar{P}_i$ agree on the ranking of schools within each similarity set.

For the $\bar{P}_i$-first P_i-disagreement pair (s,t), because s and t are in different similarity sets, changing $\bar{P}_i$ by switching (only) the positions of schools s and t does not change the ranking of any two schools in the same similarity set and it reduces the number of pairs of schools on which $\bar{P}_i$ disagrees with P_i. Therefore, we can obtain P_i from $\bar{P}_i$ through a finite sequence $\bar{P}_i = P_i^1 = ... = P_i^K = P_i$ of preference orders for i, such that P_i^{k+1} is derived from P_i^k by switching (only) the positions of the P_i^k-first P_i-disagreement pair of schools. So if we can establish that each such position switch improves or leaves unchanged student i's match under the PE mechanism, the proof of (6.9) in Case II (and hence the entire proof) will be complete.

Let (s,t) the $\bar{P}_i$-first P_i-disagreement pair. Let $\bar{P}_i^{s\leftrightarrow t}$ be the preference ordering that is obtained from $\bar{P}_i$ by switching (only) the positions of schools s and t. So $s\bar{P}_i^{s\leftrightarrow t}t$, and $[r\bar{P}_i^{s\leftrightarrow t}u \Leftrightarrow r\bar{P}_i u]$ for all sets of schools $\{r,u\} \neq \{s,t\}$. The proof of (6.9) in Case II will be complete if we can show that

$$f_i(\bar{P}_i^{s\leftrightarrow t}, \sigma_{-i}) R_i f_i(\bar{P}_i, \sigma_{-i}). \tag{6.11}$$

Let $\bar{\sigma}^{s\leftrightarrow t} = (\bar{P}_i^{s\leftrightarrow t}, \sigma_{-i})$, and let $\bar{\mu}_1^{s\leftrightarrow t}, ..., \bar{\mu}_K^{s\leftrightarrow t}$ be the output-sequence of the sEADA* algorithm in state $\bar{\sigma}^{s\leftrightarrow t}$. By Theorems 3.4 and 3.5, $\bar{\mu}_K^{s\leftrightarrow t}$ is the unique priority-efficient matching in state $\bar{\sigma}^{s\leftrightarrow t}$, i.e., $f_i(\bar{P}_i^{s\leftrightarrow t}, \sigma_{-i}) = \bar{\mu}_K^{s\leftrightarrow t}$.

Suppose we could show that the sEADA* output-sequence $\bar{\mu}_1, ..., \bar{\mu}_N$ in state $\bar{\sigma}$ is feasible in state $\bar{\sigma}^{s\leftrightarrow t}$. Then, by Theorem 3.5, $\bar{\mu}_N$ is the unique priority-efficient matching in state $\bar{\sigma}^{s\leftrightarrow t}$, i.e., $\bar{\mu}_N = f_i(\bar{P}_i^{s\leftrightarrow t}, \sigma_{-i})$, and so $f_i(\bar{P}_i, \sigma_{-i}) = \bar{\mu}_N = f_i(\bar{P}_i^{s\leftrightarrow t}, \sigma_{-i})$. Hence, (6.9) would hold. Similarly, (6.9) would hold if the sEADA* output-sequence $\bar{\mu}_1^{s\leftrightarrow t}, ..., \bar{\mu}_K^{s\leftrightarrow t}$ in state $\bar{\sigma}^{s\leftrightarrow t}$ were feasible in state $\bar{\sigma}$.

So to establish (6.9), it suffices to show that, either,

$$\bar{\mu}_1, ..., \bar{\mu}_N \text{ is feasible in state } \bar{\sigma}^{s\leftrightarrow t}, \tag{6.12}$$

or,

$$\bar{\mu}_1^{s\leftrightarrow t}, ..., \bar{\mu}_K^{s\leftrightarrow t} \text{ is feasible in state } \bar{\sigma}. \tag{6.13}$$

The remainder of the proof establishes that one of (6.12) and (6.13) must hold.[23] We consider

[23] Here we make use of the flexibility inherent in the class feasible sequences of matchings. It is generally

two subcases,

Case II.a. There is no n such that $\bar{\mu}_n(i) = s$.

Case II.b. There is n such that $\bar{\mu}_n(i) = s$.

Proof of (6.12) in Case II.a. In this case there is no n such that $\bar{\mu}_n(i) = s$. We will show that, in this case, (6.12) holds.

Since $\bar{\mu}_1(i) \neq s$, we have that for any school $r \in S$,

$$r \bar{P}_i^{s \leftrightarrow t} \bar{\mu}_1(i) \Rightarrow r \bar{P}_i \bar{\mu}_1(i). \tag{6.14}$$

The relation (6.14) has several consequences. First, because $\bar{\mu}_1$ is $\bar{\sigma}$-stable and $\bar{\mu}_1(i) \neq s$, $\bar{\mu}_1$ is $\bar{\sigma}^{s \leftrightarrow t}$-stable. Second, because $\bar{\mu}_n$ $\bar{\sigma}$-dominates $\bar{\mu}_{n-1}$ for each $n > 1$, and no $\bar{\mu}_n(i)$ is equal to s, $\bar{\mu}_n$ $\bar{\sigma}^{s \leftrightarrow t}$-dominates $\bar{\mu}_{n-1}$ for each $n > 1$. Third,, because $\bar{\mu}_N$ is Pareto-efficient in state $\bar{\sigma}$ and $\bar{\mu}_N(i) \neq s$, $\bar{\mu}_N$ is Pareto efficient in state $\bar{\sigma}^{s \leftrightarrow t}$. Finally, if for any student j and for any $n > 1$, student j is $\bar{\mu}_{n-1}$-improvable in state $\bar{\sigma}^{s \leftrightarrow t}$, then (6.14) implies that student j is $\bar{\mu}_{n-1}$-improvable in state $\bar{\sigma}$. Consequently, by feasibility in state $\bar{\sigma}$, $\bar{\mu}_n$ does not violate j's priority in state $\bar{\sigma}$. Hence, by (6.14), $\bar{\mu}_n$ does not violate i's priority in state $\bar{\sigma}^{s \leftrightarrow t}$. Hence, $\bar{\mu}_1, ..., \bar{\mu}_N$ is feasible in state $\bar{\sigma}^{s \leftrightarrow t}$ and so (6.12) holds in Case II.a.

Proof of (6.13) in Case II.b. In this case there is $\bar{n}$ such that $\bar{\mu}_{\bar{n}}(i) = s$. We will show that, in this case, (6.13) holds.

By replacing $\bar{P}_i$ with $\bar{P}_i^{s \leftrightarrow t}$ in the argument that established (6.10), we have $f_i(P_i, \sigma_{-i}) R_i \bar{\mu}_1^{s \leftrightarrow t}$. For any n, we claim that, either $f_i(P_i, \sigma_{-i})$ and $\bar{\mu}_n^{s \leftrightarrow t}(i)$ are in the same similarity set or $f_i(P_i, \sigma_{-i}) P_i \bar{\mu}_n^{s \leftrightarrow t}$.

To establish the claim, suppose that $f_i(P_i, \sigma_{-i})$ and $\bar{\mu}_n^{s \leftrightarrow t}(i)$ are in distinct similarity sets. We must show that $f_i(P_i, \sigma_{-i}) P_i \bar{\mu}_n^{s \leftrightarrow t}$. There are two cases. First, suppose that $\bar{\mu}_n^{s \leftrightarrow t}(i)$ and $\bar{\mu}_1^{s \leftrightarrow t}(i)$ are in the same similarity set. Then $f_i(P_i, \sigma_{-i})$ and $\bar{\mu}_1^{s \leftrightarrow t}(i)$ are distinct because they are in distinct similarity sets. Hence, $f_i(P_i, \sigma_{-i}) R_i \bar{\mu}_1^{s \leftrightarrow t}$ and strict preferences imply $f_i(P_i, \sigma_{-i}) P_i \bar{\mu}_1^{s \leftrightarrow t}$. But then also $f_i(P_i, \sigma_{-i}) P_i \bar{\mu}_n^{s \leftrightarrow t}$ because $\mathcal{S}$ is a similarity partition for P, which takes care of this first case. Alternatively, suppose that $\bar{\mu}_n^{s \leftrightarrow t}(i)$ and $\bar{\mu}_1^{s \leftrightarrow t}(i)$ are in distinct similarity sets. Then Corollary 6.11 implies that $\bar{\mu}_1^{s \leftrightarrow t} P_i \bar{\mu}_n^{s \leftrightarrow t}$. Since $f_i(P_i, \sigma_{-i}) R_i \bar{\mu}_1^{s \leftrightarrow t}$, transitivity implies $f_i(P_i, \sigma_{-i}) P_i \bar{\mu}_n^{s \leftrightarrow t}$, establishing the claim.

Because we are in Case II, all of the schools $f_i(P_i, \sigma_{-i}), \bar{\mu}_1(i), \bar{\mu}_2(i), ..., \bar{\mu}_N(i) = f_i(\bar{P}_i, \sigma_{-i})$ are in the same similarity set in $\mathcal{S}$. Therefore, because $\bar{\mu}_{\bar{n}}(i) = s$ (since we are in Case II.B), $f_i(P_i, \sigma_{-i})$ is in the same similarity set as s. Since t and s are in distinct similarity sets in the similarity partition $\mathcal{S}$ for P, and since $t P_i s$, every school in the same similarity set as t is preferred according to P_i to every school in same similarity set as s. In particular,

false, for example, that the output-sequence of the sEADA* algorithm in state $\bar{\sigma}$ is also the output-sequence of the sEADA* algorithm in state $\bar{\sigma}^{s \leftrightarrow t}$.

$tP_i f_i(P_i, \sigma_{-i})$.

As already established, either $f_i(P_i, \sigma_{-i})$ and $\bar{\mu}_n^{s\leftrightarrow t}(i)$ are in the same similarity set, or $f_i(P_i, \sigma_{-i})P_i\bar{\mu}_n^{s\leftrightarrow t}$. In the former case, $\bar{\mu}_n^{s\leftrightarrow t}(i)$ is in the same similarity set as s and so $tP_i\bar{\mu}_n^{s\leftrightarrow t}(i)$ (since, as already established, every school in the same similarity set as t is P_i-preferred to every school in same similarity set as s). In the latter case, $tP_i f_i(P_i, \sigma_{-i})$ and transitivity yield $tP_i\bar{\mu}_n^{s\leftrightarrow t}(i)$. So $tP_i\bar{\mu}_n^{s\leftrightarrow t}(i)$ holds in either case. In particular, $\bar{\mu}_n^{s\leftrightarrow t}(i) \neq t$.

Since n was arbitrary, we have established that there is no n such that $\bar{\mu}_n^{s\leftrightarrow t}(i) = t$, which is Case II.a but with the roles of s and t reversed. So reversing the roles of s and t in the proof for Case II.a shows that $\bar{\mu}_1^{s\leftrightarrow t}, ..., \bar{\mu}_K^{s\leftrightarrow t}$ is feasible in state $\bar{\sigma}$, which establishes (6.13) in Case II.b. Q.E.D.

References

Abdulkadiroğlu, A, P. A. Pathak, and A. E. Roth (2009), "Strategy-Proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match," *American Economic Review*, 99, 1954-1978.

Abdulkadiroğlu, A., and T. Sönmez (2003): "School Choice: A Mechanism Design Approach," *American Economic Review*, 93, 729-747.

Alcalde, J., and A. Romero-Medina (2017): "Fair Student Placement," *Theory and Decision*, 83, 293-307.

Balinski, M., and T. Sönmez (1999): "A Tale of Two Mechanisms: Student Placement," *Journal of Economic Theory* 84, 73-94.

Blair, C. (1988): "The Lattice Structure of the Set of Stable Matchings with Multiple Partners," *Mathematics of Operataions Research*, 13, 619-628.

Chwe, M. S.-Y. (1994): "Farsighted Coalitional Stability," *Journal of Economic Theory*, 63, 299-325.

Dubins, L. E., and D. A. Freedman (1981): "Machiavelli and the Gale–Shapley Algorithm," *American Mathematical Monthly*, 88, 485–494.

Ehlers, L. (2007): "von Neumann-Morgenstern Stable Sets in Matching Problems, *Journal of Economic Theory*, 134, 537-547.

Ehlers, L. (2008): "Truncation Strategies in Matching Markets," *Mathematics of Operations Research*, 33, 327-335.

Ehlers, L., and T. Morrill (2020): "(Il)legal Assignments in School Choice," *Review of Economic Studies*, 87, 1837–1875.

Erdil, A., and H. Ergin (2008), "What's the Matter with Tie-Breaking? Improving Efficiency in School Choice," *American Economic Review*, 98 669-689.

Gale, D., and L. S. Shapley (1962): "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, 69, 9-15.

Gale, D., and M. Sotomayor (1985): "Some Remarks on the Stable Matching Problem," *Discrete Applied Mathematics* 11, 223-232.

Kesten, O. (2010): "School Choice with Consent," *The Quarterly Journal of Economics*, August, 1297-1348.

Mauleon, A., V. J. Vannetelbosch, and W. Vergote (2011): "von Neumann–Morgenstern Farsightedly Stable Sets in Two-Sided Matching," *Theoretical Economics*, 6, 499–521.

Pápai, S. (2000): "Strategyproof Assignment by Hierarchical Exchange," *Econometrica*, 68, 1403- 1433.

Reny, P. J. (2020): "A Simple Sufficient Condition for a Unique and Student-Efficient Stable Matching in the College Admissions Problem," Department of Economics, University of Chicago.

Roth, A. E., and U. G. Rothblum (1999): "Truncation Strategies in Matching Markets–In Search of Advice for Participants," *Econometrica*, 67, 21-43.

Roth. A. E., and M. A. Oliveira Sotomayor (1990): *Two-Sided Matching: A Study in Game Theoretic Modeling and Analysis.* New York: Cambridge University Press.

Tang, Q., and J. Yu (2014): "A New Perspective on Kesten's School Choice with Consent Idea", *Journal of Economic Theory*, 154, 543–561.

Troyan, P., D. Delacrétaz, and A. Kloosterman (2020): "Essentially Stable Matchings," *Games and Economic Behavior*, 120, 370-390.

Von Neumann, J,. and O. Morgenstern (1944): *Theory of Games and Economic Behavior*, Princeton: Princeton University Press.