

What's the Matter with Tie-Breaking? Improving Efficiency in School Choice

By AYTEK ERDIL AND HALUK ERGIN*

In several school choice districts in the United States, the student proposing deferred acceptance algorithm is applied after indifferences in priority orders are broken in some exogenous way. Although such a tie-breaking procedure preserves stability, it adversely affects the welfare of the students since it introduces artificial stability constraints. Our main finding is a polynomial-time algorithm for the computation of a student-optimal stable matching when priorities are weak. The idea behind our construction relies on a new notion which we call a stable improvement cycle. We also investigate the strategic properties of the student-optimal stable mechanism. (JEL C78, D82, I21)

Until about a decade ago, children in the United States were assigned to public schools by the district they live in, without taking into account the preferences of their families. Such systems overlooked reallocations of seats which could Pareto improve welfare. Motivated by such concerns, several US cities, including New York City, Boston, Cambridge, Charlotte, Columbus, Denver, Minneapolis, Seattle, and St. Petersburg-Tampa, started centralized school choice programs. Typically in these programs, each family submits a preference list of schools, including those outside their district, and then a centralized mechanism assigns students to schools based on the preferences. The mechanisms initially adopted by school choice programs were ad hoc, and did not perform well in terms of efficiency, incentives, and/or stability. Atila Abdulkadiroğlu and Tayfun Sönmez (2003) brought these issues to light, which triggered an interest in the matching literature about further analysis and design of school choice mechanisms.¹

A school choice problem consists of a set of students and a set of schools where each school x has a quota q_x of available seats. Each student has a preference ranking over schools and her outside option, which corresponds to remaining unassigned or going to a private school, and each school has a priority ranking over students. The school choice model is closely related to the college admissions model of David Gale and Lloyd S. Shapley (1962). The important difference between the two models is that in school choice, the priority rankings are determined by local (state or city) laws and education policies, and do not reflect the school preferences, whereas in the college admissions model these rankings correspond to college preferences.² As a consequence, in the college admissions model, the preferences of students as well as colleges are taken into account in welfare considerations. On the other hand, in the school choice model, schools

* Erdil: University of Oxford, Department of Economics and Nuffield College, New Road, OX1 1NF, UK (e-mail: aytek.erdil@nuffield.ox.ac.uk); Ergin: Washington University in Saint Louis, Department of Economics, Campus Box 1208, 1 Brookings Drive, Saint Louis, MO 63130 (e-mail: hergin@artsci.wustl.edu). We would like to thank the two referees and John Hatfield, Phil Reny, Al Roth, and Tayfun Sönmez for helpful comments and suggestions. We are especially grateful to Özgür Sümer for his help with the computer simulations.

¹ See Abdulkadiroğlu, Parag A. Pathak, and Alvin E. Roth (2005, 2008), Abdulkadiroğlu et al. (2005, 2006), Ergin (2002), Ergin and Sönmez (2006), and Onur Kesten (2005, 2006).

² There are certain exceptions like New York City, where a number of schools determine their own priority orders. See Abdulkadiroğlu and Sönmez (2003), Michel Balinski and Sönmez (1999), and Ergin (2002) for a more detailed discussion of the relationship between the two models.

are treated as indivisible objects to be consumed by the students, and only student preferences constitute the welfare criteria.

Given a priority ranking for each school and a preference profile of the students, a matching *violates the priority of student i* , if there are a student j and a school x such that i prefers x to her current assignment, and j is assigned to x while he has less priority for school x than i . A matching is *stable* if (a) it does not violate any priorities, (b) every student weakly prefers his assigned seat to remaining unassigned, and (c) no student would rather be matched to a school that has empty seats. Stability has been a property of central interest in the college admissions model and, in general, in two-sided matching markets. In addition to the theoretical plausibility of the notion, Roth (2002) draws from both empirical and experimental evidence to show how stability has been an important criterion for a successful clearinghouse in matching markets ranging from the entry level labor market for new physicians in the United States to college sorority rush. In the context of school choice, legal and political concerns appear to strongly favor stable mechanisms. For instance, if the priority of student i for school x is violated, then the family of student i has incentives to seek legal action against the school district for not assigning her a seat at school x , and the district authorities seem to be extremely averse to such violations of priorities. Along those concerns, Boston officials decided to adopt a mechanism that always produces stable matchings at the expense of efficiency, rather than a mechanism—namely the top trading cycles mechanism—that could guarantee efficiency, yet not stability.

Gale and Shapley (1962) gave a constructive proof of the existence of a stable matching by describing a simple algorithm. This is known as the student proposing *deferred acceptance (DA) algorithm*:

At the first step, every student applies to her favorite acceptable school. For each school x , q_x applicants who have highest priority for x (all applicants if there are fewer than q_x) are placed on the hold list of x , and the others are rejected.

At step $t > 1$, those applicants who were rejected at step $t - 1$ apply to their next best acceptable schools. For each school x , the highest priority q_x students among the new applicants and those in the hold list are placed on the new hold list and the rest are rejected.

The algorithm terminates when every student is either on a hold list or has been rejected by every school that is acceptable to her. After this procedure ends, schools admit students on their hold lists, which yields the desired matching.

Gale and Shapley (1962) show that, when preferences and priorities are strict, the DA algorithm yields the unique stable matching that is Pareto superior to any other stable matching from the viewpoint of the students. Hence, the outcome of the student proposing DA algorithm is also called the *student optimal stable matching*, and the mechanism that associates the student optimal stable matching to any school choice problem is known as the *student optimal stable mechanism (SOSM)*.³ Beside the fact that it gives the most efficient stable matching, another appealing feature of the SOSM, when priorities are strict, is that it is strategy-proof, that is, no student has an incentive to misstate her true preference ranking over schools (Lester E. Dubins and David A. Freedman 1981; Roth 1982). Due to these desirable features, the DA algorithm has been adopted by the school choice programs of New York City (in 2003) and Boston (in 2005), in consultation with economists Abdulkadiroğlu, Pathak, Roth, and Sönmez.⁴

³ The SOSM has played a key role in the redesign of the US hospital-intern market in 1998. See Roth and Elliott Peranson (1999) and Roth (2003).

⁴ See Abdulkadiroğlu, Pathak, and Roth (2005), Abdulkadiroğlu et al. (2005, 2006), and *The Boston Globe* article "School assignment flaws detailed," September 12, 2003.

The DA algorithm, as described above, requires that both the preference orders and priority orders be strict for it to be deterministic and single-valued. This is because whenever a student proposes, she chooses her next best school, and a school rejects the lowest priority students among those who applied. Obviously, indifference classes would create ambiguities in those choices. In the context of school choice, it might be reasonable to assume that the students have strict preferences, but school priority orders are typically determined according to criteria that do not provide a strict ordering of all the students. Instead, school priorities are weak orderings with quite large indifference classes. For instance, in Boston there are mainly four indifference classes for each school in the following order: (a) the students who have siblings at that school (sibling) and are in the reference area of the school (walk zone), (b) sibling, (c) walk zone, and (d) all other students.⁵ Common practice in these cases is to exogenously fix an ordering of the students, chosen randomly, and break all the indifference classes according to this fixed strict ordering. Then, one can apply the DA algorithm to obtain the student optimal stable matching with respect to the strict priority profile derived from the original one. Since the breaking of indifferences does not switch the positions of any two students in any priority order, the outcome would also be stable with respect to the original priority structure.

The following example illustrates that, although tie-breaking seems like a quick solution to having indifferences in priority orders, it is not costless.

Example 1: Consider a school choice problem with three students, three schools each having one seat, and the following priority orders:

\succsim_x	\succsim_y	\succsim_z
1	2	3
2,3	1,3	1,2 .

If the ties in the priority orders are broken favoring 1 over 2 over 3, then we obtain the strict priority structure \succsim' below. Consider the following preference profile R of the students:

R_1	R_2	R_3	\succsim'_x	\succsim'_y	\succsim'_z
y	z	y	1	2	3
x	y	z	2	1	1
z	x	x	3	3	2 .

The student optimal stable matching for the preference profile R and the strict priority structure \succsim' is

$$\mu = \begin{pmatrix} 1 & 2 & 3 \\ x & y & z \end{pmatrix}.$$

⁵ There are also students who have a guaranteed priority to a given school. For a complete description, see Abdulkadiroğlu et al. (2006) or “Introducing the Boston Public Schools 2007: A Guide for Parents and Students,” available at <http://www.boston.k12.ma.us/schools/assign.asp> (accessed September 12, 2007).

However, there is another matching:

$$\nu = \begin{pmatrix} 1 & 2 & 3 \\ x & z & y \end{pmatrix},$$

which Pareto dominates μ and is stable with respect to \succeq and R .

When priorities are allowed to have indifferences, we define a *student optimal stable matching* to be a stable matching that is not Pareto dominated by any other stable matching. In Example 1, ν is a student optimal stable matching, whereas μ is not, since it is Pareto dominated by the stable matching ν . Given a profile of weak priorities, we define the *student optimal stable mechanism* (SOSM) to be the mechanism that associates the set of student optimal stable matchings with every preference profile. The SOSM may be multivalued because of the indifferences in the priority orders.

The example above showed how the arbitrariness of the tie-breaking rule may cause an efficiency loss. How can one find a stable matching that avoids such inefficiency? The computer science literature on matching mostly includes results showing how introducing ties makes various problems *computationally hard*.⁶ In contrast with those, we will be giving a strikingly fast solution to the problem of finding a student optimal stable matching, which also shows that not all problems become highly complex when indifferences exist.

Suppose, for instance, that at a stable matching, (i) student 1 would rather have a seat at school x in which student 2 has a seat; (ii) student 2 would rather have a seat at school y in which student 3 has a seat; and (iii) student 3 would rather have a seat at school z in which student 1 has a seat. It is Pareto improving to carry out the corresponding cyclic trade of seats among 1, 2, and 3, which would reallocate the seats so that 1 is assigned to x , 2 is assigned to y , and 3 is assigned to z . However, even if the matching we started with is stable, in general, there is no guarantee that the resulting matching after executing this cyclic trade will also be stable. For instance, there could be a student 4 who also desires to be assigned to x in the initial matching and has less x -priority than 2, but strictly more x -priority than 1. In this case, the exchange would violate the priority of 4 for x , since 4 prefers x to her current assignment and 1 is assigned to x while he has less priority for school x than 4.

We call a cycle consisting of (i), (ii), (iii) as the one above a *stable improvement cycle* if at the original matching: (i') 1 was among the highest x -priority students among those who desire a seat at x , (ii') 2 was among the highest y -priority students who desire a seat at y , and (iii') 3 was among the highest z -priority students who desire a seat at z . If we start at a stable matching and execute a stable improvement cycle, it is guaranteed that we will end up at a stable matching since conditions (i'), (ii'), and (iii') make sure that there are no intermediate students like 4 whose priority would be violated after carrying out the cyclic trade.

We show in our main result that if a stable matching is not in the student optimal stable set, then it admits a stable improvement cycle. Hence, starting at any stable matching, we can construct a sequence of Pareto improving stable matchings by carrying out stable improvement cycles, until we reach a student optimal stable matching. The procedure, which we call *the stable improvement cycles mechanism*, is polynomial in the number of students plus schools; hence, it is computationally very simple. Since our model is finite, exhaustive verification methods also give finite time algorithms that compute a student optimal stable matching. In contrast with our stable improvement cycles algorithm, however, we note that these exhaustive algorithms are computationally quite complex and infeasible even for school choice problems consisting of a small number of students and schools.

⁶ See for instance Magnús Halldórsson et al. (2003) and David F. Manlove et al. (2002).

TABLE 1—IMPROVEMENTS VIA STABLE IMPROVEMENTS CYCLES^a

Year	Number of assigned students	Number of improving students	Percent of improving students
2003–2004	63,749	1,693	2.7
2004–2005	76,831	1,986	2.6
2005–2006	73,378	1,818	2.5
2006–2007	73,115	1,487	2.0

^aThese numbers are from Abdulkadiroğlu, Pathak, and Roth (2008).

We can start the stable improvement cycles procedure with a random tie-breaking followed by the DA algorithm to obtain a stable matching to Pareto improve upon. This initial temporary matching is essentially what is returned in the New York City and Boston School Choice Programs. Therefore, the search for stable improvement cycles also provides an *optimality test* for these mechanisms. Furthermore, as each cycle Pareto improves the initial match, one can provide a measure of the efficiency loss due to random tie-breaking.

Using preference data from the New York City High School Match, Abdulkadiroğlu, Pathak, and Roth (2008) carry out this empirical exercise. They document that, on average, over 1,700 students would have been matched with schools higher on their preference lists without hurting others had the stable improvement cycles mechanism been applied to the reported preferences. Thus, the efficiency cost from an arbitrary tie-breaking can be significant.

Table 1 shows how many students' matches could have been improved via the stable improvement cycles in the years where the DA algorithm was used following a random tie-breaking. To complement this empirical finding, we report a number of sensitivity analyses using simulations. In particular, we investigate how the size of the Pareto improvement varies as the correlation among student preferences and the weight given to locational proximity change.

We next turn to analyze the strategic properties of the SOSM when priorities are weak. We show that, in general, there may not exist a deterministic or stochastic selection from the SOSM, which is strategy-proof. We also show that, given any such selection, it is a Nash equilibrium for all students to state their true preferences at profiles where the outcome of the SOSM is a singleton. Hence, the former negative result seems to be intimately related to multiplicities in the student optimal stable set.

Even when there is room for manipulation, this would require a particular kind of information about the preferences and the priority rankings of the other students. As a theoretical benchmark, we show that if a student has *symmetric beliefs* about two schools, then under an anonymous random selection from the SOSM, she can not benefit by a misstatement that switches the ranking of these two schools in her true preference profile. Accordingly, under low information, when it is common knowledge that everyone prefers being assigned to a school to staying unmatched, the preference revelation game induced by the stable improvement cycles mechanism has truth-telling as a Bayesian Nash equilibrium.

Throughout the paper, our welfare considerations take only student preferences into account, and we have a one-sided perspective in that sense. In Erdil and Ergin (2006), we model both sides as economic agents, such as workers versus firms. In two-sided matching literature, it has been a standard assumption that agents are not indifferent between any two members of the opposite side, despite the existence of such indifferences in various actual settings. A number of issues arise if agents are not assumed to have strict preferences on the opposite side. Most importantly, stability no longer implies Pareto efficiency, and the deferred acceptance algorithm can not be applied to produce a Pareto efficient or a worker/firm optimal stable matching. In Erdil and Ergin (2006), we allow ties in preference rankings on both sides and explore the Pareto domination relation on stable matchings, as well as the two relations defined via workers' welfare and firms'

welfare. Our structural results lead to *fast* algorithms to compute a Pareto efficient and stable matching, and a worker (or firm) optimal stable matching.

In general, insisting on stability would have efficiency costs (from the perspective of students) regardless of the tie-breaking rule. A complete characterization of priority structures for which the stability constraints do not lead to such an efficiency loss is given in Erdil (2008).

I. The Model and Preliminary Observations

Let N denote a finite set of students and X a finite set of schools. Let $q_x \geq 1$ denote the number of available seats in school x . Throughout, we will maintain the assumption that student preferences are strict: A **preference profile** is a vector of linear orders (complete, transitive, and antisymmetric relations) $R = (R_i)_{i \in N}$, where R_i denotes the preference of student i over $X \cup \{i\}$. Being assigned to oneself is interpreted as not being assigned to any school. Let P_i denote the asymmetric part of R_i . A **matching** is a function $\mu : N \rightarrow X \cup N$ satisfying: (i) $\forall i \in N : \mu(i) \in X \cup \{i\}$ and (ii) $\forall x \in X : |\mu^{-1}(x)| \leq q_x$. A **mechanism** f is a function that associates a non-empty set of matchings with every preference profile. A **random mechanism** F associates a probability distribution over matchings with every preference profile R .

A **priority structure** is a profile of weak orders (complete and transitive relations) $\succeq = (\succeq_x)_{x \in X}$, where for each $x \in X$, \succeq_x ranks students with respect to their priority for x . Let \succ_x denote the asymmetric part of \succeq_x . We say that \succeq is **strict** if for any $x \in X$, \succeq_x is antisymmetric. Let $\mathcal{T}(\succeq)$ denote the set of strict priority profiles \succeq' obtained by breaking the ties in \succeq .⁷ Given \succeq and R , the matching μ **violates the priority of i for x** , if there is a student j such that j is assigned to x , whereas i both desires x and has strictly higher priority for it, i.e., $\mu(j) = x$, $xP_i\mu(i)$, and $i \succ_x j$. The matching μ is **stable** if (i) it does not violate any priorities, (ii) $\mu(i)R_i\mu(i)$ for any i , and (iii) there do not exist i and x such that $xP_i\mu(i)$ and $q_x > |\mu^{-1}(x)|$. Let S^\succeq denote the **stable mechanism**, i.e., the mechanism that associates to each R the set of stable matchings with respect to \succeq and R .

Given R , the matching μ' **Pareto dominates** the matching μ if $\mu'(i)R_i\mu(i)$ for every $i \in N$, and $\mu'(j)P_j\mu(j)$ for some $j \in N$. Given \succeq and R , the matching μ is **constrained efficient** (or **student optimal stable**) if (i) $\mu \in S^\succeq(R)$ and (ii) μ is not Pareto dominated by any other $\mu' \in S^\succeq(R)$. Let f^\succeq denote the **SOSM**, i.e., the mechanism that associates to each R the set of constrained efficient matchings with respect to \succeq and R . Given \succeq , a mechanism f is **constrained efficient** if for any R , $f(R) \subset f^\succeq(R)$.

THEOREM 0 (Gale and Shapley 1962): *For any strict \succeq and R , $f^\succeq(R)$ consists exactly of the matching given by the DA algorithm.*

The following are well-known facts about how tie-breaking affects the stable and the student optimal stable mechanisms. Their variants can be found in Gale and Shapley (1962), Roth and Marilda Sotomayor (1990), and Lars Ehlers (2006). We explain the arguments behind them below for completeness.

OBSERVATION 1: $S^\succeq = \bigcup_{\succeq' \in \mathcal{T}(\succeq)} S^{\succeq'}$.

OBSERVATION 2: $f^\succeq \subset \bigcup_{\succeq' \in \mathcal{T}(\succeq)} f^{\succeq'}$.

⁷ Formally, $\mathcal{T}(\succeq)$ is the set of strict priority structures \succeq' such that $i \succ_x j$ implies $i \succ'_x j$ for all $x \in X$ and $i, j \in N$.

When one breaks ties in \succeq to obtain \succeq' , the constraints required by stability weakly increase. Hence, stability becomes weakly more difficult to be satisfied under \succeq' , which gives the “ \supset ” part in Observation 1. To see the “ \subset ” part of Observation 1, take any preference profile R and let μ be a stable matching with respect to R and \succeq . For each school x , break the ties among the students within each indifference class of \succeq_x by favoring those who are assigned to x under μ , if any. By stability of μ prior to the tie-breaking, if a student i strictly prefers a school x to her match, then any other student j who is assigned to x must have weakly higher \succeq_x -priority than i . Since j is matched to x under μ , she can not have less \succeq'_x -priority than i after the tie-breaking above. This shows that $\mu \in S^{\succeq'}(R)$.

To see Observation 2, take any R and let $\mu \in f^{\succeq}(R)$. From Observation 1, there exists a tie-breaking \succeq' for which μ continues to be stable. Since any other matching μ' that is stable with respect to \succeq' is also stable with respect to \succeq , constrained efficiency of μ under \succeq implies constrained efficiency of μ under the tie-breaking \succeq' . We conclude that $f^{\succeq'}(R) = \{\mu\}$. As illustrated in Example 1, applying the DA algorithm after breaking the ties in \succeq may lead to outcomes that are not constrained efficient under \succeq . Hence, the inclusion in Observation 2 might in fact be proper.

II. A Constrained Efficient Solution

Let μ be a stable matching for some fixed \succeq and R . We will say that a student i desires school x if she prefers x to her assignment at μ , that is $xP_i\mu(i)$. For each school x , let D_x denote the set of highest \succeq_x -priority students among those who desire x . We will suppress the dependence of D_x on μ .

DEFINITION 1: A *stable improvement cycle* consists of distinct students $i_1, \dots, i_n \equiv i_0$ ($n \geq 2$) such that:

- (i) $\mu(i_\ell) \in X$ (each student in the cycle is assigned to a school);
- (ii) i_ℓ desires $\mu(i_{\ell+1})$; and
- (iii) $i_\ell \in D_{\mu(i_{\ell+1})}$,

for any $\ell = 0, \dots, n - 1$.

Given a stable improvement cycle, define a new matching μ' by:

$$\mu'(j) = \begin{cases} \mu(j) & \text{if } j \notin \{i_1, \dots, i_n\} \\ \mu(i_{\ell+1}) & \text{if } j = i_\ell \end{cases}.$$

Note that the matching μ' continues to be stable and it Pareto dominates μ . We are now ready to present our main result.

THEOREM 1: Fix \succeq and R , and let μ be a stable matching. If μ is Pareto dominated by another stable matching ν , then it admits a stable improvement cycle.⁸

⁸ We could actually “squeeze in” a stable improvement cycle between any two Pareto-ranked stable matchings. Formally, we could guarantee that the new stable matching μ' , obtained from μ by applying the stable improvement cycle, lies weakly below ν in a Pareto sense.

If a stable matching is not constrained efficient, then there must exist a Pareto improvement which is still stable. Theorem 1 tells us that to find such a Pareto improvement, it is enough to look for a stable improvement cycle. By successive application of this result, we can define a procedure that produces a constrained efficient matching, which we call the **stable improvement cycles algorithm**.

Step 0. Select a strict priority structure \succeq' from $\mathcal{T}(\succeq)$. Run the DA algorithm and obtain a temporary matching μ^0 .

Step $t \geq 1$.

(*t.a*) Given μ^{t-1} , let the schools stand for the vertices of a directed graph, where for each pair of schools x and y , there is an edge $x \rightarrow y$ if and only if there is a student i who is matched to x under μ^{t-1} , and $i \in D_y$.

(*t.b*) If there are any cycles in this directed graph, select one. For each edge $x \rightarrow y$ on this cycle, select a student $i \in D_y$ with $\mu^{t-1}(i) = x$. Carry out this stable improvement cycle to obtain μ^t , and go to step (*t+1.a*). If there is no such cycle, then return μ^{t-1} as the outcome of the algorithm.

In the description above, we leave it open how the procedure should select \succeq' in step 0, and how it should select the cycle and the student in step (*t.b*). Therefore, one can think of the description above as corresponding to a *class* of algorithms, where an algorithm is determined only after we fully specify how to act when confronted with multiplicity. One can imagine these selections to be random or dependent on the earlier selections. Let F^{\succeq} denote the random mechanism induced by our algorithm when the selections are made independently and randomly with equal probabilities each time the algorithm faces a multiplicity. In Section IV, we will analyze the strategic behavior of the students under the random mechanism F^{\succeq} . Remember that given \succeq , R , and $\mu \in f^{\succeq}(R)$, there is a tie-breaking $\succeq' \in \mathcal{T}(\succeq)$ such that the DA algorithm applied to (R, \succeq') returns μ . Since each tie-breaking has positive probability of being selected at step 0 of the algorithm corresponding to F^{\succeq} , $F^{\succeq}(R)$ gives positive probability to every constrained efficient matching.

Note that Observation 2 also yields an algorithm to find a student optimal stable matching. Namely, one could apply the DA algorithm to all possible tie-breakings of the given priority structure, record the outcomes, and Pareto compare them to find a student optimal stable matching. Even with a single indifference class of only 100 students, however, this would amount to running the DA algorithm more than 10^{90} times, a computationally infeasible task. From a practical perspective, the value of our algorithm comes from its remarkably small computational complexity.⁹

Stable improvement cycles are closely related to Gale's top-trading cycles, originally introduced in Shapley and Herbert Scarf (1974) and later studied in detail by Szilvia Pápai (2000) and Abdulkadiroğlu and Sönmez (2003). At a matching μ , a **top-trading cycle** consists of students $i_1, \dots, i_n \equiv i_0$ ($n \geq 2$) such that conditions (i) and (ii) in our definition of a stable improvement cycle are satisfied, and additionally $\mu(i_{\ell+1})$ is student i_ℓ 's top ranked school for $\ell = 0, \dots, n - 1$. Suppose that matching μ is stable to start with. There are two reasons for which we could not make use of top-trading cycles in our construction. First, since condition (iii) is not required in a top-trading cycle, there is no guarantee that the matching μ' obtained after executing the

⁹In addition to the DA algorithm used in practice, it involves a repetition of cycle search in a directed graph. The latter is known to be of complexity $O(|V| + |E|)$ where V is the set of vertices and E the set of edges (Thomas H. Cormen et al. 2001). This obviously is very fast; the question, then, is how many times one has to repeat the cycle search. Notice that with every cycle, at least two students improve; therefore, each cycle brings at least two moves up with respect to the students' preferences. Since there are $|N|$ students and the student preferences involve $|X|$ schools, there could be at most $|N|(|X| - 1)$ moves up. Therefore, cycle search has to be repeated at most $\frac{1}{2}|N|(|X| - 1)$ times.

top-trading cycle will continue to be stable. Second, top-trading cycles are *too demanding* for our purposes, since even when there exist Pareto improving trading cycles that preserve stability, there may not exist such a cycle where all participating students receive their top choices.

III. Simulations

In order to compare the performance of the *stable improvement cycles algorithm* to that of the DA algorithm applied with random tie-breaking, we design a random environment and employ computer simulations. Upon consulting the experts, both the New York City Department of Education (in 2003) and the Boston Public School System (in 2005) adopted mechanisms that first randomly pick an ordering of the students according to which they break the ties between students of the same priority, and then apply the student proposing DA algorithm to the derived school choice problem with strict priority structure and preferences.¹⁰ For brevity, let us call this mechanism RDA, standing for *Randomize and apply the DA algorithm*.

Recall that our algorithm begins like RDA, but goes further, and Pareto improves on the outcome of RDA by allowing exchanges of school seats through stable improvement cycles as described in Section II. It stops once it reaches a student optimal stable matching. Whenever RDA fails to return a student optimal stable matching, our mechanism will improve on it. In comparing the performance of the two mechanisms, we will ask several questions. First of all, how significant is the improvement, that is, how many students are affected by it, and how many moves up in their preferences are achieved? And, second, how do the number of students, the number of schools, correlation in preferences, and preferences for locational proximity affect the size of the Pareto improvement?

In developing our setup for the computational experiments, we aim to account for a number of aspects of real life school choice problems. For instance, in order to ensure that certain schools are demanded more than others, we will allow for correlation in student preferences. We also account for locational preferences, since, *ceteris paribus*, a student is more likely to prefer a school located closer to her residence to schools that are located farther away. We also make sure that preferences for proximity are consistent with the priority structure in the following sense: if a student has a high priority at a school because she is in the walk zone of that school, then she should in turn be more likely to favor that school because of its locational proximity.

Let the students be indexed by $i = 1, \dots, n$, and the schools be indexed by $x = 1, \dots, m$, where, for simplicity, we assume that $q = n/m$ is an integer equal to the number of seats in each school. For each student i and school x , we represent the locations of the student and the school in a two-dimensional map by points $\ell^i, \ell^x \in \mathbb{R}^2$, respectively. Let $i = 0$ denote an artificially created dummy student with average tastes. Let Z_{ix} be i.i.d. normally distributed random variables with mean zero and variance one for $i = 0, 1, \dots, n; x = 1, \dots, m$. We define the utility of student $i = 1, \dots, n$ for school $x = 1, \dots, m$ by:

$$(1) \quad U_{ix} = -\beta d(\ell^i, \ell^x) + (1 - \beta)[\alpha Z_{0x} + (1 - \alpha)Z_{ix}],$$

where $d(.,.)$ denotes the usual Euclidean distance, and $\alpha, \beta \in [0, 1]$ are fixed common parameters. The parameter α captures the correlation in student preferences and the parameter β captures student preferences' sensitivity to locational proximity.¹¹ We assume that each student is in

¹⁰ See Abdulkadirođlu, Pathak, and Roth (2005, 2008) and Abdulkadirođlu et al. (2005).

¹¹ It is easy to see that for any two students i and j , and school x , if $\beta < 1$, then correlation between U_{ix} and U_{jx} is given by $\alpha^2/[\alpha^2 + (1 - \alpha)^2]$. Hence, correlation is strictly increasing in α ; it is 0 at $\alpha = 0$, and 1 at $\alpha = 1$. Similarly, students put zero weight on location if $\beta = 0$, they evaluate schools solely based on proximity if $\beta = 1$, and intermediate values of β qualitatively capture the weight given to proximity in student preferences.

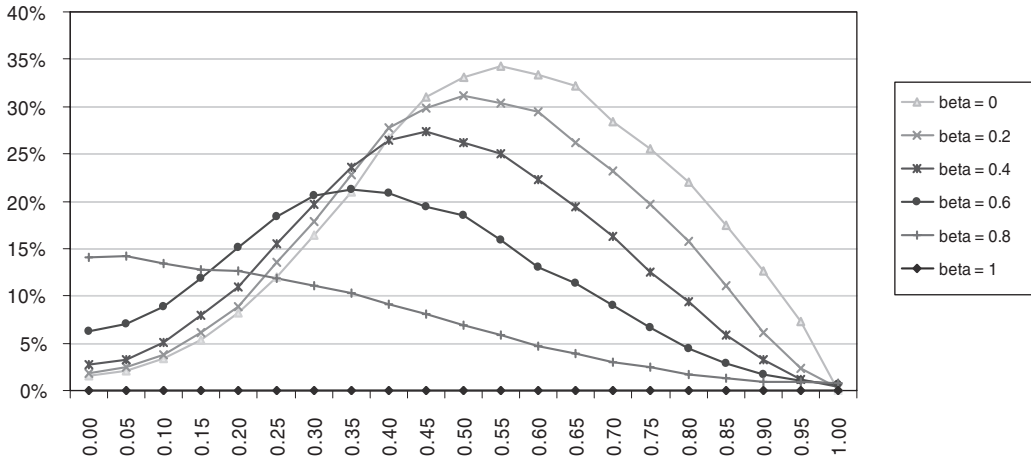


FIGURE 1. PERCENT OF IMPROVING STUDENTS AS A FUNCTION OF ALPHA
(1,000 students and 20 schools)

the walk zone of the school located closest to her residence. Each school's priority order consists of only two indifference classes determined by its walk zone.

Given parameters n , m , α , and β , in each *simulation experiment*, we go through the following steps:

1. Generate locational parameters of students and schools by using the i.i.d. uniform distribution on $[0, 1] \times [0, 1]$;
2. Determine the priority structure from the locational parameters;
3. Generate the i.i.d. normal variables Z_{ix} and determine the utility values from formula (1). Define the preference of each student i by: $xP_i y \Leftrightarrow U_{ix} > U_{iy}$, where all schools are acceptable;
4. Randomly pick an ordering of the students and apply RDA;
5. Pareto improve over RDA by using the stable improvement cycles algorithm. Compute the percentage of students who improve from RDA, as well as the average number of steps up the preference list of the improving students.

Unlike earlier computational simulations in the literature such as Yan Chen and Sönmez (2006) with 36 students and 7 schools, and Kesten (2005) with 240 students and 12 schools, we conducted experiments with relatively large sets of schools and students. We ran a number of experiments where the number of schools ranges from 10 to 100, and the number of students per school from 5 to 200. We report a few of them in Figure 1, which shows how the parameters α and β affect the percentage of improving students for a school choice problem consisting of 1,000 students and 20 schools. For each (α, β) pair, the plotted percentage improvement corresponds to an average over 200 simulation experiments.

In Figure 1, first consider the curve corresponding to $\beta = 0$. For $\alpha = 0$, we do not observe almost any stable improvement cycles. This is not surprising, because in this case, the preferences are independent and schools are symmetric; therefore students' top choices are evenly

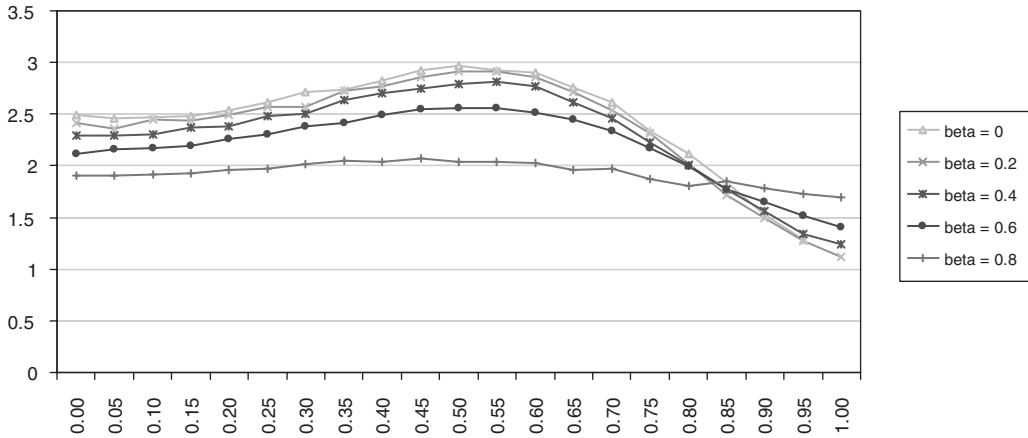


FIGURE 2. TOTAL STEPS UP/NUMBER OF IMPROVING STUDENTS AS A FUNCTION OF ALPHA
(1,000 students and 20 schools)

distributed across schools, leading to little conflict. As correlation in preferences increases, some schools are more in demand than others. As a result, how the ties are broken at these schools becomes critical, and we observe an increase in the number of students who can be improved via stable improvement cycles. Figure 1 also suggests that the efficiency loss peaks for some value of α , and starts decreasing thereafter. This is due to the fact that, for values of α close to 1, the preferences start becoming so similar that exchanging seats becomes less likely to be mutually profitable. In the extreme case when $\alpha = 1$, the preferences are identical and there is no stable improvement cycle at all.

Note that as β increases, the preferences of students who live close to each other become more similar and the preferences of those who live far from each other become less similar. Hence, β can be thought of as adding a positive correlation in the preferences of students who live close to each other, and a negative correlation in the preferences of those who live far from each other. Since priorities are also based on location, an increase in β increases the correlation in the preferences of students within the same priority class and decreases the correlation in the preferences of students within different priority classes. Figure 1 suggests a direction in which these two effects systematically play out: the single-peaked pattern in α shifts “southwest” (i.e., the peak of the graph drops and it is reached for lower values of α) as β increases.

When $\beta = 1$, students’ preferences are determined solely by locational proximity. In this case, note that in Figure 1, the percentage improvement is identically zero for every value of α . The reason is that when preference for proximity is extreme, for each school x the students in the walk zone of x top rank x . It is possible to show that this fact, along with the assumption that priorities have two indifference classes, implies that RDA is *unconstrained* Pareto efficient; hence there is no room for Pareto improvement.¹²

Figure 2 reports the average number of steps up the preference lists of the improving students, which is seen to vary between one and three in the simulations above. For a robustness check, we ran 12 simulations analogous to those that generated Figures 1 and 2, where the number of students was 200, 400, 1,000, and 2,000, and the number of schools was 10, 20, and 40. In these simulations, there were no qualitative changes in the observations above.

¹² Note, however, that when $\beta = 1$, it might not be possible to assign everyone to her district school since a school could have more students in its walk zone than the number of students it can accommodate.

IV. Strategic Properties

A single-valued mechanism f is **strategy-proof** if, for any preference profile R , student i and R'_i , we have $f_i(R_i, R_{-i})R_i f_i(R'_i, R_{-i})$. We know from Dubins and Freedman (1981) and Roth (1982) that in the case of strict priorities, the constrained efficient mechanism f^{\succsim} is strategy-proof. When we allow the priority orders to be weak, the constrained efficient set is not necessarily a singleton. In this case, it is natural to ask whether there is a single-valued mechanism $f \subset f^{\succsim}$ that is strategy-proof. The following example gives a negative answer to this question.

Example 2. Consider a school choice problem with three schools, x, y, z , each having one seat, three students, 1, 2, 3, who find all schools acceptable, and

R_1	R_2	R_3	\succsim_x	\succsim_y	\succsim_z
y	y	x	1	3	3
z	z	y	2	1,2	2
x	x	z	3		1

The constrained efficient set consists of only two matchings:

$$f^{\succsim}(R) = \left\{ \begin{pmatrix} 1 & 2 & 3 \\ y & z & x \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ z & y & x \end{pmatrix} \right\}.$$

Consider the following manipulations:

R'_1	R'_2
y	y
x	x
z	z

If student 1 announces R'_1 when the other students announce truthfully, then

$$f^{\succsim}(R'_1, R_{-1}) = \left\{ \begin{pmatrix} 1 & 2 & 3 \\ y & z & x \end{pmatrix} \right\}.$$

Similarly, if student 2 announces R'_2 when the other students announce truthfully, then

$$f^{\succsim}(R'_2, R_{-2}) = \left\{ \begin{pmatrix} 1 & 2 & 3 \\ z & y & x \end{pmatrix} \right\}.$$

Consider any single-valued mechanism $f \subset f^{\succsim}$. For the preference profile R , f has to select one of the matchings, $(1y, 2z, 3x)$ or $(1z, 2y, 3x)$. If it selects $(1y, 2z, 3x)$, then student 2 has an incentive to misrepresent her preference and submit R'_2 . On the other hand, if it selects $(1z, 2y, 3x)$,

then student 1 has an incentive to misrepresent her preference and submit R'_1 . Therefore, f is not strategy-proof.

For each student i , our model specifies only an ordinal ranking R_i over $X \cup \{i\}$. Assuming that the student is an expected utility maximizer, we need to know her cardinal (vNM) utility function $u_i : X \cup \{i\} \rightarrow \mathbb{R}$ to fully specify her risk preferences. Given two probability distributions p and q over $X \cup \{i\}$, p [strictly] first-order stochastically dominates q with respect to R_i if

$$\sum_{y \in X \cup \{i\} : y R_i z} p(y) \geq \sum_{y \in X \cup \{i\} : y R_i z} q(y)$$

for all $z \in X \cup \{i\}$ [with strict inequality for some $z \in X \cup \{i\}$]. It is a standard fact that p [strictly] first-order stochastically dominates q with respect to R_i if and only if, for any vNM utility function u_i that gives the same ordinal ranking as R_i , the expected utility of p is [strictly] weakly more than the expected utility of q . Given a random mechanism F , a preference profile R , and a student i , let $F_i(R)$ denote the random allocation of i with respect to $F(R)$. The argument in Example 2 can be adapted to conclude that the impossibility above persists even for random mechanisms.

OBSERVATION 3: *Let N , X , and \succsim be as in Example 2 and let F be any random mechanism that gives a constrained-efficient allocation with probability one in each preference profile. Then there exist R , i , and R'_i , such that $F_i(R'_i, R_{-i})$ strictly first-order stochastically dominates $F_i(R_i, R_{-i})$ with respect to R_i .*

Hence, strategy-proofness and constrained efficiency are incompatible.¹³ In the example above, the strategic manipulation was aimed at ruling out the less preferred constrained efficient allocation and, consequently, singling out the preferred one. Could a student manipulate her submitted ranking to induce a new matching where she is assigned to a school preferable to every school she could possibly be assigned to under her truthful statement? It turns out that she cannot hope for more than her best possibility in the constrained efficient set.

THEOREM 2: *For any $\nu \in f^{\succsim}(R'_i, R_{-i})$, there is $\mu \in f^{\succsim}(R)$ such that $\mu(i) R_i \nu(i)$.*

Remember the random mechanism F^{\succsim} . Even when a student has perfect knowledge of the priority structure and the preferences of all students, since the algorithm involves random selections, there is uncertainty as to what outcome will be returned. The computation of the likelihood of a particular constrained efficient solution being returned is highly involved, and when faced with such uncertainty, what would an “optimist” do? Someone who tends to base her actions on her best assignment possible among the student-optimal solutions would consider manipulating only if such strategic announcement brought her a school preferable to any school she could be assigned under her truthful revelation. So Theorem 2 could be interpreted to conclude that *an optimist would not manipulate*. Moreover, if for a particular preference profile there is only one constrained efficient matching, then no student would have any incentives to unilaterally misstate her preferences.

¹³ A related result by Abdulkadiroğlu, Pathak, and Roth (2008) leads to the same conclusion: if f is the SOSM with respect to a fixed strict priority profile, then there is no strategy-proof mechanism that Pareto dominates f at every preference profile.

COROLLARY 1: *Given \succeq , let F be any random mechanism that returns a student optimal stable matching with probability one. Then, if R , i , and R'_i are such that $\mu(i)$ is the same for all $\mu \in f^{\succeq}(R_i, R_{-i})$, then $F_i(R_i, R_{-i})$ first-order stochastically dominates $F_i(R'_i, R_{-i})$ with respect to R_i . In particular, if $f^{\succeq}(R)$ is a singleton, then at R , truth-telling is a Nash equilibrium of the preference revelation game induced by F .*

We have seen in Example 2 that every selection from the SOSM was manipulable, but student 1 needed significant information regarding the preferences of students 2 and 3 in order to be able to correctly evaluate the consequences of her switching schools x and z in her preference list.¹⁴ One may ask if a student with low information about the preferences and priorities of others would find it profitable to employ such manipulation.

As a benchmark for a low information environment, we adopt the framework of Roth and Uriel Rothblum (1999). A student's beliefs about two schools x and y are **symmetric** if when one changes the roles of x and y in the random variable interpreted as her beliefs on (\succeq, R_{-i}) , the distribution of the random variable does not change.¹⁵ When this is the case, it turns out that under our random mechanism F^{\succeq} , it is not profitable for a student to misstate her preferences by switching those two schools in her preference.

THEOREM 3: *Assume that i has symmetric beliefs about schools x and y , and that R'_i is obtained from R_i by switching the places of x and y . Then, $F_i^{\succeq}(R_i, R_{-i})$ first-order stochastically dominates $F_i^{\succeq}(R'_i, R_{-i})$ with respect to R_i .*

In the firms-workers model of Roth and Rothblum (1999) with strict preferences on both sides, it was found that under the firm-proposing DA algorithm it may be profitable for a worker to submit a truncated preference, where a **truncation** of a preference list R_i containing r acceptable firms is a list R'_i containing $r' \leq r$ acceptable firms such that the r' firms in R'_i are the top r' in R_i with the same order. Since we are analyzing the SOSM, with strict priorities a truthful statement of a student would be her dominant strategy, ruling out any manipulation, including truncation strategies. It turns out that also in the case of weak priorities, truncation strategies are never profitable for students, independent of their beliefs about preferences and priorities of others.

THEOREM 4: *If R'_i is a truncation of R_i , then for any belief of i over (\succeq, R_{-i}) , $F_i^{\succeq}(R_i, R_{-i})$ first-order stochastically dominates $F_i^{\succeq}(R'_i, R_{-i})$ with respect to R_i .*

However, another set of strategies might emerge, even when the student has almost no information allowing her to distinguish between others' priorities and preferences. Formally, an **extension** of a preference list R_i containing r acceptable schools is a list R'_i containing $r' \geq r$ acceptable schools such that the r elements of R_i are the top r in R'_i with the same order. Under F^{\succeq} , manipulation by announcing an extension strategy may be profitable even under symmetric information. We next present an example illustrating this observation.

¹⁴ It is possible that a student may have an incentive to manipulate the mechanism F^{\succeq} under an incomplete information environment without having detailed information about the others' preferences. An example is when certain schools are commonly recognized as being popular, i.e., ex ante more likely to be highly ranked by the students. In that case, a student i who has high priority at a popular school x may find it profitable to lift school x in her submitted ranking. The rationale is that she may gain if she is temporarily assigned to x at step 0 of the algorithm and if she is able to "trade" x at subsequent stages of the algorithm. Such a manipulation would be profitable only if student i does not rank x very highly but has sufficient confidence on the popularity of x . Hence, one would expect the ex ante likelihood of this manipulation to be low.

¹⁵ See Appendix A3 for a precise definition.

Example 3. Consider three students, 1, 2, and 3, and two schools, x and y , each having one seat. Suppose that every student has equal priority for all schools. Student 1’s vNM preference is given by $u_1(y) = 1, u_1(x) = 0$, and $u_1(x) = -\varepsilon$ for some $\varepsilon > 0$; hence, her true ordinal preference R_1 is such that yP_1x . Her beliefs over $\succsim_x, \succsim_y, R_2$, and R_3 are independent and uniform over the respective domains; in particular, they are symmetric for x and y . Suppose that the random mechanism F^\succsim is being used and that student 1 is contemplating manipulating her true ranking and announcing the extension R'_1 such that $yP'_1xP'_1$.

Recall the algorithm corresponding to our random mechanism and fix a realization of \succsim, R_{-1} , and $\succsim' \in \mathcal{T}(\succsim)$. Conditional on (R_{-1}, \succsim') , if student 1 submits R_1 and the algorithm assigns her to y , then this assignment must have been reached in step 0 as a result of the DA algorithm being applied to (R_1, R_{-1}, \succsim') . In this case, if she submits R'_1 , the algorithm would again assign her to y in step 0 as a result of the DA algorithm being applied to $(R'_1, R_{-1}, \succsim')$. Therefore, student 1 can lose by announcing R'_1 instead of R_1 , only if the realization (R_{-1}, \succsim') is such that she is left unassigned if she announces R_1 . Before the realization of (R_{-1}, \succsim') , this expected loss is bounded above by ε from the point of view of student 1.

On the other hand, if the realization of (R_{-1}, \succsim') is such that

R_2	R_3	\succsim'_x	\succsim'_y
x	y	1	2
y	x	3	3
2	3	2	1

then student 1 is left unassigned if she submits R_1 and she is assigned to y if she submits R'_1 . Let $p > 0$ denote the probability of the realization above. If the student’s risk preferences are such that $\varepsilon < p$, then she will prefer to announce R'_1 when her true ordinal ranking is R_1 .

From the proofs of Theorems 3 and 4, it follows that the only profitable strategic manipulation in a low information environment is to lengthen one’s list. If, in addition, it is common knowledge that all schools are acceptable for all students, then being truthful is a Bayesian Nash equilibrium of the preference revelation game, under the random mechanism F^\succsim .¹⁶

V. Conclusion

Versions of the DA algorithm are being adopted in several school choice districts in the United States, and to our knowledge, all of them employ a random tie-breaking rule when faced with indifferences. We have pointed out that this may cause a significant loss of efficiency by returning a matching that is not in the student optimal stable set with respect to the original priority structure. We introduced the *stable improvement cycles algorithm* which Pareto improves upon a stable matching whenever it is possible to do so without violating priorities, and eventually returns an element of the student optimal stable set. A strategic analysis shows that there need not be a strategy-proof selection from the SOSM. A consequence is that there is a trade-off between efficiency and strategy-proofness. We were able to show that in the benchmark symmetric information environment, being truthful is a Bayesian Nash equilibrium. Developing an

¹⁶ We thank John Hatfield for suggesting this corollary.

understanding of informational environments approximating real life situations and the kind of manipulation strategies that arise in those contexts is an important open question.

APPENDIX

A1. Proof of Theorem 1

In the case of strict priorities, it is well known that the set of school seats filled is the same at every stable outcome, as is the set of students who are assigned seats. This is no longer the case for weak priorities. However, the following lemma implies that the conclusion holds between Pareto comparable stable matchings.

LEMMA 1: *Let \succeq and R be given. Suppose that μ is a stable matching that is Pareto dominated by a (not necessarily stable) matching ν . Let N' denote the set of students who are strictly better off under ν and let $X' = \mu(N')$ be the set of schools to which students in N' are assigned under μ . Then we have:*

- (i) *Students who are not in N' have the same match under μ and ν ;*
- (ii) *The number of students in N' who are assigned to a school x are the same in μ and ν ; in particular, $X' = \nu(N')$;*
- (iii) *Each student in N' is assigned to a school (i.e., not to herself) in μ and in ν .*

PROOF:

Part (i) follows from the facts that $i \in N \setminus N'$ is indifferent between $\mu(i)$ and $\nu(i)$, and her preferences are strict.

For part (ii), let us first show that $|N' \cap \mu^{-1}(x)| \geq |N' \cap \nu^{-1}(x)|$ for any school x . Suppose, to the contrary, that $|N' \cap \mu^{-1}(x)| < |N' \cap \nu^{-1}(x)|$ for some school x . Together with part (i), this implies that the number of students in N who are assigned to x in μ is less than the number of students who are assigned to x in ν . Hence, x must have empty seats under μ . However, for any $i \in N' \cap \nu^{-1}(x)$, $x = \nu(i)P_i\mu(i)$, that is, i desires x which has empty seats under μ , a contradiction to the stability of μ .

Now, suppose that the inequality $|N' \cap \mu^{-1}(x)| \geq |N' \cap \nu^{-1}(x)|$ holds strictly for some school x^* . Summing across all schools we have:

$$\sum_{x \in X} |N' \cap \mu^{-1}(x)| > \sum_{x \in X} |N' \cap \nu^{-1}(x)|.$$

In other words, the number of students in N' who are assigned to some school in μ is more than the number of students in N' who are assigned to some school in ν . Hence, there exists a student i in N' who is assigned to a school in μ , but not in ν . Since $i = \nu(i)P_i\mu(i)$, this contradicts the stability of μ .

We already know from part (ii) that:

$$|N'| \geq \sum_{x \in X} |N' \cap \mu^{-1}(x)| = \sum_{x \in X} |N' \cap \nu^{-1}(x)|.$$

To prove part (iii), we need only show that the inequality above cannot hold strictly. Suppose for a contradiction that:

$$|N'| > \sum_{x \in X} |N' \cap \mu^{-1}(x)| = \sum_{x \in X} |N' \cap \nu^{-1}(x)|.$$

Hence, there exists a student $i \in N'$ who is unmatched in ν . Note that i has to be matched in μ ; otherwise, she would be indifferent between μ and ν , a contradiction to her being in N' . But then $i = \nu(i)P_i\mu(i)$, a contradiction to the stability of μ .

PROOF OF THEOREM 1:¹⁷

Suppose that μ and ν are stable matchings and that ν Pareto dominates μ . As in Lemma 1, let N' denote the set of students who are strictly better off under ν . Let $X' = \mu(N')$ be the set of schools to which students in N' are assigned to under μ .

For any $x \in X'$, let N'_x denote the highest \succeq_x -priority students among those in N' that desire x at μ . N'_x is non-empty since x is desired by (at least) the students in N' who are assigned to x in ν . We know such students exist since $x \in X' = \nu(N')$ by part (ii) of the Lemma above.

Fix an arbitrary student $i_x \in N'_x$ and let the school $\mu(i_x)$ point to x .¹⁸ Note that $\mu(i_x) \in X'$ by part (iii) of the Lemma above and $\mu(i_x) \neq x$, since i_x desires x . We can repeat this for each school $x \in X'$ and fix a school $y \in X' \setminus \{x\}$ that points to x .

Since each school in X' is pointed to by a different school in X' , there exists a cycle of distinct schools $x_1, \dots, x_n \equiv x_0$ in X' , where x_ℓ points to $x_{\ell+1}$ for $\ell = 0, \dots, n - 1$ and $n \geq 2$. Let $i_\ell = i_{x_{\ell+1}}$ for $\ell = 0, \dots, n - 1$. Since x_ℓ points to $x_{\ell+1}$ and $i_\ell \equiv i_{x_{\ell+1}}$, we have $\mu(i_\ell) = x_\ell$. In particular, i_1, \dots, i_n are distinct and they are all assigned to a school. Moreover, by construction, $i_\ell \equiv i_{x_{\ell+1}}$ desires $\mu(i_{\ell+1}) = x_{\ell+1}$.

Let D_x denote the set of highest \succeq_x -priority students among those who desire x at μ . Let B_x denote the set of students who desire x and are strictly below D_x with respect to \succeq_x at μ . Note that $i_\ell \in D_{\mu(i_{\ell+1})} \cup B_{\mu(i_{\ell+1})}$, which implies that $D_{\mu(i_{\ell+1})}$ is nonempty. It remains to show that $i_\ell \in D_{\mu(i_{\ell+1})}$. Suppose for a contradiction that i_ℓ lies in $B_{\mu(i_{\ell+1})}$. By construction, i_ℓ is a highest $\succeq_{\mu(i_{\ell+1})}$ -priority student among those in N' who desire $\mu(i_{\ell+1})$; hence, $D_{\mu(i_{\ell+1})}$ has an empty intersection with N' .

Choose any $j \in D_{\mu(i_{\ell+1})}$. From above, $j \notin N'$ and $j \succ_{\mu(i_{\ell+1})} k$ for any $k \in N'$ who desires $\mu(i_{\ell+1})$. Since $\mu(i_{\ell+1}) \in X' = \nu(N')$ by part (ii) of the Lemma above, there is $k \in N'$ who is matched to $\mu(i_{\ell+1})$ under ν . In particular, k desires $\mu(i_{\ell+1})$ at μ ; hence $j \succ_{\mu(i_{\ell+1})} k$. Since j desires $\mu(i_{\ell+1})$ at μ and $\mu(j) = \nu(j)$ by part (i) of the Lemma above, she also desires $\mu(i_{\ell+1})$ at ν . This contradicts the stability of ν , since j has higher $\succeq_{\mu(i_{\ell+1})}$ -priority than k , who is matched to $\mu(i_{\ell+1})$ at ν .¹⁹

A2. Proof of Theorem 2

Let x be the best school that student i can be assigned to among her assignments from $f^{\succeq}(R)$. Suppose that there is a preference ordering R'_i such that $\nu \in f^{\succeq}(R'_i, R_{-i})$ and $\nu(i)P_i x$. We know that $\nu = f^{\succeq'}(R'_i, R_{-i})$ for some tie-breaking $\succeq' \in \mathcal{T}(\succeq)$. But in the case of strict priority orders,

¹⁷ In the following footnotes, we indicate the necessary modifications in the proof in order to guarantee that the new stable matching μ' , obtained from μ by applying the improvement cycle, lies weakly below ν in a Pareto sense.

¹⁸ In order to guarantee that the new stable matching μ' , obtained from μ by applying the improvement cycle, lies weakly below ν in a Pareto sense, we need to select i_x slightly more carefully: if there is a student $i \in N'_x$ who is assigned to x in ν , fix i_x to be one of those students; otherwise fix i_x arbitrarily.

¹⁹ To show that the new stable matching μ' , obtained from μ by applying the improvement cycle, may be guaranteed to lie weakly below ν in a Pareto sense, it is enough to prove that $\nu(i_\ell)R'_i\mu(i_{\ell+1})$ ($\ell = 0, \dots, n - 1$) under the more careful selection of i_x 's indicated in the footnote above. To prove this, remember that $x_{\ell+1} = \mu(i_{\ell+1})$, $i_{x_{\ell+1}} = i_\ell$, and suppose that $x_{\ell+1}P_{i_{x_{\ell+1}}}\nu(i_{x_{\ell+1}})$. Since $\nu(i_{x_{\ell+1}}) \neq x_{\ell+1}$, there must be no student in $N'_{x_{\ell+1}}$ who is assigned to $x_{\ell+1}$ at ν . Therefore, by construction of $N'_{x_{\ell+1}}$, any student $k \in N'$ who is assigned to $x_{\ell+1}$ at ν (and hence desires $x_{\ell+1}$ at μ) has strictly less $\succeq_{x_{\ell+1}}$ -priority than $i_{x_{\ell+1}}$. This contradicts the stability of ν , since $i_{x_{\ell+1}}$ desires $x_{\ell+1}$ at ν and $\nu(k) = x_{\ell+1}$. (The existence of such a k is again guaranteed since $x_{\ell+1} \in X' = \nu(N')$ by part (ii) of the Lemma.)

we also know that the only constrained efficient assignment is the student optimal one, and it is a dominant strategy to reveal one's preferences truthfully. Hence, if $\mu' = f^{\succeq'}(R)$, then

$$\mu'(i)R_i\nu(i)P_ix.$$

Since x is the best school that student i could expect from the assignments in $f^{\succeq}(R)$, μ' must not be constrained efficient with respect to R and \succeq . It is stable since it is the outcome of the DA algorithm applied to a tie-breaking \succeq' . Therefore it must be Pareto dominated by an assignment $\nu' \in f^{\succeq}(R)$ which implies $xR_i\nu'(i)R_i\mu'(i)$ leading to a contradiction with our earlier conclusion $\mu'(i)P_ix$.

A3. Proof of Theorem 3

We will use the shorthand M for a pair (\succeq, R) . We continue to suppress the quota q_x of each school x within the notation \succeq_x . For any two schools x and y , let $R^{x \leftrightarrow y}$ denote the preference profile obtained from R by switching x and y in each student's preference. Similarly, let $\succeq^{x \leftrightarrow y}$ denote the priority structure obtained from \succeq by switching the quotas and priority orders of x and y . Let $M^{x \leftrightarrow y}$ stand for $(\succeq^{x \leftrightarrow y}, R^{x \leftrightarrow y})$. Then, $R_i^{x \leftrightarrow y}$ is the preference order of student i obtained from R_i by switching x and y , and $M_{-i}^{x \leftrightarrow y} = (\succeq^{x \leftrightarrow y}, R_{-i}^{x \leftrightarrow y})$ is obtained from M_{-i} by switching the roles of x and y in \succeq and in the preferences of all students except i .

We summarize the beliefs of student i about the stated preferences and priorities of the others by a random variable \tilde{M}_{-i} . For any two schools x and y , we say that \tilde{M}_{-i} is $\{x, y\}$ -symmetric if the distributions of \tilde{M}_{-i} and $\tilde{M}_{-i}^{x \leftrightarrow y}$ coincide. This assumption can be interpreted as saying that student i 's beliefs cannot distinguish between these two schools, from the preferences and priorities of the other students. The $\{x, y\}$ -symmetry assumption does not necessarily imply that student i has no information concerning x and y ; for instance, it is consistent with beliefs where certain students always prefer x and y to a third school z , or where with a certain probability student i has higher priority than another student j for both x and y , and so on.

Given \tilde{M}_{-i} and R_i , the probability that student i is matched to $x \in X \cup \{i\}$ is

$$p_i(R_i, \tilde{M}_{-i})(x) \equiv \sum \Pr\{\tilde{M}_{-i} = (\succeq, R_{-i})\} \Pr\{F_i^{\succeq}(R_i, R_{-i}) = x\},$$

where the sum is taken across all possible realizations (\succeq, R_{-i}) of \tilde{M}_{-i} . The following lemma states that if a student i prefers school y to x , regardless of M_{-i} , her probability of receiving y when she announces $R_i^{x \leftrightarrow y}$ is less than or equal to her probability of receiving y when she is truthful. Symmetrically, her probability of receiving x when she announces $R_i^{x \leftrightarrow y}$ is at least as much as her probability of receiving x when she is truthful.²⁰

LEMMA 2: If $x, y \in X$ and yP_ix , then

- (i) $p_i(R_i^{x \leftrightarrow y}, M_{-i})(y) \leq p_i(R_i, M_{-i})(y)$;
- (ii) $p_i(R_i^{x \leftrightarrow y}, M_{-i})(x) \geq p_i(R_i, M_{-i})(x)$.

²⁰ Theorems 3 and 4 are stated for the random mechanism F^{\succeq} . Remember that in the stable improvement cycles algorithm leading to F^{\succeq} , we have assumed that a selection is made independently and randomly with equal probabilities each time the algorithm faces a multiplicity. In proving these results, we use only the facts that these randomizations are made *independently* and *anonymously*, i.e., they do not depend on the particular labeling of students and schools. For instance, the same conclusions hold for the random mechanism G^{\succeq} , which differs from F^{\succeq} only in that the algorithm leading to it selects a cycle in step (t.b) independently and randomly with equal probabilities from among the *largest* stable improvement cycles.

PROOF:

We start by proving part (i). Recall that the algorithm associated with F^{\succeq} starts with the selection of a tie-breaking of \succeq . For a fixed tie-breaking $\succeq' \in \mathcal{T}(\succeq)$, we will compare, for R_i and $R_i^{x \leftrightarrow y}$, the probabilities with which the rest of the algorithm will return an assignment μ with $\mu(i) = y$. Let $\mu^0 = f^{\succeq'}(R)$ and $R'_i = R_i^{x \leftrightarrow y}$. We consider four cases depending on $\mu^0(i)$.

If $\mu^0(i) = z$, where zP_iyP_ix , then $f^{\succeq'}(R'_i, R_{-i})(i) = z$ due to well known properties of the DA algorithm for strict priority structures. Since the mechanism will be employing stable improving cycles dominating μ^0 , her preferences regarding schools less preferable than z will not matter at all. In particular, in this case her probability of obtaining y will be zero regardless of her switching or not switching x and y in her preference.

If $\mu^0(i) = x$, then it is not hard to see again that $f^{\succeq'}(R'_i, R_{-i})(i) = \mu^0(i) = x$. If she announces $R'_i = R_i^{x \leftrightarrow y}$, she will never obtain y by participating in a stable improvement cycle in steps $t \geq 1$ of our algorithm, since xP_iy . Therefore, her probability of eventually being assigned to y in this case is zero if she announces $R_i^{x \leftrightarrow y}$.

If $\mu^0(i) = z$ where yR_izP_ix , since the outcome of the mechanism will be a weak Pareto improvement of μ^0 with respect to the announced preferences, by announcing $R_i^{x \leftrightarrow y}$, her probability of getting school y will be zero.

Last, let $\mu^0(i) = z$ where yP_ixP_iz . At step 0 of the algorithm, the pictures are identical for her announcing R_i and $R_i^{x \leftrightarrow y}$ since she prefers both x and y to z . Similarly, at any stage of the updating of the directed graph, the pictures look alike when the temporary assignment at that stage is less preferable than both x and y . Therefore, for all those steps, the probabilities of the next step are identical for R_i and $R_i^{x \leftrightarrow y}$. Those probabilities would matter when, at a temporary assignment, she is assigned to a school w such that yR_iwR_ix . For that very step, announcing $R_i^{x \leftrightarrow y}$ would equal her probability of obtaining y to zero, as she would not appear to be desiring y anymore. And, finally, if her temporary assignment were to be wP_iy , it would not matter if she announced R_i or $R_i^{x \leftrightarrow y}$.

This argument is independent of $\succeq' \in \mathcal{T}(\succeq)$, and we have shown for each tie-breaking that her probability of receiving y by announcing $R_i^{x \leftrightarrow y}$ is less than or equal to what it would be when she announces R_i . Therefore, part (i) follows. To see part (ii), let $M' = (R_i^{x \leftrightarrow y}, M_{-i})$. Since xP_iy , by part (i),

$$p_i(R_i^{x \leftrightarrow y}, M'_{-i})(x) \leq p_i(M')(x),$$

which could be rewritten as (ii).

LEMMA 3: For any distinct $x, y, z \in X$ and R_i , if \tilde{M}_{-i} is $\{x, y\}$ -symmetric, then

$$p_i(R_i^{x \leftrightarrow y}, \tilde{M}_{-i})(z) = p_i(R_i, \tilde{M}_{-i})(z).$$

PROOF:

By the anonymous specification of the random mechanism F^{\succeq} , for any fixed M , we have

$$p_i(M^{x \leftrightarrow y})(z) = p_i(M)(z).$$

Hence, if we apply the identity above to the profile $(R_i^{x \leftrightarrow y}, M_{-i})$, we obtain

$$p_i(R_i^{x \leftrightarrow y}, M_{-i})(z) = p_i(R_i, M_{-i}^{x \leftrightarrow y})(z).$$

By $\{x, y\}$ -symmetry of beliefs, integrating each side over possible realizations of \tilde{M}_{-i} yields the desired equality.

PROOF OF THEOREM 3:

The proof follows from Lemma 2 and Lemma 3 above.

A4. Proof of Theorem 4

Suppose that x is the school at the bottom of her truncated preference list R'_i . We will prove that, for any school yR_ix ,

$$\Pr\{i \text{ receives } y | (R'_i, M_{-i})\} \leq \Pr\{i \text{ receives } y | (R_i, M_{-i})\},$$

which in turn would imply the desired stochastic dominance. As in the proof of Lemma 2, we will prove the inequality above conditional on a fixed tie-breaking selected at the beginning of the mechanism. As our argument will be independent of which tie-breaking is chosen, integrating over all possible tie-breakings would yield the inequality.

Let \succsim' be the tie-breaking, and let $f^{\succsim'}(R) = \mu^0$ and $f^{\succsim'}(R'_i, R_{-i}) = \nu^0$. If $\nu^0(i) = zR_ix$, by the DA algorithm, $\nu^0 = \mu^0$ and in particular $\mu^0(i) = z$. So the rest of our algorithm looks identical starting from μ^0 or ν^0 , implying

$$\Pr\{i \text{ receives } y | (R'_i, M_{-i}), \succsim' \text{ selected}\} = \Pr\{i \text{ receives } y | (R_i, M_{-i}), \succsim' \text{ selected}\}.$$

If $\mu^0(i) = z$ where xP_iz , then $\nu^0(i) = i$; that is, i will not be assigned to any school under ν^0 . By part (ii) of Lemma 1, she will continue to be assigned to herself for any stable improvement of ν^0 ; therefore, her probability of receiving any school is zero, verifying the desired inequality trivially.

REFERENCES

- Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth.** 2005. "The New York City High School Match." *American Economic Review*, 95(2): 364–67.
- Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth.** 2008. "Strategy-Proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match." <http://kuznets.fas.harvard.edu/~aroth/papers/MatchingWithIndifferences.NYC.pdf>.
- Abdulkadiroğlu, Atila, Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez.** 2005. "The Boston Public School Match." *American Economic Review*, 95(2): 368–71.
- Abdulkadiroğlu, Atila, Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez.** 2006. "Changing the Boston School Mechanism: Strategy-proofness as Equal Access." <http://kuznets.fas.harvard.edu/~aroth/papers/bostonMay182006.pdf>.
- Abdulkadiroğlu, Atila, and Tayfun Sönmez.** 2003. "School Choice: A Mechanism Design Approach." *American Economic Review*, 93(3): 729–47.
- Balinski, Michel, and Tayfun Sönmez.** 1999. "A Tale of Two Mechanisms: Student Placement." *Journal of Economic Theory*, 84(1): 73–94.
- Chen, Yan, and Tayfun Sönmez.** 2006. "School Choice: An Experimental Study." *Journal of Economic Theory*, 127(1): 202–31.
- Cormen, Thomas H., Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein.** 2001. *Introduction to Algorithms*. 2nd ed. Cambridge, MA: MIT Press.
- Dubins, Lester E., and David A. Freedman.** 1981. "Machiavelli and the Gale-Shapley Algorithm." *American Mathematical Monthly*, 88(7): 485–94.
- Ehlers, Lars.** 2006. "Respecting Priorities when Assigning Students to Schools." University of Montréal Department of Economics Working Paper 2006–04.
- Erdil, Aytekin.** 2008. "Efficient Assignment Based on Coarse Priorities." Unpublished.
- Erdil, Aytekin, and Haluk Ergin.** 2006. "Two-sided Matching with Indifferences." Unpublished.
- Ergin, Haluk.** 2002. "Efficient Resource Allocation on the Basis of Priorities." *Econometrica*, 70(6): 2489–97.

- Ergin, Haluk, and Tayfun Sönmez.** 2006. "Games of School Choice under the Boston Mechanism." *Journal of Public Economics*, 90(1–2): 215–37.
- Gale, David, and Lloyd S. Shapley.** 1962. "College Admissions and the Stability of Marriage." *American Mathematical Monthly*, 69(1): 9–15.
- Halldórsson, Magnús M., Robert W. Irving, Kazuo Iwama, David F. Manlove, Shuichi Miyazaki, Yasufumi Morita, and Sandy Scott.** 2003. "Approximability Results for Stable Marriage Problems with Ties." *Theoretical Computer Science*, 306(1–3): 431–47.
- Kesten, Onur.** 2005. "Student Placement to Public Schools in the US: Two New Solutions." Unpublished.
- Kesten, Onur.** 2006. "On Two Competing Mechanisms for Priority-Based Allocation Problems." *Journal of Economic Theory*, 127(1): 155–71.
- Manlove, David F., Robert W. Irving, Kazuo Iwama, Shuichi Miyazaki, and Yasufumi Morita.** 2002. "Hard Variants of the Stable Marriage." *Theoretical Computer Science*, 276(1–2): 261–79.
- Pápai, Szilvia.** 2000. "Strategyproof Assignment by Hierarchical Exchange." *Econometrica*, 68(6): 1403–33.
- Roth, Alvin E.** 1982. "The Economics of Matching: Stability and Incentives." *Mathematics of Operations Research*, 7(4): 617–28.
- Roth, Alvin E.** 1984. "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory." *Journal of Political Economy*, 92(6): 991–1016.
- Roth, Alvin E.** 2002. "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics." *Econometrica*, 70(4): 1341–78.
- Roth, Alvin E.** 2003. "The Origins, History, and Design of the Resident Match." *Journal of the American Medical Association*, 289(7): 909–12.
- Roth, Alvin E., and Elliott Peranson.** 1999. "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design." *American Economic Review*, 89(4): 748–80.
- Roth, Alvin E., and Uriel G. Rothblum.** 1999. "Truncation Strategies in Matching Markets—In Search of Advice for Participants." *Econometrica*, 67(1): 21–43.
- Roth, Alvin E., and Marilda A. Oliveira Sotomayor.** 1990. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. New York: Cambridge University Press.
- Shapley, Lloyd S., and Herbert Scarf.** 1974. "On Cores and Indivisibility." *Journal of Mathematical Economics*, 1: 23–28.

This article has been cited by:

1. Abdulkadiroğlu Atila, Pathak Parag A., Roth Alvin E.. 2009. Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match. *American Economic Review* **99**:5, 1954-1978. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
2. Y. Stephen Chiu, Weiwei Weng. 2009. Endogenous preferential treatment in centralized admissions. *The RAND Journal of Economics* **40**:2, 258-282. [[CrossRef](#)]
3. Fuhito Kojima, , Parag A. Pathak. 2009. Incentives and Stability in Large Two-Sided Matching Markets. *American Economic Review* **99**:3, 608-627. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]