07-076

Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match

Atila Abdulkadiroğlu
Parag A. Pathak
Alvin E. Roth

Copyright © 2007 by Atila Abdulkadiroğlu, Parag A. Pathak, and Alvin E. Roth

Working papers are in draft form. This working paper is distributed for purposes of comment and discussion only. It may not be reproduced without permission of the copyright holder. Copies of working papers are available from the author.

Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match*

Atila Abdulkadiroğlu† Parag A. Pathak‡ Alvin E. Roth§

This version: December 15, 2006

Abstract

The design of the New York City (NYC) High School match involved tradeoffs between incentives and efficiency, because some schools are strategic players that rank students in order of preference, while others order students based on large priority classes. Therefore it is desirable for a mechanism to produce stable matchings (to avoid giving the strategic players incentives to circumvent the match), but is also necessary to use tie-breaking for schools whose capacity is sufficient to accommodate some but not all students of a given priority class. We analyze a model that encompasses one-sided and two-sided matching models. We first observe that breaking indifferences the same way at every school is sufficient to produce the set of student optimal stable matchings. Our main theoretical result is that a student-proposing deferred acceptance mechanism that breaks indifferences the same way at every school is not dominated by any other mechanism that is strategy-proof for students. Finally, using data from the recent redesign of the NYC High School match, which places approximately 90,000 students per year, we document that the extent of potential efficiency loss is substantial. Over 6,800 student applicants in the main round of assignment could have improved their assignment in a (non strategy-proof) student optimal mechanism, if the same student preferences would have been revealed.

*We thank Susan Athey, Neil Dorosin, Aytek Erdil, Fuhito Kojima and Tayfun Sönmez for helpful comments, and participants of the Industrial Organization seminar at Harvard University. For financial support, Abdulkadiroğlu thanks the Alfred P. Sloan Foundation and NSF, Pathak thanks the Spencer Foundation, and Roth thanks the NSF. Aytek Erdil suggested a simplification of the proof of our Theorem 1.

†Duke University, email: atila.abdulkadiroglu@duke.edu, address: Department of Economics, Duke University, Durham NC, 27708.

‡Harvard University, email: ppathak@fas.harvard.edu, address: Department of Economics, Harvard University, Cambridge, MA 02138 and Harvard Business School, Boston, MA 02163.

§Harvard University, email: aroth@hbs.edu, address: Department of Economics, Harvard University, Cambridge, MA 02138 and Harvard Business School, Boston, MA 02163.

1 Introduction

In 2003-04, the authors of this paper assisted the New York City Department of Education (NYCDOE) in redesigning the student assignment mechanism used to match over 90,000 entering students to public high schools each year (Abdulkadiroğlu, Pathak and Roth 2005). A largely decentralized and congested assignment process with counterproductive incentives was replaced with a more centralized, uncongested procedure, with more straightforward incentives, based on a student-proposing deferred acceptance algorithm adapted to satisfy various constraints of the NYCDOE.[1]

School choice in New York, and in other cities including Boston (in which a new design was implemented in 2006, see Abdulkadiroğlu, Pathak, Roth, and Sönmez 2005, 2006), requires extensions of the standard models and results of matching theory found, for example, in Roth and Sotomayor (1990). Some school choice environments are *one-sided* in the sense that only the students are strategic players whose welfare and incentives need to be considered (while the school places are simply objects to be allocated, cf. Abdulkadiroğlu and Sönmez 2003). Other school choice environments are traditional two-sided matching markets. The situation in New York City is a hybrid: some schools are active strategic players that rank their students in terms of preferences, others have no preferences, and others fall in between (see Appendix 1). There are incentive and welfare considerations that affect schools as well as students.

In particular, student i and school s form a *blocking pair* if i prefers school s to the school she is assigned, and school s ranks student i ahead of another student who is assigned to school s. A matching without blocking pairs is a *stable* matching. When schools actively rank students, if there is a blocking pair, the school has an incentive to circumvent the match to enroll the students it would prefer. This was an important feature of the old system in NYC, in which some schools concealed capacity in an effort to be matched later with preferable students.[2]

Thus when schools actively rank students, it is important to obtain a stable matching, i.e. a matching such that there are no blocking pairs.[3] This was one of two principal objectives of the NYC school match design. The second was that the school match

[1] As discussed in our 2005 paper, some of these constraints, like the one limiting each student to apply to no more than 12 schools, interfered with the dominant strategy properties of the unconstrained mechanism for a minority of students. In our present theoretical treatment, we will consider the design problem without these idiosyncratic constraints.

[2] E.g. the Deputy Chancellor of Schools, quoted in the New York Times (11/19/04): "Before you might have a situation where a school was going to take 100 new children for 9th grade, they might have declared only 40 seats, and then placed the other 60 outside the process."

[3] Stable matchings may be relevant even when all schools are passive. In this case, stable matchings eliminate "justified envy." See Abdulkadiroğlu and Sönmez (2003) and Abdulkadiroğlu, Pathak, Roth and Sönmez (2006) for discussion in the context of Boston's new student assignment mechanism.

process should be *strategy-proof* for students, i.e. that it should be a dominant strategy for students to state their true preferences.

In two-sided matching problems with strict preferences, stable matchings are efficient, and there is a unique stable matching that gives every student the best match she can get at any stable matching. This stable matching is traditionally called *the* student-optimal matching, and is the outcome produced by the student-proposing deferred acceptance algorithm, which is strategy-proof for students. Thus, if all preferences were strict, there is an off-the-shelf solution (already applied in other matching problems such as for medical residents) that would achieve both objectives.[4]

However, a primary feature of school choice problems is that there are indifferences–ties–in how students are ordered by at least some schools. When school preferences involve indifferences, there is a non-empty set of stable matchings that are weakly Pareto optimal for the students. It will be useful in what follows to call any member of this set *a* student optimal stable matching.[5]

When many students are equivalent from the point of view of a school with limited space, an assignment mechanism must include a tie-breaking procedure. Random tie breaking (by assigning each student a lottery number to be compared when students are otherwise tied) preserves the ex ante equivalence of the students, and also preserves the strategy-proofness of the student proposing deferred acceptance algorithm. But random tie breaking introduces artificial stability constraints (since, after random tie breaking, schools appear to have strict preferences between students for whom they are indifferent), and these constraints can harm student welfare.[6] In other words, when the deferred acceptance algorithm is applied to the strict preferences that result from tie breaking, the outcome it produces may not in fact be a student optimal stable matching in terms of the original preferences, i.e. there may be other stable matchings that weakly Pareto dominate this outcome from the point of view of student welfare. That is, there may be other stable matchings with respect to the original preferences at which some students are better off, and no students are worse off.

Nevertheless, some forms of random tie breaking may be preferable to others. One of the first design decisions we were confronted with was whether to assign lottery numbers to each student at each school (multiple random tie breaking), or to conduct a single

[4]See Roth and Peranson (1998) on the current design of the medical resident match, and Gale and Shapley (1962) and Roth (1982, 1985) on the basic optimality and incentive results. The NYCDOE initially contacted us because they were familiar with the design of the medical match.

[5]See Milgrom (2006) for a similar new treatment of the core for auctions with non-transferable utility, and its relation to matching.

[6]From this point on, we will mostly be concerned with student welfare and incentives, and we will speak about Pareto efficiency and strategy-proofness with respect to students only. In the conclusion, we will revisit the larger welfare and incentive issues that are implicitly also included when we deal with stable matchings.

lottery among all students that would give each student a lottery number to be used for tie breaking at every school (single random tie breaking). Computations with simulated and then actual submitted preferences indicated that single tie breaking had superior welfare properties.[7] (The computations based on the revealed preferences will be discussed in Section 5, and Table 1.) The first theorem in this paper provides some theoretical insight into this difference. We show that for every student optimal stable matching there exist single-tie-breaking rules from which that matching can be obtained by the deferred acceptance algorithm, i.e. every student optimal stable matching can be obtained with single tie breaking. Therefore none of the additional outcomes that are sometimes produced by the deferred acceptance algorithm with multiple tie breaking (but not by single tie breaking) are student-optimal.[8]

However, single tie breaking can also lead to a matching which is not student-optimal. Thus there will sometimes be a potential opportunity to improve on the outcome of deferred acceptance with single tie breaking. Theorem 2, our main theoretical result, shows that there exists no strategy-proof mechanism (stable or not) that Pareto improves on the deferred acceptance algorithm with single tie breaking (even when Pareto improvements are with respect to students only).

This theorem has as corollaries two known results. The first (due to Erdil and Ergin, forthcoming) is that, contrary to the case when all preferences are strict, no strategy-proof mechanism exists that always produces a student optimal stable matching. The second is that neither serial dictatorship nor top trading cycles dominates deferred acceptance with tie breaking (since they are both strategy-proof), although neither one is dominated by deferred acceptance, since both are Pareto efficient for students. Thus, there is a tradeoff between strategy-proofness and efficiency, a tradeoff that hinges on stability. Deferred acceptance with tie breaking, a stable mechanism, is on the efficient frontier of all strategy-proof mechanisms.

This theorem does not speak to the magnitude of the tradeoff to expect in real school choice plans. We take a step towards investigating this question using data on student preferences from 2003-04 in New York City. Using the algorithm recently developed by Ergin and Erdil (forthcoming) to find student optimal matchings (starting from true preferences and the outcome of the deferred acceptance algorithm with single tie breaking), we find that the apparent efficiency loss is considerable in New York. More than 6,800

[7] Tayfun Sönmez also played an important role in some of the early discussions of this.

[8] That the manner of tie breaking has important consequences for the outcome of deferred acceptance algorithms (when stability is an issue) contrasts in a surprising way with the case of one-sided matching. Pathak (2006) shows the strategy-proof top trading cycles mechanism that produces efficient (but not stable) matchings is unaffected by the choice of multiple versus single tie breaking, when all schools are indifferent between students.

students could be matched to schools they prefer to their assignment from deferred acceptance, without harming any others, if the same preference information could be elicited. This contrasts with the similar exercise on data from Boston's new assignment system in 2005-06 (Abdulkadiroğlu, Pathak, Roth and Sönmez 2006), in which no Pareto improvements at all could be made for students. This raises a number of new questions that we discuss in the conclusion concerning when potential efficiency gains exist, and whether they can be realized.

The present paper thus has three goals. It presents some new models and theory, suitable to the kinds of hybrid matching problems that arise in school choice, in which indifferences are ubiquitous. It presents some new empirical results concerning the magnitudes of the tradeoffs we discover. And it tells an aspect of a design story that, as economists gain experience in design, has become increasingly familiar, namely that the solution to design problems often requires extensions of existing theory (cf Roth 2002).

The next section describes how the design of a school matching mechanism is different from closely related matching models and results that have been usefully applied to aspects of the design of related matching markets. Section 3 introduces the matching model with active and passive schools and reviews some basic properties of stable matchings in this domain. Section 4 presents the two theoretical results. In Section 5 we empirically compare the outcome from deferred acceptance with single tie breaking to deferred acceptance with multiple tie breaking, and to a student optimal matching, using preference data from New York City's high school match, and compare this to the similar computation for Boston.

2 Related literature

The *school choice* problem, studied as a mechanism design problem by Abdulkadiroğlu and Sönmez (2003), is closely related to *two-sided* matching and *one-sided* matching. In the models we will consider, matching is many-to-one, since schools match with many students, but students only match with one school.

In *two-sided* matching models (Gale and Shapley 1962), there are two disjoint sets of agents, and every agent is an active strategic player with preferences over the agents in the other set.[9] A matching of agents on one side to agents on the other (that respects the relevant capacity constraints) is stable if it is individually rational and there is no pair of agents who each prefer to be assigned to one another than to their allocation in the

[9] Gale and Shapley (1962) phrased their discussion in a school choice context as a "college admissions" problem and studied the set of stable matchings. Balinski and Sönmez (1999) developed the theory of two-sided matching in relation to college admissions where ordering of students at colleges are determined via students' scores at an entrance exam.

matching. Stable matchings are Pareto optimal with respect to the set of all agents, and in the core of the market whose rules are that any pair of agents on opposite sides of the market may be matched (subject to capacity constraints) if and only if they both agree (Roth and Sotomayor 1990). Empirical observation suggests that centralized matching mechanisms in labor markets are most often successful if they produce stable matchings (Roth 1984, 1990, 1991; Roth and Xing 1994; Roth and Rothblum 1999).

When preferences are strict, Gale and Shapley (1962) showed that a class of *deferred acceptance algorithms* identify a stable match that is optimal for one or the other sides of the market, in the sense that every agent on one side of the market does at least as well at their optimal stable matching as at any other stable matching. Roth (1982a, 1985) showed that there does not exist any stable mechanisms that are strategy-proof for all agents (and in particular, none are strategy-proof for schools matched to more than one student), but the mechanism that produces the student optimal stable matching is strategy-proof for the students, in the sense that it is a dominant strategy for each student to state his true preferences. When preferences are not strict, there will not in general exist a unique stable match that is weakly Pareto optimal for each side of the market, rather there will a non-empty set of stable matches that are weakly Pareto optimal for agents on that side of the market.

In contrast, *one-sided* matching problems consist of a set of agents and a set of objects to be matched with agents (Shapley and Scarf 1974, Hylland and Zeckhauser 1979). Pareto efficient mechanisms have been developed and analyzed by numerous authors (Roth and Postlewaite 1977, Roth 1982b, Ma 1994, Abdulkadiroğlu and Sönmez 1999, Papai 2000). For these kinds of problems, the top trading cycles mechanism introduced in Shapley and Scarf (1974, and attributed to David Gale) is strategy-proof as a direct mechanism (Roth 1982b).[10]

As in other areas of practical market design (cf. Roth 2002) the design of actual school choice systems raises new theoretical questions. Several papers have been written about Boston's school choice system in which all schools are passive: see Abdulkadiroğlu, Pathak, Roth and Sönmez (2005, 2006), Abdulkadiroğlu and Yasuda (2006), Chen and Sönmez (2006), Ergin and Sönmez (2006) and Pathak and Sönmez (2006). New York City motivates the study of a different set of issues related to the fact that some schools are active strategic agents, and some are passive.

This paper is most closely related to Erdil and Ergin (forthcoming), who, in an important contribution that is complementary and contemporaneous to ours, also analyze indifferences in school choice problems.[11] They present a polynomial-time algorithm,

[10] Top trading cycles played a prominent initial role in the proposal for designing clearinghouses for kidney exchange, see Roth, Sönmez and Ünver (2004). Subsequent theoretical developments were necessitated by some of the practical constraints of kidney exchange, see Roth, Sönmez and Ünver (2005a, 2005b, forthcoming).

[11] Much of the attention in the prior literature to indifferences has been in the computer science literature, with

stable-improvement cycles, for the computation of a student-optimal matching when priorities are weak orders. The idea is, starting from a stable matching, to improve it through a sequence of trades among students and schools such that after each trade in the sequence, the resulting matching is still stable. We will revisit their proposal in section 5. They also present an impossibility theorem which demonstrates in an economy with four or more students and at least three schools, there is no strategy-proof mechanism which always selects a student optimal stable matching. In this context, our main theoretical result implies that student proposing deferred acceptance with tie breaking is optimal within the class of stable and strategy-proof mechanisms. Finally, Erdil and Ergin (2006) analyze a two-sided matching model with indifferences on both sides of the market and extend the algorithm in Erdil and Ergin (forthcoming) to construct Pareto efficient matchings with respect to both sides.

3 Model

We consider a model in which a school may have a strict order of students, be indifferent among all students, or partition students into indifference classes, with each class holding one or more students. Stability continues to be a key notion, with blocking pairs always involving strict preferences for both parties.

A matching problem consists of a finite set of students (individuals) I and a finite set of schools S. The number of available seats, i.e. the capacity at $s \in S$ is q_s and $q = (q_s)_{s \in S}$ is the list of capacities. Every student $i \in I$ has a strict preference relation P_i over $S \cup \{i\}$. Let sR_is' if sP_is' or $s = s'$. Every school s has a weak preference relation R_s over $I \cup \{s\}$. Let $\succ_s$ and $\sim_s$ represent the asymmetric and symmetric parts of R_s, respectively. To simplify the analysis, we assume that either $i \succ_s s$ or $s \succ_s i$ but not $i \sim_s s$. The preference relation of a school s over subsets of students is *responsive* to R_s, i.e. a school's preferences over *groups* of students is such that, for any group of students T with $|T| < q_s$, the school prefers $T \cup i$ to $T \cup j$ if and only if $i \succ_s j$, and prefers $T \cup i$ to T if and only if $i \succ_s s$ (Roth 1985). Let $P_{I'} = (P_i)_{i \in I' \subset I}$ and $P_{-I'} = (P_i)_{i \in I - I'}$. We define $R_{S'}$ and $R_{-S'}$ similarly. We fix I, S and q throughout the paper. Therefore a problem or preference profile is given by (P_I, R_S).

When school preferences are strict, this model reduces to the college admissions model (Gale and Shapley 1962, Roth 1985); when every school is indifferent among all students, it reduces to the house allocation model where there are potentially multiple places in each house (Hylland and Zeckhauser 1979). More importantly, the model allows both types of schools, as well as schools with several indifference classes simultaneously.

the conclusion that many issues that are computationally simple with strict preferences become computationally complex in the presence of indifferences (see, for instance, Irving (1994)).

We say $s \in S$ is **acceptable for** $i \in I$ if $sP_i i$. Remaining unmatched is also acceptable for i. Similarly $i \in I$ is **acceptable for** $s \in S$ if $i \succ_s s$. If x is not acceptable for y, then x is **unacceptable** for y.

A **matching** is a correspondence $\mu : S \cup I \to S \cup I$ such that

i) $|\mu(i)| = 1$ and $\mu(i) \subset S \cup \{i\}$ for each $i \in I$;

ii) $|\mu(s)| \leq q_s$ and $\mu(s) \subset I$ for each $s \in S$;

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

If $\mu(i) = i$, i remains unmatched. If $|\mu(s)| < q_s$, some seats at s remain unfilled. For some $X \subset I$, let $\mu(X) = \{s \in S | \mu(s) \in X\}$.

A matching μ is **individually rational** if it matches every $x \in I \cup S$ with agent(s) that is(are) acceptable for x. A matching μ is **blocked by** (i, s) if i and s are not matched to each other by μ, but both prefer to be, that is $\mu(i) \neq s$, $sP_i\mu(i)$, and either ($|\mu(s)| < q_s$ and $i \succ_s s$) or ($i \succ_s i'$ for some $i' \in \mu(s)$). μ is **stable** if μ is individually rational and not blocked by any student-school pair (i, s).

A stable matching μ is **student-optimal** if there is no other stable matching υ such that $\upsilon(i)R_i\mu(i)$ for all $i \in I$ and $\upsilon(i)P_i\mu(i)$ for some $i \in I$.[12]

A strict preference relation P_s (resulting from tie breaking) is **consistent** with a weak preference relation R_s if iP_sj implies iR_sj, iP_ss implies $i \succ_s s$ and sP_si implies $s \succ_s i$. A profile of strict preferences P_S is consistent with a profile of weak preferences R_S if P_s is consistent with R_s for every $s \in S$. A tie-breaking rule τ breaks indifferences/ties at schools for each problem (P_I, R_S) and produces P_S that is consistent with R_S. We use the terms indifference and ties interchangeably throughout the paper.

For a given profile of *strict* preferences, the unique student-optimal stable matching is obtained by the following **student-proposing deferred acceptance** algorithm (Gale and Shapley 1962):

Step 1: Each student proposes to her most preferred school. Each school tentatively assigns its seats to its proposers one at a time in the order of its preference. When all of its seats are tentatively assigned, it rejects all the proposers who remain unassigned.

In general, at

Step k: Each student who was rejected in the previous step proposes to her next preferred school. Each school considers the set of students it has been holding and its new proposers. It tentatively assigns its seats to these students one at a time in the order of its preference.

[12] Recall from the introduction that in the model we develop here, with indifferences, the student-optimal stable matchings defined in this way are not unique, but that the set of stable matchings, and hence the set of student optimal stable matchings, is non-empty.

When all of its seats are tentatively assigned, it rejects all the proposers who remain unassigned.

The algorithm terminates when no student proposal is rejected. Then each student is assigned to the school she proposed to last and has not been rejected. This matching is stable (since any student who prefers to be matched to a different school has already proposed to and been rejected by that school).

Note that the deferred acceptance algorithm can be adapted to our economy with indifferences, by using any fixed tie breaking procedure to convert the preferences into strict preferences, and then applying the algorithm above. Of course the particular tie breaking rules will affect the outcome of the algorithm, which we will refer to as the deferred acceptance algorithm with a tie breaking rule.

When preferences are strict, there exists a unique student-optimal stable matching that every student likes as well as any other stable matching, and a school-optimal stable matching that every school likes as well as any other matching (Roth 1985). However, as the example below demonstrates, when there are indifferences, there may be multiple student-optimal stable matchings.

Example 1 *(There can be multiple student-optimal stable matchings) There is one school s with one seat and two students $\{i, j\}$. Both i and j are acceptable for s, and s is indifferent between i and j. Both i and j prefer s to being unmatched. There are two stable matchings, $\mu_1(i) = s$, $\mu_1(j) = j$ and $\mu_2(i) = i$, $\mu_2(j) = s$. Neither stable matching is dominated by the other.*

When preferences are strict, there is no individually rational matching (stable or not) that is preferred to the unique student-optimal stable matching by every student, but even when preferences are strict, there may be unstable matchings at which some students do better than the student-optimal stable match and no student does worse (Roth 1982). In our model, as Example 1 demonstrates, there may be multiple student-optimal matchings. This weak Pareto optimality result generalizes for every student-optimal matching.

Proposition 1 *If μ is a student-optimal matching, there is no individually rational matching υ (stable or not) such that $\upsilon(i) P_i \mu(i)$ for all $i \in I$.*

Proof. Suppose that μ is a student-optimal matching for the profile (P_I, R_S). Then construct P_S that is consistent with R_S as follows: For every i, j such that $\mu(i) \neq \mu(j)$ and $i \sim_{\mu(i)} j$, let $i P_{\mu(i)} j$. The tie breaking among other students can be done arbitrarily. Then the student proposing deferred acceptance algorithm produces μ for (P_I, P_S). That is, μ is the unique student-optimal matching for some (P_I, P_S) such that P_S is consistent with

R_S. Then, from Roth (1982), there is no matching ν (stable or not) that is individually rational under (P_I, P_S) and $\nu(i)P_i\mu(i)$ for all $i \in I$.

To the contrary, suppose that there is a matching ν that is individually rational under (P_I, R_S) and $\nu(i)P_i\mu(i)$ for all $i \in I$. Since P_S is consistent with R_S, ν is individually rational under (P_I, P_S) as well. Then $\nu(i)P_i\mu(i)$ for all $i \in I$ contradicts with μ being the student-optimal matching for (P_I, P_S). $\diamond$

4 Incentives and Optimality

4.1 Preliminaries

A direct **mechanism** φ is a function that maps every (P_I, R_S) to a matching. For $x \in I \cup S$, let $\varphi_x(P_I; R_S)$ denote x's match under φ, given (P_I, R_S).

A mechanism φ is **dominant strategy incentive compatible (DSIC)** for $i \in I$ if for every (P_I, R_S) and every P_i',

$$\varphi_i(P_I; R_S) R_i \varphi_i(P_i', P_{-i}; R_S).$$

Similarly, φ is DSIC for $s \in S$ if for every (P_I, R_S) and every R_s',

$$\varphi_s(P_I; R_S) R_s \varphi_s(P_I; R_{-s}, R_s').$$

A mechanism is DSIC for $A \subset I \cup S$ if it is dominant strategy incentive compatible for every agent in A. A mechanism is DSIC if it is DSIC for all $a \in I \cup S$. We will use the term **strategy-proof** and DSIC for all students interchangeably.

When preferences are strict, there is no DSIC mechanism (Roth 1982), and there is no mechanism that is DSIC for schools (Roth 1985).[13] These negative results generalize directly to our model.

When preferences are strict, the student proposing deferred acceptance algorithm is DSIC for every student (Dubins and Freedman (1981), Roth (1982)). It is straightforward to see how this result extends to our model. In particular, consider an arbitrary tie-breaking rule $\tau = (\tau_s)_{s \in S}$ which is a collection of $|S|$ functions which map each school's preference ordering R_s to a strict preference ordering P_s that is consistent with R_s: $\tau_s(R_s) = P_s$.[14] For a fixed tie-breaking rule τ, let the mechanism DA^τ be the student-proposing deferred acceptance algorithm acting on the preferences (P_I, P_S) to produce the

[13] Kojima and Pathak (2006) show that, as markets get large with bounded preference lists, the ability of schools to manipulate either through misstating their preferences or their capacities gets small. (See also Immorlica and Mahdian (2005) for the case of one-to-one matching in large markets, and the empirical results of Roth and Peranson (1999) for the medical labor matching market.)

[14] It is possible to define a tie-breaking rule more generally to allow how ties are broken at a school to depend on the profile of school preferences at every school.

unique student-optimal matching with respect to (P_I, P_S). Then dominant strategy incentive compatibility of the student proposing deferred acceptance mechanism for every student implies that DA^τ is strategy-proof.

For the rest of the paper we will fix R_S, and refer to a problem as P_I. We define the following notion of optimality: φ **dominates** ψ if

$$\text{for all } P_I: \ \varphi_i(P_I; R_S) R_i \psi_i(P_I; R_S) \text{ for all } i \in I, \text{ and}$$
$$\text{for some } P_I: \varphi_i(P_I; R_S) P_i \psi_i(P_I; R_S) \text{ for some } i \in I.$$

A mechanism is **optimal** if there is no other mechanism that dominates it.

We focus on **single tie-breaking** (STB) rules: Let $r : I \to \mathbb{N}$ be an ordering of students such that $r(i) = r(j) \Rightarrow i = j$. For every $s \in S$, associate with R_s a strict preference relation P_s as follows: $iP_sj \Leftrightarrow [(i \succ_s j)$ or $(i \sim_s j$ and $r(i) < r(j))]$.

4.2 Motivating Examples

In this section, we develop the main ideas via several examples.

Example 2 *(Tie-breaking is consequential.) Consider an economy with three students i_1, i_2, i_3 and three schools s_1, s_2, s_3, each with one seat. Student preferences, P, are as follows:*

$$i_1 : s_2 - s_1 - s_3$$
$$i_2 : s_1 - s_2 - s_3$$
$$i_3 : s_1 - s_2 - s_3.$$

If schools are indifferent among students, the student-optimal matchings are the following:

$$\begin{pmatrix} i_1 & i_2 & i_3 \\ s_2 & s_1 & s_3 \end{pmatrix}, \begin{pmatrix} i_1 & i_2 & i_3 \\ s_2 & s_3 & s_1 \end{pmatrix}$$
$$\begin{pmatrix} i_1 & i_2 & i_3 \\ s_3 & s_1 & s_2 \end{pmatrix}, \begin{pmatrix} i_1 & i_2 & i_3 \\ s_3 & s_2 & s_1 \end{pmatrix}.$$

These matching are produced when the ties are broken as $i_1 - i_2 - i_3$ or $i_2 - i_1 - i_3$, $i_1 - i_3 - i_2$ or $i_3 - i_1 - i_2$, $i_2 - i_3 - i_1$ and $i_3 - i_2 - i_1$, respectively. Indeed, when the ties at every school are broken the same way, DA reduces to a serial dictatorship, which is Pareto optimal.

Example 3 *(Multiple tie breaking can be inefficient) Breaking ties at every school in the same way is not required for optimality. For example, when schools are indifferent as above, if the ties are broken as follows*

$$s_1 : i_2 - i_3 - i_1$$
$$s_2 : i_2 - i_1 - i_3$$
$$s_3 : i_3 - i_1 - i_2,$$

then DA produces the first matching in Example 2. Yet, arbitrary tie breaking may cause inefficiency due to artificial stability constraints that result.[15] *For example, consider the following tie-breaking*

$$\begin{aligned} s_1 &: i_1 - i_3 - i_2 \\ s_2 &: i_2 - i_1 - i_3 \\ s_3 &: i_3 - i_1 - i_2. \end{aligned}$$

DA produces the following matching

$$\mu_1 = \begin{pmatrix} i_1 & i_2 & i_3 \\ s_1 & s_2 & s_3 \end{pmatrix},$$

which is dominated by

$$\mu_2 = \begin{pmatrix} i_1 & i_2 & i_3 \\ s_2 & s_1 & s_3 \end{pmatrix}.$$

However, for every student optimal matching, we can find a way of breaking ties such that DA results in that matching.

Example 4 *(Single tie-breaking does not always yield student-optimal stable matchings.) As Example 3 demonstrates, tie-breaking has important welfare consequences. These welfare consequences are present even with single tie-breaking. Suppose schools s_2 and s_3 have the following strict preferences:*

$$\begin{aligned} s_2 &: i_2 - i_1 - i_3 \\ s_3 &: i_3 - i_1 - i_2. \end{aligned}$$

School s_1 does not rank students, i.e. s_1 is indifferent among students. The stable matchings are the following:

$$\mu_1 = \begin{pmatrix} i_1 & i_2 & i_3 \\ s_1 & s_2 & s_3 \end{pmatrix}, \ \mu_2 = \begin{pmatrix} i_1 & i_2 & i_3 \\ s_2 & s_1 & s_3 \end{pmatrix}, \ \mu_3 = \begin{pmatrix} i_1 & i_2 & i_3 \\ s_3 & s_2 & s_1 \end{pmatrix}.$$

Note that μ_1, μ_2 and μ_3 are produced by the student proposing deferred acceptance algorithm when the indiffence in s_1's preferences is broken as $s_1 : i_1 - i_3 - i_2$, $s_1 : i_2 - i_x - i_y$ and $s_1 : i_3 - i_x - i_y$, respectively. However, μ_2 dominates μ_1 despite μ_1 being stable. That is, DA need not produce a student-optimal stable matching even if ties at schools are broken the same way.

[15] This point was first suggested in footnote 13 by Abdulkadiroğlu and Sönmez (2003).

Example 5 *(Welfare enhancement involves losing strategy-proofness.) Consider Example 4 and the student proposing deferred acceptance mechanism when the ties at s_1 are broken as $s_1 : i_1 - i_3 - i_2$. We will refer to that mechanism as DA. When the students' preference profile is P, DA produces μ_1. Suppose that there is a strategy proof mechanism φ that dominates DA, and in particular produces μ_2 under P. Now consider the following preference profile P', which we obtain by changing i_1's preferences in P:*

$$\begin{aligned} &i_1 : s_2 \\ &i_2 : s_1 - s_2 - s_3 \\ &i_3 : s_1 - s_2 - s_3. \end{aligned}$$

Under P', DA produces

$$\mu' = \begin{pmatrix} i_1 & i_2 & i_3 \\ i_1 & s_2 & s_1 \end{pmatrix}.$$

Since φ dominates DA, φ must also produce μ' under $(P'; R_S)$. Note that i_1 remains unmatched at μ'. But then i_1 can manipulate φ under P' by misrepresenting her preferences as P_{i_1}, because then she is matched with $s_2 P'_{i_1} i_1$. Therefore no such strategy-proof mechanism exists for this problem which dominates DA.

4.3 Theoretical Results

We assume throughout this section that $q_s = 1$ for each $s \in S$ without loss of generality.[16] When a single tie breaking rule is applied at the beginning of a student proposing deferred acceptance algorithm, we will refer to the resulting algorithm as a DA-STB. A deferred acceptance algorithm with multiple tie breaking rules, e.g. one for each school, will be called a DA-MTB. We now focus on comparing DA-STB and DA-MTB and investigating whether it is possible to improve on DA-STB without sacrificing strategy proofness.

Our first result states that every student-optimal stable matching can be produced by some DA-STB.

Theorem 1 *For every preference profile (P_I, R_S) and every student-optimal matching μ for that preference profile, there is a single ordering of students r such that $DA^r(P_I, R_S) = \mu$.*

Proof. Suppose that μ is a student-optimal stable matching at some (P_I, R_S). For each school s, consider the students who prefer s to their assignment in μ and are ranked

[16] For the many-to-one problem, we can construct a related problem where each school seat is a separate object to be assigned, as in the construction used in Section 5.3 of Roth and Sotomayor (1990). Since we will be considering incentives and welfare from the point of view of students (who each are matched to at most one school in any case), our conclusions will carry over to the many-to-one model also.

highest in R_s among such students,

$$B(s) = \{i : sP_i\mu(i) \quad \text{and} \quad iR_sj \quad \text{for every } j \text{ such that } sP_j\mu(j)\}.$$

Let $B(S) = \cup_s B(s)$.

Let A be the set of students in $B(S)$ who are assigned to a school under μ: $A = \{i \in B(S) : \mu(i) \in S\}$. A stable improvement cycle consists of students $\{i_1, ..., i_n\} \subset A$, $n \geq 2$, such that $i_l \in B(\mu(i_{l+1}))$ and $\mu(i_{l+1})P_{i_l}\mu(i_l)$ for $l = 1, ..., n$ where $l+1$ is replaced by 1 when $l = n$ (Erdil and Ergin, forthcoming). Since μ is student-optimal, there does not exist any stable improvement cycle (Corollary 1, Erdil and Ergin, forthcoming).

Construct a directed graph with vertices (nodes) A and a directed edge from node i to node j if $\mu(j)P_i\mu(i)$ and $jR_{\mu(j)}i$; that is, i envies j's school place and the school j is assigned to is indifferent between i and j. Since there is no stable improvement cycle, the directed graph must be acyclic.

We will construct an ordering based on this directed graph utilizing two properties of the graph. First, there is a node with no incoming edges. To see this, suppose that every node has at least one incoming edge. Then since there are a finite number of nodes, starting from any node we can always leave a node by an incoming edge until we return to a visited node, which leads to a cycle, and a contradiction. Second, after removing a node, we still have a directed acyclic graph, since if there is a cycle after removing a node, then there must be a cycle in the original graph.

Construct an ordering $\rho : A \to \{1, ..., |A|\}$ as follows: find a node with no incoming edges. Remove this node and all its outgoing edges. Set the value of ρ of this node to $|A|$. By the two properties above, when we remove this node we still have a directed acyclic graph and there will be a node with no incoming edges. From this graph, we iterate the process and set the value of ρ of the next node to $|A| - 1$, and so on.[17]

Next, construct an ordering $r : I \to \mathbb{N}$ of students as follows:[18] For every $j, k \in A$, set $r(j) < r(k)$ if $\rho(j) < \rho(k)$. For every $i \in I - B(S)$ and $j \in A$, set $r(i) < r(j)$. Finally, for every student $l \in B(S) - A$, set $r(j) < r(l)$ for all $j \in A$.

Let $\nu = DA^r(P_I, R_S)$. We will show that $\nu = \mu$. Suppose to the contrary that there exists $j \in I$ such that $\mu(j) \neq \nu(j)$. Since μ is student-optimal (and since students' preferences are strict), there exists some $i \in I$ such that $\mu(i)P_i\nu(i)$. Let $C = \{i : \mu(i)P_i\nu(i)\}$ be the set of students who prefer μ to ν. For any $i_k \in C$, let $i_{k+1} = \nu(\mu(i_k))$, or $\nu(i_{k+1}) = \mu(i_k)$. Since ν is stable, there is no blocking pair, so $i_{k+1}R_{\mu(i_k)}i_k$.

[17] Erdil and Ergin (2006) prove a closely related theorem (Theorem 4) in a two-sided matching model where efficiency is defined relative to both sides of the market, and each school has one position. Since we are concerned only with the welfare of students, we obtain a result that generalizes to a model in which schools have multiple seats.

[18] Recall that lower numbers means a student is more preferred, i.e. $r(i) = 1$ means that student i is the most preferred student.

The proof by contradiction has three steps. First, we will show that for any $i_k \in C$, the student who is matched to $\mu(i_k)$ under ν, i_{k+1}, also prefers her assignment under μ to ν and so is in C. Next, we will show that in the course of $DA^r(P_I, R_S)$, student $i \in C$ can only be displaced by some other student in C. Finally, we argue that i_k could have displaced i_{k+1}, only if i_{k+1} were displaced herself. Therefore, no member of C can be rejected first, and so C must be empty.

To show the first step, note that $\mu(i_{k+1}) \neq \nu(i_{k+1})$. Suppose that $\nu(i_{k+1}) = \mu(i_k) P_{i_{k+1}} \mu(i_{k+1})$. Then $i_k R_{\mu(i_k)} i_{k+1}$ by stability of μ so that by construction $i_{k+1} \sim_{\mu(i_k)} i_k$. But then $i_{k+1} \in B(\mu(i_k))$ so that $r(i_k) < r(i_{k+1})$. Then since $i_k \in C$, $\mu(i_k) = \nu(i_{k+1}) P_{i_k} \nu(i_k)$, which contradicts with stability of $DA^r(P_I, R_S) = \nu$. Therefore $\mu(i_{k+1}) P_{i_{k+1}} \nu(i_{k+1})$, so $i_{k+1} \in C$.

We prove the second step by contradiction. Suppose that there is some $i \in C$ and $j \in I - C$ such that $\mu(i) P_j \nu(j)$ and $j R_{\mu(i)} i$. Since $j \in I - C$, we have $\nu(j) R_j \mu(j)$ therefore $\mu(i) P_j \mu(j)$. Then stability of μ implies that $j \sim_{\mu(i)} i$, which in turn implies that $j \in B(\mu(i))$ so that $r(i) < r(j)$. Therefore, no $i \in C$ is rejected by $\mu(i)$ in $DA^r(P_I, R_S)$ in favor of any $j \in I - C$ such that $\mu(i) P_j \nu(j)$. This implies that every $i \in C$ is rejected by $\mu(i)$ in $DA^r(P_I, R_S)$ in favor of some $i' \in C - \{i\}$.

Finally, in the process of $DA^r(P_I, R_S)$, no $i_k \in C$ will be rejected by $\mu(i_k)$ before $i_{k+1} = \nu(\mu(i_k))$ is rejected by $\mu(i_{k+1})$. Therefore, no $i \in C$ will be rejected by $\mu(i)$ in $DA^r(P_I, R_S)$, so that $C = \varnothing$, i.e. $\nu(i) R_i \mu(i)$. Then optimality of μ implies $\nu(i) = \mu(i)$ for all $i \in I$. $\diamond$

Corollary 1 *For any (P_I, R_S), any matching that can be produced by some DA-MTB but that cannot be produced by any DA-STB involves inefficiency.*

This theorem provides some support for a single tie breaking rule, because any matching produced by some DA-MTB that is not produced by some DA-STB cannot be student-optimal. The result, however, only speaks on the size of the set of matchings, and says nothing about the frequency of matchings.

For any student optimal matching, there exists a single tie-breaking rule r for which DA^r yields that matching. A natural question is if there is a smaller set of tie-breaking rules that will yield any student-optimal matching.[19] To see that there is not, consider a problem in which all n students have the same set of preferences over all schools and all schools are indifferent between students. There are $n!$ student-optimal matchings, which correspond exactly to the $n!$ single tie-breaking rules. This shows that the set of single tie breaking rules is the smallest such set.

[19] A special thanks to Fuhito Kojima for asking us this question.

Corollary 2 *The set of single tie breaking rules is a minimal set of tie breaking rules which under DA yields a set of matchings that contains the set of student-optimal matchings.*

Even though every student-optimal stable matching can be generated by some single tie-breaking rule, every single tie-breaking rule can produce a stable matching which is not student-optimal for some (P_I, R_S), which follows from example 4.

In case of a matching which is not student-optimal, one can easily imagine Pareto improving the matching. Our next result states that such improvements harm incentives. Therefore one can interpret the inefficiency associated with a single tie breaking rule as the cost of providing straightforward incentives to students. The next theorem generalizes the observation of example 5 and establishes that DA with some tie-breaking rule is not dominated by any other strategy-proof mechanism.

Theorem 2 *For any tie-breaking rule τ, there is no mechanism that is DSIC for every student and that dominates DA^τ.*

Proof. We begin by establishing the following property of a matching that dominates a stable matching.

Claim 1: Suppose that ν dominates $\mu = DA^\tau(P_I; R_S)$ for a given tie-breaking rule τ. Then the same set of students are matched in both ν and μ.

If there exists a student who is assigned under μ and unassigned under ν, then $\nu(i) = iP_i\mu(i)$, which implies that μ is not individually rational, a contradiction. So every student assigned under μ is also assigned under ν. Therefore $|\nu(S)| \geq |\mu(S)|$. If $|\nu(S)| > |\mu(S)|$ then there exists some $s \in S$ and $i \in I$ such that $|\nu(s)| > |\mu(s)|$ and $\nu(i) = s \neq \mu(i)$. This implies that there is a vacancy at s under μ and i is acceptable for s. Furthermore, $sP_i\mu(i)$ since ν dominates μ. These together imply that μ is not stable, a contradiction. So $|\nu(S)| = |\mu(S)|$. Then the same set of students are matched in both ν and μ since $|\nu(S)| = |\mu(S)|$ and every student assigned under μ is also assigned under ν . $\diamond$

Fix R_S. Suppose that there exists a strategy-proof mechanism φ and tie-breaking rule r such that φ dominates DA^τ. There exists a profile P_I such that

$$\varphi_i(P_I; R_S) R_i DA_i^\tau(P_I; R_S) \text{ for all } i \in I, \text{ and}$$
$$\varphi_i(P_I; R_S) P_i DA_i^\tau(P_I; R_S) \text{ for some } i \in I.$$

We will say that the matching $\varphi(P_I; R_S)$ dominates the matching $DA^\tau(P_I; R_S)$, where $DA^\tau(P_I; R_S)$ denotes the student optimal stable matching for $(P_I; P_S^\tau)$.

Let $s_i = DA_i^\tau(P_I; R_S)$ and $\hat{s}_i = \varphi_i(P_I; R_S)$ denote i's assignment under $DA^\tau(P_I; R_S)$ and $\varphi(P_I; R_S)$, respectively. Let C be the set of students $i \in I$ for whom $\hat{s}_i P_i s_i$. For each $i \in C$, define preference compression $\hat{P}_i$ as the preference list: $\hat{s}_i \hat{P}_i s_i \hat{P}_i i$ so that $\hat{s}_i$ and s_i are acceptable and all other schools are unacceptable for i under $\hat{P}_i$. For any set of agents J, let $\hat{P}_J = (\hat{P}_j)_{j \in J}$.

Claim 2: *$DA_i^\tau(\hat{P}_C, P_{I-C}; R_S) = DA_i^\tau(P_I; R_S)$ for all $i \in C$.*

Proof of Claim 2:

Since the matching $DA^\tau(P_I; R_S)$ is stable in $(\hat{P}_C, P_{I-C})$, it follows that no student in C does worse under DA^τ when preferences are $(\hat{P}_C, P_{I-C})$, we know that (i) $DA(\hat{P}_C, P_{I-C}; R_S) \in \{\hat{s}_i, s_i\}$ for all $i \in C$. Since DA^τ is coalitionally strategy-proof (Theorem 4.10, Roth and Sotomayor 1990, Dubins and Freedman 1981), (ii) there is some $i \in C$ such that $DA_i^\tau(\hat{P}_C, P_{I-C}; R_S) = DA_i^\tau(P_I; R_S) = s_i$. By assumption, (iii) $q_s = 1$ for all $s \in S$. Also, (iv) for every $i \in C$, there is $j \in C$ such that $\hat{s}_i = s_j$. Then (i), (ii), (iii) and (iv) imply that $DA_i^\tau(\hat{P}_C, P_{I-C}; R_S) = DA_i^\tau(P_I; R_S)$ for all $i \in C$. $\diamond$

Claim 3: *If there exists P_I such that the matching $\varphi(P_I; R_S)$ dominates the matching $DA^\tau(P_I; R_S)$, then φ is not strategy-proof.*

Proof of Claim 3:

Since φ dominates DA^τ, there exists a profile P_I where the matching $\varphi(P_I; R_S)$ dominates the matching $DA^\tau(P_I; R_S)$. If C is the set of students who strictly prefer matching $\varphi(P_I; R_S)$ to $DA^\tau(P_I; R_S)$, consider the economy $(\hat{P}_C, P_{I-C})$. Claim 2 shows that $DA_i^\tau(\hat{P}_C, P_{I-C}; R_S) = DA_i^\tau(P_I; R_S)$. Dominance implies that:

$$\varphi_i(\hat{P}_C, P_{I-C}; R_S) R_i DA_i^\tau(\hat{P}_C, P_{I-C}; R_S) = DA_i^\tau(P_I; R_S), \quad \forall i \in C.$$

Therefore for each $i \in C$, $\varphi_i(\hat{P}_C, P_{I-C}; R_S) \in \{s_i, \hat{s}_i\}$.

Pick an arbitrary student $i \in C$, and consider the following alternative preference relation for $i : \hat{s}_i P_i' i$, so that only $\hat{s}_i$ is acceptable to i under P_i'. Let $\hat{P}_I' = (P_i', \hat{P}_{C-i}, P_{I-C})$. By strategy-proofness of DA^τ, $DA_i^\tau(\hat{P}_I'; R_S) = i$. Also, dominance of φ over DA^τ implies that $\varphi_i(\hat{P}_I'; R_S) \in \{\hat{s}_i, i\}$.

If $\varphi_i(\hat{P}_I'; R_S) = \hat{s}_i$, then i is unmatched under DA^τ but matched under φ, which contradicts the first claim. Thus, $\varphi_i(\hat{P}_I'; R_S) = i$. In the profile $\hat{P}_I'$, student i could benefit from submitting the preference relation: $\hat{s}_i \hat{P}_i s_i \hat{P}_i i$, for she will receive $\hat{s}_i$ and prefers it to being unmatched and so φ is not strategy-proof. $\diamond$

In NYC, where strategy-proofness of the new system was a primary criterion (see the discussion section), Theorem 2 helps justify the use of the deferred acceptance algorithm, despite the fact that it does not always produce a student optimal stable match. And Theorem 1 provides additional support for the choice of a single tie-breaking rule. Note

that Theorem 2 implies that there is no strategy-proof mechanism (stable or not) which Pareto dominates DA^{τ}. The observations by Erdil and Ergin (forthcoming) and Kesten (2004) that the efficiency enhancing algorithms they study are not strategy-proof follow as a corollaries of Theorem 2.

5 Empirical results from NYC

In this section, we investigate empirically the efficiency loss of DA-STB using data from the main round of the 2003-04 NYC high school match. We work under the assumption that the preferences we receive in the main round reflect the true preferences of students.[20]

The first question we address is the comparison between DA-STB and student proposing deferred acceptance with school specific tie breaking. The first column of Table 1 shows the number of students who submitted a preference list of a certain length. The table shows that nearly 75% of students submit a rank order list shorter than the maximal 12 schools. This suggests, for these students, the limit on the length of the preference list did not affect their preference submission strategy. The next two columns report the expected distribution of student-proposing deferred acceptance with a single (uniform random) tie breaker and multiple (school specific, independent) tie breaking . The table reports the distribution for 250 draws of the tie breaker.[21] The next column shows that under single tie breaking approximately 24.82% of students receive their top choice, while under multiple tie breaking, 23.34% of students receive their top choice, an expected difference of 1,255 students. When we compare the number of students who receive their 6th choice or higher, DA-STB outperforms school specific tie breaking, while if we compare the number of students who receive their 7th choice or higher, school specific tie breaking outperforms

[20] As noted in our 2005 report, there is a minority of students for whom the NYC mechanism, as implemented, may not make it a dominant strategy to state true preferences. These include the students who have ranked 12 choices, which is approximately 25% of the entire population. Since the length of the preference list was capped at 12, this requires that if a student actually prefers more than 12 then she must select which 12 to rank on her application form. The other type of incentive problem applies to the top 2 percent of students on the grade 7 English Language Arts exam. These students are guaranteed a seat at certain schools only if they rank that school as their first choice. Finally, since the students in New York had previously been assigned through a multiple-offers mechanism where the preference revelation game was complex, it is possible that some students were slow to react to the change in the mechanism, despite the NYCDOE's efforts to communicate the difference. Laboratory evidence in Kagel and Roth (2000) suggest that learning may require some repetitions, and Chen and Sönmez (2006) experiments suggest that in one-shot versions, even in strategy-proof mechanisms, not all students may report the truth.

[21] With 250 simulations, we find that the standard deviation in the fraction of students receiving a certain choice from either DA-STB or DA-MTB is less than 0.07% (Table 1), while the standard deviation in the number of students improving in a student optimal matching is less than 0.07% (Table 2). This suggests that even with a limited number of simulations, the numbers we report in the table are not due to simulation variance.

DA-STB. (Note in particular that neither distribution stochastically dominates the other.)

Table 2 shows the distribution of allocations resulting from a DA-STB followed by stable improvement cycles (SIC), the algorithm proposed by Erdil and Ergin (forthcoming).[22] An expected 63,795 students are matched across 250 draws of the random tie-breaker.[23] Only these students can improve their matching under stable improvement cycles, for the unassigned students cannot be part of a cycle. We find that about 6,854, or 10.5% of students assigned at this stage can improve their matching, with the largest difference in the number of students who receive their top choice. Under student proposing deferred acceptance with a single tie breaker, 24.82% receive their top choice, while in a student-optimal matching, approximately 28.00% of students receive their top choice. The last two columns of Table 2 report how much the students who receive a different assignment under a student optimal matching improve relative to their assignment from DA-STB. More than 30% of the students who improve (2,091 out of 6,854) receive a school that is one position higher on their rank order list than the school they receive from DA-STB. However a substantial number of students could move much further up their preference lists, including about 50 who could move from their twelfth to their first choice.

In the Boston data, in contrast, preliminary simulations show that in 250 different single lottery outcomes of student proposing deferred acceptance in Boston Public School's choice plan 2005-06, each matching is student optimal (Abdulkadiroğlu, Pathak, Roth and Sönmez 2006). That is, there are *no* stable improvement cycles in each of the matchings for 250 different single lottery draws for grades K0, K1, K2, 01, 06, and 09. One reason for the difference is that there is a significantly larger percentage of students who receive their top choice in 2005-06 in Boston than New York, and while in Boston there are indifferences within the priority classes (sibling-walk, sibling, walk and random), the magnitude of indifferences is smaller than New York where many schools like the formerly zoned schools are indifferent between all applicants.

6 Discussion and Conclusions

It is tempting to try to capture some of the efficiency loss we observe in the data from New York. However it is not clear either how feasible this is, or how high would be the costs from the loss of strategy proofness.

[22] The SIC procedure identifies cycles among students on the way to computing a student optimal matching, but it does not specify which cycle to process when there are multiple cycles. At each step of the procedure, our implementation arbitrarily selects a cycle when there are many cycles.

[23] This would correspond roughly to the "main round" of the NYC high school match. In NYC, students left unmatched after the main round are informed of which schools still have places, and asked to submit new preferences, which are used to match remaining schools to remaining students.

On the feasibility front, the efficiency losses are only identifiable because we have preferences that we can take as a reasonable approximation of true preferences, since they were elicited from an algorithm that is strategy proof for the large majority of students. As Theorem 2 makes clear, any algorithm that would improve on DA-STB from an efficiency point of view would not be strategy-proof. Nothing is yet known about what kinds of preferences one could expect to be strategically submitted to such a mechanism, or what their welfare consequences would be.

On the cost front, strategy-proofness, which would be lost in any attempt to improve efficiency, is important in its own right. Economists and social planners like it because it yields valuable preference data. (And in New York City, schools revealed to be unpopular have been closed.[24]) Market designers and school policy-makers like it because it allows simple advice to be given to families about how to participate in the matching system.

For instance, NYC School Chancellor Joel Klein stated (NYT 10/24/03) that the "changes are intended to reduce the strategizing parents have been doing to navigate a system that has a shortage of good high schools." Furthermore, Peter Kerr, another NYCDOE official, wrote (NYT 11/3/03): "The new process is a vast improvement... For example, for the first time, students will be able to list preferences as true preferences, limiting the need to game the system. This means that 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." In every year since 2003-04, the High School directory makes a point to advise families to express their preferences truthfully. In Boston, too, strategy-proofness was a major factor in deciding to move to a new school choice system (cf Abdulkadiroğlu, Pathak, Roth, and Sönmez 2006).

Consequently, there is room for more work on both these fronts, to further illuminate the tradeoff between efficiency and strategy-proofness. In particular, for what kinds of preferences will there be substantial efficiency loss with DA-STB (as in New York but not in Boston)? Can these efficiency losses in fact be reduced by non-strategy-proof mechanisms? To put it another way, how much of the potential efficiency gains can be actually achieved by a non-strategy-proof mechanism, and how should we formulate the issues associated with gauging "how much" strategy-proofness we can trade for how much efficiency?[25]

We have concentrated here on the welfare considerations that arise in school choice because of the fact that many students are regarded by schools as equivalent. This is an issue of importance whether or not the schools are active strategic players, and we

[24] See e.g. Gootman (2006) for a report that cites demand data in the match as a reason for the closing of South Shore high school.

[25] Erdil and Ergin (forthcoming) take some preliminary steps in this direction when they analyze strategic behavior when students have symmetric beliefs in the manner of Roth and Rothblum (1999). Symmetric beliefs are, however, a very strong assumption (cf. Ehlers 2006).

have concentrated on the welfare and incentives for students. In the particular problem confronting New York City, where some schools are active players, there are also welfare and strategic considerations that apply to schools. While we have not emphasized it here, some of these are addressed through the stability of the DA-STB mechanism adopted in New York, and some are ameliorated by the size of the system (cf. Kojima and Pathak 2006).[26]

In summary, this paper fills in some of the new theory demanded by the design of school choice mechanisms, and shows empirically that the efficiency costs of strategy-proofness need not be small. As economists are more often asked to design practical markets and allocation mechanisms, we will increasingly see two-way feedback between theory and design. When we began the design of the NYC high school match in 2003, we had a lot of highly relevant theory to draw on, but as we looked into the particular requirements of the NYC school match, we found ourselves running into problems beyond the available theory, and using simulations and examples to make design decisions for which no reliable theory yet existed. In the present paper, we develop some of the theory we would have liked to have in 2003, and provide support for some of the design decisions made in a more timely way on the basis of those early simulations and examples. In doing so, we raise some new theoretical questions, to which it would be helpful to have answers before the next major design (or redesign) of school matching systems.

[26] Under the old NYC system, which produced unstable outcomes, schools had an incentive not to reveal their full capacity so that they could match afterwards with preferred students. This motivation is addressed by the stability of the current system, but no stable mechanism completely eliminates the possibility of manipulation by withholding capacity (Sönmez 1997, 1999). However Kojima and Pathak (2006) show that these incentives become small as the market becomes large in an appropriate way.

References

[1] Atila Abdulkadiroğlu, Parag A. Pathak, and Alvin E. Roth (2005a), "The New York City High School Match," *American Economic Review*, Papers and Proceedings, May 2005.

[2] Atila Abdulkadiroğlu, Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez (2005), "The Boston Public School Match," *American Economic Review*, Papers and Proceedings, May 2005.

[3] Atila Abdulkadiroğlu, Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez (2006), "Changing the Boston Mechanism: Strategyproofness as Equal Access," Unpublished mimeo, Harvard University.

[4] Atila Abdulkadiroğlu and Tayfun Sönmez (1999), "House Allocation with Existing Tenants," *Journal of Economic Theory*, 88: 233-260.

[5] Atila Abdulkadiroğlu and Tayfun Sönmez (2003), "School choice: A mechanism design approach," *American Economic Review*, 93: 729-747.

[6] Atila Abdulkadiroğlu and Yasuke Yosuda (2006), "Mechanism Design without Monetary Transfers with an Application to School Choice," Unpublished mimeo, Duke University.

[7] Michel Balinksi and Tayfun Sönmez (1999), "A tale of two mechanisms: student placement," *Journal of Economic Theory*, 84: 73-94.

[8] Yan Chen and Tayfun Sönmez (2006), "School choice: an experimental study, " *Journal of Economic Theory*, 127: 2002-2231.

[9] Lester E. Dubins and David A. Freedman (1981), "Machiavelli and the Gale-Shapley Algorithm," *American Mathematical Monthly*, 88: 485-494.

[10] Lars Ehlers (2006), "Truncation Strategies in Matching Markets," Unpublished mimeo, CIREQ.

[11] Aytek Erdil and Haluk Ergin (forthcoming), "What's the Matter with Tie-breaking? Improving Efficiency in School Choice," forthcoming, *American Economic Review.*

[12] Aytek Erdil and Haluk Ergin (2006), "Two-Sided Matching with Indifferences," Unpublished mimeo, Harvard Business School.

[13] Elissa Gootman (2006), "Lafayette Among 5 High Schools to Close," N.Y. Times, December 14.

[14] Haluk Ergin (2002), "Efficient Resource Allocation on the Basis of Priorities, "*Econometrica*, 70: 2489-2497.

[15] Haluk Ergin and Tayfun Sönmez (2006), "Games of School Choice under the Boston Mechanism, " *Journal of Public Economics*, 90: 215-237.

[16] David Gale and Lloyd Shapley (1962), "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, 69: 9-15.

[17] David Gale and Marilda Sotomayor (1985), "Some remarks on the stable matching problem," *Discrete Applied Mathematics*, 11: 223-32.

[18] Arnaud Hylland and Richard Zeckhauser (1979), "The efficient allocation of individuals to positions," *Journal of Political Economy*, 87: 293-314.

[19] Nicole Immorlica and Mohammed. Mahdian (2005), "Marriage, honesty, and stability,"*SODA*.

[20] Robert Irving (1994), "Stable marriage and indifference," *Discrete Applied Mathematics*, vol. 48: 261-272.

[21] John Kagel and Alvin Roth (2000), "The dynamics of reorganization in matching markets: A laboratory experiment motivated by a natural experiment, " *Quarterly Journal of Economics*, 201-235.

[22] Onur Kesten (2004), "Student Placement to Public Schools in the US: Two new solutions," Unpublished mimeo, University of Rochester.

[23] Fuhito Kojima and Parag Pathak (2006), "Incentives and Stability in Large Two-Sided Matching Markets," Unpublished mimeo, Harvard University.

[24] Jipeng Ma (1994), "Strategy-proofness and strict core in a market with indivisibilities," *International Journal of Game Theory*, 23: 75-83.

[25] Paul Milgrom (2006), "Incentives in Core-Selecting Auctions," Unpublished mimeo, Stanford University.

[26] Szilvia Pápai (2000), "Strategyproof assignment by hierarchical exchange," *Econometrica*, 68: 1403-1433.

[27] Parag Pathak (2006), "Lotteries in Student Assignment," Unpublished mimeo, Harvard University.

[28] Parag Pathak and Tayfun Sönmez (2006), "Leveling the Playing Field: Sincere and Strategic Players in the Boston Mechanism," Unpublished mimeo, Boston College and Harvard University.

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

[30] Alvin E. Roth (1982b), "Incentive Compatibility in a Market with Indivisible Goods," *Economics Letters*, 9: 127-132.

[31] Alvin E. Roth (1984), "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory." *Journal of Political Economy*, 92: 991-1016.

[32] Alvin E. Roth (1985), "The College Admissions Problem is Not Equivalent to the Marriage Problem," *Journal of Economic Theory*, 36: 277-288.

[33] Alvin E. Roth (1990), "New Physicians: A Natural Experiment in Market Organization." *Science*, 250: 1524-1528.

[34] Alvin E. Roth (1991), "A Natural Experiment in the Organization of Entry Level Labor Markets: Regional Markets for New Physicians and Surgeons in the U.K." *American Economic Review*, 81: 415-440.

[35] Alvin E. Roth (2002), "The economist as engineer: game theory, experimental economics, and computation as tools of design economics." *Econometrica*, 70: 1341-1378.

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

[37] Alvin E. Roth and Andrew Postlewaite (1977), "Weak versus strong domination in a market with indivisible goods." *Journal of Mathematical Economics*, 4: 131-137.

[38] Alvin E. Roth and Uriel G. Rothblum (1999), "Truncation Strategies in Matching Markets: In Search of Advice for Participants," *Econometrica*, 67: 21-43.

[39] Alvin E. Roth, Tayfun Sönmez, and Utku Ünver (2004), "Kidney Exchange," *Quarterly Journal of Economics*, 119: 457-488.

[40] Alvin E. Roth, Tayfun Sönmez, and Utku Ünver (2005a), "Pairwise Kidney Exchange," *Journal of Economic Theory*, 125(2): 151-188.

[41] Alvin E. Roth, Tayfun Sönmez, and Utku Ünver (2005b), "A Kidney Exchange Clearinghouse in New England," *American Economic Review, Papers and Proceedings*, 95(2): 376-380.

[42] Alvin E. Roth, Tayfun Sönmez, and Utku Ünver (forthcoming), "Efficient Kidney Exchange: Coincidence of Wants in a Market with Compatibility-Based Preferences," *American Economic Review.*

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

[44] Lloyd Shapley and Herb Scarf (1974), "On cores and indivisibility," *Journal of Mathematical Economics*, 1: 23-28.

[45] Tayfun Sönmez (1997), "Manipulation via Capacities in Two-Sided Matching Markets," *Journal of Economic Theory*, 77: 197-204.

[46] Tayfun Sönmez (1999), "Can Pre-Arranged Matches be Avoided in Two-Sided Matching Markets?" *Journal of Economic Theory*, 86: 148-156.

Table 1— Tie-breaking in Student-Proposing Deferred Acceptance in the Main Round

Choice	Number Ranking	Single Tie-Breaking (250 draws)	Multiple Tie-Breaking (250 draws)
1	5,797 (6.7%)	21,038 (24.82%)	19,783 (23.34%)
2	4,315 (5.0%)	10,686 (12.61%)	10,831 (12.78%)
3	5,643 (6.6%)	8,031 (9.48%)	8,525 (10.06%)
4	6,158 (7.2%)	6,238 (7.36%)	6,633 (7.83%)
5	6,354 (7.4%)	4,857 (5.73%)	5,108 (6.03%)
6	6,068 (7.1%)	3,586 (4.23%)	3,861 (4.56%)
7	5,215 (6.1%)	2,721 (3.21%)	2,935 (3.46%)
8	4,971 (5.8%)	2,030 (2.40%)	2,141 (2.53%)
9	4,505 (5.2%)	1,550 (1.83%)	1,617 (1.91%)
10	5,736 (6.7%)	1,232 (1.45%)	1,253 (1.48%)
11	9,048 (10.5%)	1,016 (1.20%)	894 (1.05%)
12	22,239 (25.8%)	810 (0.96%)	372 (0.44%)
unassigned	-	20,952 (24.72%)	20,795 (24.54%)

Notes: The table is based on data from the main round in 2003-04 provided by the New York City Department of Education's Office of High School Admissions. The column Number Ranking reports the distribution of the length of student preference lists in the main round. The next two columns report the expected distribution of choices from 250 draws of the single tie breaker and multiple school-specific tie breakers, respectively.

Table 2— Student Optimal Matchings in the Main Round

Choice received	Deferred Acceptance (250 draws)	Student Optimal Matching (250 draws)	Improvement in Choice	Number of Students (250 draws)
1	21,038 (24.82%)	23,729 (28.00%)	+1	2,091
2	10,686 (12.61%)	11,352 (13.40%)	+2	1,373
3	8,031 (9.48%)	8,005 (9.45%)	+3	934
4	6,238 (7.36%)	5,917 (6.98%)	+4	705
5	4,857 (5.73%)	4,378 (5.17%)	+5	544
6	3,586 (4.23%)	3,151 (3.72%)	+6	412
7	2,721 (3.21%)	2,271 (2.68%)	+7	280
8	2,030 (2.40%)	1,647 (1.94%)	+8	208
9	1,550 (1.83%)	1,170 (1.38%)	+9	154
10	1,232 (1.45%)	905 (1.07%)	+10	101
11	1,016 (1.20%)	723 (0.85%)	+11	52
12	810 (0.96%)	548 (0.65%)		
unassigned	20,952 (24.72%)	20,952 (24.72%)	total	6,854

Notes: The table is based on author's calculations using data from the main round in 2003-04 provided by the New York City Department of Education's Office of High School Admissions. The column Deferred Acceptance reports the expected distribution of choices from 250 different single tie breakers in the main round, and the next column reports the expected distribution when each of these matchings is improved to a student optimal matching using the procedure described in Erdil and Ergin (forthcoming). The last column reports the expected improvement in choice obtained by students in a student optimal matching.

Appendix 1: NYC Schools

In New York City, there are three main types of high schools: 1) schools that are active players who explicitly rank students in order of preference, 2) schools that are passive players who order students based on priorities, which are set centrally by the Department of Education, and 3) schools where a fraction of seats are determined by active ranking and the remaining fraction are based on priorities. Table A summarizes the number of schools in each group by borough and by capacity. The screened, audition, and specialized high schools are strategic players, who explicitly rank students in order of preference. For example, Townsend Harris is a screened high school in Flushing, NY which evaluates students based on their test scores, attendance and punctuality. At Towsend Harris, all students are required to have a minimum 90th percentile on Math and Reading standardized tests as well as a minimum grade point average of 90 in June of 7th grade when being considered for a 9th grade seat. Various other criteria are used at screened and audition schools. The unscreened schools are passive players, and use priorities based on geographic location, current middle school, or other criteria. The academic comprehensive program at Forest Hills High School in Queens, for instance, places students who live in an attendance zone near the school in a higher priority class than students from outside the priority zone. Finally, Educational option (Ed-opt) schools are permitted to rank students for half of their positions, and are required to admit students according to priorities for the other half. Table 1 shows that nearly half of all schools are Educational option, and more than half of total district capacity is at schools who use priorities to order students.

When priorities are used at unscreened and Educational option programs in New York City, many students fall into the same priority class. For instance, at Forest Hills where were 474 seats in 2003-04, 352 (10.7%) of student applicants are from the assignment zone while the remaining 2,937 are from outside.

Table A— Number and Capacities of Programs

Panel A: Number of Programs by Borough

	Unscreened	Screened and Auditioned	Specialized	Educational Option	Total
Brooklyn	21	49	1	70	141 (26.5%)
Bronx	36	33	2	45	116 (21.8%)
Manhattan	5	53	2	79	139 (26.1%)
Queens	19	47	1	52	119 (22.4%)
Staten Island	5	6	0	6	17 (3.2%)
Total	86 (16.2%)	188 (35.3%)	5 (1.1%)	252 (47.4%)	532

Panel B: Total Capacity by Borough

	Unscreened	Screened and Auditioned	Specialized	Educational Option	Total
Brooklyn	7,260	5,385	1,000	10,238	23,833 (27.1%)
Bronx	8,632	4,388	908	5,163	19,091 (21,6%)
Manhattan	1,410	7,400	888	6,859	16,557 (18.5%)
Queens	9,390	7,463	150	8,121	25,124 (28.5%)
Staten Island	2,860	526	0	185	3,571 (4.0%)
Total	29,552 (33.5%)	25,162 (28.5%)	2,946 (3.3%)	30,566 (34.5%)	88,226

Notes: Constructed from data provided by the New York City Department of Education Office of High School Admissions.