Expanding "Choice" in School Choice

Atila Abdulkadiroğlu* Yeon-Koo Che† Yosuke Yasuda‡

May 31, 2007

Very Preliminary; Please do not circulate beyond this seminar.

ABSTRACT: Truthful revelation of preferences has emerged as a desideratum in the design of school choice programs. The Gale-Shapely's deferred acceptance algorithm achieves this strategyproofness but limits students' abilities to communicate their preference intensities, which entails an ex ante inefficient allocation when schools are indifferent among students with the same ordinal preferences. We propose a new deferred acceptance procedure in which students are allowed, via signaling of their preferences, to influence how they are treated in a tie for a school. This new procedure preserves strategyproofness of ordinal preferences and yields a more desirable allocation.

KEYWORDS: Gale-Shapley's deferred acceptance algorithm, choice-augmented deferred acceptance, tie breaking, ex ante Pareto efficiency.

*Duke University.
†Columbia University; email: yk.che@columbia.edu.
‡Princeton University.

1 Introduction

School choice has been a subject of intense research and public debate in recent years. Its goal of expanding one's freedom to choose a public school beyond his/her residence area has a broad public support, as a number of states has recently adopted some form of choice programs. Yet, how best to implement such a program remains actively debated. The center of the debate is a proposal made by Abdulkadiroğlu and Sönmez (2003), in conjunction with the programs Abdulkadiroğlu, Pathak, Roth and Sönmez helped develop for Boston and the New York City (Abdulkadiroğlu et. al, 2005a,b; 2006a,b).

Their proposal argues against a popular method, called "Boston" Mechanism, in favor of the Gale-Shapley's *Student-Proposing Deferred Acceptance* algorithm (henceforth, *DA algorithm*). The Boston mechanism assigns to a school those students who "ranked" that school as their top choice, and assigns the others only when there are remaining seats. Under this mechanism, families may not rank their most preferred school as their top choice, fearing that oversubscription of that school may bump them into a school even worse than their second-best choice, say. This incentive to "game the system" would be absent in their proposed DA algorithm, for its deferred acceptance feature means that a family need not sacrifice its chance at a less preferred school when picking even a popular school as its top choice (Dubins and Freedman, 1981; Roth, 1982).[1]

This so-called *strategy-proofness* is practically important, for the parents participating in the program can be advised to behave truthfully and non-strategically (see Abdulkadiroğlu et. al, 2006b). Further, the DA algorithm has a desirable welfare property: If the schools have strict preferences, it implements the student optimal outcome (Gale and Shapley, 1962; Roth, 1982, 1985). By contrast, again given strict priorities, any stable outcome may arise in (full-information) Nash equilibrium of the Boston mechanism (Ergin and Sönmez, 2006).[2] To illustrate, suppose there are three students

[1] The process of the (student-proposing) DA algorithm is described in the next section.

[2] To be precise, the same is true with the DA algorithm. Assume that there are a finite number of students and a finite number of schools each with one seat to fill. All agents have strict preferences. Fix any stable matching ξ. Suppose under the DA algorithm that students adopt a strategy profile wherein each student i submits $\xi(i)$ as her top choice (with the remaining parts of the list specified arbitrarily). Consider any deviation by student i which lists school s as the top choice. For this deviation to be profitable, $s \succ_i \xi(i)$ and $i \succ_s \xi(s)$, but this contradicts the stability of ξ. Nevertheless, the strategy-

$I = \{i_1, i_2, i_3\}$ to be assigned to three schools $S = \{s_1, s_2, s_3\}$ each with one seat to fill. The students' preferences are

$$P(i_j) = s_1 - s_2 - s_3, \; j = 1, 3$$
$$P(i_2) = s_2 - s_1 - s_3.$$

The schools' priorities are

$$P(s_1) = i_2 - i_1 - i_3,$$
$$P(s_2) = i_1 - i_2 - i_3,$$
$$P(s_3) = i_3 - i_1 - i_2.$$

There are two stable matchings:

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

Clearly, ξ is Pareto superior to ξ', for the students. Under the DA algorithm, it is weakly dominant for student i_j to submit $P(i_j)$, so ξ arises. The Boston mechanism admits both ξ and ξ' as equilibria, and the equilibrium strategies supporting ξ' are undominated and robust to trembles.[3]

The DA algorithm is not as well-justified, however, when the schools do not have strict priorities so that many students with the same priority standing compete for limited seats in a school. Indeed, ties of this sort are endemic to school choice since students often have similar ordinal preferences, and schools have very coarse priority rules based on a few categories, such as "walk zone" and "siblings", causing many students to fall into the same priority class. How to break them can have a significant efficiency consequence. Not only is there no tie-breaking rule that guarantees maximum student welfare even among stable matchings,[4] there is no obvious procedure even within

proofness means that it is a weak dominant strategy for each student to submit his/her preferences truthfully.

[3]To see how ξ' can be sustained, suppose student i_j, $j = 1, 3$, lists $s_2 - s_1 - s_3$, and student i_2 lists $s_1 - s_2 - s_3$. If student i_1 deviates unilaterally to s_1, her most preferred school, she will lose a seat at s_1 to i_2 (who has the highest priority at s_1); meanwhile, the seat at s_2 will have been assigned to i_3, so she will be assigned to s_3 in the next round. Likewise, i_2 is strictly worse off by deviating unilaterally.

[4]Without strict preferences, no strategy-proof procedure can implement the student-optimal stable matching in a general environment (Erdil and Ergin, forthcoming).

random procedures: As Abdulkadiroğlu et. al (2006a) shows, even the most sensible random tie-breaking rule can leave a considerable efficiency loss.

In particular, if students have common ordinal preferences over schools, the DA algorithm assigns students randomly over schools, whenever they are tied according to a school's priority order. If the students differ in their relative preferences, this may not be efficient at least from the ex ante perspective. To see this, suppose three students, $I = \{i_1, i_2, i_3\}$, to three schools, $S = \{s_1, s_2, s_3\}$, each with one seat to fill. The schools are indifferent to all three students, and students' preferences are represented by von-Neumann Morgenstern (henceforth, vNM) utility value v_j^i, $i = 1, 2, 3$:

	v_j^1	v_j^2	v_j^3
$j = 1$	4	4	3
$j = 2$	1	1	2
$j = 3$	0	0	0

If the DA algorithm is used with ties broken randomly, then all three submit true (ordinal) preferences, and they will be assigned to the schools with equal probabilities. Hence, the students obtain expected utilities of $EU_1 = EU_2 = EU_3 = \frac{5}{3}$. It is easy to see that this assignment is Pareto-dominated by the following assignment: *Assign student 3 to s_2, and students 1 and 2 randomly between s_1 and s_3*, which yields expected utilities of $EU_1' = EU_2' = EU_3' = 2 > \frac{5}{3}$. Remarkably, Boston mechanism implements this latter matching: 1 and 2 will rank s_1 as their top choice; and, given that, 3 will choose s_2 as her top choice.[5] This example illustrates one advantage the Boston mechanism has over the DA algorithm: Unlike the latter, the former is responsive to the agents' cardinal preferences, or their relative preference intensities over choices.

Is there a way to harness this "responsiveness to cardinal preferences" without sacrificing strategy-proofness? We propose an algorithm that will accomplish this goal.

[5]This does not contradict Ergin and Sönmez (2006)'s finding that the Boston mechanism is (weakly) Pareto dominated by the DA, which relies on strict preferences by the schools.

2 Choice-Augmented DA Algorithm: Illustration

Consider an environment with a finite number of students and schools each endowed with a finite number of seats. The students have strict preferences over schools, and the schools have priorities characterized by strict preferences over students. Consider the student-proposing DA algorithm (Gale and Shapley 1962): At the first step, each student applies to her most preferred school. Every school tentatively admits its applicants in the order of its priority order of the students. When all of its seats are assigned, it rejects the remaining applicants. In general, each student who was rejected in the previous step applies to her next preferred school. Each school considers the set of students it has tentatively admitted and the new applicants. It tentatively admits from these students in the order of priority. When all of its seats are assigned, it rejects the remaining applicants. The process terminates when no student proposal is rejected.

When schools priorities are characterized by weak preferences, DA with single tie breaking (DA-STB) first generates a random ordering of students, then breaks the ties at school priorities according to this ordering and then it applies the (student-proposing) DA with the induced strict school priorities. On the other hand, DA with multiple tie breaking (DA-MTB) first generates an ordering of students at every school independently randomly, then breaks the ties at a school according to its associated ordering and then applies the (student-proposing) DA with the induced strict school priorities.

Consider a general environment with any finite number of students and schools each endowed with a finite number of seats. The students are assumed to have strict preferences over schools, represented by von-Neumann Morgenstern utilities, and the schools have priorities characterized by weak preferences.

Now consider the following modification of the DA-STB algorithm labeled *Choice-Augmented Deferred Acceptance* (henceforth, CADA):[6]

[6]The proposed mechanism is similar in spirit to the proposal by Sönmez and Ünver (2003) to augment "course bidding" with the DA algorithm. These two proposals differ in the application as well as in the nature of the inquiry: We are interested in the *ex ante* efficiency benefit of augmenting the "choice" element to the DA algorithm, whereas their interest is in the benefit of adding the DA feature to the course bidding (i.e., the choice element), which is to select an *ex post* efficient outcome among market outcomes. Our proposal to improve efficiency is also related to Casella (2005) and Jackson and

- All students are asked to submit ordinal preferences, plus an "auxiliary message" naming of one's "target" school.
- Ties at schools' priorities are broken in the following way: First, a random ordering of students, just like the standard DA-STB. Next, at every indifference class at every school's priority order, ties are broken in favor of those students who selected that school as their "target" and then according to the random order.
- The students are then assigned via the DA algorithm, respecting the schools' original priority order, but using the "choice"-adjusted priority list whenever there are ties.

Clearly, the deferred acceptance feature preserves the incentive to reveal the ordinal preferences truthfully; the gaming aspect is limited to manipulating the outcome of tie-breaking. This limited introduction of "choice signaling" can however improve upon the DA rule in a significant way. In the above example, the CADA implements the Pareto superior matching: All students will submit the ordinal preferences truthfully, but 1 and 2 will choose s_1 as their target, and 2 will choose s_2 as her target. In this case, the CADA resembles the Boston mechanism.

In general, the CADA is different from the Boston mechanism. In fact, if schools have many priorities (so their preferences are almost "strict"), then the auxiliary message would have little bite; thus the CADA will very much resemble the DA. For instance, in the second example above, the schools have strict priorities, so the CADA coincides with DA, implementing the student optimal stable matching ξ. In a sense, the CADA delivers the best of both worlds, implementing the outcome of the Boston mechanism only when it is known to clearly dominate the DA, without sacrificing the strategy-proofness.

Sonnenschein (forthcoming) in that how a student is treated in a tie at one school is "linked" to how she is treated in a tie at another school.

3 Model

3.1 Primitives

There are $n \geq 2$ schools, $S = \{1, ..., n\}$, each with a unit mass of seats to fill. There are mass n of students who are indexed by vNM values $\mathbf{v} = (v_1, ..., v_n) \in \mathcal{V} = [0,1]^n$ they attach to the n schools. The set of student types, $\mathcal{V}$, is equipped with a (Lebesgue) measure μ. We assume that μ is absolutely continuous with strictly positive density in the interior of $\mathcal{V} = [0,1]^n$ and that the values are distinct for μ-a.e. $\mathbf{v}$. That is, $\mu(\{\mathbf{v} \in \mathcal{V} | v_i = v_j \text{ for some } i \neq j\}) = 0$.

Let τ^k be any ordered list of k schools, and let Π^k be the set of all such ordered list of k schools. Let $\pi^k(\mathbf{v})$ be the type $\mathbf{v}$-student's k most preferred schools (listed in the descending order of her preferences). Let $\Pi = \cup_{k=1}^n \Pi^k$. Let $m_{\tau^k} := \mu(\{\mathbf{v} \in \mathcal{V} | \pi^k(\mathbf{v}) = \tau^k\}$. By the full support assumption, $m_{\tau^k} > 0$ for each $\tau^k \in \Pi^k \; \forall \; k \leq n$. Hence, for each $i \in S$, $m_i = \mu(\{\mathbf{v} \in \mathcal{V} | \pi^1(\mathbf{v}) = i\})$ represents the measure of students whose most preferred school is i. It is useful to define the set

$$S^{**} = \{i \in S | m_i \geq m_j \forall j \in S\}$$

of *most popular* schools.

Finally, let $\mathfrak{m} := \{m_\tau\}_{\tau \in \Pi^n}$ be a profile of measures of all ordinal types. Let $\mathfrak{M} := \{\{m_\tau\}_{\tau \in \Pi^n} | \sum_{\tau \in \Pi^n} m_\tau = n\}$ be the set of all possible measure profiles. The set $\mathfrak{M}$ has $n! - 1$ dimensions. We say a result holds for a *generic* $\mathfrak{m}$ if the result holds for a subset (of $\mathfrak{M}$), whose closure has the same Lebesque measure as $\mathfrak{M}$.

An *assignment*, denoted by $\mathbf{x}$, is a probability distribution over S; and the n-dimensional simplex, denoted Δ^n, represents the set of all possible assignments. An *allocation* is a measurable function $\phi := (\phi_1, ..., \phi_n) : \mathcal{V} \mapsto \Delta^n$ mapping from the student types to the set of assignments, such that $\int \phi_i(\mathbf{v}) d\mu(\mathbf{v}) = 1$ for each $i \in S$, with the interpretation that student $\mathbf{v}$ is assigned by mapping $\phi = (\phi_1, ..., \phi_n)$ to school i with probability $\phi_i(\mathbf{v})$. Let $\mathcal{X}$ denote the set of entire allocations.

3.2 Ex Ante Welfare Standards

We say that allocation $\tilde{\phi} \in \mathcal{X}$ *weakly Pareto-dominates* allocation $\phi \in \mathcal{X}$ if

$$\mathbf{v} \cdot \tilde{\phi}(\mathbf{v}) \geq \mathbf{v} \cdot \phi(\mathbf{v}) \text{ for } \mu - a.e.\, \mathbf{v}, \tag{1}$$

and that $\tilde{\phi} \in \mathcal{X}$ *Pareto-dominates* allocation $\phi \in \mathcal{X}$ if (1) holds and there exists a set $A \subset N$ with $\mu(A) > 0$ such that

$$\mathbf{v} \cdot \tilde{\phi}(\mathbf{v}) > \mathbf{v} \cdot \phi(\mathbf{v}) \text{ for all } \mathbf{v} \in A.$$

Definition 1. *Allocation* $\mathbf{x} \in \mathcal{X}$ *is* ***Pareto optimal*** *if there is no other allocation in* $\mathcal{X}$ *that Pareto-dominates* $\mathbf{x}$*.*

It is useful for our purpose to introduce a weaker property. Fix an allocation $\phi \in \mathcal{X}$. For a set of schools $K \subset S$, an assignment $\mathbf{x} \in \Delta^n$ is said to be a *within* K *reassignment of* $\phi(\mathbf{v})$ if $x_j = \phi_j(\mathbf{v})$ for each $j \in S \backslash K =: J$. Let $\Delta_{\phi_J}(\mathbf{v})$ be the set of all within K reassignments of $\phi(\mathbf{v})$. We then call an allocation $\tilde{\phi} \in \mathcal{X}$ a within K reallocation of $\phi(\mathbf{v})$ if $\tilde{\phi}(\mathbf{v}) \in \Delta_{\phi_J}(\mathbf{v})$ for all $\mathbf{v} \in \mathcal{V}$. Let

$$\mathcal{X}_{\phi_J} := \{\tilde{\phi} \in \mathcal{X} | \tilde{\phi}(\mathbf{v}) \in \Delta_{\phi}(\mathbf{v}), \forall \mathbf{v} \in \mathcal{V}\}$$

be the set of all such allocations. In words, a within K reallocation of ϕ involves possible trading of students' assignment probability shares over schools within K.

Definition 2. *For any* $K \subset S$*, an allocation* $\phi \in \mathcal{X}$ *is* ***Pareto optimal within*** K *if there is no within* K *reallocation of* ϕ *that Pareto dominates* ϕ*.*

If an allocation $\phi \in \mathcal{X}$ is *Pareto optimal within* K, then it leaves no Pareto improving opportunities for trading assignment shares within K. We make some obvious observations.

Lemma 1. *If* $\phi \in \mathcal{X}$ *is Pareto optimal within* K'*, then it is Pareto optimal within* $K \subset K'$*. (ii) Every allocation is Pareto optimal within* K *with* $|K| = 1$*.*

4 Description of the Mechanisms and Equilibrium

In this section, we give a heuristic description of the three mechanisms with continuum of students, discuss equilibrium behavior and state the existence result. We defer a more rigorous treatment of this section to the appendix.

A tie breaker for school i is a measurable function $F_i : \mathcal{V} \to [0, n]$ of students. Accordingly, student $\mathbf{v}$ draws a better random number at school i when $F_i(\mathbf{v})$ is smaller. That is, students with smaller random numbers are ranked higher at school's priority order. We refer $\omega = F_i(\mathbf{v})$ as $\mathbf{v}$'s random draw at school i.

Given a list of tie breaker $\{F_i\}$, it is easy to generalize DA: At the first step, each student applies to her most preferred school. Every school i ranks its applicants according to F_i. School i tentatively admits up to unit mass from its applicants in the order of its priority order of F_i. If the measure of the applicants exceeds 1, it rejects the remaining applicants. In general, each student who was rejected in the previous step applies to her next preferred school. Each school considers the set of students it has tentatively admitted and the new applicants. It tentatively admits up to unit mass from these students in the order of priority. When all of its seats are assigned, it rejects the remaining applicants. The process converges when the set of students that are rejected is null. Although DA might not converge in finite time, it converges in the limit (Proposition 4). Furthermore, it is a (weak) dominant strategy for each student to submit her ordinal preferences truthfully to DA (Proposition 5).

DA-STB and *DA-MTB*

DA-STB generates a random ordering of students, that is $F_i(\mathbf{v}) = F_j(\mathbf{v})$ a.e. $\mathbf{v}$ for all $i, j \in S$. Given μ, and that students report their ordinal preferences truthfully, the equilibrium induces a cutoff $c_i \in L$ for each school i such that a student who ever subscribes to school i gets assigned if and only if his draw ω is less than c_i. The n cutoffs can be listed in an ascending order as $(\hat{c}^1, ..., \hat{c}^n)$, with $\hat{c}^i \leq \hat{c}^j$ for $i < j$. (The order may be weak since the cutoff may be the same for several schools, in which case we simply repeat the same number.) As will be shown in the next lemma, these cutoffs are uniquely determined by $\mathfrak{m} := \{m_\tau\}_{\tau \in \Pi^n}$ and all distinct generically:

Lemma 2. *The cutoffs for the schools are under DA-STB are uniquely determined by* $\mathfrak{m}$*, and satisfies* $\hat{c}^1 > 0$ *and* $\hat{c}^n = n$*. For a generic* $\mathfrak{m}$*, the cutoffs are all distinct.*

DA-MTB draws a new tie breaker for every school. The MTB rule can be similarly characterized as follows. Suppose each student draws a number ω_i independently and uniformly from an interval $L := [0, n]$ for all $i \in S$ such that a student with a lower draw has a higher priority than one with a higher draw, at school $i \in S$. Given μ, and students' reporting truthfully about their ordinal preferences, the equilibrium induces a cutoff $c_i \in L$ for each school i such that a student who ever subscribes to school i gets assigned if and only if his draw ω_i is less than c_i. While the cutoffs under STB rule can be different from those under MTB, we must have at least one school whose cut off is equal to n.

CADA

CADA can be generalized similarly:

- All students are asked to submit ordinal preferences, formally a permutation $\pi : S \mapsto S$ of schools, plus an "auxiliary message" naming of one's "target" school $s \in S$.
- Ties at schools' priorities are broken in the following way: A random ordering of students is drawn. At every school's priority order, ties are broken first in favor of those students who selected that school as their "target" and then according to the random ordering.
- The students are then assigned via the DA algorithm using the "choice"-adjusted random priority list.

Let Π be the set all possible permutations of schools. Then, a pure strategy is $s = (\pi, i) \in \Pi \times S$. Consider a mixed strategy $\sigma \in \Sigma$, where Σ is a set of all probability distributions over $\Pi \times S$. The profile of strategies simply a mapping $\sigma : \mathcal{V} \mapsto \Sigma$, giving a mixed strategy $\sigma(\mathbf{v})$ for student $\mathbf{v}$. Each profile σ induces an allocation ϕ^σ. Consider a strategy profile $(\sigma, \sigma_{-\mathbf{v}}))$. A Nash equilibrium is a profile σ^* such that for a.e. $\mathbf{v} \in N$, $\mathbf{v} \cdot \phi^{\sigma^*(\mathbf{v})}(\mathbf{v}) \geq \mathbf{v} \cdot \phi^{(\sigma, \sigma^*_{-\mathbf{v}}))}(\mathbf{v})$ for all $\sigma \in \Sigma$.

The deferred acceptance feature of CADA preserves strategy-proofness:

Proposition 1. *It is (weak) dominant strategy for each student to submit her ordinal preferences truthfully under CADA*

Let $\pi_{[\cdot]}(\mathbf{v})$ denote a permutation of S indicating type $\mathbf{v}$ student's ordinal preferences. That is, $\pi_i(\mathbf{v})$ indicates the i-th preferred school of that student. Clearly, $v_{\pi_i(\mathbf{v})} > v_{\pi_j(\mathbf{v})}$ if and only if $i < j$. Following Proposition 1, we focus on the strategy choice in auxiliary message, and use $\nu = (\nu_1, ..., \nu_n) \in \Delta$ to denote a mixed strategy just on S, the set of schools. We focus on equilibrium that is robust to an arbitrarily small measure of students tremble with the reporting of their ordinal preferences. Formally, we require an equilibrium ν^* to be a limit as $\epsilon \to 0$, of sequence $\{\nu^\epsilon\}$ of equilibrium for which there is an additional $\epsilon > 0$ measure of students randomizing over all schools uniformly in both ordinal list and auxiliary message.

Theorem 1. *There exists an equilibrium in mixed strategies.*

We focus on equilibrium that is robust to an arbitrarily small measure of students tremble with the reporting of their ordinal preferences. Formally, we require an equilibrium ν^* to be a limit as $\epsilon \to 0$, of sequence $\{\nu^\epsilon\}$ of equilibrium for which there is an additional $\epsilon > 0$ measure of students randomizing over all schools uniformly in both ordinal list and auxiliary message. We say that a student *applies to school* i if she is rejected by all schools she lists ahead of i in her (truthful) ordinal list. We say that a student *subscribes to school* $i \in S$ if she picks school i as its target *and* applies to that school during the DA process. (The latter event depends on where she lists school i in her ordinal list and the other students' strategies as well as the outcome of tie-breaking). Let $\bar{\nu}_i^*(\mathbf{v})$ be the probability that a student $\mathbf{v}$ subscribes to school i in equilibrium. We say a school $i \in S$ is *oversubscribed* if $\int \bar{\nu}_i^*(\mathbf{v}) d\mu(\mathbf{v}) \geq 1$ and *undersubscribed* if $\int \bar{\nu}_i^*(\mathbf{v}) d\mu(\mathbf{v}) < 1$.

Lemma 3. *A student whose most preferred school is undersubscribed must be choosing that school as her target in equilibrium. In equilibrium, $\nu^*(\mathbf{v}) = \bar{\nu}^*(\mathbf{v})$ for μ-a.e. $\mathbf{v}$.*

In light of this lemma, we shall refer to "picks a school i as target" simply as "subscribes to school i."

5 Welfare Analysis of Alternative Procedures

5.1 Welfare Properties of DA-STB and DA-MTB

It is useful to establish the following lemma, which will be used in the subsequent theorems.

Lemma 4. *An allocation ϕ is Pareto optimal within schools $\{i, j\} \subset S$, if $\phi_i(\mathbf{v}) = 0$ for μ-a.e. $\mathbf{v}$ with $v_j > v_i$.*

Consider first the DA-STB.

Theorem 2. *The allocation ϕ^S arising from DA-STB has the following properties:*

(i) Seats at the most popular school $i \in S^{**}$ are assigned only to the students who prefer that school the most.

(ii) The allocation ϕ^S is Pareto optimal within any two schools.

(iii) There exists no $K \subset S$ with $|K| > 2$ such that ϕ^S is Pareto optimal within K.

Now consider the DA-MTB. Let school $w \in S$ be the worst school if its cutoff under DA-MTB is n. There exists only one worst school for a generic $\mathfrak{m}$.

Theorem 3. *For a generic $\mathfrak{m}$, the allocation ϕ^M from the DA-MTB satisfies the following results:*

(i) For $n \geq 3$, a positive measure of seats at each school are assigned to some students who do not prefer that school the most.

(ii) The allocation ϕ^M is Pareto optimal within $\{i, w\}$ for each $i \in S \backslash \{w\}$.

(iii) There exists no $K \subset S \backslash \{w\}$ with $|K| > 1$ such that ϕ^M is Pareto optimal within K.

(iv) There exists no $K \subset S$ with $|K| > 2$ such that ϕ^M is Pareto optimal within K.

5.2 Welfare Properties of CADA

Let K and J be the set of oversubscribed and undersubscribed schools, respectively. Then, we have the following result.

Theorem 4. *The equilibrium allocation ϕ^* arising from CADA satisfies the following properties:*

(i) It is Pareto optimal within set K of oversubscribed schools.

(ii) If all but one $(n-1)$ schools are oversubscribed, then the equilibrium allocation is Pareto optimal.

(iii) The allocation from CADA (which is based on STB as pre-adjusted tie-breaking rule) is Pareto optimal within any $\{i, j\} \subset S$.

The proof (in Appendix) parallels the argument for the First Welfare Theorem, and this observation is suggestive also of the benefits associated with the "choice signaling." Essentially, students' signaling of their choices creates "competitive market" in a limited extent, in which the congestion serves as a price that sorts out the students with different marginal rates of substitution efficiently.

Remark 1. *Even though the resulting allocation may not be fully Pareto optimal, the fact that Pareto efficiency is achieved across oversubscribed schools seems appealing, since it is the sorting across popular schools that are accomplished by CADA.*

Although Theorem 2 shows that adding one undersubscribed school to oversubscribed schools preserves Pareto efficiency, as the following counter example shows, it cannot be generalized when more than one school is undersubscribed.

Example 1. *There are four schools, $S = \{1, 2, 3, 4\}$, and two types of students $\mathcal{V} = \{\mathbf{v}^1, \mathbf{v}^2\}$, with $\mu(\mathbf{v}^1) = 3 - \varepsilon$ and $\mu(\mathbf{v}^2) = 1 + \varepsilon$ where ε is a small number.*

	v_j^1	v_j^2
$j = 1$	10	10
$j = 2$	3	5
$j = 3$	1	4
$j = 4$	0	0

In this case, type 1 *students submit school* 1 *and type* 2 *students submit school* 2*. More specifically, the allocation* ϕ^* *has* $\phi^*(\mathbf{v}^1) = (\frac{1}{3-\varepsilon}, 0, \frac{2-\varepsilon}{2(3-\varepsilon)}, \frac{2-\varepsilon}{2(3-\varepsilon)})$ *and* $\phi^*(\mathbf{v}^2) = (0, \frac{1}{1+\varepsilon}, \frac{\varepsilon}{2(1+\varepsilon)}, \frac{\varepsilon}{2(1+\varepsilon)})$.[7] *While schools* 1 *and* 2 *are oversubscribed, this allocation is not Pareto optimal within* $\{1, 2, 3\}$ *since type* 1 *students can trade probability shares of school* 1 *and* 3 *in exchange for probability share at* 2*, with type* 1 *students, e.g., decreasing* $\phi_1^*(\mathbf{v}^1)$ *by* δ *and* $\phi_3^*(\mathbf{v}^1)$ *by* 4δ *while increasing* $\phi_2^*(\mathbf{v}^1)$ *by* 5δ *for some small* δ*.*

Theorem 4 refers to the endogenous property of an equilibrium, namely the set of over/under-subscribed schools. We provide some characterization on these sets below. Let

$$S^* := \{i \in S | m_i \geq 1\}$$

be the set of *popular* schools which each cannot accommodate all the students who prefer them the most. It is useful to notice, for a later purpose, that a most popular school must be popular, i.e., $S^{**} \subset S^*$.

Proposition 2. *Every allocation arising from CADA is Pareto optimal within* S^**.*

Proposition 2 provides a lower bound for the set of oversubscribed school.

Clearly, Proposition 2 is not very useful if the students have similar preferences. In particular, if all students have the same ordinal preferences, then $|S^*| = 1$, so Proposition does not have bite. Yet, the set of oversubscribed schools can be much bigger than S^* even in this case. We can provide some insight into this question, by introducing more structure into the preferences.

Suppose all students have the uniform ordinal preferences, with the schools indexed by the uniform ranking. Let $\mathcal{V}^U$ be the set of vNM values such that $v_1 > ... > v_n$ for all $\mathbf{v} \in \mathcal{V}_U$. Define

$$\mathcal{V}_2^U := \left\{ \mathbf{v} \in \mathcal{V}^U \middle| \frac{\sum_{i=1}^n v_i}{n} < v_2 \right\}.$$

Proposition 3. *The allocation resulting from CADA is Pareto optimal within* $\{1, 2, ..., k\}$ *for some* $k \geq 2$*, if* $\mu(\mathcal{V}_2^U) \geq 1$*.*

[7] This allocation ϕ^* is derived by CADA-MTB. While CADA-STB induces different allocation, the following argument still holds.

5.3 Comparison of Procedures

First, we show that CADA is superior to DA-STB and DA-MTB in the sense that the former is more likely to achieve Pareto efficiency than the latter.

Corollary 1. *The allocation resulting from CADA is Pareto optimal if:*

(i) All but one schools are popular, or

(ii) $n = 3$ and all students have the same ordinal preferences and $\mu(\mathcal{V}_2^U) \geq 1$ holds.

However, for a generic $\mathfrak{m}$, allocations from DA-STB or DA-MTB cannot be Pareto optimal for $n > 2$.

The following two observations describe the sense in which CADA is superior to DA-STB and the sense in which the latter is superior to DA-MTB.

Corollary 2. *For each set $K \subset S$ such that the allocation from DA-STB is Pareto optimal within K, there exists $K' \supset K$ such that the allocation resulting from CADA is Pareto optimal within K', where $|K'| > |K|$ if $|S^*| > 2$.*

Comparison of Theorem 2 and 3 is summarized as follows.

Corollary 3. *A most popular school is more likely to be assigned to those who prefer the most than the others under DA-STB than under DA-MTB. The allocation arising from DA-STB is Pareto optimal within any 2 schools, whereas the allocation from DA-MTB generically fails to be Pareto optimal within two schools unless they contain a worst school.*

Example 2. *There are three schools, $S = \{1, 2, 3\}$, and three types of students $\mathcal{V} = \{\mathbf{v}^1, \mathbf{v}^2, \mathbf{v}^3\}$, each with $\mu(\mathbf{v}^i) = 1$.*

	v_j^1	v_j^2	v_j^3
$j = 1$	5	4	1
$j = 2$	1	2	5
$j = 3$	0	0	0

Notice that $S^{**} = \{1\}$ and $S^{*} = \{1, 2\}$. It follows from Corollary 1 that the allocation from CADA is Pareto optimal. More specifically, the equilibrium allocation is $\phi^*(\mathbf{v}^1) = \phi^*(\mathbf{v}^2) = (\frac{1}{2}, 0, \frac{1}{2})$ and $\phi^*(\mathbf{v}^3) = (0, 1, 0)$.

The allocation from DA-STB is Pareto optimal within any pair of two schools: $\phi^S(\mathbf{v}^1) = \phi^S(\mathbf{v}^2) = (\frac{1}{2}, \frac{1}{6}, \frac{1}{3})$ and $\phi^S(\mathbf{v}^3) = (0, \frac{2}{3}, \frac{1}{3})$.[8] *This allocation is not Pareto optimal since student 1 can trade probability shares of schools 1 and 3 in exchange for probability share at school 2, with student 2.*

The allocation from DA-MTB is $\phi^M(\mathbf{v}^1) = \phi^M(\mathbf{v}^2) \approx (0.392, 0.274, 0.333)$ and $\phi^M(\mathbf{v}^3) \approx (0.215, 0.451, 0.333)$.[9] *This is not Pareto optimal within $\{1, 2\}$.*

Example 3. *There are three schools, $S = \{1, 2, 3\}$, and two types of students $\mathcal{V} = \{\mathbf{v}^1, \mathbf{v}^2\}$, with $\mu(\mathbf{v}^1) = 2$ and $\mu(\mathbf{v}^2) = 1$.*

	v_j^1	v_j^2
$j = 1$	3	3
$j = 2$	1	2
$j = 3$	0	0

In this case, $S^* = S^{**} = \{1\}$. Yet, it follows from Corollary 1 that the allocation from CADA is Pareto optimal. More specifically, the allocation ϕ^* has $\phi^*(\mathbf{v}^1) = (\frac{1}{2}, 0, \frac{1}{2})$ and $\phi^*(\mathbf{v}^2) = (0, 1, 0)$.

DA-STB and DA-MTB entail the same allocation $\phi^{DA}(\mathbf{v}^1) = \phi^{DA}(\mathbf{v}^2) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, which is Pareto optimal within any pair of two schools. (The result of Theorem 3 does not hold for DA-MTB, because its assumption of full-support.) This allocation is not Pareto optimal since type 1 students can trade probability shares of school 1 and 3 in exchange for probability share at 2, with type 2 students.

Example 4. *There are three schools, $S = \{1, 2, 3\}$, and two types of students $\mathcal{V} =$*

[8] Assuming that each student has a single uniform draw from $[0, 3]$, the cutoff for school 1 is $c_1 = 1.5$, the cutoff for school 2 is $c_2 = 2$, and the one for school 3 is 3.

[9] Again, assuming that each student has a uniform draw from $[0, 3]$ for each school separately, the cutoff for school 1 is $c_1 = \frac{5-\sqrt{7}}{2} \approx 1.177$, the cutoff for school 2 is $c_2 \approx 1.354$, and the one for school 3 is 3.

$\{\mathbf{v}^1, \mathbf{v}^2\}$, *each with* $\mu(\mathbf{v}^1) = 2$ *and* $\mu(\mathbf{v}^2) = 1$.

	v_j^1	v_j^2
$j = 1$	10	10
$j = 2$	1	2
$j = 3$	0	0

In this example, the allocation arising from CADA is not Pareto optimal. All students subscribe to school 1 in equilibrium, so the allocation ϕ^* *is* $\phi^*(\mathbf{v}^1) = \phi^*(\mathbf{v}^2) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, *just as with DA-STB and DA-MTB.*

In fact, if $\mathcal{V} = \mathcal{V}^U$, we can show that the CADA Pareto-dominates the DA. An (ex ante) assignment for student $\mathbf{v}$ is an element of Δ^n. An allocation ϕ is a mapping $\phi : \mathcal{V} \mapsto \Delta^n$ such that $\int \phi_i(\mathbf{v}) d\mu(\mathbf{v}) = 1$ for each school i. Let $\mathcal{X}$ be the set of all allocations and $\phi^{DA} \in \mathcal{X}$ be the allocation induced by the DA rule under a well-defined random tie-breaking rule. Since every student submits the same ranking order over schools, we have

$$\phi^{DA}(\mathbf{v}) = (\frac{1}{n}, ..., \frac{1}{n}) \text{ for all } \mathbf{v}.$$

Consider now CADA with any random tie-breaking rule. Let $\nu(\mathbf{v}) \in \Delta^n$ be the equilibrium mixed strategy adopted by type $\mathbf{v}$. Let

$$\alpha_i := \int \nu_i(\mathbf{v}) d\mu(\mathbf{v})$$

be the measure of students who pick $i \in S$ as their target. Suppose that the CADA induces an allocation $\psi^{CADA} : S \mapsto \Delta^n$ in equilibrium, whereby a student is assigned to school j with probability ψ_j^{CADA} if she declares i as her target.

Given the design of CADA, we must have

$$\sum_{i \in S} \alpha_i \psi_j^{CADA}(i) = 1, \, \forall j \in S.$$

In equilibrium, each student must play its best response. Hence, for a student with $\mathbf{v}$, if $j \in \text{supp}(\nu(\mathbf{v}))$, then

$$j \in \arg\max_{i \in S} \sum_k v_k \psi_k^{CADA}(i).$$

Suppose now she randomizes over her auxiliary message by choosing school i with probability

$$y_i := \frac{\alpha_i}{\sum_j \alpha_j} = \frac{\alpha_i}{n}.$$

Then, she will get the expected payoff of

$$\sum_k v_k \left(\sum_j y_j \psi_k^{CADA}(j) \right)$$
$$= \sum_k \frac{v_k}{n},$$

which is precisely what she will get under the DA algorithm. Since the randomization is not necessarily (and generally) optimal, we arrive at the following conclusion.

Theorem 5. *Suppose all students have the same ordinal preferences. For a DA with any random tie-breaking rule, there exists a CADA which makes all students weakly better off than they are from the DA.*

Corollary 4. *Assume that the schools have no priorities. Any allocation resulting from CADA Pareto-dominates the allocation arising from DA if* $|S^*| > 2$.

6 Simulations

Although the theoretical results in the previous sections are in favor of CADA, they do not speak to the magnitude of efficiency gains or loses in each mechanism. In this section, we numerically investigate these questions via simulations. Our simulations also help highlight the sources of efficiency gains and loses.

In our numerical model, we have 5 schools each with a capacity of 20 seats and 100 students. Student i's vNM value for school j, v_{ij}, is given by

$$v_{ij} = \alpha u_j + (1 - \alpha) u_{ij}$$

where $\alpha \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ and u_j are common across students and u_{ij} is specific to student i and school j. For each α, we construct an experiment as follows: We draw $\{u_j\}$ and $\{u_{ij}\}$ uniformly randomly to construct student preferences. We run four experiments for each preference profile:

1. DA-STB: A single random ordering of students is generated uniformly randomly. Then the assignment is found by DA-STB. This is repeated 2,000 times.

2. DA-MTB: A random ordering of students is generated for every school independently and uniformly randomly. Then the assignment is found by DA-MTB. This is repeated 2,000 times.

3. CADA: The complete information Nash equilibrium is computed. Given other students' target school choices, a student's best response for target school choice is computed by running CADA with 2,000 different random tie breakings. Each time, a single random ordering is used to break ties at schools. An equilibrium is found when no student updates her best response.

4. CADA with naive play (NAIVE): Each student's target school is set to her first choice. CADA is run with 2,000 different random tie breakings. Each time, a single random ordering is used to break ties at schools.

We calculate the averages from 64 iterations of each experiment in figures 1-8. Welfare generated in an iteration of an experiment is the sum of the expected utilities of students in that iteration, which runs 2,000 independent randomizations for tie breaking. Then we compute the welfare generated by a mechanism by averaging the welfare generated in all iterations of the associated experiment (Figures 1-4). Similarly, the probability that a sample student gets her first choice is the average of the probability of getting first choice for all students and all iterations (Figures 5,6). The average number of oversubscribed schools in CADA is computed by averaging the number of oversubscribed schools in all iterations of that experiment (Figure 7). The average number of students selecting their k-th choice as her target school in equilibrium in CADA is computed by averaging the number of students selecting their k-th choice as her target school in all iterations of that experiment (Figure 8).

Figures 1-4 compare welfare across the four mechanisms. Our benchmark in these figures is DA-STB and we calculate a mechanism's percentage welfare gain in comparison to DA-STB. Figures 3-6 aim to uncover numerically the source of the differences in welfare across the mechanism. Finally, Figures 7 and 8 show equilibrium behavior in CADA.

If a student's vNM value for a school increases, the likelihood of the student selecting that school as target in an equilibrium of CADA, therefore the likelihood of the student getting that school, are expected to increase. This feature of CADA contributes to welfare. Figure 1 show significant welfare gains in CADA in comparison to DA-STB. The percentage gain raises from 0.53% for $\alpha = 0$ to 3.62% for $\alpha = 0.7$, then drops to %2 for $\alpha = 0.9$. The correlation among preferences is smaller for smaller values of α. Therefore, competition for a school takes place among a smaller number of students for small values of α. As a result, the welfare gain in CADA from allocating schools to students with higher vNM values is smaller for smaller values of α below 0.7. The gain increases monotonically as α increases to 0.7. At the very extreme where $\alpha = 1$, i.e. every student has the same vNM values, every random allocation would be exante efficient, therefore we would not have any welfare gain in CADA. Figure 1 verifies that by showing a monotone decrease in percentage welfare gain as α increases from 0.7 to 0.9. It is worthwhile to note that the gain is still 2% for $\alpha = 0.9$, which suggests that convergence to zero welfare gain as α goes to 1 is slow, i.e. CADA continues to do its job even for higher values of α.

Since CADA is more complicated to play for students than DA-STB is, one natural question is whether students would be harmed if they did not play their equilibrium strategies in CADA. Without any theoretical justification we can offer, it seems natural to us that a naive player would simply select her first choice as her target school, as selecting any other school definitely signals some thinking. Therefore, in order to investigate the question we post above, we assume that every student plays naively and select her first choice as her target school. We still see welfare gains with naive play in CADA, ranging from 0.2% for $\alpha = 0$ to 0.94% for $\alpha = 0.6$. This result conforms with the principle of "first no harm" of market design.

Finally, the welfare loss in DA-MTB in comparison to DA-STB in Figure 1 is almost the inverse image of the gain associated with CADA, reaching its deep at $\alpha = 0.5$.

In order to understand the sources of welfare gains/losses numerically, we focus on specific sets of schools next. In Figure 2, we look at the welfare gains conditional on receiving an oversubscribed school. As this figure demonstrates, most of the welfare gain in CADA is associates with allocating oversubscribed schools more efficiently. In particular, the welfare gain in CADA is, on average, 40% higher in Figure 2 than in Figure 1. The welfare gain in the naive play of CADA is almost doubled on average in Figure 2. Surprisingly, the welfare loss associated with DA-MTB is also doubled on average in Figure 2. We repeat the same exercise for popular schools in Figure 3 and for most popular schools in Figure 4. In particular, we observe gains in the equilibrium and naive plays of CADA and loss in DA-MTB. The gain in CADA over popular schools (Figure 3) is 16% less on average than the overall gain (Figure 1), and the gain in CADA over most popular schools (Figure 4) is 42% less on average than the overall gain (Figure 1). This further confirms that the more efficiency gain is associated with the allocation of oversubscribed schools.

Next we explore the probability of getting the first choice school in Figure 5 an getting one the first and second choices in Figure 6. Figure 5 reveals that the equilibrium of CADA performs slightly better than DA-STB in assigning students to their first choices. The naive play of CADA performs the best. This is expected since every student gets priority in her first choice in the naive play. Finally, DA-MTB performs worst among the four. Figure 6 shows that all four perform equally well in assigning students to one of their first and second choices. Figure 1 and Figure 6 together suggest that although CADA and DA-STB assign similar number of students to their first and second choices, CADA does a better job in allocating these schools.

A comparison of the equilibrium play of CADA with naive play of CADA sheds further light on the workings of CADA in improving welfare. The welfare gain is always larger in equilibrium in Figure 1. In particular, for $\alpha = 0.9$, the gain is 2% in equilibrium and it is 0.8% in the naive play of CADA. However, in Figure 5, the probability of getting the first choice is higher in the naive play (0.236) than in the equilibrium play (0.224). This suggest that some of the welfare gain associated with the equilibrium play of CADA comes from allocating first choice schools more efficiently. As we mention above, the same is true for first and second choice schools as well.

Finally, Figure 7 shows that the average number of oversubscribed schools varies between 2.5 and 3.1. As Figure 8 shows, when $\alpha = 0$, almost all of the students (96.8) select their fist choice as target, a few (3.1) select her second choice and almost zero (0.1) select her third choice as target. None select fourth choice as target. Note that it is never optimal to select the fifth (last) choice as target. As α increases, the number of students selecting their fist choice as target decreases monotonically and the number of students selecting any other choice increases monotonically. In particular, for $\alpha = 0.9$, on average 57.1 students select first choice, 30.2 select second choice, 11.8 select third choice and 0.9 select fourth choice as target. This monotone pattern in behavior can be explained by the extend of competition over schools. As α, therefore correlation among preferences, increases, more people rank the same school as first choice. Therefore competition for one's first choice becomes more intense. This gives incentives to students to compete for their second, third and even fourth choices by selecting them as their target school in CADA.

7 Extension: Enriching the Auxiliary Message

The auxiliary message can be expanded to include more than one school, perhaps at the expense of becoming less practical. In general, the auxiliary message can include a rank order of schools up to $k \leq n$, with the tie broken in the lexicographic fashion according to this rank order: A student is reordered to be ahead of another one at the priority list of school $i \in S$ if and only if the former ranks it higher than the latter in the auxiliary message. We call the associated CADA a ***CADA of degree*** k.

In fact, the CADA can be designed to coincide with the Boston mechanism if the schools have no priorities and *all students have the same ordinal preferences.* One can simply augment the auxiliary message to include an entire separate preference list of schools; and they can be used to construct the priority lists for schools. It is easy to see how the resulting CADA implements the Boston mechanism.[10] In the outcome implemented, no school admits a student that ranks it lower ahead of the one that ranks it higher in the auxiliary message. Such an enriching of the auxiliary message does not alter the qualitative feature of CADA (albeit perhaps making it more complicated and less practical), in particular rendering Theorem 5 applicable. We thus have the following rather conclusion.

[10]In general, when every student has the same ordinal preferences, the CADA of degree k coincides with the Boston mechanism in which the students can submit up to k schools.

Remark 2. *If all students have the same ordinal preferences and the schools have no priority, then the Boston mechanism weakly Pareto dominates the DA algorithm.*

Expanding the auxiliary message will certainly complicate the deliberation on the part of students participating in the school choice program and becomes practically cumbersome. The beauty of CADA is that the auxiliary message can be kept as simple as practically manageable, if necessary, to $k = 1$ as has been considered before. We explore what additional benefit may be obtained by adding more schools in the auxiliary message.

The nature of welfare gain it may bring can be understood with the following examples. First, we can note that enriching the message does not generally guarantee full Pareto efficiency. Consider Example 4 again. Allowing the students to include the second message does not make any difference: All students will pick school 1 as their first target and school 2 as their second target, and the precisely the same allocation will arise in equilibrium (which also coincides with one arising from DA-STB).

The enriching of message can have a second-order effect which can make a difference in the following two examples. The first example illustrates how the additional signaling made possible by a richer message improves the allocation.

Example 5. *There are 4 schools, $S = \{1, 2, 3, 4\}$, and two types of students $\mathcal{V} = \{\mathbf{v}^1, \mathbf{v}^2\}$, each with $\mu(\mathbf{v}^1) = 3$ and $\mu(\mathbf{v}^2) = 1$.*

	v_j^1	v_j^2
$j = 1$	20	20
$j = 2$	4	3
$j = 3$	1	2
$j = 4$	0	0

With CADA of degree 1, all students subscribe to school 1, so the allocation is completely random with $\phi^*(\mathbf{v}^j) = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, $j = 1, 2$. With CADA of degree 2, all students pick school 1 as their first target; but type 1 students pick school 2 as their second target whereas type 2 students pick school 3 as their second target. Consequently, the allocation becomes $\phi^{**}(\mathbf{v}^1) = (\frac{1}{4}, \frac{1}{3}, \frac{1}{12}, \frac{1}{3})$ and $\phi^{**}(\mathbf{v}^2) = (\frac{1}{4}, 0, \frac{3}{4}, 0)$. This allocation ϕ^{**} Pareto *dominates ϕ^*, although the former is not Pareto optimal (since the two types could benefit further from trading shares between school 1 and school 3).*

A richer message need not be always better. A richer message space clearly generates more opportunity for a student to self select at different tiers of schools. But the alternative opportunities may work as substitutes and militate each other. For instance, an opportunity to self select at a lower tier of schools may reduce a student's incentive to self select at a higher tier of schools, even though the latter kinds of self selection may be more important from the social welfare perspective. This kind of "crowding out" arises in the next example.

Example 6. *(Crowding Out) There are 4 schools, $S = \{1, 2, 3, 4\}$, and two types of students $\mathcal{V} = \{\mathbf{v}^1, \mathbf{v}^2\}$, with $\mu(\mathbf{v}^1) = 3$ and $\mu(\mathbf{v}^2) = 1$.*

	v_j^1	v_j^2
$j = 1$	12	8
$j = 2$	2	4
$j = 3$	1	3
$j = 4$	0	0

Consider first CADA of degree 1. Here, all type 1 students choose school 1 as their target, and all type 2 students choose school 2 as their target. In other words, the latter type of students self select into the second popular school. The resulting allocation is $\phi^*(\mathbf{v}^1) = (\frac{1}{3}, 0, \frac{1}{3}, \frac{1}{3})$ and $\phi^*(\mathbf{v}^2) = (0, 1, 0, 0)$. The expected utilities are $EU^1 = 4.33$ and $EU^2 = 4$. In fact, this allocation is Pareto optimal.

Suppose now CADA of degree 2 is used. In equilibrium, type 1 students choose school 1 and 2 as their first and second targets, respectively. Meanwhile, type 2 students choose school 1 (instead of school 2!) as their first target and school 3 as their second target. Here, the opportunity for type 2 students to self select at a lower tier school (school 3) blunts their incentive to self select at a higher tier school (school 2). The resulting allocation is thus $\phi^{**}(\mathbf{v}^1) = (\frac{1}{4}, \frac{1}{3}, \frac{1}{12}, \frac{1}{3})$ and $\phi^{**}(\mathbf{v}^2) = (\frac{1}{4}, 0, \frac{3}{4}, 0)$, which yield expected utilities of $\overline{EU}^1 = 3.75$ and $\overline{EU}^2 = 4.25$. This allocation is not Pareto optimal since type 2 *students can trade probability shares of school* 1 *and* 3 *in exchange for probability share at* 2*, with type* 2 *students.*

Even though ϕ^* does not Pareto dominate ϕ^{**}, the former is Pareto optimal whereas the latter is not. Further, the former is superior to the latter in the Utilitarian sense (recall that students' payoffs are normalized so that they aggregate to the same value for both types): the former gives aggregate utilities of 17 *as opposed to* 15.5 *the latter gives.*

The last example suggests that the benefit from enriching the message space is not unambiguous. This is a potentially important point. In practice, expanding a message space adds a burden on the parents to be strategically more sophisticated. Hence avoiding such a demand for strategic sophistication is an important quality for a procedure to succeed. This makes the simple CADA (i.e., of degree 1) quite appealing. That this practical benefit may not even involve a welfare sacrifice is reassuring about the simple CADA.

8 Conclusion

In this paper, we propose a new deferred acceptance procedure in which students are allowed, via signaling of their preferences, to influence how they are treated in a tie for a school. This new procedure, choice-augmented DA algorithm (CADA), makes the most of two existing procedures, the Gale-Shapely's deferred acceptance algorithm (DA) and the Boston mechanism. While the DA achieves the strategyproofness, a desideratum in the design of school choice programs, it limits students' abilities to communicate their preference intensities, which entails an ex ante inefficient allocation when schools are indifferent among students with the same ordinal preferences. The Boston mechanism, on the other hand, is responsive to the agents' cardinal preferences and may achieve more efficient allocation than the DA, but fails to satisfy strategyproofness. We show that by creating a channel of revealing students' cardinal preference the CADA implements a more efficient allocation than the DA without sacrificing the strategyproofness of ordinal preferences.

References

[1] Abdulkadiroğlu, A., Pathak, P., and Roth, E. (2005a), "The New York City High School Match," *American Economic Review*, Papers and Proceedings, 95, 364-367.

[2] Abdulkadiroğlu, A., Pathak, P., and Roth, E. (2006a), "Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match," Unpublished manuscript, Duke University.

[3] Abdulkadiroğlu, A., Pathak, P., Roth, E., and Sönmez, T. (2005b), "The Boston Public School Match," *American Economic Review*, Papers and Proceedings, 95, 368-371.

[4] Abdulkadiroğlu, A., Pathak, P., Roth, E., and Sönmez, T. (2006b), "Changing the Boston Mechanism: Strategyproofness as Equal Access," Unpublished manuscript, Harvard University.

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

[6] Casella, A. (2005), "Storable Votes," *Games and Economic Behavior*, 51, 391-419.

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

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

[9] Ergin, H., and Sönmez, T. (2006), "Games of School Choice under the Boston Mechanism," *Journal of Public Economics*, 90, 215-237.

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

[11] Jackson, M., and Sonnenschein, P. (forthcoming), "Overcoming Incentive Constraints by Linking Decisions," *Econometrica*.

[12] Mas-Colell, A. (1984), "On a Theorem of Schmeidler," *Journal of Mathematical Economics*, 13, 201-206.

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

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

[15] Sönmez, T., and Ünver, M.U. (2003), "Course Bidding at Business Schools," Unpublished manuscript, Koc University.

9 Appendix: The Mechanisms and Proofs

The DA algorithm can be generalized to the case with continuum of students. We need to introduce the following additional notation for this purpose. A strict ordering of students at school i is a bounded measurable function $F_i : \mathcal{V} \to \mathbb{R}$ such that $\mu(\{\mathbf{v}' : F_i(\mathbf{v}') = F_i(\mathbf{v})\}) = 0$ for a.e. $\mathbf{v} \in \mathcal{V}$. Let $F = \{F_i : i \in S\}$ denote a profile of strict orderings of students at schools.

Given a strict ordering of students F_i at school i, define a measurable choice function Ch_i over subsets of $\mathcal{V}$ as the set of best ranked students for school $i \in S$ according to F_i from a given set up to the capacity. Formally, for any measurable $X \subset \mathcal{V}$, (i) $Ch_i(X) \subset X$, that is $Ch_i(X)$ is chosen from X; (ii) $\mu(Ch_i(X)) \leq 1$, that is the chosen set does not exceed the capacity; (iii) $F_i(\mathbf{v}) < F_i(\mathbf{v}')$ for all $\mathbf{v} \in Ch_i(X)$ and $\mathbf{v}' \in X \backslash Ch_i(X)$, that is better ranked students are chosen.

A profile of students' ordinal preferences is given by a measurable function $P : \mathcal{V} \to \Pi$, where $P(\mathbf{v}) \in \Pi$ denotes $\mathbf{v}$'s (not necessarily true) ordinal preferences. Define the DA (deferred acceptance) mapping as follows: For every $P : \mathcal{V} \to \Pi$, $P' = DA(P) \in \Pi$ is determined as follows: Every student $\mathbf{v}$ applies to her most preferred school in $P(\mathbf{v})$. Every school i (tentatively) admits from its applicants in the order of F_i. If all of its seats are assigned, it rejects the remaining applicants. If a student $\mathbf{v}$ is rejected by i, $P'(\mathbf{v})$ is obtained from $P(\mathbf{v})$ by deleting i in $P(\mathbf{v})$. If a student $\mathbf{v}$ is not rejected, then $P'(\mathbf{v}) = P(\mathbf{v})$. More formally, let $T_i(P) = \{\mathbf{v} \in \mathcal{V} : i \text{ is ranked first in } P(\mathbf{v})\}$ be the set of students that rank i as first choice. Note that $T_i(P)$ is measurable. Then each school i admits students in $Ch_i(T_i(P))$ and rejects students in $T_i(P) \backslash Ch_i(T_i(P))$. If $\mathbf{v} \in T_i(P) \backslash Ch_i(T_i(P))$ for some $i \in S$, then $P'(\mathbf{v})$ is obtained from $P(\mathbf{v})$ by deleting i from the top of $P(\mathbf{v})$; otherwise $P'(\mathbf{v}) = P(\mathbf{v})$. Since P is a measurable function, P' is also measurable.

Repeated application of the DA mapping gives us the DA algorithm. That is, given a problem (P, F), let $P^0 = P$ and define $P^t = DA(P^{t-1})$ for $t > 0$. Then P^t converges almost everywhere to some measurable P^* (Proposition ? below). The matching generated by the DA algorithm is found by assigning every $\mathbf{v}$ to her top choice in $P^*(\mathbf{v})$. Denote that matching by $\lambda(P, F)$.

Existence

Proposition 4. *For every (P, F), P^t converges almost everywhere to some measurable $P^* : \mathcal{V} \to \Pi$.*

Proof. Define the set of rejected students as $R^t = \{\mathbf{v} : \mathbf{v} \in T_i(P^t) \backslash Ch_i(T_i(P^t))$ for some $i \in S\}$. Then $\mu(R^t)$ goes to zero as t goes to infinity. Otherwise, if $\mu(R^t) \geq \kappa > 0$ for all t, all the schools in every student's preference would be deleted in finite time because of finiteness of the number of schools, which in turn would imply that $\mu(R^t)$ goes to zero, a contradiction. Therefore, P^t converges almost everywhere to some P^*. Since every P^t is measurable, P^* is also measurable. ▌

Now we are ready to define our mechanisms. Before that let us give the following general result:

Ordinal Strategy-proofness

Let $P_{-\mathbf{v}} : \mathcal{V} \backslash \{\mathbf{v}\} \to \Pi$ denote the profile of ordinal preferences of all students but $\mathbf{v}$ determined by P. Let $P^{true}(\mathbf{v}) \in \Pi^n$ be the preference relation induced by $\mathbf{v}$, that is $iP^{true}(\mathbf{v})j$ if and only if $\mathbf{v}_i > \mathbf{v}_j$.

Proposition 5. *For every (P, F), it is a (weak) dominant strategy for every student to submit her ordinal preferences truthfully to DA, that is, for all $\mathbf{v} \in \mathcal{V}$, $P_{-\mathbf{v}} : \mathcal{V} \backslash \{\mathbf{v}\} \to \Pi$ and $\tilde{P} \in \Pi$,*

$$\mathbf{v}_{\lambda(P^{true}(\mathbf{v}), P_{-\mathbf{v}}, F)} \geq \mathbf{v}_{\lambda(\tilde{P}, P_{-\mathbf{v}}, F)}$$

Proof. Recall that $P(\mathbf{v}) \in \mathcal{P}$ does not need to be equal to, $P^{true}(\mathbf{v})$, the preference relation that is induced by $\mathbf{v}$. Also one needs only ordinal information for DA. Therefore in this proof we will treat $\mathbf{v}$ as a name index that keeps track of the ordinal information. Given any $K \in \mathbb{N}_+$, construct a discretization of (P, F) for $\mathbf{v}$ as follows: For every $i = 1, ..., n$ and $z = 0, ..., K$, set $v_{i,z} = zh_i$ where $h_i = \frac{1}{K}$. Now consider the following sets: $V_{z_1 \ldots z_n} = \{\mathbf{v} : v_{i,z_i} \leq v_i \leq v_{i,z_i+1}\}$ for $z_i = 0, ..., K-1$ and $i = 1, ..., n$. Let $\mu_{K,\min} = \min_{V_{z_1 \ldots z_n}} \mu(V_{z_1 \ldots z_n})$. For any $V_{z_1 \ldots z_n}$, let $\#V_{z_1 \ldots z_n}$ denote the integer part of $\frac{\mu(V_{z_1 \ldots z_n})}{\mu_{K,\min}}$. Pick $\mathbf{v}$. In addition, pick $\#V_{z_1 \ldots z_n}$ students in total from every set $V_{z_1 \ldots z_n}$ without repetition. Let $\{\mathbf{v}_l\}$ denote the set of students that are picked. If $\frac{|\{\mathbf{v}_l\}|}{n}$ is not an integer, pick additional students from the larger sets until obtaining an integer $\frac{|\{\mathbf{v}_l\}|}{n}$. Note that the number of additional students to be picked this way is less than n and n is fixed, therefore this will be negligible in the limit as K goes to infinity. Construct a problem with $|\{\mathbf{v}_l\}|$ students and n schools each with capacity $\frac{|\{\mathbf{v}_l\}|}{n}$. Each student $\mathbf{v}_l$'s strict ordinal preference relation is given by $P(\mathbf{v}_l)$. Student orderings at every school i are given by $\{F_i(\mathbf{v}_l)\}$. Recall that $\mathbf{v}_l$ is ranked higher than $\mathbf{v}'_l$ at i if $F_i(\mathbf{v}_l) < F_i(\mathbf{v}'_l)$. Denote this problem by $(P, F)_K$. As K goes to infinity, $(P, F)_K$ approximate the functions (P, F) arbitrarily closely. Suppose to the contrary that there is a problem (P, F) and $\mathbf{v} \in \mathcal{V}$ such that $P(\mathbf{v}) \neq P^{true}(\mathbf{v})$ and $\mathbf{v}_{\lambda(P,F)} > \mathbf{v}_{\lambda(P^{true}(\mathbf{v}),P_{-\mathbf{v}},F)}$, that is $\mathbf{v}$ prefers her DA matching under (P, F) to her DA matching when she submits her true preference relation. Equivalently, $P^{true}(\mathbf{v})$ does not solve $\mathbf{v}$'s optimization problem

$$\max_{\tilde{P}} \mathbf{v}_{\lambda(\tilde{P},P_{-\mathbf{v}},F)}$$

Then, since the (P, F) functions can be approximated arbitrarily closely by a $(P, F)_K$, there exists a finite problem $(P, K)_K$, a discretization of (P, F) for $\mathbf{v}$, such that $P^{true}(\mathbf{v})$ does not solve $\mathbf{v}$'s optimization problem in that finite problem. This contradicts with the well-know result that in every finite problem, submitting true preferences to the student-proposing deferred acceptance mechanism is a dominant strategy for every student (Dubins and Friedman 1981, Roth 1982). ▌

9.1 Deferred Acceptance with Single Tie-Breaking (STB)

The STB rule can be easily characterized as follows. Students are ranked at every school by a uniformly randomly drawn strict ordering $\tilde{F} : \mathcal{V} \to [0, n]$ which is common to every school.[11] Then the DA-STB assignment is given the by corresponding DA algorithm, $\lambda(P, \{F_i = \tilde{F} \text{ all } i\})$. We refer $\omega = \tilde{F}(\mathbf{v})$ as $\mathbf{v}$'s random draw.

We prove Lemma 2 next.

Proof of Lemma 2. Consider the first cutoff $\hat{c}^1$. Suppose this is the cutoff for school i. Take any student. If the student's top choice is not i. Then, if she ever gets to school i — meaning she is turned down by schools she lists ahead of i — then it means that

[11] More formally, let $\mathcal{F} = \{F |\ F : \mathcal{V} \to [0, n],\ F$ is measurable, $\mu(\{\mathbf{v}' : F(\mathbf{v}') = F(\mathbf{v})\}) = 0$ for a.e.$\mathbf{v} \in \mathcal{V}\}$. A random tie breaker F is drawn from $\mathcal{F}$ uniformly randomly.

her draw $\omega > c^i$, so she will never get into school i. This means that only students whose most preferred school is i can only get assigned to school i. It then follows that $m_i \cdot \frac{\hat{c}_i^1}{n} = 1$, so $\hat{c}_i^1 = \frac{n}{m_i}$. For $\hat{c}^1$ to be the lowest cutoff, we must have $\hat{c}^1 = \frac{n}{\max_{i \in S} m_i}$. Hence, a most popular school has the lowest cutoff. We conclude that $\hat{c}^1$ is uniquely determined by $\mathfrak{m}$ (more specifically, by $\max\{m_i\}$, which is a function of $\mathfrak{m}$). Since $\max_{i \in S} m_i \geq 1$, the $\hat{c}^1 \in (0, n]$.

We work recursively to define the rest of the cutoffs. Suppose that cutoffs $\{\hat{c}^j\}$, $j < k$, are uniquely determined by $\mathfrak{m}$ such that $\hat{c}^i \leq \hat{c}^j \leq n$ for all $1 \leq i < j \leq k-1$. Let the cutoff $\hat{c}^j$, $j \leq k-1$, be the cutoff of school $\kappa(j) \in S$, where $\kappa : \{1, ..., k-1\} \mapsto S$ is one-to-one. We now determine $\hat{c}^k$ uniquely as a function of $\mathfrak{m}$ and establish $\hat{c}^k \geq \hat{c}^{k-1}$. Let $S^{k-1} := \{j | j = \kappa(j') \text{ for some } j' \leq k-1\}$ be the associated set. Suppose that the cutoff $\hat{c}^k$ determines the cutoff of school $i \in S \backslash S^{k-1}$. Then, arguing as before, a student who prefers a school $j \in S \backslash (S^{k-1} \cup \{i\})$ to each school in $S^{k-1} \cup \{i\}$ never stands a chance to get in i. (Clearly, κ, S^{k-1} and $\hat{c}^j$ all depend on $\mathfrak{m}$, which we suppress for convenience.)

For any nonempty subset $S' \subset S^{k-1}$, let $\hat{\Pi}(S')$ be the set of all permutations of S'. Let $\chi(S') := \{j \in S' | \kappa^{-1}(j) \geq \kappa^{-1}(j'), \forall j' \in S'\}$ be the school which has the highest index in S^{k-1}, meaning that $\chi(S')$ will be the school that has the largest cutoff among S' (yet still has a lower cutoff than $\hat{c}^k$. Then, for school i to have cutoff $\hat{c}^k$, the cutoff must be $\hat{c}^k = \hat{c}_i^k$, where $\hat{c}_i^k$ satisfies

$$m_i \cdot \frac{\hat{c}_i^k}{n} + \sum_{S' \subset S^{k-1}} \left[\sum_{\tau \subset \hat{\Pi}(S')} m_{(\tau,i)} \left(\frac{\hat{c}_i^k - \hat{c}^{\chi(S')}}{n} \right) \right] = 1,$$

or

$$\hat{c}_i^k = \frac{n + \sum_{S' \subset S^{k-1}} \left[\hat{c}^{\chi(S')} \left(\sum_{\tau \subset \hat{\Pi}(S')} m_{(\tau,i)} \right) \right]}{m_i + \sum_{S' \subset S^{k-1}} \left(\sum_{\tau \subset \hat{\Pi}(S')} m_{(\tau,i)} \right)}.$$

Let $\hat{c}^k := \min\{n, \min_{j \in S \backslash S^{k-1}} \hat{c}_j^k\}$, and $i = \kappa(k) := \arg\min_{j \in S \backslash S^{k-1}} \hat{c}_j^k$. Note that this definition conforms to the case of $k = 1$. We must have

$$m_i \cdot \frac{\hat{c}^k}{n} + \sum_{S' \subset S^{k-1}} \left[\sum_{\tau \subset \hat{\Pi}(S')} m_{(\tau,i)} \left(\frac{\hat{c}^k - \hat{c}^{\chi(S')}}{n} \right) \right] \leq 1, \tag{2}$$

where the inequality holds with equality if $\hat{c}^k < n$. We next show that $\hat{c}^k \geq \hat{c}^{k-1}$. Suppose to the contrary that $\hat{c}^k < \hat{c}^{k-1} \leq n$. We can rewrite (2) (with equality) as

$$m_i \cdot \frac{\hat{c}^k}{n} + \sum_{S' \subset S^{k-2}} \left[\sum_{\tau \subset \hat{\Pi}(S')} m_{(\tau,i)} \left(\frac{\hat{c}^k - \hat{c}^{\chi(S')}}{n} \right) \right] +$$
$$\sum_{S' \subset S^{k-1}, S' \not\subset S^{k-2}} \left[\sum_{\tau \subset \hat{\Pi}(S')} m_{(\tau,i)} \left(\frac{\hat{c}^k - \hat{c}^{\chi(S')}}{n} \right) \right] = 1.$$

Hence,

$$m_i \cdot \frac{\hat{c}^k}{n} + \sum_{S' \subset S^{k-2}} \left[\sum_{\tau \subset \hat{\Pi}(S')} m_{(\tau,i)} \left(\frac{\hat{c}^k - \hat{c}^{\chi(S')}}{n} \right) \right] \geq 1,$$

from which it follows that $\hat{c}_i^k \leq \hat{c}^k < \hat{c}^{k-1}$, contradicting the definition of $\hat{c}^{k-1}$.

Let $l \in S \backslash S^{n-1}$ be the last school left. We prove $\hat{c}_l^n = n$. Recall

$$\hat{c}_l^n = \frac{n + \sum_{S' \subset S^{n-1}} \left[\hat{c}^{\chi(S')} \left(\sum_{\tau \subset \hat{\Pi}(S')} m(\tau, l) \right) \right]}{m_l + \sum_{S' \subset S^{n-1}} \left(\sum_{\tau \subset \hat{\Pi}(S')} m(\tau, l) \right)}. \tag{3}$$

The denominator of (3) measures of all students, so it equals n.[12] The second term in the numerator of (3) becomes, when divided by n,

$$\sum_{S' \subset S^{n-1}} \left[\frac{\hat{c}^{\chi(S')}}{n} \left(\sum_{\tau \subset \hat{\Pi}(S')} m_{(\tau,l)} \right) \right],$$

which measures all students who are assigned to S^{n-1} and thus equals the sum of all terms on the left side of (2) across $k = 1, ..., n-1$. It thus follows that

$$\sum_{S' \subset S^{n-1}} \left[\frac{\hat{c}^{\chi(S')}}{n} \left(\sum_{\tau \subset \hat{\Pi}(S')} m_{(\tau,l)} \right) \right] \leq n - 1. \tag{4}$$

Substituting (4) into (3) gives

$$\hat{c}_l^n \leq \frac{n + n(n-1)}{n} = n.$$

[12] The denominator consists of measures of all students whose most preferred school is l, and of all the student whose second preferred school is l, and so on and so forth, thus telescoping to the sum of all students.

To prove $\hat{c}_l^n = n$, suppose $\hat{c}_l^n < n$. Then, by monotonicity, $\hat{c}^k \leq \hat{c}^n < \hat{c}_l^n \leq n$, so (2) must hold with equality for all $k = 1, ..., n-1$, which means that (4) must hold with equality. Therefore, $\hat{c}_l^n = n$, a contradiction. We conclude that $\hat{c}^n = \hat{c}_l^n = n$.

Although it is possible for $\hat{c}^k = \hat{c}^{k+1}$ for some $i = 1, ..., n-1$, it is easy to see that this is not generic. If $\hat{c}^k = \hat{c}^{k+1}$, this means that there are $i \neq j$ such that $\hat{c}_i^k = \hat{c}_j^{k+1}$, which entails a loss of dimension for $\mathfrak{m}$ within $\mathfrak{M}$. Hence, the Lebesque measure of the set of $\mathfrak{m}$'s involving such a restriction is zero. It thus follows that $\hat{c}^i < \hat{c}^j$ if $i < j$ for a generic $\mathfrak{m}$.

9.2 Deferred Acceptance with Multiple Tie-Breaking

The MTB rule can be similarly characterized as follows. Students are ranked at every school by an independently uniformly randomly drawn strict ordering $F_i : \mathcal{V} \to [0, n]$. Then the DA-MTB assignment is given the by corresponding DA algorithm, $\lambda(P, \{F_i\})$. We refer $\omega_i = F_i(\mathbf{v})$ as $\mathbf{v}$'s random draw at school i.

9.3 Choice Augmented Deferred Acceptance

In CADA, in addition to ordinal preferences over schools, each student sends an auxiliary message of a target school, $s \in S$. Then ties at school priorities are broken as in DA-STB, then at each school, students who select that school as their target are elevated in the randomly generated ranking. Formally, an auxiliary message profile is given by a measurable function $s : \mathcal{V} \to S$, where $s(\mathbf{v}) \in S$ denotes $\mathbf{v}$'s submitted "target" school. Given a tie breaker $\tilde{F}$, and an auxiliary message profile s, generate a measurable function $F_i : \mathcal{V} \to [0, n]$, a strict ordering of students at i as follows:

$$F_i(\mathbf{v}) = \begin{cases} \tilde{F}(\mathbf{v}) & \text{if } s(\mathbf{v}) = i \\ n + \tilde{F}(\mathbf{v}) & \text{if } s(\mathbf{v}) \neq i \end{cases}$$

That is, under F_i, ties are broken first in favor of students who report i as their target school and then according to $\tilde{F}$. F_i is a measurable function since $\tilde{F}$ and s are measurable. Then the CADA assignment is given the by corresponding DA algorithm, $\lambda(P, \{F_i\})$.

Let $\Pi \times S$ denote the set of pure strategies $s = (P, i) \in \Pi \times S$, and Σ the set of (Borel) probability measures on $\Pi \times S$ endowed with the weak convergence topology.

We refer a member of Σ as an action profile (distribution). For any player $\mathbf{v} \in \mathcal{V}$ and any action profile $\sigma \in \Sigma$, let $\sigma_{\mathbf{v}}$ denote the action of $\mathbf{v}$ and $\sigma_{-\mathbf{v}}$ denote the action profile of players $\mathcal{V} \backslash \{\mathbf{v}\}$ induced by σ. For any action profile $\sigma \in \Sigma$, let ϕ^{σ} be the allocation induced by CADA under σ, where $\phi^{\sigma}(\mathbf{v}) = (\phi_i^{\sigma}(\mathbf{v}))_{i \in S}$ and $\phi_i^{\sigma}(\mathbf{v})$ is the probability that player $\mathbf{v}$ is assigned $i \in S$.

Proof of Theorem 1. The proof is a direct application of Theorem 1 and 2 in Mas-Colell (1984).

9.4 Remaining Proofs

Proof of Lemma 3. Consider throughout a student of type $\mathbf{v}$, whose values are all distinct. There are μ-a.e. such $\mathbf{v}$. Suppose her most-preferred school $\pi_1(\mathbf{v}) =: i$ is undersubscribed. It is then her best response to pick i as her target, since doing so can guarantee assignment to i for sure, whereas choosing some other school as the target may result in assignment to some other school. In fact, away from the limit, choosing $j \neq i$ as target will result in positive probability of non-assignment to i. Hence, the student must be choosing i as her target in equilibrium.

Consider any $\mathbf{v}$ (with distinct values). Suppose first $\nu_i^*(\mathbf{v}) > 0$ for some oversubscribed school i. It follows from the above observation that her most preferred school must be an oversubscribed school (not necessarily i). Given the distinct values, she must strictly prefer school i to all undersubscribed schools. Hence, she lists i ahead of all undersubscribed schools in her ordinal list. Whenever she picks i, she will fail to place in any oversubscribed schools other than i that she may list ahead of i, so she will apply to school i with probability one.

Suppose $\nu_j^*(\mathbf{v}) > 0$ for some undersubscribed school j. Then, the student must prefer j to all other undersubscribed schools, so she will apply to j with probability one whenever she fails to place in any oversubscribed school she may list ahead of j in the ordinal list. Whenever she picks j as her target, she is surely rejected by all oversubscribed schools she may list ahead of j, so she will apply to j with probability one.

We thus conclude that $\nu^*(\mathbf{v}) = \bar{\nu}^*(\mathbf{v})$ for μ-a.e. $\mathbf{v}$.

Proof of Lemma 4. Let $J_{ij} := S \backslash \{i, j\}$, and consider any within $\{i, j\}$ reallocation of ϕ, $\tilde{\phi} \in \mathcal{X}_{\phi_{J_{ij}}}$, that weakly Pareto dominates ϕ. Let $\mathcal{V}^j := \{\mathbf{v} \in \mathcal{V} | v_j > v_i\}$. Then, we must have $\tilde{\phi}_i(\mathbf{v}) \geq \phi_i(\mathbf{v})$ for μ-a.e. $\mathbf{v} \in \mathcal{V} \backslash \mathcal{V}^j$ (since any shifting of probability shares away from i to j for such a student make her worse off). Since $\phi_i(\mathbf{v}) = 0$ for each $\mathbf{v} \in \mathcal{V}^j$, we have

$$1 \geq \int_{\mathcal{V} \backslash \mathcal{V}^j} \tilde{\phi}_i(\mathbf{v}) d\mu(\mathbf{v}) \geq \int_{\mathcal{V} \backslash \mathcal{V}^j} \phi_i(\mathbf{v}) d\mu(\mathbf{v}) = 1,$$

which, combined with above fact, implies that

$$\tilde{\phi}_i(\mathbf{v}) = \phi_i(\mathbf{v}) \text{ and } \tilde{\phi}_j(\mathbf{v}) = \phi_j(\mathbf{v}) \text{ for } \mu - a.e. \, \mathbf{v} \in \mathcal{V} \backslash \mathcal{V}^j, \tag{5}$$

which in turn implies (since $\tilde{\phi} \in \mathcal{X}_{\phi_{J_{ij}}}$) that

$$\int_{\mathcal{V}^j} \tilde{\phi}_j(\mathbf{v}) d\mu(\mathbf{v}) = \int_{\mathcal{V}^j} \phi_j(\mathbf{v}) d\mu(\mathbf{v}). \tag{6}$$

The same argument as above establishes that $\tilde{\phi}_j(\mathbf{v}) \geq \phi_j(\mathbf{v})$ for each $\mathbf{v} \in \mathcal{V}^j$, which together with (6) implies that

$$\tilde{\phi}_i(\mathbf{v}) = \phi_i(\mathbf{v}) \text{ and } \tilde{\phi}_j(\mathbf{v}) = \phi_j(\mathbf{v}) \text{ for } \mu - a.e. \, \mathbf{v} \in \mathcal{V}^j. \tag{7}$$

Since any $\tilde{\phi} \in \mathcal{X}_{\phi_{J_{ij}}}$ that weakly Pareto dominates ϕ satisfies (5) and (7), there exists no $\tilde{\phi} \in \mathcal{X}_{\phi_{J_{ij}}}$ that Pareto dominates ϕ.

Proof of Theorem 2:

Part (i): By definition, the most popular school has the lowest cutoff say $\hat{c}^1$. No student whose most preferred school is different from the most popular school will never be assigned to that school, since whenever she fails to get into a more preferred school than the most popular school, she must have a draw $\omega > \hat{c}^1$.

Part (ii): Pick any two schools $i, j \in S$. Without loss, the cutoff of school i is no greater than the cutoff of j. Let $\mathcal{V}^j := \{\mathbf{v} \in \mathcal{V} : v_i < v_j\}$. Then, any type $\mathbf{v} \in \mathcal{V}^j$ has $\phi_i^S(\mathbf{v}) = 0$, since whenever such a student fails to get in school j, his draw must be larger than the cutoff of school j which is larger than the cutoff of school i. Then, by Lemma 4, ϕ^S is Pareto optimal within $\{i, j\}$.

Part (iii): By Lemma 2, $\hat{c}^1 < ... < \hat{c}^n$ for a generic $\mathfrak{m}$. Fix any such $\mathfrak{m}$. Define a bijection $\bar{\kappa} : S \mapsto S$ such that $\hat{c}^i = \hat{c}^i_{\kappa^*(i)}$, and let $\bar{\tau} = (\bar{\kappa}(1), ..., \bar{\kappa}(n))$ be the resulting list of schools. Now consider the set of $\mathcal{V}_{\bar{\tau}} := \{\mathbf{v} \in \mathcal{V} | v_{\bar{\kappa}(i)} > v_{\bar{\kappa}(j)} \text{ if } i < j\}$ student types that have the same preference order as $\bar{\tau}$. We first prove that any such student type has positive probability of assigning to every school in DA-STB: $\phi_i^S(\mathbf{v}) > 0$ for each $i \in S$ if $\mathbf{v} \in \mathcal{V}_{\bar{\tau}}$. This can be shown as follows. By the full support assumption, $m_{\bar{\tau}} = \mu(\mathcal{V}_{\bar{\tau}}) > 0$. For any school $i \in S$, a student type $\mathbf{v} \in \mathcal{V}_{\bar{\tau}}$ will be assigned to that school whenever her draw ω lies in between $\hat{c}^{\bar{\kappa}^{-1}(i)-1}$ and $\hat{c}^{\bar{\kappa}^{-1}(i)}$ (let $\hat{c}^0 \equiv 0$), since her draw will not be low enough to get in any school she prefers but low enough to get in i. Hence, her probability of getting assigned to school i is $\frac{\hat{c}^{\bar{\kappa}^{-1}(i)} - \hat{c}^{\bar{\kappa}^{-1}(i)-1}}{n} > 0$ since the cutoff orders are strict.

Take any $K = \{k_1, k_2, k_3\} \subset S$. We show that ϕ^S fails to be Pareto optimal within K. Consider again the student type $\mathbf{v} \in \mathcal{V}_{\bar{\tau}}$. Without loss, $v_{k_1} > v_{k_2} > v_{k_3}$ for any such type. By the above observation, $\phi_{k_i}^S(\mathbf{v}) > 0$ for $i = 1, 2, 3$. By the full support assumption, there is a Pareto improving within K reallocation of ϕ^S among the student types in $\mathcal{V}_{\bar{\tau}}$. Simply those with a high v_{k_2} relative to (v_{k_1}, v_{k_3}) sell shares at (k_1, k_3) in exchange for an increased share at k_2, with those with a low v_{k_2} relative to (v_{k_1}, v_{k_3}). We thus conclude that there exists no K with $|K| = 3$ such that ϕ^S is Pareto optimal within K. The statement holds then by Lemma 1.

Proof of Theorem 3:

Part (i): For generic $\mathfrak{m}$, there are at least two schools, say i and j, whose cutoffs are strictly below n, so a positive measure of those who prefer either of these schools the most are not assigned to that school. A positive measure of those who prefer i the most but not assigned to i have $k \in S \backslash \{i\}$ as the second most preferred school and are assigned to it with positive probability. Similarly, a positive measure of those who prefer j the most but not assigned to j have i as the second most preferred school and are assigned to it with positive probability. Hence, a positive measure of seats at every school are assigned to those who not prefer that school the most.

Part (ii): If a student prefers school w to school i, then she must rank w higher than i in equilibrium. Since school w is the worst school, the probability that she is assigned to school i becomes 0. Hence, by Lemma 4, ϕ^M is Pareto optimal within $\{i, w\}$ for any $i \in S \backslash \{w\}$.

Part (iii): Choose any two schools $\{i, j\}$ for $i, j \in S \backslash \{w\}$. Generically, there are positive measure of students whose best school is i and the second best is j, and whose best school is j and the second best is i, respectively. Since neither school i nor j is a worst school, each type of students are assigned to both schools with strictly positive probabilities in equilibrium. Obviously, trading probability shares between different types can increase the expected utilities of both types' students.

Part (iv): This part follows directly from Part (iii) since any three schools include two schools that are not the worst school.

Proof of Theorem 4:

Part (i): Let $\nu^*(\cdot)$ be an equilibrium, and let $\phi^*(\cdot)$ be the allocation induced by $\nu^*(\cdot)$. Let K and J be the sets of over- and under-subscribed schools. For any $\mathbf{v} \in \mathcal{V}$, consider an optimization problem:

$$[P(\mathbf{v})] \qquad \max_{\mathbf{x} \in \Delta_{\phi_J^*}(\mathbf{v})} \sum_{i \in S} v_i x_i$$

subject to

$$\sum_{i \in K} p_i x_i \le \sum_{i \in K} p_i \phi_i^*(\mathbf{v}),$$

where $p_i \equiv \int \nu_i^*(\tilde{\mathbf{v}}) d\mu(\tilde{\mathbf{v}}) \ge 1$.

We first prove that $\phi^*(\mathbf{v})$ solves $[P(\mathbf{v})]$. To this end, consider any $\mathbf{x} \in \Delta_{\phi_J^*}(\mathbf{v})$ satisfying the constraint of $[P(\mathbf{v})]$. Consider now the original CADA game and suppose a type $\mathbf{v}$-student faces all others playing their parts of the equilibrium strategies ν^*. Consider a strategy called $s_i, i \in K$ in which she picks school $i \in S$ as her target in her auxiliary message and submits it as her top choice in her ordinal list, and but submits truthful ordinal list otherwise. If type $\mathbf{v}$ plays strategy s_i, then she will be assigned to school i with probability

$$\frac{1}{\int \nu_i^*(\tilde{\mathbf{v}}) d\mu(\tilde{\mathbf{v}})} = \frac{1}{p_i}.$$

With the remaining probability, she will be assigned to the same way as she would if she had played $\nu^*(\mathbf{v})$ and finds herself being assigned to one of schools in J. Let $\bar{\phi}_i^*(\mathbf{v})$ be the conditional assignment probability to $i \in J$ (satisfying $\sum_{i \in J} \bar{\phi}_i^*(\mathbf{v}) = 1$) that would arise if a student with $\mathbf{v}$ reaches in the DA process without choosing any school in J as a target. Hence, when playing s_i, she will be assigned to school $i \in J$ with probability

$$\left(1 - \frac{1}{p_i}\right) \bar{\phi}_i^*(\mathbf{v}).$$

Suppose now the type $\mathbf{v}$ student randomizes by choosing "strategy s_i" with probability $y_i := p_i x_i$, for each $i \in K$, and with probability

$$y_j := \nu_j^*(\mathbf{v}) + \left[\sum_{i \in K} p_i(\phi_i^*(\mathbf{v}) - x_i) \left(1 - \frac{1}{p_i}\right)\right] \bar{\phi}_j^*(\mathbf{v}),$$

for each $j \in J$. Observe $y_j \geq 0$ for all $j \in S$. This is obvious for $j \in K$. For $j \in J$, this follows since the terms in the square brackets are nonnegative:

$$\begin{aligned}
\sum_{i \in K} p_i(\phi_i^*(\mathbf{v}) - x_i) \left(1 - \frac{1}{p_i}\right) &= \left[\sum_{i \in K} p_i(\phi_i^*(\mathbf{v}) - x_i)\right] - \left[\sum_{i \in K} (\phi_i^*(\mathbf{v}) - x_i)\right] \\
&= \sum_{i \in K} p_i(\phi_i^*(\mathbf{v}) - x_i) \\
&\geq 0,
\end{aligned}$$

where the second equality holds since $\mathbf{x} \in \Delta_{\phi^*}(\mathbf{v})$ and the inequality follows from the fact that $\mathbf{x}$ satisfies the constraint of $[P(\mathbf{v})]$ and that $\sum_{j \in J} p_j x_j = \sum_{j \in J} p_j \phi_j^*(\mathbf{v})$ (since $\mathbf{x} \in \Delta_{\phi_J^*}(\mathbf{v})$). Further,

$$\begin{aligned}
\sum_{i \in S} y_i &= \sum_{i \in K} p_i x_i + \sum_{i \in J} \left[\nu_i^*(\mathbf{v}) + \left[\sum_{i \in K} p_i(\phi_i^*(\mathbf{v}) - x_i) \left(1 - \frac{1}{p_i}\right)\right] \bar{\phi}_j^*(\mathbf{v})\right] \\
&= \sum_{i \in K} p_i x_i + \sum_{i \in J} \nu_i^*(\mathbf{v}) + \sum_{i \in K} \left[p_i(\phi_i^*(\mathbf{v}) - x_i) \left(1 - \frac{1}{p_i}\right)\right] \left(\sum_{i \in J} \bar{\phi}_j^*(\mathbf{v})\right) \\
&= \sum_{i \in K} p_i x_i + \sum_{i \in J} \nu_i^*(\mathbf{v}) + \sum_{i \in K} \left[p_i(\phi_i^*(\mathbf{v}) - x_i) \left(1 - \frac{1}{p_i}\right)\right] \\
&= \sum_{i \in K} p_i \phi_i^*(\mathbf{v}) + \sum_{i \in J} \nu_i^*(\mathbf{v}) + \sum_{i \in K} (\phi_i^*(\mathbf{v}) - x_i) \\
&= \sum_{i \in K} p_i \frac{\nu_i^*(\mathbf{v})}{p_i} + \sum_{i \in J} \nu_i^*(\mathbf{v}) \\
&= \sum_{i \in S} \nu_i^*(\mathbf{v}) = 1.
\end{aligned}$$

The third equality holds since $\sum_{i\in J}\bar{\phi}_i^*(\mathbf{v}) = 1$, and the fifth follows since $\mathbf{x} \in \Delta_{\phi^*}(\mathbf{v})$, (which implies $\sum_{i\in K} x_i = \sum_{i\in K} \phi^*(\mathbf{v})$) and since $\phi_i^*(\mathbf{v}) = \frac{\bar{\nu}_i^*(\mathbf{v})}{p_i} = \frac{\nu_i^*(\mathbf{v})}{p_i}$ (where the last equality follows from Lemma 3).

By playing the mixed strategy $(y_1, ..., y_n)$, the student is assigned to school $i \in K$ with probability

$$\frac{y_i}{p_i} = x_i,$$

and to each school $j \in J$ with probability

$$\begin{aligned}
& y_j + \left[\sum_{i\in K} y_i \left(1 - \frac{1}{p_i}\right)\right] \bar{\phi}_j^*(\mathbf{v}) \\
&= \nu_j^*(\mathbf{v}) + \left[\sum_{i\in K} p_i(\phi_i^*(\mathbf{v}) - x_i)\left(1 - \frac{1}{p_i}\right)\right] \bar{\phi}_j^*(\mathbf{v}) + \left[\sum_{i\in K} p_i x_i \left(1 - \frac{1}{p_i}\right)\right] \bar{\phi}_j^*(\mathbf{v}) \\
&= \nu_j^*(\mathbf{v}) + \left[\sum_{i\in K} p_i \phi_i^*(\mathbf{v}) \left(1 - \frac{1}{p_i}\right)\right] \bar{\phi}_j^*(\mathbf{v}) \\
&= \nu_j^*(\mathbf{v}) + \left[\sum_{i\in K} \nu_i^*(\mathbf{v}) \left(1 - \frac{1}{p_i}\right)\right] \bar{\phi}_j^*(\mathbf{v}) \\
&= \phi_j^*(\mathbf{v}) = x_j.
\end{aligned}$$

In other words, the student $\mathbf{v}$ can replicate any $\mathbf{x} \in \Delta_{\phi_J^*}(\mathbf{v})$ that satisfies $\sum_{i\in K} p_i x_i \le \sum_{i\in K} p_i \phi_i^*(\mathbf{v})$ by playing a strategy available in the CADA game. Since $\phi^*(\cdot)$ solves the CADA game and is still feasible in more constrained problem $[P(\mathbf{v})]$, it must solve $[P(\mathbf{v})]$. Moreover, since μ is atomless and $[P(\mathbf{v})]$ is linear objective function on a convex set, $\phi^*(\mathbf{v})$ must be the unique solution to $[P(\mathbf{v})]$ for μ-a.e. $\mathbf{v}$.

We prove the statement of the theorem by contradiction. Suppose to the contrary that there exists an allocation $\phi(\cdot) \in \mathcal{X}_{\phi_J^*}$ that Pareto dominates $\phi^*(\cdot)$. Then, for μ-a.e. $\mathbf{v}$, $\phi(\mathbf{v})$ must either solve $[P(\mathbf{v})]$ or violate the constraint. For μ-a.e. $\mathbf{v}$, the solution to $[P(\mathbf{v})]$ is unique and coincides with $\phi^*(\mathbf{v})$. Therefore, we must have

$$\sum_{i\in K} p_i \phi_i(\mathbf{v}) \ge \sum_{i\in K} p_i \phi_i^*(\mathbf{v}), \tag{8}$$

for μ-a.e. $\mathbf{v}$. Further, there must exist a set $A \subset \mathcal{V}$ with $\mu(A) > 0$ such that each student $\mathbf{v} \in A$ must strictly prefer $\phi(\mathbf{v})$ to $\phi^*(\mathbf{v})$, which must imply (since $\phi^*(\mathbf{v})$ solves

[$P(\mathbf{v})$])

$$\sum_{i\in K} p_i\phi_i(\mathbf{v}) > \sum_{i\in K} p_i\phi_i^*(\mathbf{v}),\ \forall \mathbf{v}\in A. \tag{9}$$

Combining (8) and (9), we get

$$\begin{aligned}&\int \sum_{i\in K} p_i\phi_i(\mathbf{v})d\mu(\mathbf{v}) > \int \sum_{i\in K} p_i\phi_i^*(\mathbf{v})d\mu(\mathbf{v})\\ \Leftrightarrow &\sum_{i\in K} p_i \int \phi_i(\mathbf{v})d\mu(\mathbf{v}) > \sum_{i\in K} p_i \int \phi_i^*(\mathbf{v})d\mu(\mathbf{v}).\end{aligned} \tag{10}$$

Now since $\phi(\cdot)\in\mathcal{X}$, for each $i\in S$,

$$\int \phi_i(\mathbf{v})d\mu(\mathbf{v}) = 1 = \int \phi_i^*(\mathbf{v})d\mu(\mathbf{v}).$$

Multiplying both sides by p_i and summing over K, we get

$$\sum_{i\in K} p_i \int \phi_i(\mathbf{v})d\mu(\mathbf{v}) = \sum_{i\in K} p_i \int \phi_i^*(\mathbf{v})d\mu(\mathbf{v}),$$

which contradicts (10). We thus conclude that ϕ^* is Pareto optimal within K.

Part (ii): Consider the following maximization problem for every $\mathbf{v}\in\mathcal{V}$:

[$\overline{P}(\mathbf{v})$]
$$\max_{\mathbf{x}\in\Delta} \sum_{i\in S} v_i x_i$$

subject to

$$\sum_{i\in K} p_i x_i \le 1. \tag{11}$$

When we have only one undersubscribed school, called school n, the allocation x_n is completely pinned down by the allocation among $n-1$ oversubscribed schools, that is,

$$x_n = 1 - \sum_{i\in K} x_i.$$

Therefore, an allocation $\mathbf{x}\in\Delta$ is feasible in CADA game if (and only if) (11) holds.

Now consider the following maximization problem:

[$\overline{P}'(\mathbf{v})$]
$$\max_{\mathbf{x}\in\Delta} \sum_{i\in S} v_i x_i$$

subject to

$$\sum_{i \in K} p_i x_i \leq \sum_{i \in K} p_i \phi_i^*(\mathbf{v}). \tag{12}$$

Since $\phi^*(\cdot)$ solves less constrained problem $[\overline{P}(\mathbf{v})]$ and is still feasible in $[\overline{P}'(\mathbf{v})]$, it must be an optimal solution for $[\overline{P}'(\mathbf{v})]$. The rest of the proof is shown by the same argument as in Part (i).

Part (iii): Choose two schools $i, j \in S$. If both schools are oversubscribed, i.e., $i, j \in K$, then the result is implied by Part (i).

Next suppose $i \in K$ and $j \in J$. We first show that $\phi_i = 0$ for any $\mathbf{v}$ with $v_i < v_j$. Suppose not. Then this student must subscribe to school i with strictly positive probability. However, since $v_i < v_j \leq \max_{j \in J} v_j$, subscribing to the best undersubscribed school gives her strictly higher utility. Therefore, $\phi_i(\mathbf{v}) = 0$ must hold for any $\mathbf{v}$ with $v_i < v_j$. Hence, the result follows from Lemma 4.

Finally, suppose $i, j \in J$. Without any loss, assume that the cutoff for school i is no greater than the cutoff for school j, where the cutoffs were created by a STB, prior to adjustment based on the auxiliary message. Fix any student of type $\mathbf{v}$ with $v_i < v_j$. Such a student never chooses i as a target unless $\phi_i^*(\mathbf{v}) = 0$. Either way, such a student will never be assigned to i, since either she chooses j as target in which case she will be assigned to j with probability 1 or she will apply to i only if she fails to place in j, which means that her draw in STB is not good enough for her to get in i. The result follows then by Lemma 4.

Proof of Proposition 2. We show that every school in S^* is oversubscribed in equilibrium. Suppose to the contrary that school $i \in S^*$ is undersubscribed. Then, by Lemma 3, every student with $\mathbf{v}$ with $\pi_1(\mathbf{v}) = i$ must subscribe to i, a contradiction. Since each school in S^* is oversubscribed, the result follows from Theorem 4.

Proof of Proposition 3. We can show that, if a school $j > 1$ is oversubscribed, then school $j - 1$ is oversubscribed. (Those who pick j as target should have picked j, giving a contradiction.)

It then suffices to show that at least schools $\{1, 2\}$ are oversubscribed. Suppose not. Then, only school 1 is oversubscribed in equilibrium. Suppose mass $m_2 < 1$ of students pick school 2 as their target; and all other mass $n - m_2$ pick school 1 as their target.

(No student picks school $j > 2$, since picking school 2 will guarantee enrollment, which dominates choosing any school $j > 2$.) Pick any student with $\mathbf{v}$ such that $\frac{\sum_{i=1}^n v_i}{n} < v_2$. If the student picks school 2, she can guarantee the payoff of v_2. If the student picks school 1, she can get

$$v_1 \frac{1}{n - m_2} + v_2 \frac{1 - m_2}{n - m_2} + (\sum_{i=3}^{n} v_i) \frac{1}{n - m_2} = (\sum_{i=1}^{n} v_i) \frac{1}{n - m_2} - v_2 \frac{m_2}{n - m_2},$$

which is less than v_2. Hence, all such students must be choosing 2 as their target. Since there is more than unit mass of such students, school 2 cannot be undersubscribed, which contradicts the hypothesis that only school 1 is oversubscribed.