Expanding “Choice” in School Choice

Atila Abdulkadiroğlu
Duke University

Yeon-Koo Che
Columbia University

Yosuke Yasuda
National Graduate Institute for Policy Studies

November 28, 2008

ERID Working Paper Number 20

Economic Research Initiatives at Duke
WORKING PAPERS SERIES

Expanding "Choice" in School Choice

Atila Abdulkadiroğlu, Yeon-Koo Che, and Yosuke Yasuda*

November 28, 2008

Abstract: Truthful revelation of preferences has emerged as a desideratum in the design of school choice programs. Gale-Shapley's deferred acceptance mechanism is strategy-proof for students but limits their ability to communicate their preference intensities. This results in ex-ante inefficiency when ties at school preferences are broken randomly. We propose a variant of deferred acceptance mechanism which allows students to influence how they are treated in ties. It maintains truthful revelation of ordinal preferences and supports a greater scope of efficiency.

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

1 Introduction

Public school choice has been a subject of intense research and policy debate in recent years. The idea of expanding one's choice of school beyond his/her residence area has broad public support, as exemplified by the number of districts that offer parental choice

*Abdulkadiroğlu: Department of Economics, Duke University, Durham, NC 27708 (email: atila.abdulkadiroglu@duke.edu); Che: Department of Economics, Columbia University, New York City, NY 10027 (email: yk.che@columbia.edu); Yasuda: National Graduate Institute for Policy Studies, Tokyo 106-8677, Japan (email: yyasuda@grips.ac.jp). We are grateful to Francis Bloch, Olivier Compte, Haluk Ergin, Matt Jackson, Navin Kartik, Fuhito Kojima, David Levine, Paul Milgrom, Andrew Postlewaite, Al Roth, Ilya Segal, Ron Siegel, and other seminar participants at Brown University, Caltech, Duke University, GREQAM, Harvard Market Design Workshop, Hong Kong University, Northwestern University, Paris School of Economics, Princeton University (IAS) Stanford University, Washington University, Yonsei University, and Universities of Pompeu Fabra, and Rochester.

over public schools.[1] Yet, how to operationalize the idea of school choice remains actively debated.

This debate, initiated by Abdulkadiroğlu and Sönmez (2003), is centered around a popular method, the "Boston" mechanism, which was used by Boston Public Schools (BPS) until the 2004-2005 school year to assign K-12 pupils to the city schools. Under the Boston mechanism, each school assigns its seats in the order students rank that school during registration. Specifically, each school accepts first those who rank it the first, and accepts those who rank it the second only when the seats are available, and so forth. Under this system, a student's ranking of a school matters crucially for her chance of assignment at that school. This feature may engender strategic behavior in the families' application. For instance, a family may not list their most preferred school as the top choice if that school is very popular among others: Ranking it at the top will not improve their chance with that school appreciably, but it may rather jeopardize their shot at their second, or even less, preferred school, which could have been available to them if they had ranked it at the top instead. This incentive to "game the system" raises difficulties for families and administrators alike.[2]

In 2005, BPS replaced the Boston mechanism with the student-proposing deferred acceptance (henceforth DA) mechanism. Originally proposed by David E. Gale and Lloyd S. Shapley (1962), the DA has students apply to schools in the order they rank them, but schools select the students based *solely on their priorities with each school*. Specifically, in the first round students apply to their to top-ranked schools, and the schools select from them according to their priorities, up to their capacities but only *tentatively*, and reject the others. In the second round, those rejected by their top choice apply to their second-ranked schools, and schools reselect from those held from the first round and from new applicants up to their capacities (only based on the school's ranking of them), again tentatively, and reject the others. This process continues until no students are rejected, at which point the tentative assignment becomes final. A crucial difference relative to the Boston is that a student's ranking of a school does not affect her chance of assignment at that school, once she applies to that school in the process. This means that the families

[1]Government policies that allow parents to choose schools for their children take various forms, including interdistrict and intradistrict public school choice, offered widely across the US, as well as open enrollment, tax credits and deductions, education savings accounts, publicly funded vouchers and scholarships, private voucher programs, contracting with private schools, home schooling, magnet schools, charter schools and dual enrollment. See an interactive map at http://www.heritage.org/research/Education/SchoolChoice/SchoolChoice.cfm for a comprehensive list of choice plans throughout the US.

[2]It also raises a fairness issue since not all families may be equally sophisticated at strategizing (Abdulkadiroğlu, Pathak, Roth, and Sönmez, 2006).

have dominant strategies to report truthfully about their rankings, a property known as "strategyproofness" (Dubins and Freedman, 1981; Roth, 1982). For instance, top-ranking a very popular school will not jeopardize one's chance at less preferred schools in case she fails to get into the top school.

Besides strategyproofness, the DA mechanism is well-justified in terms of student welfare, *if student and school preferences do not involve indifferences.* Given strict preferences on both sides, the DA algorithm produces the so-called *student optimal stable matching* — a matching that is most preferred by every student among all stable matchings (Gale and Shapley, 1962).[3] By contrast, any stable matching may arise in a full-information Nash equilibrium of the Boston mechanism (Ergin and Sönmez, 2006).

In practice, however, schools do not have strict preferences over students. For instance, the Boston public schools prioritize applicants based on whether students have siblings attending a given school or whether they live within its walk zone. This leaves many students in the same priority class. The resulting indifferences in school preferences present challenges in attaining the dual objectives of strategyproofness and welfare, for no strategyproof mechanism implements a student-optimal stable matching for every preference profile (Erdil and Ergin, 2008). Furthermore, one has to break ties at school preferences in order to adopt the DA mechanism, and any inefficiency associated with a realized tie-breaking cannot be removed *ex post* without harming students' incentives ex ante (Abdulkadiroğlu, Pathak and Roth, 2008).[4] These papers provide a solution either in the way of a random tie breaking procedure or of a method for the students to "trade" assignments *ex post*.

The existing literature on school choice, including these papers, is primarily concerned with the students' *ex post ordinal* welfare, namely, how well a given procedure assigns students based on their preference *orderings* for *realized* tie-breaking. Such a perspective does not capture how well a procedure does in terms of students' *ex ante welfare* — i.e., on average across all realizations of tie-breaking (not just under some realization) — and how well it does in resolving students' conflicting interests based on their relative preference intensities over schools.

To illustrate why ex ante welfare matters, suppose there are three students, $\{1, 2, 3\}$, to be assigned to three schools, $\{s_1, s_2, s_3\}$, each with one seat. All students are ranked into the same priority class at every school, and students' preferences are represented by the following von-Neumann Morgenstern (henceforth, vNM) utility values, where v_j^i is student i's vNM utility value for school s_j:

[3] A matching is *stable* if no student or school can do strictly better by breaking off current matching either unilaterally or by rematching with some other partner without making it worse off.

[4] Also see Ehlers (2006) on matching with indifferences.

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

Every feasible matching that assigns each student to a school is stable due to the indifferences at school preferences. Since the students have the same ordinal preferences, any such assignment is also *ex post* Pareto efficient. Therefore, there is no basis for comparing different procedures based on the ex post welfare criterion. In particular, the *stable improvement cycles* algorithm (Erdil and Ergin, 2008), which finds a student-optimal stable matching for every preference profile, has no bite in this example. Yet, how the students' conflicting interests are resolved matters greatly for their ex ante welfare.

To see this, suppose first 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 $EU_1 = EU_2 = EU_3 = \frac{5}{3}$.

This assignment is *ex ante* 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 $EU_1' = EU_2' = EU_3' = 2 > \frac{5}{3}$. Intuitively, starting from the random assignment, this latter assignment executes a trading of probability shares of schools by transferring student 3's share of schools 1 and 3 to students 1 and 2 in exchange of the latter students' shares of school 2. Such a trade is beneficial for all parties given their preference intensities.

Surprisingly, this latter, Pareto dominating, assignment arises as the unique equilibrium of the Boston mechanism.[5] Students 1 and 2 have a dominant strategy of ranking the schools truthfully, and student 3 finds it her best response to strategically rank s_2 as her first choice.[6] The feature of the Boston mechanism crucial for this outcome is that a student can increase her probability of getting a school simply by ranking that school higher in her choice list. It is thus perhaps not surprising that some parents regarded this ability to influence the assignment as a merit of Boston mechanism, not as its shortcoming. At a public hearing by the BPS School Committee, a parent argued:

> I'm troubled that you're considering a system that takes away the little power that parents have to prioritize... what you call this strategizing as if strategizing is a dirty word... (Recording from Public Hearing by the School Committee, 05-11-04).

[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.

[6] In equilibrium, student 2 will be assigned to s_2, and students 1 and 2 will be assigned between s_1 and s_3 with equal probabilities, for these students will have lower priority than student 3 at school 2.

This ability to influence one's treatment in a competition is suppressed in the DA, for a school never discriminates its applicants based on where they rank that school in their choice lists. However, it is this latter property — nondiscrimination of applicants based on choice rankings — that yields the incentives for truthful revelation of preferences in DA. This suggests that there is a tradeoff between incentives and ex ante efficiency.

Clearly, the DA is extreme in resolving this tradeoff; it guarantees truthful revelation of preferences but denies students any "say" over how they should be treated by each school. Some parents seem to have found this feature of DA as troublesome, as one parent put it as follows:

> ... if I understand the impact of Gale Shapley, and I've tried to study it and I've met with BPS staff... I thought I understood that in fact the random number ... [has] preference over your choices... (Recording from the BPS Public Hearing, 6-8-05).

The current paper provides some welfare justifications to these sentiments expressed by the parents, and suggests that there is a potentially better way to balance the tradeoff than either DA or Boston mechanism. Appreciable welfare gain can be obtained by offering students *simple* and *practical* ways to signal their preference intensities, with no sacrifice on (ordinal) strategy-proofness. We propose a procedure that accomplishes this goal and a new efficiency notion that enhances our understanding and ability to compare various assignment mechanisms on the efficiency ground. The next section illustrates our proposal.

2 Choice-Augmented DA Algorithm: Illustration

We already described how the DA algorithm works when the schools' priories involve strict preferences over students.[7] Suppose now schools' priorities are characterized by weak preferences. Then, ties must be broken to *generate* strict school preferences for the DA algorithm to be employed. There are two common methods of breaking ties. *Single tie-breaking* randomly assigns every student a single lottery number to break ties at every school, whereas *multiple tie-breaking* randomly assigns a distinct lottery number to each student at every school. Clearly, a DA algorithm is well defined with respect to the strict priority list generated by either method. We refer the DA algorithms using single and multiple tie-breaking by *DA-STB* and *DA-MTB*, respectively.

We propose an alternative way to break a tie, one that allows students to influence its outcome based on their communication. The associated DA algorithm, which we refer as *Choice-Augmented Deferred Acceptance* (henceforth, CADA), is described as follows:

[7] For a more detailed description, see Gale and Shapley (1962).

- **Step 1:** All students submit ordinal preferences, plus an "auxiliary message, "naming one's "target" school. If a student names a school for a target, she is said to have "targeted" the school.

- **Step 2:** The schools' strict priorities over students are generated based on their *inherent priorities* and the students' auxiliary messages, as follows. First, each student is independently randomly assigned two lottery numbers. Call one *target lottery number* and the other *regular lottery number*. Each school's *strict priority list* is then generated as follows: (i) First consider the students in the school's highest priority group. Within that group, rank at the top those who name the school as their target. List them in the order of their target lottery numbers, and list below them the rest (who didn't name that school for target) according to their regular lottery numbers. (ii) Move to the next highest priority group, and list them below in the same fashion, and repeat this process until all students are ranked in a strict order.

- **Step 3:** The students are then assigned to schools via the DA algorithm, using *each student's ordinal preferences* from Step 1 and *each school's strict priority list* compiled in Step 2.

To illustrate Step 2, suppose there are five students $N = \{1, 2, 3, 4, 5\}$ and two schools $S = \{A, B\}$, neither of which has inherent priority ordering over the students. Suppose students $1, 3$ and 4 targeted A and 2 and 5 targeted B, and that students are ordered according to their target and regular lottery numbers as follows:

$$\mathbf{T}(N) : 3 - 5 - 2 - 1 - 4; \quad \mathbf{R}(N) : 3 - 4 - 1 - 2 - 5.$$

Then the priority list for school A first reorders students $\{1, 3, 4\}$, who targeted that school, based on $\mathbf{T}(N)$, to $3 - 1 - 4$, and reorders the rest, $\{2, 5\}$, based on $\mathbf{R}(N)$ to $2 - 5$, which produces a complete list for A:

$$\mathbf{P}_A(N) = 3 - 1 - 4 - 2 - 5.$$

Similarly, the priority list for B is determined as

$$\mathbf{P}_B(N) = 5 - 2 - 3 - 4 - 1.$$

The process of compiling the priority lists resembles the STB in that the same lottery is used by different schools, but only within each group. Unlike STB, though, different lotteries are used across different groups. This ensures that a student who has a bad draw at her target school gets a "new lease of life" with another independent draw for the other schools.[8]

Clearly, the deferred acceptance feature preserves stability and the incentives 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 target s_1, and 3 will target s_2. In this case, the CADA resembles the Boston mechanism.

In general, 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. Furthermore, CADA delivers a more efficient matching without sacrificing strategy-proofness. The rest of the paper makes this sense precise. That is, we demonstrate the nature of welfare benefits that CADA will have relative to the DA algorithms, when there are sufficiently numerous students and numerous school seats.

Specifically, we consider a model of a "large" economy populated by a continuum of students and a finite number of schools each with a continuum of capacities. We then compare alternative procedures, DA-STB, DA-MTB and CADA, in terms of the scope of efficiency achieved under different procedures. To illustrate our approach, suppose a procedure determines for each student the probabilities of her getting assigned to alternative schools. Call these her *shares* of schools. Then the "scope of efficiency" can be measured by the set of schools whose shares cannot be traded among students in a way that benefits all the students. The bigger this set is, arguably the more efficient the outcome is, with the outcome being fully Pareto efficient if the set coincides with the entire set of schools.

Our main result is then stated in this term: *The CADA mechanism supports a greater scope of efficiency than the DA mechanisms with either tie-breaking procedure.* Specifically, the DA mechanisms support efficient allocation of at most a pair of schools, whereas CADA supports efficient allocation of a (weakly) bigger set of schools. In particular, CADA entails efficient allocation among schools that are relatively popular—in the sense of being oversubscribed by students in their target choice. The economics of this property closely resembles that of competitive markets. Essentially, the students participating in CADA can be seen as making purchasing decisions on the shares of schools. For instance, a student can raise her share of a school, say A, by targeting it, but that lowers her priority standing in

[8]Although the primary reason for our choice is technical, this choice also has additional benefit of allaying a similar concern about the STB raised in the wake of the NYC redesign. One criticism against STB was that if a student has a bad draw, then she will not have a low priority with just one or two schools, but with *every* school she applies to. CADA mitigates this problem.

other schools, say B and C, thus reducing her shares of those schools. The exact tradeoffs faced by a student are determined by how many other students are targeting A, rather than B or C. If there are many such students, then raising a share of A is "expensive," for it requires giving up large amounts of shares of B and C. In other words, relative degrees of congestion at different schools act as "prices" that regulate individuals' decisions. In a "large" economy, students become price takers, so the resulting allocation resembles that of competitive markets, which, as is well known, yields an efficient allocation (among the oversubscribed schools).

In addition to showing the benefit of CADA, we also argue that DA-STB is more desirable than DA-MTB from an ex ante welfare perspective. In particular, we show that the former supports greater scope of efficiency than the latter. The choice between single versus multiple tie-breaking has proven to be an important policy choice in high school admissions in New York City.[9] Our finding informs the choice between DA-STB and DA-MTB in favor of the former.

The idea of CADA appears similar to the proposal by Sönmez and Ünver (2003) to imbed the DA algorithm in "course bidding" employed by some business schools. These two proposals differ in the application, however, as well as in the nature of the inquiry: We are interested in studying the benefit of adding a "signalling" element to the DA algorithm. By contrast, their interest is in studying the effect of adding ordinal preferences and the DA feature to the course bidding. In a broader sense, our paper is an exercise of mechanism design without monetary transfers, and in fact it is closer in nature to the recent ideas of "storable votes" (Casella, 2005) and "linking decisions" (Jackson and Sonnenschein, 2007).[10] Just like them, CADA "links" how a student is treated in a tie at one school to how she is treated in a tie at another school, and this linking makes communication credible. Clearly, applying the idea in a centralized matching is novel and differentiates the current paper. There is a further difference. Jackson and Sonnenschein (forthcoming) demonstrated the efficiency of linking when (linkable) decisions tend to infinity, relying largely on the logic of the law of large numbers. To our knowledge, the current paper is the first to characterize the precise welfare benefit of linking a fixed (finite) number decisions (albeit with continuum of agents).

The rest of the paper is organized as follows. We present the formal model and welfare criterion in Section 3, provide welfare comparison across the three alternative procedures in Section 4. Section 5 presents simulation to quantify the welfare benefits of CADA. Section 6 then considers the implication of enriching the message used in the CADA and the robustness of our results to some students not behaving in a strategically sophisticated

[9] See Abdulkadiroğlu, Pathak and Roth (forthcoming) for a detailed discussion.

[10] See also Che and Gale (1998, 2000) for the effect of budgetary limits in mechanism design.

way. Section 7 concludes.

3 Model and Basic Analysis

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 measure μ. We assume that μ is absolutely continuous with strictly positive density in the interior of $\mathcal{V}$. The assumptions that the aggregate measure of students equal aggregate capacities of schools and that all students find every school acceptable are made for convenience and will not affect our main results (see Subsection 6.5).

The students' vNM values induce their ordinal preferences. Let $\pi := (\pi_1, ..., \pi_n) : \mathcal{V} \to S^n$ be such that $\pi_i(\mathbf{v}) \neq \pi_j(\mathbf{v})$ if $i \neq j$ and that $v_{\pi_i(\mathbf{v})} > v_{\pi_j(\mathbf{v})}$ implies $i < j$. In other words, $\pi(\mathbf{v})$ lists the schools in the descending order of the preferences for a student with $\mathbf{v}$, with $\pi_i(\mathbf{v})$ denoting her i-th preferred school. Let Π denote the set of all ordered lists of S. Then, for each $\tau \in \Pi$,

$$m_\tau := \mu(\{\mathbf{v} | \pi(\mathbf{v}) = \tau\})$$

represents the measure of students whose ordinal preferences are τ. By the full support assumption, $m_\tau > 0$ for each $\tau \in \Pi$. Finally, let $\mathfrak{m} := \{m_\tau\}_{\tau \in \Pi}$ be a profile of measures of all ordinal types. Let $\mathfrak{M} := \{\{m_\tau\}_{\tau \in \Pi^n} | \sum_{\tau \in \Pi} m_\tau = n\}$ be the set of all possible measure profiles. We say a property holds *generically* if it holds for a subset of $\mathfrak{m}$'s that has the same Lebesque measure as $\mathfrak{M}$.

An *assignment*, denoted by $\mathbf{x}$, is a probability distribution over S, and this is an element of a simplex, $\Delta := \{(x_1, ..., x_n) \in \mathbb{R}^n_+ | \sum_{i \in S} x_i = 1\}$. We are primarily interested in how a procedure determines the assignment for each student ex ante prior to conducting the lottery. To this end, we define an *allocation* to be a measurable function $\phi := (\phi_1, ..., \phi_n) : \mathcal{V} \mapsto \Delta$ 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 all allocations.

3.2 Welfare Standards

We begin with two standard notions of ex ante welfare. To begin, we say allocation $\tilde{\phi} \in \mathcal{X}$ *weakly Pareto-dominates* allocation $\phi \in \mathcal{X}$ if, for almost every $\mathbf{v}$,

$$\sum_{i \in S} v_i \tilde{\phi}_i(\mathbf{v}) \geq \sum_{i \in S} v_i \phi_i(\mathbf{v}), \tag{1}$$

and that $\tilde{\phi}$ *Pareto-dominates* ϕ if the former weakly dominates the latter and the inequality of (2) is strict for a positive measure of $\mathbf{v}$'s. We also say $\tilde{\phi} \in \mathcal{X}$ *ordinally-dominates* $\phi \in \mathcal{X}$ if the former has higher chance of assigning each student to her more preferred school than the latter in the sense of first-order stochastic dominance: for a.e. $\mathbf{v}$,

$$\sum_{i=1}^{k} \tilde{\phi}_{\pi_i(\mathbf{v})}(\mathbf{v}) \geq \sum_{i=1}^{k} \phi_{\pi_i(\mathbf{v})}(\mathbf{v}), \ \forall k = 1, ..., n-1, \tag{2}$$

with the inequality being strict for some k, for a positive measure of $\mathbf{v}$'s.

Definition 1. *(i) An allocation $\phi \in \mathcal{X}$ is* ***Pareto efficient (PE)*** *if there is no other allocation in $\mathcal{X}$ that Pareto-dominates ϕ.*

(ii) An allocation $\phi \in \mathcal{X}$ is ***ordinally efficient (OE)*** *if there is no other allocation in $\mathcal{X}$ that ordinally-dominates ϕ.*

For our purpose, it is useful to introduce additional welfare notions, those relating to the scope of efficiency. To begin, fix an assignment $\mathbf{x} \in \Delta$, and a subset $K \subset S$ of schools. An assignment $\tilde{\mathbf{x}} \in \Delta$ is said to be a *within K reassignment of* $\mathbf{x}$ if $\tilde{x}_j = x_j$ for each $j \in S \backslash K$, and let $\Delta^K_{\mathbf{x}} \subset \Delta$ be the set of all such reassignments. Then, a *within K reallocation of* an allocation $\phi \in \mathcal{X}$ is an element of a set

$$\mathcal{X}^K_\phi := \{\tilde{\phi} \in \mathcal{X} | \tilde{\phi}(\mathbf{v}) \in \Delta^K_{\phi(\mathbf{v})}, \text{ a.e. } \mathbf{v} \in \mathcal{V}\}.$$

In words, a within-K reallocation of ϕ represents an outcome of students trading their shares of schools only within K.

Definition 2. *(i) For any $K \subset S$, an ex ante allocation $\phi \in \mathcal{X}$ is* ***Pareto efficient (PE) within*** *K if there is no within K reallocation of ϕ that Pareto dominates ϕ.*

(ii) For any $K \subset S$, an ex ante allocation $\phi \in \mathcal{X}$ is ***ordinally efficient (OE) within*** *K if there is no within K reallocation of ϕ that ordinally dominates ϕ.*

(iii) An allocation is ***pairwise PE*** *(resp.* ***pairwise OE****) if it is PE (resp. OE) within every $K \subset S$ with $|K| = 2$.*

These welfare criteria are quite intuitive. Suppose the students are initially endowed with ex ante shares ϕ of schools but they can trade these shares amongst them. Can they trade mutually beneficially if the trading is restricted to the shares of K? The answer is no if allocation ϕ is *PE within* K. In other words, the size of the latter set represents the restriction on the trading technologies and thus determines the scope of markets within which efficiency is realized. The bigger the set is, the less restricted the agents are in the scope of trading, so it means a more efficient allocation. Clearly, if an allocation is Pareto efficient within the set of all schools, then it is fully Pareto efficient. In this sense, we can view the size of such set as a measure of efficiency.

A similar intuition holds with respect to ordinal efficiency. In particular, ordinal efficiency can be characterized by the inability to form a cycle of traders who beneficially swap their probability shares of schools. Formally, let $\rhd^\phi$ be the binary relation on S defined by

$$i \rhd^\phi j \iff \exists A \subset \mathcal{V}, \mu(A) > 0, \text{ s.t. } v_i > v_j \text{ and } \phi_j(\mathbf{v}) > 0, \forall \mathbf{v} \in A,$$

and say that ϕ *admits a trading cycle within* K if there exist $i_1, i_2, ... i_l \in K$ such that $i_1 \rhd^\phi i_2, ..., i_{l-1} \rhd^\phi i_l$, and $i_l \rhd^\phi i_1$. The next lemma is then adapted from Bogomolnaia and Moulin (2001).

Lemma 1. *An allocation ϕ is OE within $K \subset S$ if and only if ϕ does not admit a trading cycle within K.*

Before proceeding further, we observe how different notions relate to one another.

Lemma 2. *(i) If an allocation is PE (resp. OE) within K', then it is PE (resp. OE) within $K \subset K'$;*

(ii) An allocation is OE within $K \subset S$ if it is PE within K.

(iii) For any K with $|K| = 2$, if an allocation is OE within K, then it is PE within K.

(iv) If an allocation is OE, then it is pairwise PE.

Part (i) follows since a Pareto improving within-K reallocation constitutes a Pareto improving within-K' reallocation for any $K' \supset K$. Likewise, a trade cycle within any set forms a trade cycle within its superset. Part (ii) follows since if an allocation is not ordinally efficient within K, then it must admit a trading cycle within K, which easily implies there being a Pareto improving reallocation. Part (iii) follows since, whenever there exist an allocation that is not Pareto efficient within a pair of schools, there must be two groups of agents who would benefit from swapping their probability shares of these schools, from which the allocation must admit a trade cycle within that pair. Part (iv) then follows from Part (iii).

These characterizations are tight. The converse of Part (iii) does not hold for any K with $|K| > 2$. In the example from the introduction, the DA allocation is OE but not PE. Likewise, an allocation that is PE within K need not be OE within any $K' \supsetneqq K$. To see this, imagine a situation in which an allocation is Pareto improvable *upon* only via a trade cycle that includes a school in $K' \backslash K$. In that case, the allocation may be PE within K yet it will not be OE within K'.

3.3 Alternative School Choice Procedures

We consider three alternative procedures for assigning students to the schools: (1) Deferred Acceptance with Single Tie-breaking (DA-STB), (2) Deferred Acceptance with Multiple Tie-Breaking (DA-MTB), and (3) Choice-Augmented Deferred Acceptance (CADA). These procedures, introduced earlier, can be extended to the continuum of students in a natural way.

The alternative procedures differ only by the way the schools break ties. The tie-breaking rule is well-defined for DA-STB and DA-MTB, and it follows Step 2 of Section 2 in the case of CADA, except that these rules must be extended to our continuous economy model. The formal descriptions are provided in Appendix A. Here, we offer the following heuristic descriptions:

- **DA-STB:** The mechanism draws a single random number $\omega \in [0,1]$ for each student, and a agent with a lower number has a higher priority than the ones with higher numbers for each school.

- **DA-MTB:** For each student, the mechanism draws n independent random numbers $(\omega_1, ..., \omega_n)$ from $[0,1]^n$. The i-th component, ω_i, of student's random draw then determines her priority at school i, with a lower number draw having a higher priority than a higher number.

- **CADA:** The mechanism draws two random numbers $(\omega_T, \omega_R) \in [0,1]^2$ for each student. School i then ranks those students who targeted that school, based on their values of ω_T and then ranks the others based on the values of $1 + \omega_R$ (with a lower number having a higher priority in both cases). In other words, those who didn't target the school receive penalty score of 1.

For each procedure, the DA algorithm is readily defined using the appropriate tie-breaker and the students' ordinal preferences as inputs. Appendix A provides a precise algorithm, which is sketched here. At the first step, each student applies to her most preferred school. Every school i tentatively admits up to unit mass from its applicants in the order of its priority order, and reject the rest if there is any. 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 its priority, and rejects the rest. The process converges when the set of students that are rejected has zero measure. Although this process might not complete in finite time, it converges in limit, and the allocation in the limit is well defined (see Theorem 10 of Appendix A). Further, each of the procedure is ordinally strategy proof:

Theorem 1. (Ordinal strategy-proofness) *In each of the three procedures, it is a (weakly) dominant strategy for each student to submit her ordinal preferences truthfully.*

Proof: The proof follows from Theorem 11 in Appendix A.

3.4 Characterization of Equilibria

□ *DA-STB* and *DA-MTB*

In either form of DA algorithm, the resulting allocation is conveniently characterized by the "cutoff" of each school — namely, the highest lottery number that would get students admitted by each school. Specifically, the DA-STB process induces a cutoff $c_i \in [0, 1]$ for each school i such that a student who ever applies to school i gets admitted by that school if and only if her (single) draw ω is less than c_i. We first establish that these cutoffs are well defined and generically distinct.

Lemma 3. *There exists a unique set of cutoffs $\{c_i\}_{i \in S}$ for the schools under DA-STB and satisfies $c_i > 0$. There exists a school with cutoff 1. For a generic $\mathfrak{m}$, the cutoffs are all distinct.*

Importantly, these cutoffs pin down the allocation of all students. To see this, consider any student with $\mathbf{v}$ and a school i with a cutoff c_i. Suppose school j has the highest cutoff among those the student prefers over i, and its cutoff is c_j. If $c_j > c_i$. Then, the student will never receive school i since whenever she has a draw $\omega < c_i$ good enough for i, she will get into school j or better. Suppose now $c_j < c_i$. Then, she will get into school i if and only if she receives a draw $\omega \in [c_j, c_i]$. The probability of this event is precisely the distance between the two cutoffs, $c_i - c_j$. Formally, let $S(i, \mathbf{v}) := \{j \in S | v_j > v_i\}$ denote the set of schools more preferred to i by type-$\mathbf{v}$ students. Then, the allocation ϕ^S arising from DA-STB is given by

$$\phi_i^S(\mathbf{v}) := c_i - \max_{j \in S(i,\mathbf{v})} c_j, \forall \mathbf{v}, \forall i \in S,$$

where $c_\emptyset := 0$.

DA-MTB is similar to DA-STB, except that each student has independent draws $(\omega_1, ..., \omega_n)$, one for each school. The DA process again induces a cutoff $\tilde{c}_i \in [0,1]$ for each school i such that a student who ever applies to school i gets assigned to it if and only if her draw for school i, ω_i, is less than $\tilde{c}_i$. These cutoffs are well defined.

Lemma 4. *There exists a unique set of cutoffs $\{\tilde{c}_i\}_{i\in S}$ under DA-MTB, such that $\tilde{c}_i > 0$ for all $i \in S$. There exists a "worst" school w such that $\tilde{c}_w = 1$. For a generic $\mathfrak{m}$, there is single worst school.*

Given the cutoffs $\{\tilde{c}_i\}_{i\in S}$, a type $\mathbf{v}$-student receives school i whenever she has rejectable draws $\omega_j > \tilde{c}_j$ for all school $j \in S(i, \mathbf{v})$ she prefers to i and an acceptable draw $\omega_i < \tilde{c}_i$ at school i. Formally, the allocation ϕ^M from DA-MTB is determined by:

$$\phi_i^M(\mathbf{v}) := \tilde{c}_i \prod_{j\in S(i,\mathbf{v})} (1 - \tilde{c}_j), \forall \mathbf{v}, \forall i \in S,$$

with the convention $\tilde{c}_\emptyset := 0$.

□ *CADA*

As with the two other procedures, given the students' strategies on their messages, the DA process induces cutoffs for the schools, one for each school in $[0, 2]$. Of particular interest is the equilibrium in the students' choices of messages. Given Theorem 1, the only nontrivial part of the students' strategy concerns her "auxiliary message." Let $\sigma = (\sigma_1, ..., \sigma_n) : \mathcal{V} \mapsto \Delta$ denote the students' mixed strategy, whereby a student with $\mathbf{v}$ targets i with probability $\sigma_i(\mathbf{v})$. We first establishes existence of equilibrium.

Theorem 2. (Existence) *There exists an equilibrium σ^* in pure strategies.*

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 targets school i *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 $\bar{\sigma}_i^*(\mathbf{v})$ be the probability that a student $\mathbf{v}$ subscribes to school i in equilibrium. We say a school $i \in S$ is *oversubscribed* if $\int \bar{\sigma}_i^*(\mathbf{v}) d\mu(\mathbf{v}) \geq 1$ and *undersubscribed* if $\int \bar{\sigma}_i^*(\mathbf{v}) d\mu(\mathbf{v}) < 1$. In equilibrium, there will be at least (generically, exactly) one undersubscribed school which anybody can get admitted to (that is, even when she fails to get in any other schools she listed ahead of that school). Formally, a school $w \in S$ is said to be "worst" if its cutoff on $[0, 2]$ equals precisely 2. Then, we have the following lemma.

Lemma 5. *(i) Any student who prefers the worst school the most is assigned to that school with probability 1 in equilibrium. (ii) If her most preferred school is undersubscribed but not the worst school, then she targets that school in equilibrium. (iii) For almost every student with* $\mathbf{v}$ *such that* $\pi_1(\mathbf{v}) \neq w$, $\sigma^*(\mathbf{v}) = \bar{\sigma}^*(\mathbf{v})$ *in equilibrium.*

In light of Lemma 5-(iii), we shall refer to "targets a school i " simply as "subscribes to school i."

4 Welfare Analysis of Alternative Procedures

We compare ex ante welfare of three alternative procedures in this section. We begin with DA-STB.

Theorem 3. (DA-STB) *(i) The allocation* ϕ^S *from DA-STB is OE, and is thus pairwise PE.*

(ii) For a generic $\mathfrak{m}$*, there exists no* $K \subset S$ *with* $|K| > 2$ *such that* ϕ^S *is PE within* K*.*

Proof. The results can be proven with the aid of Figure 1.

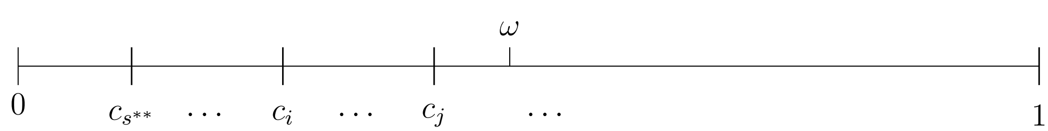

Figure 1: Ordinal efficiency of DA-STB

Following Lemma 3, the cutoffs of the schools are deterministic, as depicted in the figure. Suppose $i \rhd^{\phi^S} j$. Then, there must exist a positive measure of students who prefer school i to school j but are assigned to j with positive probability. It must then follow that $c_i < c_j$. Or else, any students who prefer school i can never be assigned to j. This is because any such student will rank i ahead of j (by strategyproofness), so if she is rejected by i, her draw must be $\omega > c_i \geq c_j$, not good enough for j. Since the cutoffs are (at least weakly) ordered, this means that ϕ^S cannot admit a trading cycle. Hence, it is OE (and thus pairwise PE).

To prove Part (ii), recall from Lemma 3 that the schools' cutoffs are generically distinct. Take any set $\{i, j, k\}$ with $c_i < c_j < c_k$.

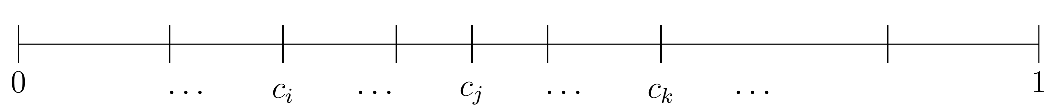

Figure 2: Ex ante Pareto inefficiency of DA-STB within $\{i, j, k\}$.

Then, by the full support assumption, there exists a positive measure of $\mathbf{v}$'s satisfying $v_i > v_j > v_k > v_l$ for all $l \neq i, j, k$. These students will then have the positive chance of being assigned to each school in $\{i, j, k\}$, for their draws will land in the intervals, $[0, c_i]$, $[c_i, c_j]$ and $[c_j, c_k]$, with positive probabilities. Again, given the full support assumption, such students will all differ in their marginal rate of substitution among the three schools. Then, just as with the motivating example, one can construct a mutually beneficial trading of shares of these schools among these students. ▮

The result can be summarized as saying that DA-STB can yield an ordinally efficient allocation in the large economy, but this is the most that can be expected from DA-STB, in the sense that the scope of efficiency is generically limited to (sets of) three schools.

Remark 1. *With finite students, the allocation from DA-STB is ex post Pareto efficient but is not OE, but as the number of students and school seats grow to our limiting model, the DA-STB allocation becomes arbitrarily close to being OE. This is an implication of Che and Kojima (2008), who show that the random priority rule (which coincides with our DA-STB) becomes indistinguishable from the probabilistic serial mechanism (which is known to be OE) as the economy grows large. When the schools have intrinsic priorities, the DA-STB is not even ex post Pareto efficient (Abdulkadiroglu, Pathak and Roth (forthcoming)).*

Next, consider the DA-MTB. Let school $w \in S$ be the worst school if its cutoff under DA-MTB is 1. There exists only one worst school for a generic $\mathfrak{m}$.

Theorem 4. (DA-MTB) *(i) The allocation ϕ^M from DA-MTB is PE within $\{i, w\}$ for each $i \in S \setminus \{w\}$.*

(ii) Generically, there exists no $K \subset S \setminus \{w\}$ with $|K| > 1$ such that ϕ^M is OE within K.

(iii) Generically, there exists no $K \subset S$ with $|K| > 2$ such that ϕ^M is OE within K.

Proof. The results (ii) and (iii) follow from the failure of ex post Pareto efficiency in the DA-MTB. Take any two schools $\{i, j\}$, neither of which is a worst school. There is a positive measure of students whose first and second preferred schools are i and j, respectively (call them "type-i"). Likewise, there is a positive measure of so-called "type-j" students first and second preferred schools are j and i, respectively. A positive measure of type-i students draw (ω_i, ω_j) such that $\omega_i > \tilde{c}_i$ and $\omega_j < \tilde{c}_j$; and positive measure of type-j students draw (ω_i', ω_j') with $\omega_i' < \tilde{c}_i$ and $\omega_j' > \tilde{c}_j$ (see Figure 3 below). Clearly, the former type students are assigned to j and the latter to i, so both types of students will benefit from swapping their assignments.

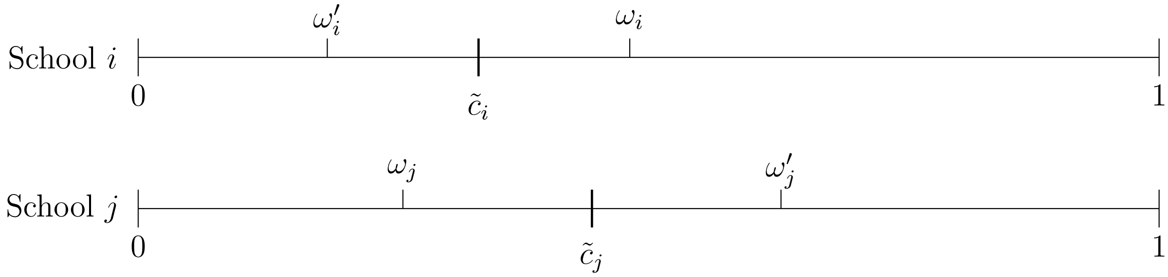

Figure 3: Pareto inefficiency within $\{i, j\}$ under DA-MTB.

Hence, the allocation from DA-MTB is not pairwise OE. Parts (ii) and (iii) then follow given that generically there is only one worst firm (Lemma 4).

To prove Part (i), suppose school $j = w$ is the worst school. Then, $c_j = 1$. Hence, any students who prefer j over i can never be assigned to i. Hence, the allocation does not admit any trading cycle within $\{i, j\}$, and is thus OE within $\{i, j\}$ (Lemma 1). The allocation is then PE by Lemma 2-(iii). ∎

Theorems 3 and 4 have an obvious implication.

Corollary 1. *Assume $n \geq 3$. Then, for a generic $\mathfrak{m}$, a PE allocation never arises from the DA algorithm with either tie-breaking procedure.*

We next turn to the CADA algorithm. The welfare properties of its allocation are characterized as follows.

Theorem 5. (CADA) *(i) An equilibrium allocation ϕ^* of CADA is OE, and is thus pairwise PE.*

(ii) An equilibrium allocation of CADA is PE within the set of oversubscribed schools.

(iii) If all but one schools are oversubscribed, then the equilibrium allocation of CADA is PE.

Theorem 5-(ii) and (iii) showcase the ex ante efficiency benefit associated with CADA. As mentioned earlier, the benefit parallels that of a competitive market. Essentially, CADA supports "competitive markets" for oversubscribed schools. Each student is given a "budget" of unit probability she can allocate across alternative schools for targeting. A given unit probability can buy different amounts of shares for different schools, depending on how many others name those schools. If a mass $z_i \geq 1$ students applies to school i, allocating a unit budget can only buy a share $1/z_i$. Hence, the relative congestion at alternative schools, or their relative popularity, serves as relative "prices" for these schools. In a large economy, individual students take these prices as given, so the prices play the usual role of allocating resources efficiently. It is therefore not surprising that the proof follows the First Welfare Theorem.

Why are competitive markets limited only to oversubscribed schools? Why not undersubscribed schools? The reason has to do with that one can get into an undersubscribed school in two different ways: She can target it, in which case she gets assigned to it for sure; alternatively, she can name an oversubscribed school but the school rejects her, in which case she may still get assigned to the undersubscribed school via the usual DA channel. This means that no single price system regulates the students' assignments to the undersubscribed schools. Furthermore, a spill-over from the oversubscribed schools accounts for assignment of some students to these schools. Consequently, competitive markets do not extend to them.

Finally, Part (i) asserts ordinal efficiency for CADA. At first glance, this feature may be a little surprising in light of the fact that different priority lists are used by different schools. As is clear from DA-MTB, this feature is susceptible to ordinal inefficiency. This is not the case, however, in the equilibrium of CADA. To see this, observe first that any student who is assigned to an oversubscribed school with positive probability must strictly prefer it to any undersubscribed school (or else she should have secured assignment to the latter school by targeting it). This means that we cannot have $j \rhd^{\phi^*} i$ if school j is undersubscribed and school i is oversubscribed. This means that if the allocation admits any trading cycle, it must be within oversubscribed schools or within undersubscribed schools. The former is ruled out by Part (ii) and the latter by the same argument as Theorem 3-(i).

The characterization of Theorem 5 is tight in the sense that there is generally no bigger set that includes all oversubscribed schools *and some undersubscribed school* that supports Pareto efficiency.[11]

[11] To see this, suppose there are four schools, $S = \{1,2,3,4\}$, and four types of students $\mathcal{V} = \{\mathbf{v}^1, \mathbf{v}^2, \mathbf{v}^3, \mathbf{v}^4\}$, with $\mu(\mathbf{v}^1) = \frac{3-\varepsilon}{2}$, $\mu(\mathbf{v}^2) = \frac{1+\varepsilon}{2}$, $\mu(\mathbf{v}^3) = \frac{3-\varepsilon}{2}$, and $\mu(\mathbf{v}^4) = \frac{1+\varepsilon}{2}$ where ε is a small number.

	v_j^1	v_j^2	v_j^3	v_j^4
$j=1$	10	10	20	20
$j=2$	3	5	9	8
$j=3$	1	4	8	1
$j=4$	0	0	0	0

In this case, type 1 and 3 students subscribe to school 1, and type 2 and 4 students subscribe to school 2. More specifically, the allocation ϕ^* has $\phi^*(\mathbf{v}^1) = \phi^*(\mathbf{v}^3) = (\frac{1}{3-\varepsilon}, 0, \frac{2-\varepsilon}{2(3-\varepsilon)}, \frac{2-\varepsilon}{2(3-\varepsilon)})$ and $\phi^*(\mathbf{v}^2) = \phi^*(\mathbf{v}^4) = (0, \frac{1}{1+\varepsilon}, \frac{\varepsilon}{2(1+\varepsilon)}, \frac{\varepsilon}{2(1+\varepsilon)})$. Although schools 1 and 2 are oversubscribed, this allocation is not PE 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. The allocation is not PE within $\{1,2,4\}$ either, since type 3 students can trade probability shares of school 1 and 4 in exchange for probability share at 2, with type 4 students. Therefore $\{1,2\}$ is the largest set of schools that support Pareto efficiency.

Theorem 5 refers to an endogenous property of an equilibrium, namely the set of over/under-subscribed schools. We provide a sufficient condition for this property. For each school $i \in S$, let $m_i^* := \mu(\{\mathbf{v} \in \mathcal{V} | \pi_1(\mathbf{v}) = i\})$ be the measure of students who prefer i the most. We then say a school i is **popular** if $m_i^* \geq 1$, namely, the size of the students whose most preferred school is i is as large as its capacity.

It is easy to see that every popular school must be oversubscribed in equilibrium. Suppose to the contrary that a popular school i is undersubscribed. Then, by Lemma 5-(ii), every student with $\mathbf{v}$ with $\pi_1(\mathbf{v}) = i$ must subscribe to i, a contradiction. Since every popular school is oversubscribed, the next result follows from Theorem 5.

Corollary 2. *Any equilibrium allocation of CADA is PE within the set of popular schools.*

Corollary 2 provides a sufficient condition for a school to be oversubscribed. But it is quite possible that a non-popular school can be oversubscribed in equilibrium. In particular, if all students have the same ordinal preferences, then there is only one popular school, so Corollary 2 has no 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. Letting $\mathcal{V}^U := \{\mathbf{v} \in \mathcal{V} | v_1 > ... > v_n\}$, the students will have the same ordinal preferences if $\mu(\mathcal{V}^U) = \mu(\mathcal{V})$. Define

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

Lemma 6. *Assume $\mu(\mathcal{V}^U) = \mu(\mathcal{V})$, then at least two schools are oversubscribed in the CADA equilibrium if $\mu(\mathcal{V}_2^U) \geq 1$.*

Full Pareto efficiency may be achieved in some cases.

Corollary 3. *The equilibrium allocation of CADA is PE if (i) all but one schools are popular, or if (ii) $n = 3$ and all students have the same ordinal preferences and $\mu(\mathcal{V}_2^U) \geq 1$ holds.*

4.1 Comparison of Procedures

A three-way comparison emerges from the preceding analysis. It provides a formal sense in which the CADA yields a better outcome than DA-STB, which in turn yields a better outcome than DA-MTB. In particular, if the allocation from DA-MTB is PE within $K \subset S$, then so is the allocation from DA-STB, although the converse does not hold; and if the allocation from DA-STB is PE within $K' \subset S$, then so is the allocation from CADA, although the converse does not hold.

Specifically, between the two DA algorithms, the allocation arising from DA-STB is OE, and thus PE within any two schools, whereas the allocation from DA-MTB generically fails to be PE within two schools unless they contain a worst school.

Meanwhile, the CADA allocation is Pareto efficient within a strictly bigger set of schools than the allocations from DA algorithms, if there are more than two popular schools. The following examples illustrate comparisons further.

Example 1. *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

It follows from Corollary 3-(i) that the allocation from CADA is Pareto efficient. 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 PE 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 $\phi^S(\mathbf{v}^3) = (0, \frac{2}{3}, \frac{1}{3})$.[12] *This allocation is not Pareto efficient 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 $\phi^M(\mathbf{v}^3) \approx (0.215, 0.451, 0.333)$.[13] *This is not PE within $\{1, 2\}$.*

Example 2. *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

It follows from Corollary 3-(ii) that the allocation from CADA is Pareto efficient. More specifically, the allocation ϕ^* has $\phi^*(\mathbf{v}^1) = (\frac{1}{2}, 0, \frac{1}{2})$ and $\phi^*(\mathbf{v}^2) = (0, 1, 0)$.

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

[13] Again, assuming that each student has a uniform draw from $[0, 1]$ for each school separately, the cutoff for school 1 is $c_1 = \frac{5-\sqrt{7}}{6} \approx 0.3923$, the cutoff for school 2 is $c_2 \approx 0.4513$, and the one for school 3 is 1.

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 PE within any pair of two schools. (The result of Theorem 4-(ii) does not hold for DA-MTB because the full-support assumption does not hold here.) This allocation is not PE 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 3. *There are three schools, $S = \{1, 2, 3\}$, and two types of students $\mathcal{V} = \{\mathbf{v}^1, \mathbf{v}^2\}$, 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 PE. 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 all students have the uniform ordinal preferences (i.e., $\mu(\mathcal{V}) = \mu(\mathcal{V}^U)$), then we can show that the CADA (weakly) Pareto-dominates the DA.

Theorem 6. *Suppose all students have the same ordinal preferences. The equilibrium allocation of CADA (weakly) Pareto dominates the allocation arising from DA with any random tie-breaking rule.*

5 Simulations

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

In our numerical model, we have 5 schools each with 20 seats and 100 students. Student i's vNM value for school j, $\tilde{v}_{ij}$, is given by

$$\tilde{v}_{ij} = \alpha u_j + (1 - \alpha) u_{ij}$$

where $\alpha \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, u_j is common across students and u_{ij} is specific to student i and school j. For each α, we draw $\{u_j\}$ and $\{u_{ij}\}$ uniformly randomly and independently from the interval $[0, 1]$ to construct student preferences. Then we normalize the level and the scale of each student's vNM utilities as follows:

$$v_{ij} = \zeta_j(\tilde{v}_{i1}, ..., \tilde{v}_{i5})) := \frac{\tilde{v}_{ij} - \min_{j'} \tilde{v}_{ij'}}{\max_{j'} \tilde{v}_{ij'} - \min_{j'} \tilde{v}_{ij'}}$$

Under this normalization, the values of schools range from zero to one, with the least preferred school worth zero and the most preferred one. Any profile of vNM values can be represented in this way; one can simply normalize according to this formula. Further, this normalization is invariant to affine transformation in the sense that fore each i and j, $\zeta_j(\tilde{v}_{i1}, ..., \tilde{v}_{i5}) = \zeta_j(a\tilde{v}_{i1} + b, ..., a\tilde{v}_{i5} + b)$, for any $a \in \mathbb{R}_{++}, b \in \mathbb{R}$.

The students' preferences become similar both ordinally and cardinally as α gets closer to 1. In one extreme case with $\alpha = 0$, students' preferences are completely uncorrelated; in the other with $\alpha = 1$, students have the same cardinal preferences. Given a cardinal utility profile, we simulate DA-STB and DA-MTB, compute a complete information Nash equilibrium of CADA and the resulting CADA outcome. We repeat that 100 times by drawing a new set of vNM utility values for each α. In addition, we solve for a first-best solution, which is a utilitarian maximum for each set of vNM utility values. We then compute average welfare under each mechanism, i.e., the total expected utilities realized under a given mechanism averaged over 100 drawings.[14]

In Figure 4, we compare the three mechanisms to the first best solution, the utilitarian efficient allocation. We plot the welfare of each mechanism as the percentage of the welfare at the first best solution. Two observations emerge from this figure. First, the welfare generated by each mechanism follows a U-shaped pattern. Second, CADA outperforms DA-STB, which outperforms DA-MTB at every value of α and the gap in performance between CADA and the other mechanisms is bigger for larger values of α. All three mechanisms perform almost equally well and produce about 96% of the first-best welfare when $\alpha = 0$. In this case, students have virtually no conflicts of interests, and each mechanism more or less assigns students to their first choice schools. The welfare gain of CADA increases as α increases. This is due to the fact that competition for one's first choice increases as α increases. In those instances, who gets her first choice matters. While DA-STB and DA-MTB determine this purely randomly, CADA does so based on students' messages. Intuitively, if a student's vNM value for a school increases, the likelihood of the student targeting that school in an equilibrium of CADA, therefore the likelihood of her getting that school, increases. This feature of CADA contributes to its welfare gain. DA-STB and DA-MTB start catching up with CADA at $\alpha = 0.9$. In this case, students have almost the same cardinal preferences, so any matching is close to being ex ante efficient. At $\alpha = 0.9$, CADA achieves 95.5% of the first best welfare, whereas DA-STB achieves 92.2%.[15]

Figure 5 gives further insight into the workings of the mechanisms. It shows the per-

[14] See Appendix C for a detailed explanation for the simulations and the computation of the numbers for the figures.

[15] At the extreme case of $\alpha = 1$, preferences are the same so every matching is efficient and the welfare generated by each mechanism is equal to the first best welfare.

centage of students getting their first choices under each mechanism. First, DA-MTB assigns significantly smaller numbers to first choices. This is due to the artificial stability constraints created by multiple tie breaking, which also explains the bigger welfare loss associated with DA-MTB. The patterns for CADA and DA-STB are more revealing. In particular, both assign almost the same number of students to their first choices for each value of α. That is, whereas the poor welfare performance of DA-MTB is explained by the low number of students getting their first choices, the difference between the other two is explainable not by *how many* students, but rather by *which students* are assigned to their first choices.

This is illustrated more clearly by Figure 6, which shows the the ratio of the mean utility of those who get their k-th choice under CADA to the mean utility of those who get their k-th choice under DA-STB at the realized matchings, for $k = 1, 2, 3$. Specifically, those who get their k-th choice achieve a higher utility under CADA than under DA-STB for each $k = 1, 2, 3$. The utility gain is particularly more pronounced for the receivers of their second or third choices. This simply reflects the feature of CADA in assigning students based on their preference intensities: under CADA, those who have less to lose from the second- or third-best choices are more likely to target those schools, and are thus more likely to comprise such assignments. Figure 6 shows that the number of oversubscribed schools is larger on average than the number of popular schools. Note that the average number of oversubscribed schools is larger than 2 at all values of α. Recalling our Theorems 3 and 5, DA-STB is never Pareto efficient within a set of more than 2 schools, whereas CADA is Pareto efficient within the set of oversubscribed schools. Figure 6 thus shows the scope of efficiency achieved by CADA can be much higher than is predicted by Corollary 2. It is also worth noting that the average number of oversubscribed schools exceeds 3 for $\alpha \leq 0.4$. This implies that there are often 4 oversubscribed schools. At those instances, CADA achieves full Pareto efficiency.

The equilibrium behavior of students in Figure 7 shows that more students target their lower ranked schools as α gets closer to 1. This monotone pattern in behavior can be explained by the extent of competition over schools. As α increases, more students have the same school as first choice. Therefore competition for one's first choice becomes more intense. This gives students incentive to target their second, third and even fourth choices in CADA.[16] This also explains the widening gap between the number of popular schools and the number of oversubscribed schools as α goes to 1 in Figure 7.

As mentioned in the introduction, some schools have (non-strict) priorities. We numerically investigate such an environment. To this end, we introduce schools priorities as follows: Each school has two priority classes, high priority and low priority. For each pref-

[16]Note that it is never optimal to target the fifth (last) choice.

erence profile above, we assume that 50 students have high priority in their first choice and low priority in their other choices, 30 students have high priority in their second choice and low priority in their other choices, and 20 students have high priority in their third choice and low priority in their other choices.[17] For convenience, if a student has high priority at a school, we refer to that school as that student's neighborhood school.

We simulate both DA-STB and CADA in this environment. Schools' priorities do not coincide with students' preferences, so both mechanism fail to produce efficient outcome or even student optimal allocation in the presence of school priorities. Erdil and Ergin (2008) proposed a way to attain constrained efficiency subject to respecting school priorities, via performing so-called stable improvement cycles after an initial DA assignment. We thus simulate this algorithm, referred to as DASTB+SIC, to see how it compares with the CADA.

In Figure 9, we compare CADA, DA-STB and DA-STB+SIC again measured as percentage of first-best welfare (ignoring priorities). Again, CADA outperforms DA-STB for all values of α. Since DA-STB+SIC is designed to attain a student optimal outcome, while CADA and DA-STB are not, the former is expected to outperform the latter. This is indeed the case up to $\alpha = 0.4$. When the students' preferences are (ordinally) more similar, CADA catches up with DASTB+SIC (at around $\alpha = 0.5$) and outperforms it as α gets large. In the latter case, the ordinal efficiency becomes less relevant, so cardinal efficiency matters more. In particular, when α is close to 1, every matching is close to being student optimal so that the stable improvement cycles algorithm has little bite. Hence, CADA allocates schools more efficiently than DASTB+SIC does for higher values of α.

DASTB+SIC assigns more students to their first choices than CADA does for low values of α. The difference vanishes for large values of α. The equilibrium behavior of students under CADA is similar; more and more students target their lower ranked schools as α gets bigger.[18] However, Figure 10 shows that more students utilize their target choice for their neighborhood schools. The intuition behind this result is subtle. As more and more students target their lower ranked choices in equilibrium, it becomes tougher to compete for lower ranked schools especially for students who do not have high priority in their lower ranked schools. In turn, those students pick their first choice more frequently, which is their neighborhood school. However, this increases the competition at first choice schools for students whose first choices are not their neighborhood schools. In turn, they target their lower-ranked neighborhood schools. In equilibrium, more students target their neighborhood schools for larger values of α.

In summary, some similarity among preferences is expected in real-life school choice

[17]This assumption is in line with empirical observation in Boston.

[18]Additional graphs are available from the authors upon request.

programs. In those instances, student optimality, therefore DASTB+SIC, has little bite in improving ex ante efficiency. CADA allocates schools more efficiently from an ex ante standpoint even though its outcome may be inefficient ex post. CADA achieves this efficiency gain without harming ordinal incentives, whereas ex post student optimality necessarily implies the loss of strategy-proofness.

6 Discussion

6.1 Enriching the Auxiliary Message

The auxiliary message can be expanded to include more than one school, perhaps at the expense of some practicality. 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.

It is worth noting that the CADA of degree n coincides with the Boston mechanism if the schools have no priorities and if all students have the same ordinal preferences. Such an enriching of the auxiliary message does not alter the qualitative features of CADA. In particular, an argument analogous to that of Theorem 6 applies to CADA of any degree, which has a rather surprising implication:

Theorem 7. *If all students have the same ordinal preferences and the schools have no priorities, then the Boston mechanism weakly Pareto dominates the DA algorithm.*

Expanding the auxiliary message may complicate the deliberation on the part of students and may be 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 assumed before. What are the benefits from adding more schools in the message? Some observations are easy to make. First, enriching the message does not generally guarantee full Pareto efficiency. Consider Example 3 again. Allowing the students to include the second message, or even a third 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, though. The first example illustrates the benefit side.

Example 4. (More is better) *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 efficient.

A richer message need not be always better. A richer message space generates more opportunities 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 5. (More is worse) *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 PE since type* 2 *students can trade probability shares of school* 1 *and* 3 *in exchange for probability share at* 2*, with type* 1 *students.*

Even though ϕ^* *does not Pareto dominate* ϕ^{**}*, the former is PE 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*, the highest possible level, whereas the latter gives* 15.5 *(which is* 0.5 *above the level that would arise from random assignment).*

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.

6.2 Strategic Naivety

Since CADA involves some "gaming" aspect, albeit limited to tie-breaking, a natural concern is that not all families may be strategically competent. This concern has arisen in the context of the Boston mechanism. It has been observed that some significant percentage of families have played suboptimal strategies, for instance, wasting their second top choices to schools that are so popular that students can get in those schools only by listing them as top choices. Such mistakes may arise because of the lack of knowledge about how the system works or the difficulty with assessing how popular schools are. The same concern may arise with respect to CADA, in that some families may not understand well the role the auxiliary message plays in the system and/or they may not judge accurately how over/undersubscribed various schools will turn out.

It is thus important to investigate how the CADA will perform when some families are not strategically sophisticated. To this end, we consider students who are "naive" in the sense that they always target their most preferred schools in the auxiliary message. Targeting the most preferred school appears to be a simple, but reasonable, choice when she/he is unsure about the popularity of alternative schools or unclear about the role the auxiliary message plays in the assignment. Such a strategy will indeed be a best response for many situations, particularly if the first choice is distinctively better than the rest of the choices, so it could be a good approximation. We assume that there is a positive measure of students who are naive in this way, and the others know the presence of these students and their behavior, and respond optimally against them. Surprisingly, the presence of naive students do not affect the main welfare results in a qualitative way.

Theorem 8. *In the presence of naive students, the equilibrium allocation of CADA satisfies the following properties:*

(i) The allocation is OE, and is thus pairwise PE.

(ii) The allocation is PE within the set K of oversubscribed schools.

(iii) If every student is naive, then the allocation is PE within $K \cup \{l\}$ for any undersubscribed school $l \in J := S \backslash K$.

Theorem 8-(i) and (ii) are qualitatively the same as the corresponding parts of Theorem 5, except for Part (iii) of Theorem 5.[19] Of course, the set of oversubscribed schools need not be the same when some fraction of students are naive, so (even) these results do not admit direct comparison between the case of fully rational students and the current case. In particular, Theorem 8-(iii) does not mean that the Pareto efficient set of schools is larger when all students are naive than when there are no naive students. When every student is naive, the set of oversubscribed schools coincides with the set of popular schools. When no students are naive, however, the former is always weakly larger than the latter and can be strictly larger (Recall Examples 1 and 3).

Nevertheless, the efficiency statement is very similar. In particular, Lemma 5-(ii) remains valid in the current context, implying that any popular schools must be oversubscribed here as well. Hence, the same conclusion as Corollary 2 holds.

Corollary 4. *In the presence of naive students, the equilibrium allocation of CADA is PE within the set S^* of popular schools.*

[19]Suppose there are three schools, $S = \{1, 2, 3\}$, and three types of students $\mathcal{V} = \{\mathbf{v}^1, \mathbf{v}^2, \mathbf{v}^3\}$, with $\mu(\mathbf{v}^1) = \mu(\mathbf{v}^2) = \mu(\mathbf{v}^3) = 1$.

	v_j^1	v_j^2	v_j^3
$j = 1$	10	10	10
$j = 2$	1	8	9
$j = 3$	0	0	0

Suppose type 1 and 2 students are strategically sophisticated while type 3 students are all naive. In this case, type 1 and 3 students submit school 1, and type 2 students target school 2. Then, the resulting ex ante allocation has $\phi^*(\mathbf{v}^1) = \phi^*(\mathbf{v}^3) = (\frac{1}{2}, 0, \frac{1}{2})$ and $\phi^*(\mathbf{v}^2) = (0, 1, 0)$. Although schools 1 and 2 are oversubscribed, this allocation is not PE since it will be Pareto improving for type 2 students to trade probability share of school 2 in exchange for probability shares at schools 1 and 3, with type 3 (naive) students. Therefore, Theorem 5-(iii) does not extend to the case in which we have *both* strategically sophisticated and naive students.

We also investigate numerically the impact of naive players on average welfare via simulations. To this end, we assume that a certain number of students play naively and target their first choice schools. Other students play their best responses in a complete information Nash equilibrium. We run this simulation with 50 naive students, with 75 naive students and with 100 (all) naive students. Figure 11 reports the average percentage welfare with respect to the first best with zero, 50, 75 and 100 (all) naive students. The welfare patterns are similar. A bigger number of strategic players yields a more efficient outcome. It is worth noting that CADA continues to outperform DA-STB even when all players are naive.

6.3 CADA with "Safety Valve"

The preceding subsection has seen that the main welfare property of CADA remains valid even when an arbitrary proportion of the student population behaves naively. This does not mean, however, that naive students are not disadvantaged by the others who may make strategic use of the message. It thus makes sense to provide an extra safeguard to those who may feel unsure about how to play the CADA game. This can be done by augmenting the message space to allow for an "exit option" and to treat those who invoke such an option *as if* they are participating in the standard DA algorithm. Specifically, the CADA can be modified as follows.

- All students submit ordinal preferences and auxiliary messages. In the auxiliary message, a student can name a target school or say "Opt Out."
- Random ordering of students are generated according to the standard method (e.g., STB). Then, run DA-STB using this priority list. (If a school has inherent priorities, the random list is used only to break a tie within the same priority class.) Assign those who have picked "Opt Out" according to this procedure.
- Assign the remaining students to the remaining seats, using the CADA algorithm. Specifically, construct the choice-augmented priority list for each school, as described before (using two random lists). Then, the assignment is made via the DA algorithm using the choice-augmented priority.

Clearly, this modified algorithm gives each student the option of achieving precisely the same lottery of assignments as she receives from the DA-STB. But she can choose to send an active signal and do better.

Theorem 9. *The CADA with the safety option makes every student (weakly) better off than she is from the DA-STB.*

6.4 Dynamic Implementation

As noted, the welfare benefit of CADA originates from the competitive markets it induces. Unlike the usual markets where there are explicit prices, however, in the CADA-generated markets, students' beliefs about the relative popularity of schools act as the prices. Hence, for the CADA to have the desirable welfare benefit, their beliefs must be reasonably accurate. In practice, the students' preferences tend to reflect the reputations that schools have developed; thus, as long as the school reputations are stable, they can serve as reasonably good proxies for the prices. Nevertheless, the students may not share the same beliefs and the beliefs may not be accurate, in which case CADA procedure will not implement the CADA equilibrium precisely.

The CADA equilibrium can be implemented more precisely by making the (shadow) prices more explicit and by facilitating students' ability to respond to them. Suppose, after submitting their ordinal preferences (which is an once and for all decision), the students can be asked to submit their auxiliary message in a dynamic fashion.

In Round 1, the students submit the names of their target schools. At the end of Round 1, the students are allowed to see the population distribution of target school choices.[20] In Round $m \geq 2$, the students are allowed to change their auxiliary messages. If the number of students who have changed their auxiliary messages from the previous round is less than some (pre-specified) threshold, then their choices in Round m become final, and CADA is run, just as before, to produce a matching. If the number exceeds the threshold, then the population distribution of choices is announced, and they move on to Round $m + 1$.

Under this dynamic mechanism, a student's choice matters only when most of the other students do not alter their choices. It is thus optimal for students to simply best respond to the announced distribution of population choices in the previous round. Clearly, the resulting best-response dynamics will lead to a Nash equilibrium (i.e., a CADA equilibrium discussed earlier), whenever the process converges. Although our dynamic process likely converges in practice, activity rules can be added to facilitate the convergence. For instance, limit cycles can be eliminated by preventing some small fraction of randomly selected students from returning back to their original choices after deviating from them once or twice.

[20] Alternatively, the clearinghouse may compute and display the probabilities of assignment to different schools that would arise (from the CADA) from each choice, assuming that no other student alters her auxiliary message.

6.5 Excess Capacities and Outside Options

Thus far, we have made simplifying assumptions that the aggregate measure of students equal the aggregate capacities of public schools and that all students find each public school acceptable. These assumptions may not hold in reality. While public schools must guarantee seats to students, all the seats need not be filled. And some students may find outside options, such as home or private schooling, better than some public schools. One can relax these assumptions by letting the aggregate capacities to be (weakly) greater than n and by endowing each student an outside option with value drawn from $[0, 1]$.[21] Extending the model in this way entails virtually no changes in the main tenet of our paper. All theoretical results continue to hold in this relaxed environment. A subtle difference arises since, with excess capacities, there may be more than one worst school under DA-MTB, so its allocation may become PE within more pairs of schools. Nevertheless, Theorems 1-9 remain valid. For instance, the DA-STB allocation is ordinally efficient; and the CADA allocation is ordinally efficient, and Pareto efficient within oversubscribed, and thus popular, schools.

6.6 Intrinsic Priorities

The theoretical part of the paper has assumed that schools have no intrinsic priorities. Assuming that schools are indifferent to all students enabled us to focus on the role of school choice mechanisms in resolving conflicts and to obtain a clear welfare characterizations across mechanisms. In practice, schools do have (coarse) priorities. Although accommodating priorities in our model seems beyond the scope of the current paper, our simulation shows that the benefit of CADA extends to such an environment especially if the students' preferences have sufficient commonality. How the theoretical results will extend to this environment remains an interesting open question.[22]

[21] This modeling approach implicitly assumes the outside options to have unlimited capacities, which may not accurately reflect the scarcity of outside option such as private schooling.

[22] At least one part of the result seems easily extendable with a slight modification of CADA. Suppose students can be partitioned into separate "priority" groups so that each group has the same intrinsic priority profile; that is, students in each priority group have the same intrinsic priority at *every* school. One can then apply the CADA tie-breaking with respect to each priority group. In other words, tie-breaking standings of students within each priority group are determined by their auxiliary messages. This modified CADA ensures that the targeting by the members of each priority group has no "externalities" on the tie-breaking of students in a different group. It is straightforward to check that Theorem 6 generalizes as follows: *The (modified) CADA allocation weakly Pareto dominates the DA allocation if students within each priority group have the same ordinal preferences.*

7 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, an important property 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 allowing students to influence tie-breaking via additional communication, CADA implements a more efficient ex ante allocation than the standard DA algorithms, without sacrificing the strategyproofness of ordinal preferences.

References

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

[2] Abdulkadiroğlu, A., Pathak, P., and Roth, E. (forthcoming), "Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match," *American Economic Review*

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

[4] Abdulkadiroğlu, A., Pathak, P., Roth, E., and Sönmez, T. (2006), "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] Bogomolnaia, A., and Moulin, H., (2001), "A New Solution to the Random Assignment Problem," *Journal of Economic Theory*, 100, 295-328.

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

[8] Che, Y.-K, and Gale, I. (1998) "Caps on Political Lobbying", *American Economic Review*, 88, 643-651.

[9] Che, Y.-K., and Gale, I. (2000) "The Optimal Mechanism for Selling to Budget-Constrained Buyer", *Journal of Economic Theory*, 92, 198-233.

[10] Che, Y.-K., and Kojima, F. (2008) "Asymptotic Equivalence of Random Priority and Probabilistic Serial Mechanisms, Unpublished mimeo, Columbia and Yale Univerisities.

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

[12] Ehlers, L. (2006) "Respecting Priorities when Assigning Students to Schools", working paper, CIREQ

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

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

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

[16] Jackson, M., and Sonnenschein, P. (2007), "Overcoming Incentive Constraints by Linking Decisions," *Econometrica*, 75, 241-257.

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

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

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

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

Appendix A: Extensions of Algorithms to the Continuous Environment

It is convenient to explicitly model the randomizing device used to break the ties. For our purpose, it is sufficient to consider a vector $\omega = (\omega_1, ..., \omega_n) \in [0, n]^n =: \Omega$ of uniformly and independently generated numbers. (The vector of ω will be sufficiently rich enough to model the procedures we study.) Formally, we augment the type space by incorporating the random draw to $\mathcal{V} \times \Omega =: \Theta$, with its generic element denoted $\theta := (\mathbf{v}, \omega)$, and endow it with a product measure $\eta = \mu \times \xi_1 \times ... \times \xi_n$, where ξ_i is a uniform measure satisfying $\xi([0, \omega_i]) = \omega_i$ for each $\omega_i \in [0, 1]$. This formalism avoids appealing to the law of large numbers (on the continuum of agents), by ensuring that a fraction ω_i of the student mass draws ω_i or less on each i-th random variable. A student of type $\theta = (\mathbf{v}, \omega)$ is then interpreted as having values $\mathbf{v}$ and drawing a vector ω. The student never observes ω, so her action required by the procedure will be measurable with respect to only $\mathbf{v}$; whereas (part or all of) ω component is "discovered" by the schools for their use in tie-breaking.

An *ex post allocation* is a measurable function $\psi := (\psi_1, ..., \psi_n) : \Theta \mapsto \Delta$ such that $\psi_i(\theta) \in \{0, 1\}$ and that $\int \psi_i(\theta)\eta(d\theta) = 1$ for each $i \in S$. Namely, ψ assigns a student with $\mathbf{v}$ to school j upon drawing ω such that $\psi_j(\mathbf{v}, \omega) = 1$. Let $\mathcal{Y}$ be the set of all ex post allocations. Later, we shall describe how each procedure generates an ex post allocation. Some procedures may not use the entire vector of ω, so the ex post allocation they produce may be measurable with respect to only some components of ω.

We define the alternative DA procedures here.

Ordinal preferences. In any DA algorithm, every student submits a ranking of schools. Formally, students' ordinal preferences are represented by a measurable function $P : \Theta \to \Pi^n$, where $P(\mathbf{v}, \omega) \in \Pi^n$ is an ordered list of n schools (ordered not necessarily according to true preferences). Since the ω is unobserved by the students (at least at the time of submitting the ordinal preferences), we require that $P(\mathbf{v}, \omega) = P(\mathbf{v}', \omega')$ whenever $\mathbf{v} = \mathbf{v}'$. We say a DA algorithm is *ordinally strategy-proof* if it is a (weak) dominant strategy for each student with $\mathbf{v}$ to choose $P(\mathbf{v}, \omega) = \pi^n(\mathbf{v})$.

School priorities (tie-breaking rules). We introduce a tie-breaker function which determines the priority of each student for each school as a function of the random draw (as well as their auxiliary message in the case of CADA), in the event of a tie. Formally, *tie-breaker function for school i* is a bounded measurable function $F_i : \Theta \mapsto \mathbb{R}$, such that a student θ' is interpreted as having a higher priority than student θ if $F_i(\theta') < F_i(\theta)$. A *tie-breaker* is a profile $\mathcal{F} = \{F_i : i \in S\}$ of tie-breaker functions. Specifically, the tie-breakers for DA-STB, DA-MTB, and CADA are determined as follows:

- *DA-STB:* The STB rule uses the same tie-breaker function for all schools. This is modeled by a tie-breaker with

$$F_i(\mathbf{v}, \omega_1, ..., \omega_n) = \omega_1,$$

$\theta = (\mathbf{v}, \omega_1, ..., \omega_n)$, for *every* school $i \in S$. In other words, a draw's draw ω_1 serves as a priority number for all schools. Heuristically, a real number ω_1 is drawn randomly from an interval $[0, 1]$, for each student, which then serves as her priority score.[23]

- *DA-MTB:* The MTB rule produce a randomly and independently drawn priority list for each school. This is modeled by a tie-breaker, with

$$F_i(\mathbf{v}, \omega_1, ..., \omega_n) = \omega_i,$$

for $i \in S$ and for each $\theta = (\mathbf{v}, \omega_1, ..., \omega_n)$. In other words, for each student, a vector $(\omega_1, ..., \omega_n)$ of independent draws determines her priority scores at different schools.

- *CADA:* In CADA, each student sends an auxiliary message of a target school (in addition to their ordinal preferences over schools). Given a (measurable) strategy profile $s : \mathcal{V} \to S$ determining the auxiliary message for each intrinsic type $\mathbf{v}$, the tie-breaker function for school i is given by

$$F_i(\mathbf{v}, \omega_1, ..., \omega_n) = \begin{cases} \omega_1 & \text{if } s(\mathbf{v}) = i \\ 1 + \omega_2 & \text{if } s(\mathbf{v}) \neq i \end{cases}$$

That is, under F_i, ties are broken first in favor of students who target i, within them according to the random draw ω_1, and then ties among the rest are broken according to a random draw $\omega_2 + n$ (where n act as a "penalty score" n). Clearly, F_i is a measurable function since ω_1 and s are measurable.

Definition of DA algorithms: Given ordinal preferences P and a tie-breaker $\mathcal{F} = \{F_i : i \in S\}$, a DA algorithm is defined as follows. First, we define a measurable function Ch_{F_i} over subsets of Θ 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 $X \subset \Theta$, let

$$Ch_{F_i}(X) := \sup\{Y \subset X | \eta(Y) \leq 1, F_i(\theta) < F_i(\theta'), \forall \theta \in Y, \theta' \in X \backslash Y\}$$

denote the set of students chosen from X such that the set does not exceed the capacity and that the chosen students have a higher priority than those not chosen.

[23] This heuristics invokes a law of large numbers, but our formal method does not rely on it for we assume a well-behaved randomization device.

Next, we define the $DA_{\mathcal{F}}$ (deferred acceptance) mapping. Consider first a mapping $Q : \Theta \to \Pi$, where $Q(\theta)$ is an ordered list of any $k \leq n$ schools. (Recall $P(\theta)$ is a special case involving the full set of schools.) The DA mapping, $Q' = DA_{\mathcal{F}}(Q) \in \Pi$ is determined as follows. Every student with θ applies to her most preferred school in $Q(\theta)$. 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 θ is rejected by i, $Q'(\theta)$ is obtained from $Q(\theta)$ by deleting i in $Q(\theta)$. If a student θ is not rejected, then $Q'(\theta) = Q(\theta)$. More formally, let $T_i(Q) = \{\theta \in \Theta : i \text{ is ranked first in } Q(\theta)\}$ be the set of students that rank i as first choice. Note that $T_i(Q)$ is measurable. Then each school i admits students in $Ch_{F_i}(T_i(Q))$ and rejects students in $T_i(Q) \backslash Ch_{F_i}(T_i(Q))$. If $\theta \in T_i(Q) \backslash Ch_{F_i}(T_i(Q))$ for some $i \in S$, then $Q'(\theta)$ is obtained from $Q(\theta)$ by deleting i from the top of $Q(\theta)$; otherwise $Q'(\theta) = Q(\theta)$. Since Q is a measurable function, Q' is also measurable.

Repeated application of the $DA_{\mathcal{F}}$ mapping gives us the DA algorithm. That is, given a problem $(P, \mathcal{F})$, let $Q^0 = P$ and define $Q^t = DA_{\mathcal{F}}(Q^{t-1})$ for $t > 0$. Then Q^t converges almost everywhere to some measurable Q^* (Theorem 10 below). The matching can be then found by assigning θ to its top choice of $Q^*(\theta)$. Formally, define a mapping $\psi^{(P,\mathcal{F})} : \Theta \mapsto \Delta$ such that $\psi_i^{(P,\mathcal{F})}(\theta) = 1$ if i is the top choice of $Q^*(\theta)$, and $\psi_i^{(P,\mathcal{F})}(\theta) = 0$ otherwise. Since the schools' capacities are respected in each round and also in the limit, the mapping must be an ex post allocation.

We present two main results:

Well-definedness of the Procedure. The existence of $\psi^{(P,\mathcal{F})}$ follows from the next theorem.

Theorem 10. *For every $(P, \mathcal{F})$, $DA^t_{\mathcal{F}}(P)$ converges almost everywhere to some measurable $Q^* : \Theta \to \Pi$.*

Proof. Define the set of rejected students as $R^t = \{\theta : \theta \in T_i(Q^t) \backslash Ch_{F_i}(T_i(Q^t)) \text{ for some } i \in S\}$. Then $\eta(R^t)$ goes to zero as t goes to infinity. Otherwise, if $\eta(R^t) \geq \kappa > 0$ 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 $\eta(R^t)$ goes to zero, a contradiction. Therefore, $DA^t_{\mathcal{F}}(P)$ converges almost everywhere to some Q^*. Since every $Q^t = DA^t_{\mathcal{F}}(P)$ is measurable, Q^* is also measurable. ∎

Ordinal Strategy-proofness. Fix arbitrary ordinal preferences P. Let $P_{-\mathbf{v}} : \mathcal{V} \backslash \{\mathbf{v}\} \to \Pi^n$ denote the ordinal preferences of all students but $\mathbf{v}$ determined by P. Recall that $\pi^n(\mathbf{v}) \in \Pi^n$ represents the truthful ordinal preference induced by $\mathbf{v}$, that is $\pi^n(\mathbf{v})$ lists i before j if and only if $v_i > v_j$. To simplify the notation, let $\psi^P := \psi^{(P,\mathcal{F})}$, with $\mathcal{F}$ suppressed, and let $\psi^* := \psi^{(\pi^n, P_{-[\cdot]}, \mathcal{F})}$ denote the matching outcome for any given type when it submits its ordinal preferences truthfully and the others report P. When students report P, a student with type $\mathbf{v}$ receives expected utility of

$$\mathbb{E}_\omega \left[v \cdot \psi^P(\mathbf{v}, \omega) \right].$$

Theorem 11. *For every $(P, \mathcal{F})$, 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,*

$$\mathbb{E}_\omega \left[\mathbf{v} \cdot \psi^*(\mathbf{v}, \omega) \right] \geq \mathbb{E}_\omega \left[\mathbf{v} \cdot \psi^P(\mathbf{v}, \omega) \right].$$

Proof. It suffices to show that, for all $\theta = (\mathbf{v}, \omega)$,

$$\mathbf{v} \cdot \psi^*(\mathbf{v}, \omega) \geq \mathbf{v} \cdot \psi^P(\mathbf{v}, \omega).$$

Suppose to the contrary that

$$\mathbf{v} \cdot \psi^*(\mathbf{v}, \omega) < \mathbf{v} \cdot \psi^P(\mathbf{v}, \omega), \tag{3}$$

for some $\theta = (\mathbf{v}, \omega)$ and P. We show that there exists a finite many-to-one matching problem for which a DA algorithm fails strategy-proofness, which will then constitute a contradiction to the standard strategy-proofness result (Dubins and Friedman, 1981; Roth, 1982).

To begin, fix any $K \in \mathbb{N}_+$, and construct a discretization of (P, F) for θ as follows: For every $\mathbf{z} = (z_1, ..., z_n)$ and $\mathbf{y} = (y_1, ..., y_n)$ where $z_i, y_j \in \{0, ..., K\}$, consider a set

$$\Theta_{\mathbf{z},\mathbf{y}} = \left\{ (\tilde{\mathbf{v}}, \tilde{\omega}) \in \Theta : \frac{z_i}{K} \leq \tilde{v}_i \leq \frac{z_i + 1}{K}, \frac{n y_j}{K} \leq \tilde{\omega}_j \leq \frac{n(y_j + 1)}{K}, i, j \in S \right\}.$$

Let $\eta_{K,\min} = \min_{\Theta_{\mathbf{z},\mathbf{y}}} \eta(\Theta_{\mathbf{z},\mathbf{y}})$, and let $\#\Theta_{\mathbf{z},\mathbf{y}}$ be the integer part of $\frac{\eta(\Theta_{\mathbf{z},\mathbf{y}})}{\eta_{K,\min}}$.

Pick $\#\Theta_{\mathbf{z},\mathbf{y}}$ students in total from every set $\Theta_{\mathbf{z},\mathbf{y}}$ at random without repetition. Let $\{\theta^l\}$ denote the set of students that are picked. If $\frac{|\{\theta^l\}|}{n}$ is not an integer, pick additional students from the larger sets until obtaining an integer $\frac{|\{\theta^l\}|}{n}$. 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. Now consider the problem in which the set $\{\theta^l\}$ of students are to be assigned to a set S of schools each with capacity $\frac{|\{\theta^l\}|}{n}$. Each student $\theta^l = (\mathbf{v}^l, \omega^l)$'s strict ordinal preference is given by $P(\theta^l)$. The schools' strict preferences are given by $\mathcal{F}$. Denote this problem by $(\{\theta^l\}, S, P, \mathcal{F})_K$, and the associated ex post allocation ψ^P_K. As K goes to infinity, $(\{\theta^l\}, S, P, \mathcal{F})_K$ approximate $(\Theta, S, P, \mathcal{F})$ arbitrarily closely. Hence, $\psi^P_K \to_{a.e} \psi^P$ and $\psi_K^{\pi^n(\mathbf{v}), P_{-\mathbf{v}}} \to_{a.e.} \psi^*$ as $K \to \infty$. Hence, if (3) holds, then there exists K such that

$$\mathbf{v} \cdot \psi^*_K(\mathbf{v}, \omega) < \mathbf{v} \cdot \psi^P_K(\mathbf{v}, \omega).$$

This contradicts the fact 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). ∎

Appendix B: Proofs of the Main Results

Proof of Lemma 3. For any $S' \subset S$ and $i \in S'$, let $m_i(S') := \mu(\{\mathbf{v}|v_i \geq v_j, \forall j \in S'\})$ be the measure of students who prefer i the most among S'. The cutoffs of the schools are then defined recursively as follows. Let $\hat{S}^0 = S$ $\hat{c}^0 = 0$, and $\hat{x}_i^0 = 0$ for every $i \in S$. Given $\hat{S}^0, \hat{c}^0, \{\hat{x}_i^0\}_{i \in S}, \ldots, \hat{S}^{t-1}, \hat{c}^{t-1}, \{\hat{x}_i^{t-1}\}_{i \in S}$, and for each $i \in S$ define

$$\hat{c}_i^t = \sup\left\{c \in [0,1] \,\middle|\, \hat{x}_i^{t-1} + m_i(\hat{S}^{t-1})\left(c - \hat{c}^{t-1}\right) < 1\right\}, \tag{4}$$

$$\hat{c}^t = \min_{i \in \hat{S}^{t-1}} \hat{c}_i^t, \tag{5}$$

$$\hat{S}^t = \hat{S}^{t-1} \setminus \{i \in \hat{S}^{t-1} | \hat{c}_i^t = c^t\}, \tag{6}$$

$$\hat{x}_i^t = \hat{x}_i^{t-1} + m_i(\hat{S}^{t-1})\left(\hat{c}^t - \hat{c}^{t-1}\right). \tag{7}$$

Each recursion step t determines the cutoff of school(s) given cutoffs $\{\hat{c}^0, ..., \hat{c}^{t-1}\}$. Students with draw $\omega > \hat{c}^{t-1}$ can never be assigned to schools $S \setminus S^{t-1}$. For each school $i \in S^{t-1}$ with remaining capacity, a fraction $\hat{x}_i^{t-1}$ is claimed by students with draws less than $\hat{c}^{t-1}$, so only fraction $1 - \hat{x}_i^{t-1}$ of seats can be assigned to students with draws $\omega > \hat{c}^{t-1}$. If school i has the next highest cutoff, $\hat{c}^t$, then the remaining capacity $1 - \hat{x}_i^{t-1}$ must equal the measure of those students who prefer i the most among S^{t-1} and have drawn numbers in $[\hat{c}^{t-1}, \hat{c}^t]$. This, and the fact that school i has cutoff $\hat{c}^t$, imply (4) and (5). The recursion definition implies (6) and (7).

The recursive equations uniquely determine the set of cutoffs $\{\hat{c}^0, ..., \hat{c}^k\}$, where $k \leq n$. The cutoff for school $i \in S$ is then given by $c_i := \{\hat{c}^t | \hat{c}_i^t = \hat{c}^t\}$. It clearly follows from (4) and (5) for $t = 1$ that $\hat{c}^1 > 0$. It also easily follows that $\hat{c}^k = 1$. Obviously $\hat{c}^k \leq 1$. We also cannot have $\hat{c}^k < 1$, or else there will be positive measure of students unassigned, which cannot occur since every student prefers each school to being unassigned and the measure of all students coincide with the total capacity of schools.

Although it is possible for more than one school to have the same cutoff, this is not generic. If there are schools with the same cutoff, we must have $i \neq j \in \hat{S}^{t-1}$ for some t and S^{t-1} such that $\hat{c}_i^t = \hat{c}_j^t$, 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 generically no two schools have the same cutoff. ▌

Proof of Lemma 4. For each $i \in S$ and any $S' \subset S \setminus \{i\}$, let

$$m_i^{S'} := \mu(\{\mathbf{v} \in \mathcal{V} | v_j \geq v_i \geq v_k, \forall j \in S', \forall k \in S \setminus (S' \cup \{i\})\})$$

be the measure of those students whose preference order of school i follows right after schools in S'. (Note that the order of schools within S' does not matter here.) We can then define the conditions for cutoffs $\{\tilde{c}_1, ..., \tilde{c}_n\}$ under DA-MTB as the following system of simultaneous equations. Specifically, for any school $i \in S$, we must have

$$\tilde{c}_i \left(m_i^{\emptyset} + \sum_{S' \subset S \setminus \{i\}} m_i^{S'} \left[\prod_{j \in S'} (1 - \tilde{c}_j) \right] \right) = 1. \tag{8}$$

The LHS has the measure of students admitted to school i, which consists of those who prefer i the most and have a good draw for school i (i.e., $\omega_i \leq \tilde{c}_i$), and those who prefer schools S', for any $S' \subset S \setminus \{i\}$, ahead of i but have bad draws in those schools and have good draw for school i. In equilibrium, the cutoffs must be such that these aggregate measures must equal one (the capacity of school i).

To show that there exists a set $\{\tilde{c}_1, ..., \tilde{c}_n\}$ of cutoffs satisfying the system of equations (8), let $\Upsilon := (\Upsilon_1, ..., \Upsilon_n) : [0,1]^n \to [0,1]^n$ be a function whose i's component is defined as:

$$\Upsilon_i(\tilde{c}_1, ..., \tilde{c}_n) = \min \left\{ \frac{1}{m_i^{\emptyset} + \sum_{S' \subset S \setminus \{i\}} m_i^{S'} \left[\prod_{j \in S'} (1 - \tilde{c}_j) \right]}, 1 \right\},$$

where we adopt the convention that $\min\{\frac{1}{0}, 1\} = 1$.

Observe that self mapping $\Upsilon(\cdot)$ is a monotone increasing on a nonempty complete lattice. Hence, by the Tarski's fixed point theorem, there exists a largest fixed point $\mathbf{c}^* = (c_1^*, ..., c_n^*)$ such that $\Upsilon(c_1^*, ..., c_n^*) = (c_1^*, ..., c_n^*)$, and $\mathbf{c}^* \geq \tilde{\mathbf{c}}^*$ for any fixed point $\tilde{\mathbf{c}}^*$.

We now show that at any such fixed point $\tilde{\mathbf{c}}^*$,

$$\frac{1}{m_i^{\emptyset} + \sum_{S' \subset S \setminus \{i\}} m_i^{S'} \left[\prod_{j \in S'} (1 - \tilde{c}_j^*) \right]} \leq 1, \tag{9}$$

for each $i \in S$. Suppose this is not the case for some i. Then, by the construction of the mapping, we must have $\tilde{c}_i^* = 1$. This means that all students are assigned to some schools. Therefore, by pure accounting,

$$\sum_{i \in S} \tilde{c}_i^* \left(m_i^{\emptyset} + \sum_{S' \subset S \setminus \{i\}} m_i^{S'} \left[\prod_{j \in S'} (1 - \tilde{c}_j^*) \right] \right) = n. \tag{10}$$

Yet, since (9) fails for some school,

$$\begin{aligned}
&\sum_{i \in S} \tilde{c}_i^* \left(m_i^{\emptyset} + \sum_{S' \subset S \setminus \{i\}} m_i^{S'} \left[\prod_{j \in S'} (1 - \tilde{c}_j^*) \right] \right) \\
&< \sum_{i \in S} \left(\frac{1}{m_i^{\emptyset} + \sum_{S' \subset S \setminus \{i\}} m_i^{S'} \left[\prod_{j \in S'} (1 - \tilde{c}_j^*) \right]} \right) \left(m_i^{\emptyset} + \sum_{S' \subset S \setminus \{i\}} m_i^{S'} \left[\prod_{j \in S'} (1 - \tilde{c}_j^*) \right] \right) \\
&= n,
\end{aligned}$$

where the strict inequality follows since, for school l for which (9) holds, $\tilde{c}_l^* = \frac{1}{m_l^\emptyset + \sum_{S' \subset S \setminus \{l\}} m_l^{S'} \left[\prod_{j \in S'} (1-\tilde{c}_j^*)\right]}$ and, for school i for which (9) does not hold, $\tilde{c}_i^* = 1 < \frac{1}{m_i^\emptyset + \sum_{S' \subset S \setminus \{i\}} m_i^{S'} \left[\prod_{j \in S'} (1-\tilde{c}_j^*)\right]}$. This inequality contradicts (10). Since (9) holds for each $i \in S$, the fixed point $(\tilde{c}_1^*, ..., \tilde{c}_n^*)$ solves the system of equations (8). By the analogous logic, there must exist a worst school w with $\tilde{c}_w = 1$. As before, it follows that the solutions to (8) are generically distinct, which implies that, for generic $\mathfrak{m}$, there is a single worst school.

To establish uniqueness, suppose to the contrary $\mathbf{c}^* > \tilde{\mathbf{c}}^*$: $c_j^* \geq \tilde{c}_j^*$ for all j and $c_i^* > \tilde{c}_i^*$ for some i. Let us denote the worst school under $\tilde{\mathbf{c}}^*$ as w, i.e., $\tilde{c}_w^* = 1$. By monotonicity and the definition of cutoffs, w must be a worst school under $\mathbf{c}^*$ as well, i.e., $c_w^* = 1$. Since (8) must be satisfied for w under both cutoffs, we have

$$\left(m_w + \sum_{S' \subset S \setminus \{w\}} m_w^{S'} \left[\prod_{j \in S'} (1 - c_j) \right] \right) = \left(m_w + \sum_{S' \subset S \setminus \{i\}} m_w^{S'} \left[\prod_{j \in S'} (1 - \tilde{c}_j) \right] \right) = 1,$$

which holds if and only if $c_j = \tilde{c}_j$ for all j. ▌

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

Proof of Lemma 5. Part (i) follows trivially since such a student can target that school and get assigned to it with probability one. To prove part (ii) consider any 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 and not a worst school. It is then her best response to target i, since doing so can guarantee assignment to i for sure, whereas targeting some other school may result in assignment to some other school. Hence, the student must be targeting i in equilibrium.

To prove part (iii), consider any $\mathbf{v}$ (with distinct values), such that $\pi_1(\mathbf{v}) \neq w$. Suppose first $\sigma_i^*(\mathbf{v}) > 0$ 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 next $\sigma_j^*(\mathbf{v}) > 0$ 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 targets j, 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 $\sigma^*(\mathbf{v}) = \bar{\sigma}^*(\mathbf{v})$ for μ-a.e. $\mathbf{v}$. ▌

Proof of Theorem 5: Proof of Part (i) builds on that of part (ii), so it will appear last. Throughout, we let K and J be the sets of over- and under-subscribed schools.

Part (ii): Let $\sigma^*(\cdot)$ be an equilibrium, and let $\phi^*(\cdot)$ be the ex ante allocation induced by $\sigma^*(\cdot)$. For any $\mathbf{v} \in \mathcal{V}$, consider an optimization problem:

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

subject to

$$\sum_{i \in K} p_i x_i \leq \sum_{i \in K} p_i \phi^*_i(\mathbf{v}),$$

where $p_i \equiv \max\{\int \overline{\sigma}^*_i(\tilde{\mathbf{v}}) d\mu(\tilde{\mathbf{v}}), 1\}$.

We first prove that $\phi^*(\mathbf{v})$ solves $[P(\mathbf{v})]$. This is trivially true for any $\mathbf{v}$ with $\pi_1(\mathbf{v}) = w$ since $x_i = \phi^*_i(\mathbf{v}) = 0, i \in K$ must hold by Lemma 3 (ii). Based on this observation, in the rest of the proof, we will restrict attention to students whose most preferred school is not the worst school.

Consider now any $\mathbf{x} \in \Delta^K_{\phi^*(\mathbf{v})}$ satisfying the constraint of $[P(\mathbf{v})]$, and suppose a type $\mathbf{v}$-student faces all others playing their parts of the equilibrium strategies σ^* under the original CADA game. Consider a strategy called s_i in which she targets school $i \in S$ 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

$$\frac{1}{\max\{\int \overline{\sigma}^*_i(\tilde{\mathbf{v}}) d\mu(\tilde{\mathbf{v}}), 1\}} = \frac{1}{p_i}.$$

If $i \in J$, this probability is one. If $i \in K$, then she will be rejected by school i with positive probability. In that event, she will pass through the DA process according to her true ordinal preferences, and will be assigned based on her non-target draw of score ω_2. Since she will never be assigned to any other schools in K, she will only be assigned to a school in J. Which school in J she is assigned to is determined solely by ω_2 (holding fixed the student's ordinal rankings), and her draw of ω_2 is independent of her draw of ω_1 (which determined her assignment to i). Hence, the conditional probability of a student getting assigned to $j \in J$, is the same, regardless of which oversubscribed school $i \in K$ turned him down. Note let that conditional probability be $\bar{\phi}^*_j(\mathbf{v})$. Obviously, $\sum_{j \in J} \bar{\phi}^*_j(\mathbf{v}) = 1$.

In summary, when playing $s_i, i \in K$ she will be assigned to school $j \in J$ with probability

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

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

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

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:

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

where the first inequality is implied by Lemma 3-(iii), the third equality holds since $\mathbf{x} \in \Delta^K_{\phi^*(\mathbf{v})}$, and the last inequality follows from the fact that $\mathbf{x}$ satisfies the constraint of $[P(\mathbf{v})]$. Further,

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

The third equality holds since $\sum_{i\in J}\bar{\phi}_i^*(\mathbf{v})=1$, the fourth is implied by Lemma 5-(iii), and the fifth follows since $\mathbf{x}\in\Delta^K_{\phi^*(\mathbf{v})}$, (which implies $\sum_{i\in K}x_i=\sum_{i\in K}\phi^*(\mathbf{v})$).

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

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

In other words, the student $\mathbf{v}$ can replicate any $\mathbf{x} \in \Delta^K_{\phi^*(\mathbf{v})}$ that satisfies $\sum_{i \in K} p_i x_i \leq \sum_{i \in K} p_i \phi_i^*(\mathbf{v})$ 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 $[P(\mathbf{v})]$ has a 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 $\phi(\cdot) \in \mathcal{X}^K_{\phi^*}$ 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

$$
\sum_{i \in K} p_i \phi_i(\mathbf{v}) \geq \sum_{i \in K} p_i \phi_i^*(\mathbf{v}), \tag{11}
$$

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

$$
\sum_{i \in K} p_i \phi_i(\mathbf{v}) > \sum_{i \in K} p_i \phi_i^*(\mathbf{v}), \ \forall \mathbf{v} \in A. \tag{12}
$$

Combining (11) and (12), we get

$$
\begin{aligned}
& \int \sum_{i \in K} p_i \phi_i(\mathbf{v}) d\mu(\mathbf{v}) > \int \sum_{i \in K} p_i \phi_i^*(\mathbf{v}) d\mu(\mathbf{v}) \\
\Leftrightarrow & \sum_{i \in K} p_i \int \phi_i(\mathbf{v}) d\mu(\mathbf{v}) > \sum_{i \in K} p_i \int \phi_i^*(\mathbf{v}) d\mu(\mathbf{v}).
\end{aligned} \tag{13}
$$

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 (13). We thus conclude that ϕ^* is Pareto optimal within K.

Part (iii): 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

$$\sum_{i\in K} p_i x_i \leq 1. \tag{14}$$

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) (14) 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

$$\sum_{i\in K} p_i x_i \leq \sum_{i\in K} p_i \phi_i^*(\mathbf{v}). \tag{15}$$

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 (ii).

Part (i): The argument in the text already established that the allocation cannot admit a trading cycle that includes both oversubscribed and unsubscribed schools. It cannot admit a trading cycle comprising only oversubscribed schools, since the allocation is PE within these schools, by Part (ii), making it OE within the schools, by Lemma 2-(ii). It cannot admit a trading cycle comprising only undersubscribed schools, since the logic of Theorem 3-(i) implies that it is OE within undersubscribed schools. Since the allocation cannot admit any trading cycle, it must be OE. ▌

Proof of Lemma 6. We can show that, if a school $j > 1$ is oversubscribed, then school $j-1$ is oversubscribed. (Those who targeted j 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 target school 2; and all other mass $n - m_2$ target school 1. (No student targets school $j > 2$, since targeting school 2 will guarantee enrollment, which dominates targeting 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 targets 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 targeting 2. 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. ▮

Proof of Theorem 6: Consider first a DA algorithm with any random tie-breaking. Since all students submit the same rank order over schools, they all must be assigned to each school with the same probability. In other words, the allocation must be

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

Consider now CADA algorithm. Let $\sigma^*(\mathbf{v}) \in \Delta$ be the equilibrium mixed strategy adopted by type $\mathbf{v}$. Then, a measure

$$\alpha_i^* := \int \sigma_i^*(\mathbf{v}) d\mu(\mathbf{v})$$

of students target $i \in S$ in equilibrium. The equilibrium induces a mapping $\varphi^* : S \mapsto \Delta$, whereby a student is assigned to school j with probability $\varphi_j^*(i)$ if she targets i.

Since in equilibrium, the capacity of each school is filled, we must have, for each $j \in S$,

$$\sum_{i \in S} \alpha_i^* \varphi_j^*(i) = 1. \tag{16}$$

That is, a measure α_i^* of students target i, and a fraction $\varphi_j^*(i)$ of those is assigned to school j. Summing the product over all i then gives the measure of students assigned to j, which must equal its capacity, 1.

Consider a student with any arbitrary $\mathbf{v} \in \mathcal{V}$. Suppose she randomizes in her auxiliary message by choosing school i with probability

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

Then, the probability that she will be assigned to any school k is

$$\sum_j y_j \varphi_k^*(j) = \sum_j \frac{\alpha_j^*}{n} \varphi_k^*(j) = \frac{1}{n},$$

where the second equality follows from (16). That is, she can replicate the same ex ante assignment with the randomization strategy as $\phi^{DA}(\mathbf{v})$. Hence, the student must be at least weakly better from CADA. ▌

Proof of Theorem 8: Part (i) is precisely the same as Part (i) of Theorem 5 and is the consequence of Part (ii) below and Part (ii) of Lemma 2 (which does not depend on whether the students are naive or not). Hence it is omitted. To prove Part (ii), it is useful to establish the following lemma. As before, let ϕ^* denote the ex ante allocation arising from the CADA game and let K and $J = S \backslash K$ be respectively the sets of oversubscribed and undersubscribed schools in equilibrium.

Lemma 7. *Any reassignment of $\phi^*(\mathbf{v})$ within K will make a naive student with $\mathbf{v}$ strictly worse off, for almost every $\mathbf{v}$.*

Proof: Consider a naive student with $\mathbf{v}$. Assume without loss of generality that she prefers i strictly over all other schools (i.e., $v_i > v_j$, $\forall j \neq i$). (This is without loss of generality since the values are distinct for almost every student type.) Since the student is naive, she subscribes to school i with probability 1. If school i is undersubscribed, then the result is trivial since $\phi_k^*(\mathbf{v}) = 0$ for all $k \in K$. Hence, suppose school i is oversubscribed. Then, any reassignment $\mathbf{x} \in \Delta_{\phi^*(\mathbf{v})}^K$ must satisfy

$$\sum_{j \in K} x_j = \phi_i^*(\mathbf{v})$$

Since $v_i > v_j$ $\forall j \neq i$, for any $\mathbf{x} \in \Delta_{\phi^*(\mathbf{v})}^K$, $\mathbf{x} \neq \phi_i^*(\mathbf{v})$, we must have

$$\sum_{j \in K} x_j v_j < \sum_{j \in K} x_j v_i = \phi_i^*(\mathbf{v}) v_i,$$

which implies that the student must be strictly worse off from any such reassignment. ‖

We are now ready to prove Parts (ii) and (iii):

Part (ii): We make use of the proof of Theorem 5. By Lemma N, a type-$\mathbf{v}$ naive student's assignment from the CADA, $\phi^*(\mathbf{v})$, is a unique solution to $[P(\mathbf{v})]$, for a.e. $\mathbf{v}$, even without the constraint

$$\sum_{i \in K} p_i x_i \leq \sum_{i \in K} p_i \phi_i^*(\mathbf{v}). \tag{17}$$

Since $\phi^*(\mathbf{v})$ is feasible under (17), this must be a unique solution to $[P(\mathbf{v})]$.

For a non-naive student with a.e. $\mathbf{v}$, the proof of Theorem 5 follows directly, so $\phi^*(\mathbf{v})$, is also a unique solution of $[P(\mathbf{v})]$. Since the equilibrium assignment of both types solves $[P(\mathbf{v})]$, the rest of the argument in the proof of Theorem 5 applies, proving that we ϕ^* is PE within K. $\|$

Part (iii): Again let ϕ^* be the ex ante allocation arising from CADA. Suppose to the contrary that there exists a within-$K \cup \{l\}$ reallocation $\tilde{\phi}$ of ϕ^* that Pareto dominates ϕ^*. By Part (ii), ϕ^* is PE within K, so $\tilde{\phi}_l(\mathbf{v}) \neq \phi^*_l(\mathbf{v})$ for a positive measure of $\mathbf{v}$, which in turn implies that there exists a set $A \subset \mathcal{V}$ with $\mu(A) > 0$ such that $\tilde{\phi}_l(\mathbf{v}) > \phi^*_l(\mathbf{v})$ for each $\mathbf{v} \in A$. Since $\tilde{\phi}(\mathbf{v}) \in \Delta^{K\cup\{l\}}_{\phi^*(\mathbf{v})}$, $\sum_{j\in K\cup\{l\}} \tilde{\phi}_j(\mathbf{v}) = \sum_{j\in K\cup\{l\}} \phi^*_j(\mathbf{v})$, so

$$\sum_{j\in K} \tilde{\phi}_j(\mathbf{v}) < \sum_{j\in K} \phi^*_j(\mathbf{v}) \quad \text{for all } \mathbf{v} \in A.$$

Assume without loss that $\mathbf{v}$ satisfies $v_i > v_j$ for all $i \in K$ and for all $j \neq i$. ($i \neq l$ since $\tilde{\phi}_l(\mathbf{v}) > \phi^*_l(\mathbf{v})$ is impossible if $i = l$.) Then, the type-$\mathbf{v}$ student's expected payoff from $\tilde{\phi}$ is

$$\begin{aligned}
\sum_{j\in S} \tilde{\phi}_j(\mathbf{v})v_j &= \sum_{j\in K} \tilde{\phi}_j(\mathbf{v})v_j + \tilde{\phi}_l(\mathbf{v})v_l + \sum_{j\in J\setminus\{l\}} \phi^*_j(\mathbf{v})v_j \\
&< \sum_{j\in K} \tilde{\phi}_j(\mathbf{v})v_i + \left(\tilde{\phi}_l(\mathbf{v}) - \phi^*_l(\mathbf{v})\right) v_i + \phi^*_l(\mathbf{v})v_l + \sum_{j\in J\setminus\{l\}} \phi^*_j(\mathbf{v})v_j \\
&= \left(\sum_{j\in K\cup\{l\}} \tilde{\phi}_j(\mathbf{v}) - \phi^*_l(\mathbf{v})\right) v_i + \sum_{j\in J} \phi^*_j(\mathbf{v})v_j \\
&= \sum_{j\in S} \phi^*_j(\mathbf{v})v_j.
\end{aligned}$$

Since this inequality holds for almost every $\mathbf{v} \in A$, and since $\mu(A) > 0$, $\tilde{\phi}$ cannot Pareto dominate ϕ^*. ▌

Appendix C: Simulations

There are 5 schools each with a capacity of 20 seats and 100 students. Fix α. We independently draw 100 sets of vNM values for students. Let $\{\tilde{v}^s_{ij}\}$ denote a draw of vNM values, where superscript s denote the draw and $\tilde{v}^s_{ij}$ denotes student i's vNM value for school j. We normalize the level and the scale of each student's vNM utilities as follows:

$$v^s_{ij} = \frac{\tilde{v}^s_{ij} - \min_{j'} \tilde{v}^s_{ij'}}{\max_{j'} \tilde{v}^s_{ij'} - \min_{j'} \tilde{v}^s_{ij'}}$$

Given a normalized draw $\{v_{ij}^s\}$, fix the mechanism, define the following: p_{ij}^s is the probability that student i is assigned school j under the mechanism. $\pi_k^s(i)$ is the school that is ranked k-th in i's preference list. P^s is the set of popular schools. O^s is the set of oversubscribed schools in an equilibrium of CADA with no naive players.

A first best or utilitarian maximum solves

$$\overline{v}_{FB}^s = \frac{1}{100} \max_{\{\hat{p}_{ij}^s\}} \sum_i \sum_j \hat{p}_{ij}^s v_{ij}^s$$

Let $\{\overline{p}_{ij}^s\}$ denote a solution to the first best. There may be multiple solutions, we arbitrarily pick one.

Furthermore, we calculate

$$\delta_1^s = \frac{1}{100} \sum_i \sum_j p_{ij}^s \cdot \mathbf{1}(\pi_1^s(i) = j)$$

where δ_1^s is the average probability of assigning a student to her first choice.

In the CADA experiments with naive players, we divide the set of students into two: N is the set of naive players who always target their first choice, and S is the set of strategically sophisticated players who play their best response strategies given others' strategies. We calculate utilitarian welfare as before. We also compute the number of students targeting their k-th choice in equilibrium, which we denote by T_k^s, $k \in \{1, 2, 3, 4\}$.

Given a draw $\{v_{ij}^s\}$, the set P^s is determined trivially. Next we describe how the other numbers are computed.

A *single tie breaker* is a list of 100 randomly drawn lottery numbers, one for each student. Under DA-STB the ties at a school are broken according to students' single random numbers. In CADA, we draw two single tie breakers, one to be used to break ties at one's target school, the other to be used at one's other schools. A *multiple tie breaker* is a list of $100 \times 5 = 500$ randomly drawn numbers, one for each student at each school. Under DA-MTB, the ties at a school are broken according to students' tie breaker numbers at that school.

For each draw $\{v_{ij}^s\}$, we independently draw 2,000 single tie breakers for DA-STB, and an additional set of 2,000 single tie breakers for CADA, and 2,000 multiple tie breakers for DA-MTB. Then p_{ij}^s for a mechanism is computed by

$$\frac{\text{Number of tie breakers at which } i \text{ is assigned } j}{2,000}.$$

The equilibrium of CADA is computed with single tie breakers being fixed. Given the strategies of other students, a student's best response is found by computing that student's expected utility over those tie breakers. Then O^s, the set of oversubscribed schools, is found by using students's equilibrium target schools. In experiments with naive players, naive players' target schools are fixed at their first choice.

Note that we are approximating the equilibrium by drawing (two sets of) 2,000 independent tie-breakers. The exact numbers are computed by considering 100! single tie-breakers and $(100!)^5$ multiple tie breakers, which is beyond the capabilities of our computational resources. Any further increase in the number of tie breakers beyond 2,000 does not increase the precision of our computations significantly.

For each $z^s \in \{\overline{v}^s, \overline{v}^s_{FB}, \pi^s_1, T^s_1, T^s_2, T^s_3, T^s_4, |P^s|, |O^s|\}$, we compute the average of z^s by

$$z = \frac{1}{100}\sum_{s=1}^{100} z^s.$$

Note that we drop all "s" from a variable to denote its mean over 100 iterations of an experiment. We report $100\frac{\overline{v}}{\overline{v}_{FB}}$ in our welfare figures.

Figure 4: Welfare as Percentage of First Best

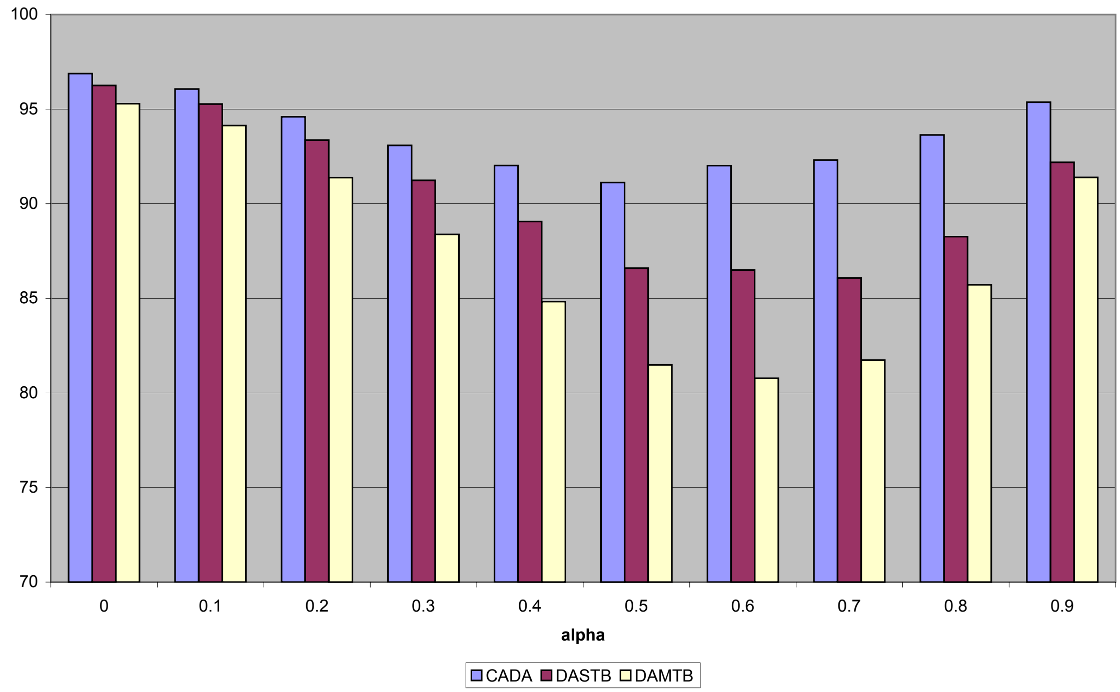

Figure 5: Percentage of Students Getting Their First Choice

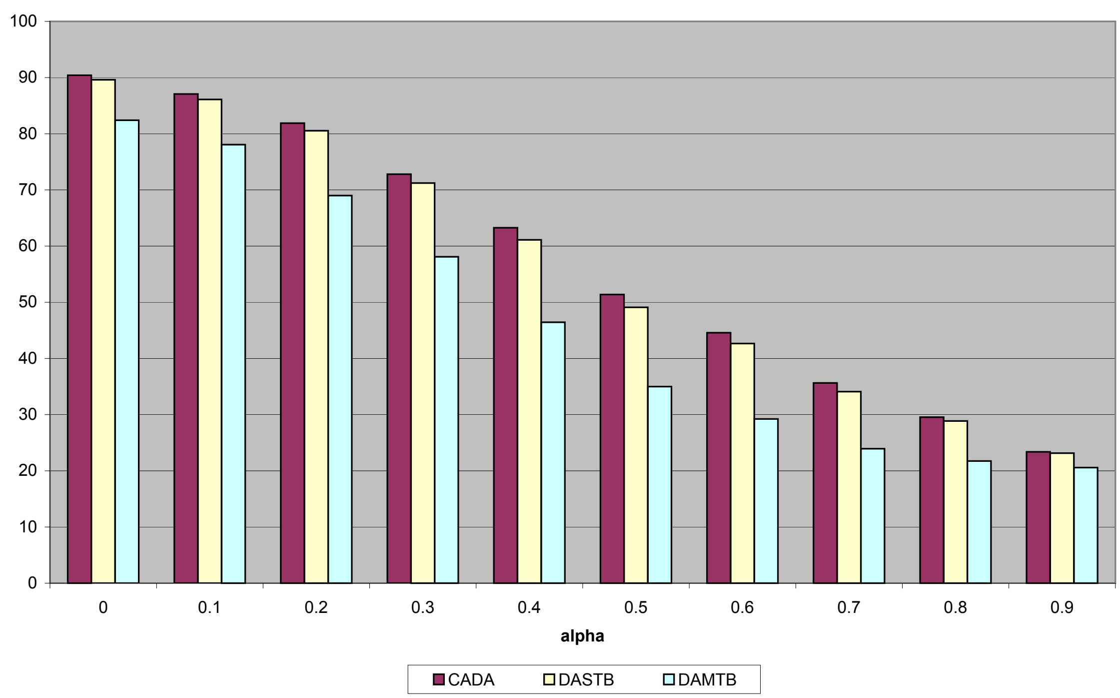

Figure 6: Average Utility of Receivers of kth Choice, CADA vs DASTB

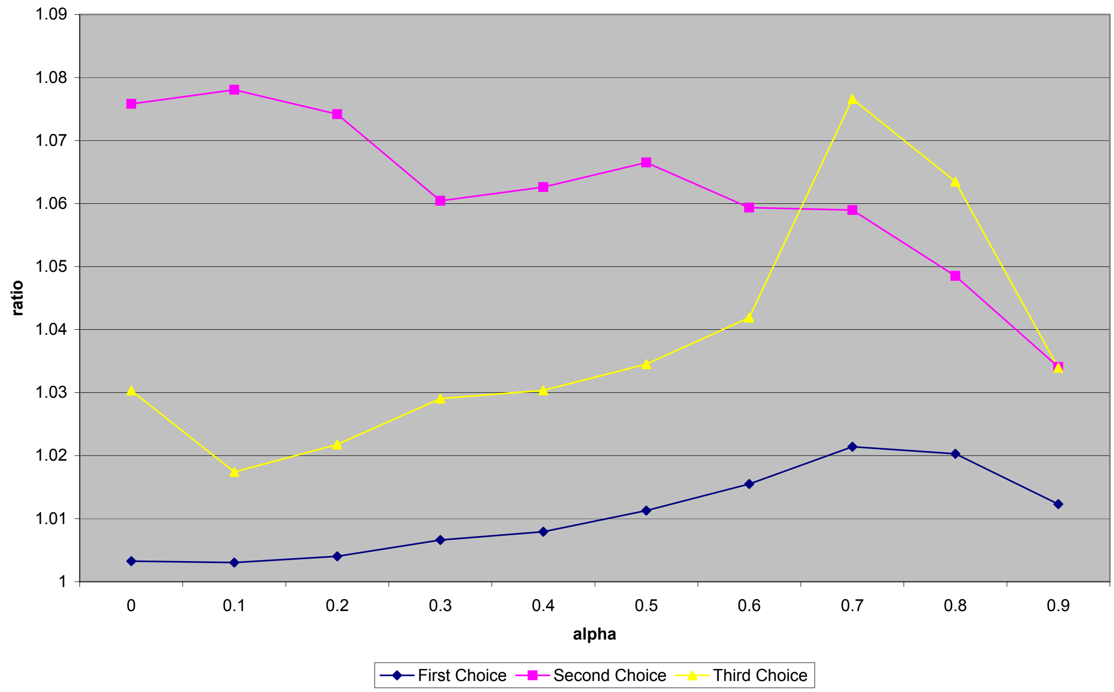

Figure 7: Average Number of Popular Schools and Oversubsribed Schools

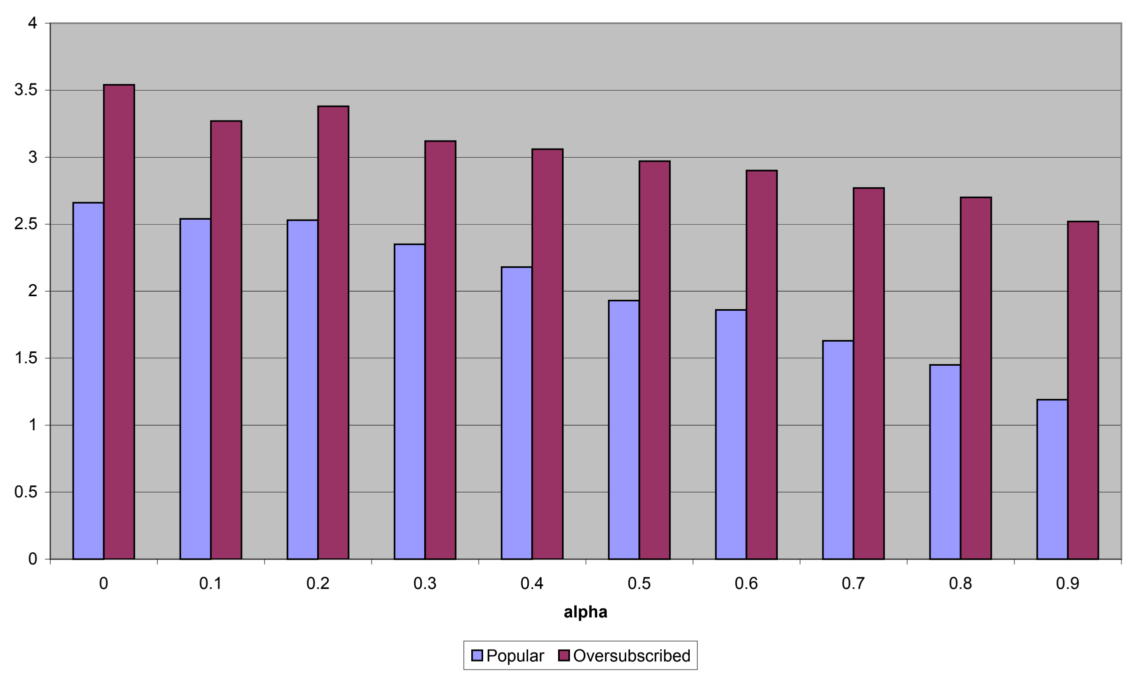

Figure 8: Average Number of Students Selecting kth Choice as Target in CADA Equilibrium

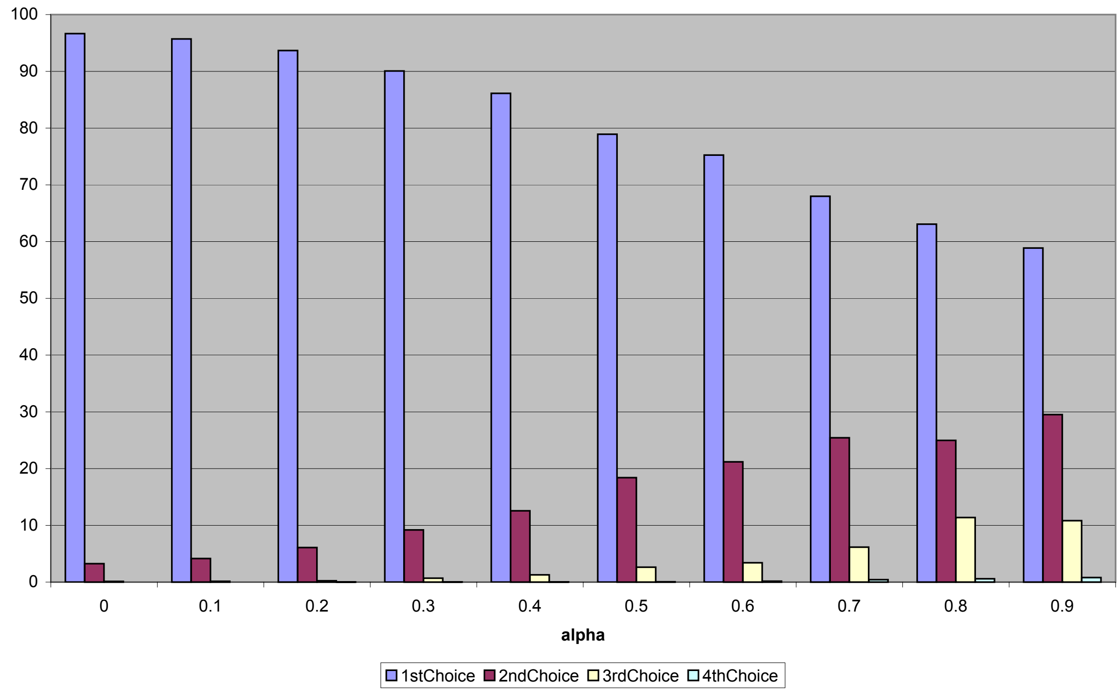

Figure 9: Welfare as Percentage of First Best - with priorities

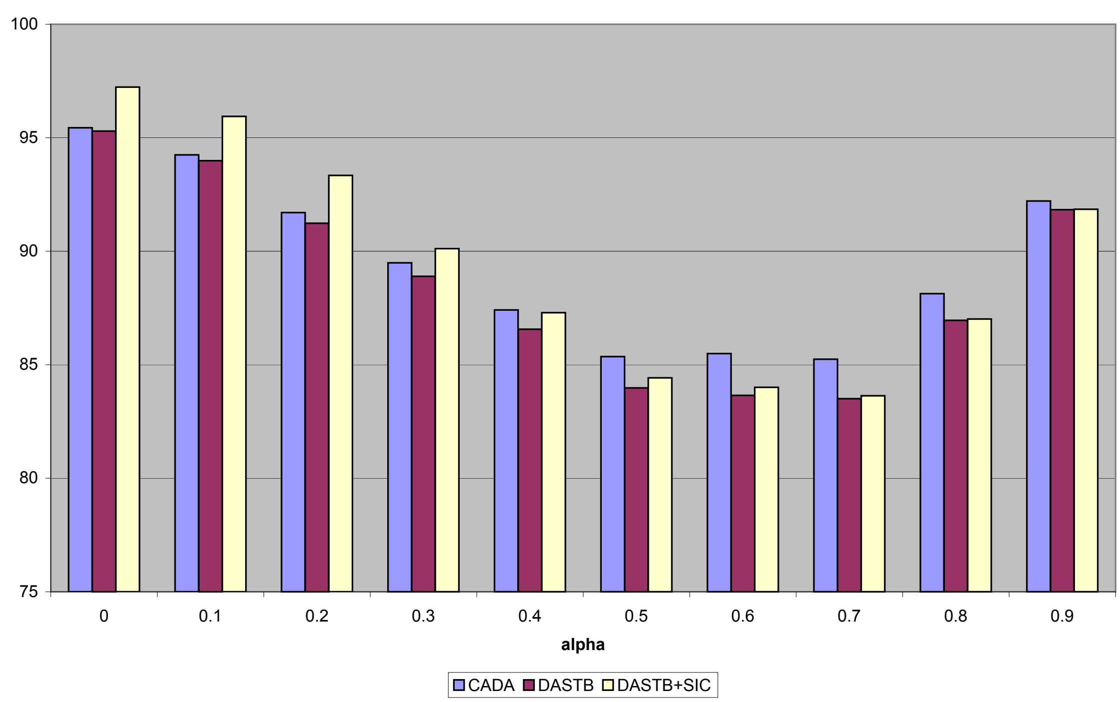

Figure 10: Percentage of Students Selecting Their Neighborhood School as Target

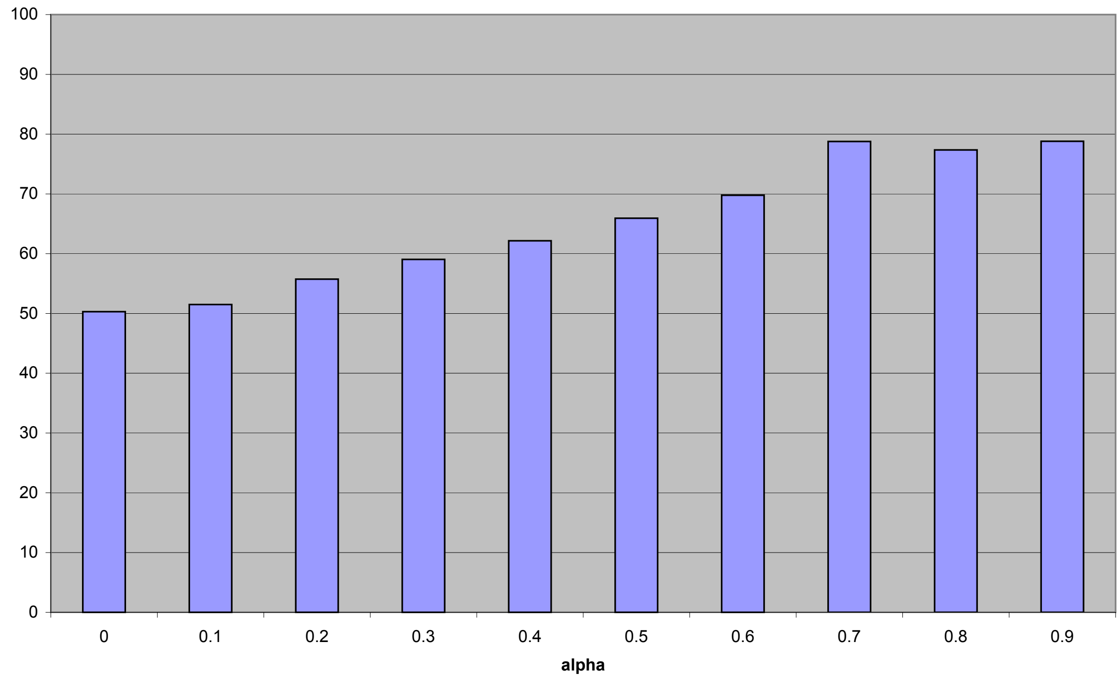

Figure 11: Welfare with Naive Players as Percentage of First Best