Improving on Strategy-proof School Choice Mechanisms: An Experimental Investigation

Clayton R. Featherstone* Muriel Niederle†

Abstract

While much of the school choice literature advocates strategyproofness, recent research has aimed to improve efficiency using mechanisms that rely on non-truthtelling equilibria. We address two issues that arise from this approach. We first show that even in simple environments with ample feedback and repetition, agents fail to reach non-truthtelling equilibria. We offer another way forward: implementing truth-telling as an ordinal Bayes-Nash equilibrium rather than as a dominant strategy equilibrium. We show that this weaker solution concept can allow for more efficient mechanisms in theory and provide experimental evidence that this is also the case in practice. In fact, truth-telling rates are basically the same whether truthtelling is implemented as an ordinal Bayes Nash equilibrium or a dominant strategy equilibrium. This provides a proof-of-concept that ordinal Bayes-Nash design might provide a middle path, achieving efficiency gains over strategy-proof mechanisms without relying on real-life agents playing a non-truth-telling equilibrium.

Keywords: School Choice Mechanisms, Market Design, Experimental Economics, Matching Theory.

*The Wharton School, University of Pennsylvania

†Stanford University and NBER

1 Introduction

Former Boston Public Schools superintendent Thomas Payzant described the rationale for switching away from a manipulable choice mechanism quite succinctly: "A strategy-proof algorithm 'levels the playing field' by diminishing the harm done to parents who do not strategize or do not strategize well."[1,2] This quote highlights two main assumptions implicit in much of the school choice literature. The first is that a significant fraction of parents will strategize poorly if faced with a situation where truth-telling is suboptimal. The second is that if truth-telling is not a dominant strategy, then parents might fail to submit preferences truthfully. In this paper, we present an experiment that investigates both of these assumptions in the context of two common[3] school choice mechanisms: the strategy-proof Deferred Acceptance mechanism ("DA") and the non-strategy-proof, "immediate acceptance" mechanism once used in Boston ("Boston").

The Boston mechanism is quite popular, presumably because it maximizes the number of students that receive their reported first choice.[4] However, it also provides participants with incentives to manipulate their preference reports, and in fact, a large number of experiments have confirmed that students do indeed deviate from truth-telling (Chen and Sönmez 2006, Pais and Pintér 2008, Calsamiglia, Haeringer and Klijn 2010)[5]. These papers mainly focus on determining what rules-of-thumb participants use when choosing their strategically manipulated preference reports. While Boston Public Schools has replaced the Boston mechanism with strategy-proof DA, several recent papers (Abdulkadiroğlu, Che and Yasuda 2008, 2011, Miralles 2009) have demonstrated that equilibrium manipulations under the old Boston mechanism

[1] A school choice mechanism maps students' reported preferences over schools to lotteries over possible allocations of students to schools. A strategy-proof school choice mechanism admits truthful preference revelation as a dominant strategy. A non-strategy-proof mechanism is also called manipulable.

[2] For more about Boston Public Schools' decision to adopt a deferred acceptance assignment system, see Abdulkadiroğlu et al. (2005) and Abdulkadiroğlu et al. (2006).

[3] DA is currently used in New York City (Abdulkadiroğlu, Pathak and Roth 2005, 2009) and Boston (Abdulkadiroğlu et al. 2005, 2006); Boston, or a close variant, is or has been used in in Charlotte-Mecklenburg (North Carolina), Lee County (Florida), Minneapolis, and Seattle (Abdulkadiroğlu and Sönmez 2003). These are, of course, non-exhaustive lists.

[4] Mechanisms like Boston are also popular in two-sided markets, see e.g. Roth (1991). While the Boston mechanism maximizes the number of students that receive the school they ranked as their first choice, it does not necessarily maximize the number of students that receive their first or second choice. For an investigation on mechanisms that seek to directly optimize the distribution of ranks received by students, see Featherstone (2013).

[5] See Calsamiglia, Haeringer and Klijn (2011) and Chen and Sönmez (2011) for comments on the data analysis in Chen and Sönmez (2006).

can yield significant efficiency gains. For example, Abdulkadiroğlu, Che and Yasuda (2011), show in a stylized environment that the Boston mechanism can sustain an equilibrium that dominates the DA equilibrium by allowing students to better express their cardinal preferences.[6] This raises a new question that current experiments fail to address, namely whether students would play close enough to the equilibrium prediction in a Boston mechanism for these gains to be realized.[7] Existing experiments cannot address this question, as they generally did not give participants sufficient information to calculate the equilibrium.[8]

The first part of our experiment tries to close this gap by considering a very simple incomplete information environment that has a unique, non-truth-telling equilibrium under the Boston mechanism.[9] As theory predicts, students truth-tell under DA. Under Boston, we confirm former results that students do not play truth-telling strategies. However, even in our simple environment with feedback and repetition, students fail to play the theoretical equilibrium and in general are far from playing a best response. As such, we confirm the worry of Boston Public Schools superintendent Thomas Payzant that equilibria that entail untruthful reporting may be difficult to successfully implement in actual school choice problems.

Payzant's second assumption is that strategy-proofness may be neccessary to "level the playing field". However, dominant-strategy incentive compatibility, while appealing, imposes a strong constraint on design, and it has been shown empirically that this constraint can come with costs.[10] If we could successfully implement truth-telling in a weaker way, then we would open the door to the possibility of leveling the playing field with non-strategy-proof designs that are customized to induce truth-telling in particular environments.

In the second part of the experiment, we try to show that this could work by considering an environment in which truth-telling is an ordinal Bayes-Nash equilibrium

[6] Troyan (2012) suggests an extension that deals with more complex priority structures.

[7] Abdulkadiroglu et al (2006) provide suggestive evidence that some participants in Boston may not not have manipulated their preference list in an optimal way.

[8] In Pais and Pintér (2008), students can calculate the equilibrium under the full information treatment, but not under their treatments with less information. The paper has however not addressed to what extent students play equilibrium.

[9] In our experiments, students know their own preferences, but only the distribution from which other students' preferences are drawn. They also only know the distribution of the randomization device that is used to choose an allocation from the lottery over allocations outputted by the school choice mechanism. In other words, they don't know the lottery tie-breakers, but they know their distribution.

[10] Abdulkadiroğlu, Pathak and Roth (2009) show that gains from the improvements suggested by Erdil and Ergin (2008) can be significant, but come at the cost of strategy-proofness. See also Kesten and Ünver (2013).

under the non-strategy-proof Boston (and of course, a dominant-strategy equilibrium under DA).[11] In this environment, we show that participants truth-tell when it is an equilibrium to do so, whether it is implemented as an ordinal Bayes-Nash equilibrium or a dominant strategy equilibrium. Indeed, truth-telling rates turn out to be high and not statistically different across the two mechanisms. Furthermore, we empirically observe that the Boston outcome stochastically dominates the DA outcome. This does not imply that Boston is a good school choice mechanism for all environments, but rather that it is a good mechanism for our particular environment. It does suggest, however, that in more complicated environments, we might be able to design better mechanisms that still "level the playing field" in Payzant's sense without being formally strategy-proof (i.e. dominant-strategy incentive-compatible). In the penultimate section of this paper, we investigate this idea further by providing theoretical intuition for why Boston performs so well in our experiment and by considering what sorts of mechanisms might outperform DA in other environments.

The efficiency gains of Boston over DA that are realized in our experiment are quite different from those mentioned in the other papers we've discussed. Our gains come out of a truth-telling equilibrium and are realized from the interim perspective, where students know their own preferences, but only the distribution over others'. Our gains are also not dependent on assumptions about cardinal utility, as the Boston outcomes dominate the DA outcomes in the first-order stochastic sense. Truth-telling is also a best-response, regardless of the underlying von Neumann-Morgenstern utilities. This stands in contrast to research that considers environments where truth-telling isn't a dominant strategy and where gains are achieved through non-truth-telling equilibria that depend on cardinal preferences, e.g. Abdulkadiroğlu, Che and Yasuda (2011).

Before moving on, we highlight two theoretical points made by this paper. First, we investigate both efficiency and equilibrium from the interim information set, where agents know their own preferences, but have only incomplete information about others' preferences. While this view is more common in the two-sided matching literature (Roth and Rothblum 1999, Kojima and Pathak 2009), it is less standard in school choice.[12] The interim approach may, however, be more realistic when considering the behavior of students. Furthermore, it may be the right way to think

[11] An ordinal Bayes-Nash equilibrium is a Bayes-Nash equilibrium in which each player's equilibrium strategy yields an outcome distribution that first-order stochastically dominates the outcome distributions of all other potential strategies (when all other players are playing equilibrium strategies). See Ehlers and Massó (2007).

[12] Notable recent exceptions include Ergin and Sönmez (2006), Erdil and Ergin (2008), Abdulkadiroğlu, Che and Yasuda (2011), and Troyan (2012).

about the objective of a school board that wants to maximize efficiency over several cohorts of students and not simply a specific year (and hence a specific set of realized preferences). Second, we focus on ordinal Bayes-Nash incentive compatibility, which we feel to be the natural solution concept for non-strategy-proof design. Although non-strategy-proof implementation requires some assumption about underlying preferences, ordinal Bayes-Nash equilibrium only requires assumptions about underlying ordinal preferences, which can be easily estimated from reports submitted to a strategy-proof mechanism such as DA. Bayes-Nash equilibrium, although an even weaker solution concept, requires knowledge of the underlying cardinal preferences of students, which are not as easily ascertained.

The rest of the paper is organized as follows. Section 2 introduces the mechanisms and environments used in the experiment and then presents the experimental design. Sections 4 and 5 discuss the results of the experiment. In Section 6, we discuss the intuition of how and when Boston can implement truth-telling, and briefly extend our findings to less symmetric environments. Section 7 concludes.

2 The experimental setup

2.1 Two school choice mechanisms

Formally, a school choice mechanism is a mapping from students' reported preferences to a lottery over allocations of students to schools. This mapping induces a preference revelation game for the students. In the mechanisms that we consider, we generate the outcome lotteries by randomizing the inputs to a deterministic algorithm; specifically, each school draws a strict priority ordering over the students from some distribution. In school choice mechanisms, this random process is usually constructed by starting with a weak ordering based on student characteristics and breaking indifferences with a lottery.[13] Note that we could either draw the lottery independently for each school (multiple lotteries) or just once, using the same lottery for all schools (single lottery).[14] While the priorities of the schools are codified by law or administrative procedure, students are strategic players.[15]

We use two mechanisms in the experiment. The first is based on the Deferred

[13]For example, in Boston, the weak ordering is based on whether a student is within walking distance or has a sibling at the school in question (Abdulkadiroğlu and Sönmez 2003).

[14]This distinction was first made in Abdulkadiroğlu, Pathak and Roth (2005).

[15]A notable exceptions to this is school choice in New York City, where school principals can strategically report their preferences over students (Abdulkadiroğlu, Pathak and Roth 2005).

Acceptance algorithm (Gale and Shapley 1962) and has been introduced in Boston and New York City schools (Abdulkadiroğlu, Pathak and Roth 2005, Abdulkadiroğlu et al. 2005).

The Deferred Acceptance Algorithm (DA)

Step 1**:** Students apply to their first choice school. Schools reject the lowest-ranking students in excess of their capacity. All other offers are held *temporarily*.

Step t**:** If a student is rejected in Step $t-1$, he applies to the next school on his rank-order list. If he has no more schools on his list, he applies nowhere. Schools consider both new offers and the offers held from previous rounds. They reject the lowest ranked students in excess of their capacity. All other offers are held *temporarily*.

STOP: The algorithm stops when no rejections are issued. Each school is matched to the students it is holding at the end.

Two well known results concerning the revelation game induced by DA are that truth-telling is a dominant strategy (Dubins and Freedman 1981, Roth 1982a,b) and that the outcome is stable relative to reported preferences (Gale and Shapley 1962).[16] The second mechanism we consider is based on a specific priority algorithm, first described as the "Boston Algorithm" by Abdulkadiroğlu and Sönmez (2003). This mechanism is commonly used in American public school choice (Abdulkadiroğlu et al. 2005) and can be thought of as an *immediate acceptance* algorithm.

The Boston Algorithm

Step 1*:* Students apply to their first choice school. Schools reject the lowest-ranking students in excess of their capacity. All other offers are *immediately accepted* and become *permanent* matches. School capacities are adjusted accordingly.

Step t**:** If a student is rejected in Step $t-1$, he applies to the next school on his rank-order list. If he has no more schools on his list, he applies nowhere. Schools reject the lowest ranked students in excess of their capacity. All other offers become *permanent* matches. School capacities are adjusted accordingly.

[16] Relative to strict preferences for both schools and students, an outcome is called *stable* if it is individually rational and there is no *blocking pair* (a student and a school who each prefer each other to their assigned match). See the Appendix for a formal definition.

All schools prioritize...		All students prefer...			
Top	$\succ$ Average	Best	$\succ$ Second	$\succ$ Third	$\succ$ Unmatched
(Ties broken by lottery)		$u = 100$	67	25	0
3 students	*2 students*	*2 seats*	*1 seat*	*1 seat*	∞ *seats*

Figure 1: The aligned preference environment

STOP: The algorithm stops when no rejections are issued.

Note that in contrast to DA, where applications are tentatively held and acceptance is deferred until the end, applications in the Boston algorithm are *immediately accepted or rejected* at each step. The chance that one might be able to lock in a school over a higher priority student by ranking that school earler in the algorithm is intuitively why Boston, unlike DA, fails to be strategy-proof.

2.2 Two preference environments

2.2.1 The aligned preference environment

The first environment we consider is the **aligned preference environment**. In it, there are three schools – "Best", "Second", and "Third" – and five students who are either "Top" or "Average". Best has two seats, while the other schools only have one, and three of the students are Top, while the other two are Average. Each school prefers a Top student to an Average one, and all students prefer Best to Second to Third. All students also have the same von Neumann-Morgenstern payoffs over schools: 100 for Best, 67 for Second, 25 for Third, and 0 for being unmatched. Each school has an independently drawn uniform lottery to order students of the same priority class. This environment is summarized in Figure 1.

The matching that results under DA in equilibrium is the unique stable matching: the two Top students with highest lottery draws at Best get Best, the other Top gets Second, and the Average with the highest lottery draw at Third gets Third, leaving the remaining Average unmatched.

Proposition 1. *In the aligned preference environment, all Bayes-Nash equilibria of the revelation game induced by DA assign the three Top students to Best and Second, one Average student to Third, and leave the other Average unmatched. Furthermore, all Bayes-Nash equilibria achieve this outcome through strategies of the following form:*

- *Top students submit preferences that rank Best first and Second second.*

- *Average students submit preferences that declare Third acceptable.*

Note that truth-telling is a dominant strategy under DA. Under Boston, the unique stable outcome is not achievable in equilibrium when lottery draws are not known;[17] in fact, when participants are risk-neutral, the pure-strategy Bayes-Nash equilibrium outcome is unstable.

Proposition 2. *In the aligned preference environment, when students are risk-neutral, there exist pure-strategy Bayes-Nash equilibria of the revelation game induced by Boston. All such equilibria assign the three Top students to Best and Third, one Average student to Second, and leave the other Average unmatched. Furthermore, all pure-strategy Bayes-Nash equilibria achieve this outcome through strategies of the following form:*

- *Top students submit preferences that rank Best first and Third second.*

- *Average students submit preferences that rank Second first.*

The equilibrium of Proposition 2 entails two types of manipulations. Top students need to misreport their second choice school, submitting Third instead of Second, that is, they must "skip the middle". Average students need to misreport their first ranked school, submitting Second instead of Best, that is, they must "skip the top". Note that, in equilibrium, relative to DA, the Average students are in expectation better off, while the Top students are, in expectation, worse off.

2.2.2 The uncorrelated preference environment

The second environment we consider is the **uncorrelated preference environment**. In it there are five students and four one-seat schools. Student preferences are drawn independently from the uniform distribution over all possible orderings that find all schools acceptable, and likewise, each school has a strict priority based on an independent draw from the uniform distribution over all orderings of the students. In the experiment the payoffs are such that a seat at the highest ranked school yields 110 points, the second highest 90, the third highest 67 and the fourth 25. Being unmatched yields 0 points.

[17] Per Ergin and Sönmez (2006), if the lottery draws at each school are common knowledge, then the equilibrium outcome with Boston is the same as with DA. One set of supporting strategies is as follows: the two Top students with the best lottery numbers at Best rank it first, the third Top student ranks Second first, the Average student who has the better lottery number at Third ranks it first, and the last Average student submits arbitrary preferences (as he will not receive any school).

In this environment, both mechanisms admit all agents truth-telling as a Nash equilibrium, but while truth-telling is a dominant strategy under DA, it is only an ordinal Bayes-Nash equilibrium strategy under Boston.[18] Under DA, truth-telling is the unique equilibrium in weakly undominated strategies, and under Boston, truth-telling is the unique equilibrium in *weakly undominated, anonymous strategies*. A strategy is anonymous if the reported rank of a school onlys depends on its true rank (and not its name).[19] Note that equilibrium predictions in the uncorrelated preference environment are not sensitive to the cardinal values. In the Appendix, we prove propositions that admit the following corollaries.

Corollary (to Proposition 5 and Lemma 2)**.** *In the uncorrelated environment, the revelation game induced by DA admits truth-telling by all students as an ordinal Bayes-Nash equilibrium; furthermore, this is the unique Bayes-Nash equilibrium in weakly undominated strategies.*

Corollary (to Proposition 8)**.** *In the uncorrelated environment, the revelation game induced by Boston admits truth-telling by all students as an ordinal Bayes-Nash equilibrium; furthermore, this is the unique Bayes-Nash equilibrium in anonymous, weakly undominated strategies.*

Loosely, the propositions in the Appendix show that there is an entire family of environments in which we expect truth-telling under Boston. These environments exhibit significant symmetry in the preference distributions, which can either be interpreted as preference idiosyncrasy, or as uncertainty in the beliefs that students hold about other students' submitted preferences.

Table 2a shows boot-strapped outcome distributions under the Boston and DA truth-telling equilibria.[20,21] Note that these distributions are interim, that is, they are from the information set of a student who knows his own preference, but only the distribution of other students' preferences and the lottery. The cumulative columns make it clear that the outcome distribution under Boston first-order stochastically

[18] For a definition of an an ordinal Bayes-Nash equilibrium, see Footnote 11.

[19] Another name for anonymous strategies could be "label-free". See the Appendix for a more formal definition.

[20] Actually, the outcome distribution under single lottery DA can be calculated in closed form, and a recursion exists for calculation the outcome distribution under Boston. Deriving these formulae would take us too far afield from the task at hand, so it is easier to think of these numbers as coming from a bootstrap, as they did for multiple lottery DA, which does not have a neat solution.

[21] There is only one column for Boston because that mechanism does not yield a different distribution over the two lottery structures in this environment.

dominates that of DA.[22] Figure 2b shows the equilibrium outcome CDFs in this environment. Note that in this representation, a distribution first-order stochastically dominates another when it is completely *above* it. In Section 5 we provide intuition and further investigate why the Boston outcome first-order stochastically dominates outcomes under DA when students submit preferences truthfully.

2.3 Experimental design

To provide an empirical test of whether agents manipulate their preference reports under the Boston algorithm, we ensure that students always have sufficient information to compute the ordinal Bayes-Nash equilibrium by informing them about the distribution from which preferences and lotteries are drawn.[23] This allows us to do three key analyses. First, we can compare submitted strategies to the equilibrium prediction. Second, we can test whether some participants manipulate their reports to maximize their payoffs given others' reports. Third, we can gain some intuition concerning which deviations from equilibrium may be more common or easy to learn. Students play the experimental game 15 times to allow for learning.[24]

Our treatments form a 2×2 factorial design: (DA, Boston)$\times$(Aligned preference environment, Uncorrelated preference environment). The environments and payoffs were discussed in Section 2.2. Payoffs were chosen to provide sufficient incentive for students to play equilibrium.[25] In all environments, each school independently draws its own lottery. Note that under DA (but not under Boston), using a single lottery instead of multiple lotteries may have an effect, which we will adress in the analysis of Section 4.2.

[22] We focus on ordinal preferences, as students, in general, submit ordinal rankings. When cardinal preferences are considered, and efficiency is measured as the ex post sum of student welfares, then Boston may dominate DA in some environments, generally by way of a non-truth-telling equilibrium (Abdulkadiroğlu, Che and Yasuda 2008, Miralles 2009).

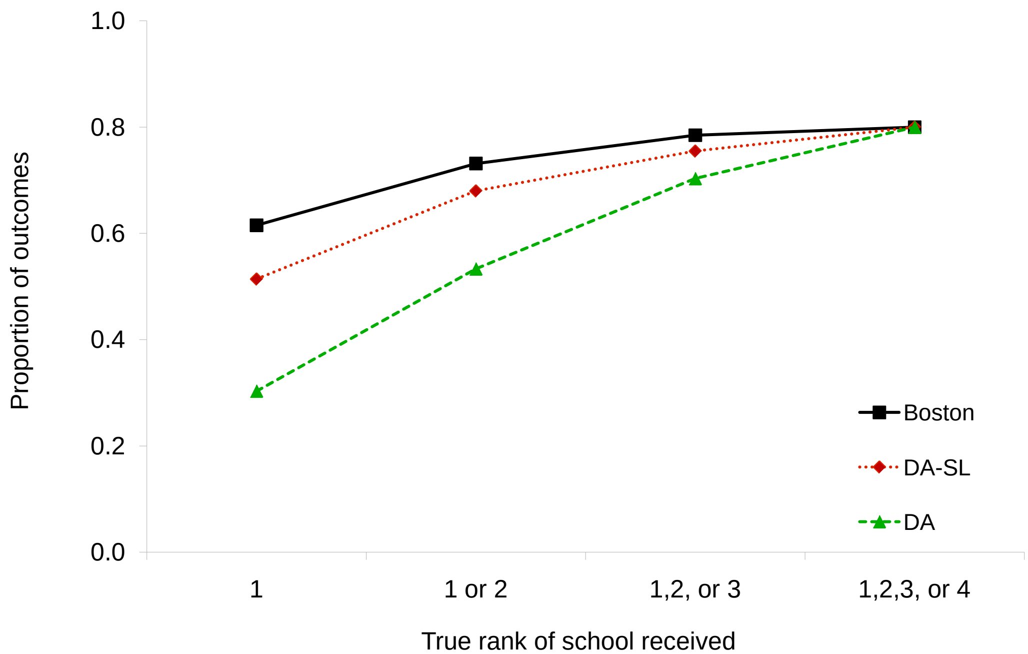
[23] Testing whether and how agents manipulate reports under the Boston algorithm is difficult to address in the field since the preferences of students are unobserved. Abdulkadiroğlu et al. (2006) analyzes submitted preferences and find some evidence of potentially sub-optimal behavior. Barros, Featherstone and Salcedo (2010) use a survey to assess preferences of students who then participate in a Boston-like mechanism and compare preference reports. This method is also used by Budish and Cantillon (2012).

[24] Repetition allows us to see whether behavior remains stable and whether participants may eventually learn to manipulate in a Boston mechanism and report truthfully in a DA mechanism. While parents in general participate in school choice mechanisms only once, they often draw from the experiences of other parents, and many districts have school choice mechanisms at several points in a child's education. In using multiple rounds, we also follow the tradition of two-sided matching experiments (Kagel and Roth 2000, Ünver 2001, McKinney, Niederle and Roth 2005).

[25] As a reminder, the cardinal payoffs only matter for equilibrium predictions in one treatment group: when we look at behavior under the Boston mechanism in the aligned preference environment.

Mechanism	Boston		DA (single lottery)		DA (multiple lotteries)	
1st choice	0.610		0.500		0.328	(0.328)
2nd choice	0.117	(0.727)	0.167	(0.667)	0.222	(0.550)
3rd choice	0.055	(0.782)	0.083	(0.750)	0.150	(0.700)
4th choice	0.018	(0.800)	0.050	(0.800)	0.100	(0.800)
No school	0.200	(1.000)	0.200	(1.000)	0.200	(1.000)

(a) Outcome distributions (CDF in parentheses)



(b) CDF plots (DA-SL is single lottery; DA is multiple)

Figure 2: Equilibrium outcomes in the uncorrelated environment

We ran four sessions under DA and seven sessions under Boston. In each session, five Stanford undergraduate students played for 15 periods in the aligned preference environment, during which players kept their role, as either a Top or an Average student. Then, after a pause to learn about the new environment, they played another 15 periods in the uncorrelated preference environment, where preferences for all students and lotteries for all schools are redrawn in each period.

In our design, participants always see the aligned preference environment before they see the uncorrelated preference environment. This may lead to less truth-telling in the Boston-uncorrelated treatment, which would actually work *against* what we are trying to show. The reason we chose the within-subjects design is that it captures the flavor of the likely strategic transition that students face when a district changes its school choice mechanism: they have to learn to truth-tell after being conditioned to strategically manipulate.

Within a session, the mechanism was held constant, and each subject participated in only one session. The experiment was conducted on computers, using z-Tree (Fischbacher 2007). At the start of a session, after reading instructions concerning the environment and the mechanism, we checked each player's understanding by having them solve the outcome of a complete information test environment, where participants were given submitted preferences and had to determine the outcome of the relevant mechanism. We repeatedly checked understanding by correcting and explaining outcomes through each subsequent step of the algorithm. Participants earned 1.5 cents for every point and were paid based on their cumulative earnings over all 30 periods of the experiment. In addition participants received a seven dollar show-up payment, bringing average payments to about twenty-five dollars for an average of 60 minutes in the lab.

3 Results from the aligned preference environment

In the aligned preference treatment, theory predicts truth-telling under DA and equilibrium deviation from truth-telling under Boston. While theory is born out under DA, under Boston, we see deviation from truth-telling, but not according to equilibrium and not in a way easily ascribed to best-responses. To assess the behavior of participants relative to equilibrium play we focus the analysis of this section on the last five periods, and hence on periods where we expect that ample learning opportunities may have helped participants converge to equilibrium (or at least some sort of stable play). The results are qualitatively the same, though slightly worse for the

Boston equilibrium prediction, when we focus on all 15 periods instead.

3.1 Strategic behavior under Boston

In the aligned environment, under the Boston mechanism, we did not observe a single period in which every player used risk-neutral Bayes-Nash equilibrium strategies.[26] What's more, the strategic deviation from equilibrium leads to an outcome that significantly deviates from equilibrium, as can be seen in Table 1. Even in this simple environment, participants do not seem to play equilibrium, which casts doubt on the notion that non-truth-telling equilibria might predict behavior in actual school choice settings.

Given that participants do not play equilibrium, we investigate two alternative theories of behavior. The first is that agents are simply truth-telling. This is not the case: the truth-telling rate is only 1.43% for Average students, confirming existing experimental results (Chen and Sönmez 2006, Pais and Pintér 2008, Calsamiglia, Haeringer and Klijn 2010).[27] Investigating the distribution of submitted first choices, we see in Table 2 that the vast majority of strategies submitted by Average students rank some school other than Best as their first choice. Not ranking unachievable schools seems easy to learn, and Average students do so almost immediately.[28] For Top students, the equilibrium response to this behavior is to report Third as their second choice. Combining the facts that 1) Tops report Best as their favorite school 92% of the time, and 2) Tops truth-tell 65.7% of the time, it seems that the Tops are having a much harder time learning to "skip the middle" than Average students did

School	Best	Second	Third	No school
Top	0.67	0.11	0.05	0.17
	$\left(\frac{2}{3}\right)$	(0)	$\left(\frac{1}{3}\right)$	(0)
Average	0.00	0.33	0.43	0.24
	(0)	$\left(\frac{1}{2}\right)$	(0)	$\left(\frac{1}{2}\right)$

Table 1: Empirical distribution of outcomes under Boston (equilibrium predictions in parentheses)

[26] This includes all 15 periods in all 7 sessions.

[27] We find that 13 out of 14 Average students deviate from truth-telling 80% of the time or more. Eleven (of 14) Average students skip school Best 100% of the time.

[28] Even in the first five rounds, Average student truth-telling was at 6%.

School	Best	Second	Third
Top	0.92	0.07	0.01
Average	0.06	0.67	0.27

Table 2: Empirical distribution of first choices under Boston

learning to "skip the top".[29]

An alternative theory of behavior is that some Top students are failing to best-respond, but that other, more sophisticated Tops are best-responding to the non-equilibrium behavior of the others. To test this, one option would be to look, round by round, for ex-post best-responses; however, this seems too stringent a test, given that participants did not have sufficient information to compute the ex-post best response. Instead, we apply a weaker test by identifying instances in the data where behavior can be unequivocally classified as suboptimal.

One of these is when Top players rank Third as their first choice, but only one player ever does this, and only in one period. This clearly sub-optimal behavior does not seem difficult for Tops to avoid. There are also other, more frequent instances in the data where we can clearly identify sub-optimal behavior. For instance, if all Tops rank Best first, and both Averages rank Second as their first choice and Third as their second choice, then Tops are constrained to best-respond by ranking Third as their second choice. To investigate whether Tops might be best-responding to each other, we will look for instances in sessions where the Tops could best-respond with the same strategy for each of the last 5 periods, that is, we ignore sessions in which the best-response for Tops is still fluctuating in the last 5 rounds.[30] The details of this analysis are in Appendix A; we summarize the results here.

In the one session in our analysis where the Tops' best-response is truth-telling in the 5 final rounds, all three truth-tell 80% of the time or more. In contrast, in the four sessions of our analysis where the Tops' best-response is "skipping the middle" only 2 of 12 do so 80% of the time or more. Of those that remain, 7 truth-tell 80% of the time, while the rest fail to play a consistent strategy. So even if we can't explicitly rule out some "sophisticated" Tops who best-respond to the suboptimal behavior of the others, we see that they are neither a large part of our experimental population

[29] In fact, a Wilcoxon signed-rank test comparing truth-telling rates for Top students across the aligned and uncorrelated environments indicates that the difference is not significant ($p = 0.61$).

[30] These limits are not overly restrictive: we are able to look at 5 of the 7 sessions in the Boston-Aligned treatment.

nor are they able to steer outcomes closer to equilibrium. It is worth noting that, when analyzing reports submitted to a Boston mechanism in the field, it is a lack of manipulation of this sort ("skipping the middle") that may be the most identifiable, even if true preferences of students are not known. Our experimental results thus dovetail nicely with Abdulkadiroğlu et al. (2006), which finds suggestive evidence that students sometimes rank schools below their first choice that are expected to be filled in the first step of the algorithm, that is, that many students fail to "skip the middle".[31]

In short, under Boston, participants are not playing equilibrium, even when we focus only on the last five periods of play, after participants had ample feedback and repetition. This suggests that a non-truth-telling equilibrium might not be predictive of behavior in the field, and hence, gains from such an equilibrium might fail to materialize.

3.2 Truth-telling under DA

The outcome in the last five periods under the DA mechanism is given by Table 3, which shows for each student type, the fraction of matches at each school. The outcome corresponds exactly to the equilibrium: the three Top students receive seats at Best and Second, while Average students either receive a seat at Third or remain unmatched. Note that under DA, compared to Boston, Top students are better off and Average students are worse off.

We begin our investigation of strategies with the submitted first choices (Table 4). Sufficient conditions for the outcome under DA to yield the stable match constrain Top students to submit preferences $(Best, Second, x)$, while only constraining Average students to rank Third as acceptable. In spite of the fact that truth-telling is not a *strict* best response, we observe that 92% of Top and 63% of Average student strategies are truth-telling.[32,33]

To compare truth-telling rates between Boston and DA, we use Mann-Whitney tests on session averages from our 7 Boston and 4 DA sessions. Running the test for

[31] If a school is filled in the first step of a Boston mechanism, then no student who has not ranked the school first has a chance of matching to it. If a student believes that a school will fill in the first step of the algorithm, then it is a clear mistake to rank it lower than first.

[32] Note that under DA, 11 out of 12 Top and 4 out of 8 Average students truth-tell 80% of the time or more. In contrast, under Boston the numbers were 13 out of 21 for Top students and 0 out of 14 for Average students.

[33] When truth-telling is a dominant strategy, overall truth-telling rates are not significantly different as we change the environment. They are 80% in the aligned and 66% in the uncorrelated environment, $(p = 0.14)$ using a Wilcoxon signed rank test.

School	Best	Second	Third	No school
Top	0.67	0.33	0.00	0.00
	$\left(\frac{2}{3}\right)$	$\left(\frac{1}{3}\right)$	(0)	(0)
Average	0.00	0.00	0.50	0.50
	(0)	(0)	$\left(\frac{1}{2}\right)$	$\left(\frac{1}{2}\right)$

Table 3: Empirical outcome distribution under DA (equilibrium predictions in parentheses)

School	Best	Second	Third
Top	1.00	0.00	0.00
Average	0.70	0.05	0.25

Table 4: Empirical distribution of first choices under DA

the two student types separately, in the aligned environment, we find that both Top ($p = 0.07$) and Average students ($p < 0.01$) are significantly more likely to use truth-telling strategies under DA than under Boston. This shows that the manipulations of both Average and Top players under the Boston mechanism are not due to the environment alone, but rather are driven by the combination of environment and mechanism.

4 Results from the uncorrelated preference environment

In the uncorrelated environment, under the Boston mechanism, everyone truth-telling is an ordinal Bayes-Nash equilibrium, while under DA, everyone truth-telling is a dominant strategy equilibrium. If students truthfully reveal their preferences, we expect the Boston outcome to first-order stochastically dominate that of DA. In the previous section, we looked at the last 5 periods to allow for ample feedback and repetition, effectively stacking the cards against our finding that the equilibrium is not predictive under Boston. In this section, we hope to show that truth-telling is an easy equilibrium to find, even without much feedback and repetition. To this end, we will stack the cards against ourselves by considering all rounds of play. Furthermore,

		True Preferences							
		Boston				DA			
		1st	2nd	3rd	4th	1st	2nd	3rd	4th
Submitted preferences	1st	0.76	0.22	0.01	0.01	0.74	0.19	0.07	0.00
	2nd	0.16	0.61	0.13	0.10	0.14	0.69	0.15	0.02
	3rd	0.06	0.07	0.80	0.07	0.07	0.11	0.77	0.05
	4th	0.02	0.10	0.06	0.82	0.05	0.01	0.01	0.93

Table 5: Empirical distribution of submitted ranks in the uncorrelated environment

the theoretical gains of Boston over DA arise in expectation *across* realizations of student preferences, so analyzing all rounds provides us with more data for these gains to take effect.

4.1 Strategies

To compare strategies between Boston and DA, note that basically all submitted strategies rank all schools.[34] The proportion of truth-telling strategies is 58% under Boston, compared with 66% under DA.[35] This difference is not significant: a Mann-Whitney test across mechanisms, comparing mean truth-telling rates in each session, yields a p-value of 0.70 ($n = 11$).[36]

To address what manipulations are submitted, we check truth-telling rates at each ranked position, that is, how often a participant's submitted k^{th}-ranked school corresponds to his true k^{th}-ranked school (see Table 5). Note that there is little discernible difference across mechanisms, as predicted by equilibrium, even though the equilibria are implemented according to different solution concepts. A plausible explanation is that truth-telling holds special sway as a focal strategy.[37]

[34]Only one student, for three rounds at the beginning of the uncorrelated preference environment, failed to rank all schools.

[35]That the students truth-tell at a much higher rate in the uncorrelated environment for Boston tells us that they are strategically responding to the environments. A Wilcoxon signed-rank test comparing truth-telling rates for Average students across the aligned and uncorrelated environments indicates that the difference is highly significant ($p = 0.02$).

[36]Truth-telling rates declined slightly from the first five periods to the last. The rates in the first five periods were 74% under DA and 61% under Boston.

[37]This idea is not a new one; Pathak and Sönmez (2008) look at a model in which some player are optimally strategic and others sub-optimally truth-tell.

4.2 Outcomes

Before we begin our analysis, note that although the DA mechanism we ran in the lab used multiple lotteries, recent work by Abdulkadiroğlu, Pathak and Roth (2009) provides simulations that indicate that DA with a single lottery may produce a better outcome in some environments. While this is the case in our environment, Boston still theoretically dominates DA with either multiple or single lottery tiebreakers. To make sure that Boston is outperforming the single lottery DA (that we did not use in the lab) as well as the multiple lottery DA (that we did use in the lab), we compute the counter-factual outcome had we used a single lottery in the laboratory as a further test.[38] We do this by taking the participants' submitted preferences and using the lottery draw for one of the schools as the universal lottery. Since we have four such lotteries (one per school) each period, we use each of them as a universal lottery and then average the resulting outcome distributions. In the figures, we denote this by "DA-SL". DA with multiple lotteries, as was used in the actual experiment, is denoted simply by "DA".

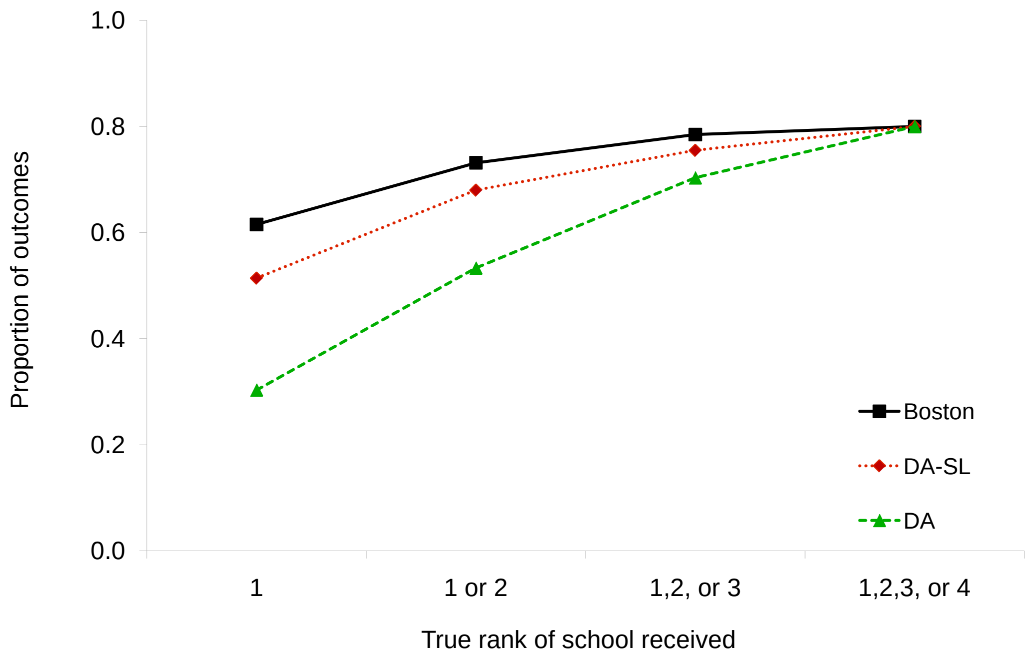
We begin our analysis by recalling Figure 2b, which shows just how much gain over DA is possible if all subjects truth-tell, averaged across all possible preference and lottery draws. Of course, in the lab, it is possible that we get "bad draws" which undermine our potential to recreate the theoretical gains. To show that this is not the case, we construct a truth-telling counterfactual, calculating the outcome distribution given the preference and lottery draws realized in the lab, but constraining the truth-telling rate to be 100%. The CDF for this counterfactual is in Figure 3a. Fortunately, the gains are still quite large and are essentially identical to those that theory predicts: participants would have a 61.5% chance to receive their first choice under Boston, which is significantly larger than the 30.3% under DA with multiple lotteries ($p < 0.01$) and the 51.4% under DA with a single lottery ($p < 0.01$).[39] When we compare the proportion of students who receive either their true first or second choice, once more Boston has significantly higher proportions than either DA or DA-SL ($p < 0.01$ in both cases).[40]

Now, we move to the actual outcome distribution observed in the lab. Even

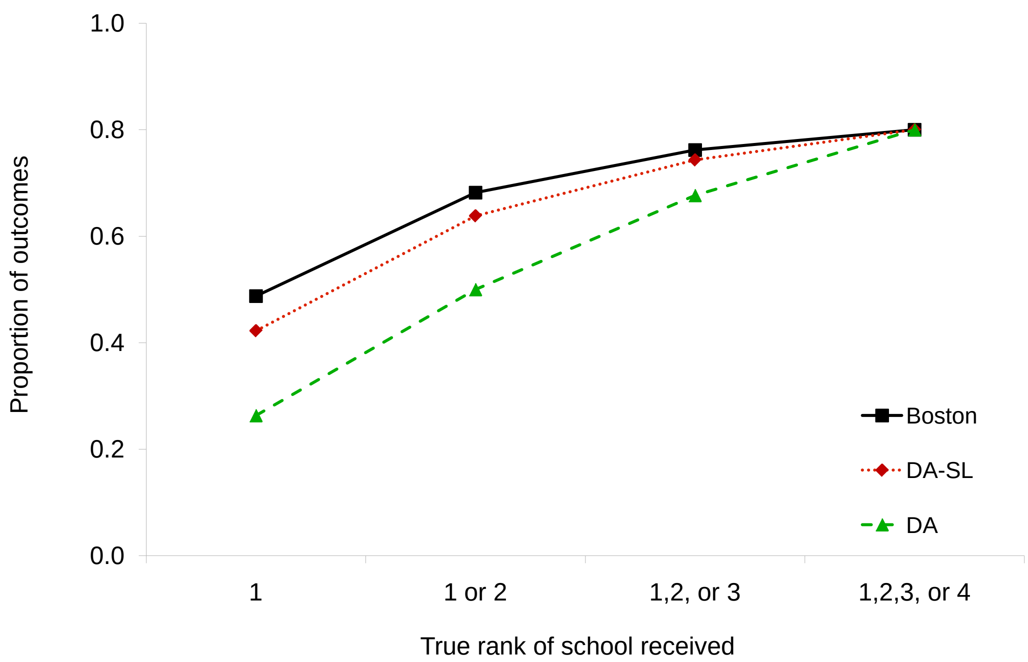
[38] It is not necessary to consider multiple versus single lottery setups under Boston: both yield the same distribution over outcomes.

[39] We use Mann-Whitney tests, where the session mean is a data point, that is, we have 7 data points for Boston, and 4 for DA and DA-SL. When we only consider the last five periods, the p-values are 0.01 and less than 0.01, respectively.

[40] When we look at the last five rounds only, the difference between Boston and DA is still significant at $p < 0.01$, while the one between Boston and DA-SL is not ($p = 0.12$).



(a) Lab outcomes under truth-telling counterfactual



(b) Actual lab outcomes

Figure 3: Outcome CDFs with preferences realized in lab (DA-SL is the single lottery counter-factual)

with a less than perfect truth-telling rate, we observe a significantly larger fraction of students receive their true first choice with Boston than with DA ($p < 0.01$), or DA-SL ($p = 0.06$).[41] When we compare the proportion of students who receive either their first or second choice, Boston once more significantly outperforms both DA ($p < 0.01$) and DA-SL ($p = 0.09$).[42] These findings can be seen graphically in Figure 3b.

Hence, it seems that the theoretical advantages of Boston over DA can indeed be realized in practice, at least in our specialized environment. This provides a proof-of-concept that non-strategy-proof mechanisms might perform well in the field if they are ordinal Bayes-Nash incentive-compatible.

5 Why did Boston succeed?

5.1 The art and science schools example

To understand why Boston outperforms DA in the uncorrelated environment, it helps to simplify. Suppose there are three students and two one-seat schools – an art school and a science school. Each student is equally likely to be an artist (`a`) who prefers the art school or a scientist (`s`) who prefers the science school. All students find both schools acceptable, and a common lottery determines priorities at both schools. In this environment, corollaries to the same Appendix B propositions mentioned in Section 2.2 tell us that truth-telling is the unique equilibrium under both Boston and DA.

We will consider the outcomes of the two mechanisms both from the interim and ex ante points of view. States of the world from the ex ante perspective can be represented as a string of three `a`'s or `s`'s; for instance, `aas` represents the state where the students with the highest two lottery draws are both artists, and the remaining student is a scientist.[43] For interim states,[44] we will use a capital letter to denote the student whose perspective we are considering; for instance, `aaS` represents the state

[41] When we only consider the last five periods, Boston still gives a higher fraction of participants their first choice. The difference is significant when compared to DA ($p < 0.01$), though not when compared to DA-SL, ($p = 0.18$). Note that a one-sided test would yield significance in all the two-sided tests that we have discussed.

[42] When we consider only the last five periods, the p values are < 0.01 and 0.03 respectively.

[43] Since we will only consider anonymous mechanisms, there is no need to keep track of whether the first artist is Student 1, 2, or 3. This simplifies the state space considerably without losing any of the intuition.

[44] That is, states from the perspective of a student who knows her own preferences, but not those of others and not the lottery draw.

aaa	$\boxed{\texttt{aas}}$	ass	sss
	asa	sas	
	saa	$\boxed{\texttt{ssa}}$	

(a) The 8 (equally likely) states of the world from the ex ante perspective. Boxed cells indicate cells where Boston and DA give different overall matches.

S aa	a S a	$\boxed{\texttt{aa S}}$
S as	a S s	as S
S sa	$\boxed{\texttt{s S a}}$	sa S
S ss	s S s	ss S

(b) The 12 (equally likely) states of the world from the (interim) perspective of scientist S. Boxed cells indicate that Boston and DA give different schools to S.

Figure 4: States of the world: ex ante and interim.

of the world we just mentioned, from the perspective of the scientist with the lowest lottery draw.

The ex ante perspective is that of a social planner (the "superintendent") who thinks about the quality of overall matches in terms of the distribution of how students rank their assigned schools.[45] In this environment, such social preferences are equivalent to maximizing the probability that an agent gets his first choice.[46] When comparing Boston and DA, the superintendent only cares about the two states in which the two mechanisms yield different matches: aas and ssa (see Figure 4a). Intuitively, these are states where DA gives a school to a student who ranks it second, even though there is an unmatched student who ranked it first. Note that Boston never does this *in any environment*.[47] So, assuming truth-telling, from the ex ante perspective, Boston is either the same as or better than DA in every state of the world. This is the precise way in which Boston outperforms DA.

Instead of Boston, however, consider running DA with the addition of a simple ex post reallocation rule: when the state of the world is aas, take the science school from the second artist and give it to the scientist. Assuming truth-telling, this clearly makes the superintendent better off, as it is the same as or better than DA in all states of the world. While ex ante students are better off, from the interim perspective the science student S is better off while an art student a is worse off. This may give an artist who is almost indifferent between the art school and the science school an incentive to masquerade as a scientist, since such an artist is more worried about being

[45] This is a common and natural way to aggregate welfare with limited information. See Featherstone (2013) for more.

[46] Everyone is unmatched one-third of the time by symmetry, so the ex ante probability of getting a first choice is sufficient for the entire distribution of ranks received.

[47] Kojima and Ünver (2010) show that Boston "respects preference rankings", which implies the statement above. Intuitively, this is because, under Boston, if a student is rejected by his k^{th} ranked school, it must have filled up by being assigned students who ranked it k^{th} or higher.

unmatched when the state is `aas` than about sometimes getting the science school instead of the art school. Analogously, we can also consider the “mirror image” of our `aas`-contingent reallocation: an `ssa`-contingent rearrangement in which we take the art school from the second scientist and give it to the artist. By symmetry, this would make the superintendent and the artist better off, but would hurt the scientist and potentially give her incentive to misreport her preferences. It is not surprising that trying to improve upon DA ex post could cause incentive problems (Erdil and Ergin 2008, Abdulkadiroğlu, Pathak and Roth 2009, Featherstone 2013). What is surprising in our example is that if we combine *both* of our problematic reallocation rules, we return to implementing truth-telling, although as a Bayes-Nash equilibrium instead of a dominant strategy equilibrium. In other words, Boston is equivalent to running DA and reallocating in the “bad” states of the world (`aas` and `ssa`).

To get a better look at why this works, consider the perspective of a scientist (capital-`S`) considering her preference report when both the `aas` and `ssa` reallocation rules are in effect. She cares most about the interim states `sSa` and `aaS`, as in those states, one of the reallocation rules changes the school she gets (see Figure 4b). Although both `sSa` and `aaS` are equally likely from her perspective, the help she gets in `aaS` (moving from unmatched to a first choice) is bigger than the hurt she gets in `sSa` (moving from second choice to unmatched). So, she is interim better off under the new rules. Moreover, she has no incentive to misrepresent herself as an artist: if she did so, the hurt (having a first choice taken away because it was *declared* as a second choice) would be bigger than the help (moving from unmatched to a second choice because it was *declared* as a first choice). By symmetry, an artist also prefers to truthfully report his preference.

Interpreting Boston as DA plus two ex post reallocation rules shows both why Boston dominates and why it admits a truth-telling equilibrium. However, much of this analysis depends on the fact that the two reallocation rules are equally likely to be triggered from the interim perspective. In the next section, we consider generalizing to a less symmetric environment.

5.2 Generalizing to less symmetric environments

Let the ex ante probability that a student is a scientist be $k > \frac{1}{2}$. Assuming truth-telling, in this new environment, Boston would still dominate DA state by state from the ex ante perspective; the fact that $k > \frac{1}{2}$ only changes the relative likelihood of those states. Because of this, however, Boston will no longer admit a truth-telling equilibrium. This is because a scientist, $\mathtt{S}$, might not be pleased that both rules are being added to DA, since from his perspective, the state in which he is hurt, $\mathtt{sSa}$, is more likely ($p_{\mathtt{sSa}} = \frac{1}{3} \cdot k \cdot (1-k)$) than the state in which he is helped, $\mathtt{aaS}$ ($p_{\mathtt{aaS}} = \frac{1}{3} \cdot (1-k)^2$). A simple calculation shows that a scientist who just barely prefers the science school would prefer to masquerade as an artist in order to increase the probability that the $\mathtt{ssa}$ reallocation rule would save him from being unmatched.[48]

If we could equalize the likelihood that the two rules change a scientist's assignment, however, he would again prefer to truthfully report his preferences, by the same logic as the previous section. This can be accomplished by only implementing the $\mathtt{ssa}$ reallocation rule with probability $\frac{1-k}{k}$, since $p_{\mathtt{sSa}} \cdot \frac{1-k}{k} = p_{\mathtt{aaS}}$. Intuitively, the superintendent is tempted to reallocate 100% of the time, but he must commit himself to only do so probabilistically in order to induce truthful preference revelation.

Proposition 3. *Consider the extension of the arts and science school example in this section, where each student is independently a scientist with probability $k > \frac{1}{2}$. Suppose we run DA and then apply the following two ex post reallocation rules:*

- *If the state is* $\mathtt{aas}$*, then give the science school to the scientist, leaving the artist with the lower lottery draw unmatched.*
- *If the state is* $\mathtt{ssa}$*, then with probability $1 - \frac{1-k}{k}$ do nothing, and with probability $\frac{1-k}{k}$, give the art school to the artist, leaving the scientist with the lower lottery draw unmatched.*

Then, all students truth-telling is an ordinal Bayes-Nash equilibrium. Further, all students interim[49] *prefer the equilibrium outcome of this mechanism to the outcome of the truth-telling equilibrium under DA.*

When $k = \frac{1}{2}$, the second rule in the theorem reallocates with probability 1, that is, the mechanism in the theorem becomes Boston. Figure 5 shows just how much mechanism of Proposition 3 dominates DA.[50]

[48] If the scientist tells the truth, then his gain from the reallocation rules is $\frac{1}{3} \cdot k \cdot (1-k) \cdot [u(\emptyset) - u(a)] + \frac{1}{3} \cdot (1-k)^2 \cdot [u(s) - u(\emptyset)]$ which approaches $\frac{1}{3} \cdot (1-k) \cdot (2k-1) \cdot [u(\emptyset) - u(s)]<0$ as $u(a) \to u(s)$. If the scientist misrepresents himself as an artist, his gain is $\frac{1}{3} \cdot k \cdot (1-k) \cdot [u(\emptyset) - u(a)] + \frac{1}{3} \cdot k^2 \cdot [u(s) - u(\emptyset)]$ which approaches $\frac{1}{3} \cdot (1-k) \cdot (2k-1) \cdot [u(s) - u(\emptyset)] > 0$ as $u(a) \to u(s)$.

[49] That is, evaluated from the information set where students know their own preference realization, but not the distribution over others' preferences and not the lottery.

[50] It is sufficient to show only how much more likely a student is to receive his first choice school, because truth-telling would not be an ordinal Bayes-Nash equilibrium if artists and scientists were unmatched differentially, contradicting Proposition 3.

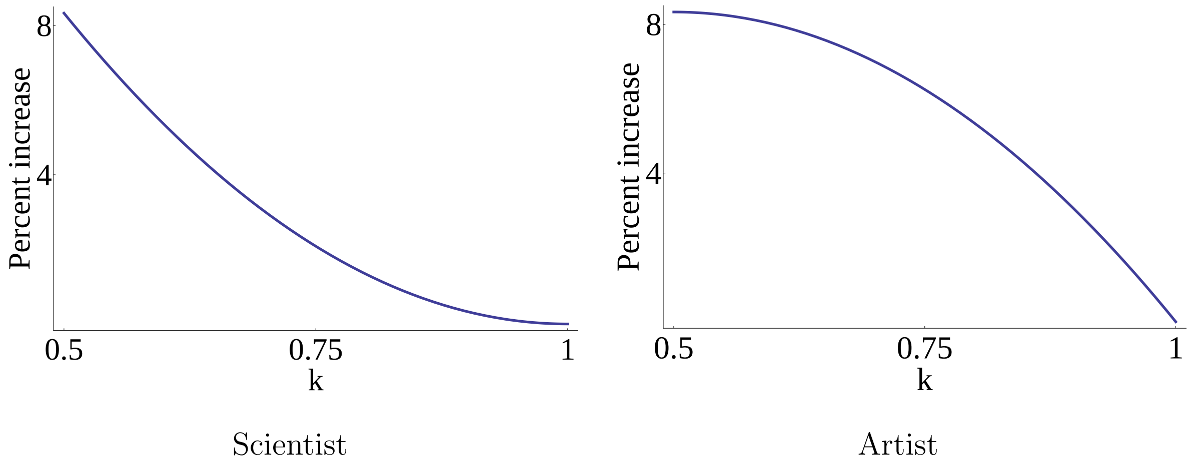


Figure 5: Increase in chance to receive the first choice school from DA to the Boston-like new mechanism

While the example is only suggestive, the intuition of probabilistically randomizing over whether a school is allowed to ex post reallocate seems generalizable and is a possible area of future research.

6 Conclusion

We had two goals for our experiment. The first concerns the new and active literature on mechanisms that rely on equilibrium manipulation of preference reports to implement outcomes superior to those of DA (Abdulkadiroğlu, Che and Yasuda 2008, Miralles 2009, Troyan 2012). To assess the viability of these approaches for practical market design, it is important to determine whether participants are likely to at least come close to equilibrium play. Empirical work suggests that this may not be the case for at least a few participants (Abdulkadiroğlu et al. 2005). Existing experiments have focused on whether participants submit preferences truthfully in non-strategy-proof mechanisms, and while these papers confirm that participants would manipulate their preferences, they in general do not provide participants with sufficient information to calculate the equilibrium (Chen and Sönmez 2006, Pais and Pintér 2008, Calsamiglia, Haeringer and Klijn 2010).

The first result of the present paper is that non-truth-telling equilibria might be hard to implement in the real world. In the aligned preference environment, while DA elicited truth-telling, Boston elicited sub-optimal manipulation of preference reports. This was the case even in the last five periods of play, where participants had ample experience and received precise feedback in a very simple environment. This leaves two potential ways to move the "improving on strategy-proof mechanisms" literature forward. The first is to give serious thought as to how students can be pushed in the direction of equilibrium. Inquiry into this question seems like a fruitful and understudied line of experimental research, but is not what we pursue in this paper. The second is to weaken the implementation of the truth-telling equilibrium.[51]

The second part of our paper shows that truth-telling equilibria that are not implemented in dominant strategies do have the potential to succeed, both in inducing truthful reporting and in yielding efficiency gains. We think of our results as showing that there is *potential* for the incentive-compatible-but-not-strategy-proof research agenda to succeed. Further theoretical development is clearly required, but, as we hope our experiment has demonstrated, such a research agenda could potentially have a large impact on student welfare.

In addition to simply pointing out that Boston can dominate DA in some environments, we have also tacitly made two suggestions for non-strategy-proof design. The first is to implement truth-telling as an ordinal Bayes-Nash equilibrium, as it requires information about the distribution of ordinal preferences (which can be observed under a strategy-proof mechanism), but not about the cardinal preferences, which are more elusive. The second is to analyze efficiency and incentives from the interim information set, in which a student know his own preferences, but only the distribution of others'. Not only does this seem like the natural assumption for what real-world students know, but from an efficiency standpoint, it seems like the sort of "best outcome averaged over many years" sort of calculus that seems natural for a school board.

References

Abdulkadiroğlu, Atila, and Tayfun Sönmez. 2003. "School Choice: A Mechanism Design Approach." *The American Economic Review*, 93(3): pp. 729–747.

Abdulkadiroğlu, Atila, Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez. 2005. "The Boston Public School Match." *The American Economic Review*, 95(2): pp. 368–371.

Abdulkadiroğlu, Atila, Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez. 2006. "Changing the Boston School Choice Mechanism." *working paper.*

[51] Barros, Featherstone and Salcedo (2010) runs an experiment with a decision aid that pursues this line of inquiry.

Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth. 2005. "The New York City High School Match." *American Economic Review*, 95(2): 364–367.

Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth. 2009. "Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match." *The American Economic Review*, 99(5): 1954–1978.

Abdulkadiroğlu, Atila, Yeon-Koo Che, and Yosuke Yasuda. 2008. "Expanding "Choice" in School Choice." *working paper*.

Abdulkadiroğlu, Atila, Yeon-Koo Che, and Yosuke Yasuda. 2011. "Resolving Conflicting Preferences in School Choice: The "Boston" Mechanism Reconsidered." *American Economic Review*, 101(1): 399–410.

Barros, Rodrigo, Clayton R. Featherstone, and Alejandrina Salcedo. 2010. "Matching university students to majors: analysis of an old mechanism and redesign of the new." *Working Paper*.

Budish, Eric, and Estelle Cantillon. 2012. "The Multi-unit Assignment Problem: Theory and Evidence from Course Allocation at Harvard." *American Economic Review*, 102(5): 2237–2271.

Calsamiglia, Caterina, Guillaume Haeringer, and Flip Klijn. 2010. "Constrained School Choice: An Experimental Study." *American Economic Review*, 100(4): 1860–74.

Calsamiglia, Caterina, Guillaume Haeringer, and Flip Klijn. 2011. "A comment on "School Choice: An Experimental Study" [J. Econ. Theory 127 (1)(2006) 202-231]." *Journal of Economic Theory*, 146(1): 392–396.

Chen, Yan, and Tayfun Sönmez. 2006. "School Choice: An Experimental Study." *Journal of Economic Theory*, 127(1): 202–231.

Chen, Yan, and Tayfun Sönmez. 2011. "Corrigendum to "School choice: An experimental study" [J. Econ. Theory 127 (1)(2006) 202-231]." *Journal of Economic Theory*, 146(1): 397–399.

Dubins, Lester E., and David A. Freedman. 1981. "Machiavelli and the Gale-Shapley Algorithm." *The American Mathematical Monthly*, 88(7): pp. 485–494.

Ehlers, Lars, and Jordi Massó. 2007. "Incomplete information and singleton cores in matching markets." *Journal of Economic Theory*, 136(1): 587–600.

Erdil, Aytek, and Haluk Ergin. 2008. “What’s the Matter with Tie-Breaking? Improving Efficiency in School Choice.” *The American Economic Review*, 98(3): 669–689.

Ergin, Haluk, and Tayfun Sönmez. 2006. “Games of School Choice under the Boston Mechanism.” *Journal of Public Economics*, 90(1-2): 215–237.

Featherstone, Clayton R. 2013. “Rank Efficiency: Investigating a Widespread Ordinal Welfare Criterion.” *working paper.*

Fischbacher, Urs. 2007. “z-Tree: Zurich Toolbox for Ready-Made Economic Experiments.” *Experimental Economics*, 10(2): 171–178.

Gale, David, and Lloyd S. Shapley. 1962. “College Admissions and the Stability of Marriage.” *American Mathematical Monthly*, 69(1): 9–15.

Kagel, John H., and Alvin E. Roth. 2000. “The Dynamics of Reorganization in Matching Markets: A Laboratory Experiment Motivated by A Natural Experiment.” *Quarterly Journal of Economics*, 115(1): 201–235.

Kesten, Onur, and M. Utku Ünver. 2013. “A Theory of School Choice Lotteries.” *working paper.*

Kojima, Fuhito, and M. Utku Ünver. 2010. “The “Boston” School-Choice Mechanism.” *working paper.*

Kojima, Fuhito, and Parag A. Pathak. 2009. “Incentives and Stability in Large Two-Sided Matching Markets.” *The American Economic Review*, 99(3): 608–627.

McKinney, C. Nicholas, Muriel Niederle, and Alvin E. Roth. 2005. “The Collapse of a Medical Labor Clearinghouse (and why such failures are rare).” *American Economic Review*, 878–889.

Miralles, Antonio. 2009. “School Choice: The Case for the Boston Mechanism.” *Auctions, Market Mechanisms and Their Applications*, 58–60.

Pais, Joana, and Ágnes Pintér. 2008. “School Choice and Information: An Experimental Study on Matching mechanisms.” *Games and Economic Behavior*, 64(1): 303–328.

Pathak, Parag A., and Tayfun Sönmez. 2008. "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism." *The American Economic Review*, 98(4): 1636–1652.

Roth, Alvin E. 1982*a*. "Incentive Compatibility in a Market with Indivisible Goods." *Economics Letters*, 9(2): 127–132.

Roth, Alvin E. 1982*b*. "The Economics of Matching: Stability and Incentives." *Mathematics of Operations Research*, 7(4): 617–628.

Roth, Alvin E. 1989. "Two-sided matching with incomplete information about others' preferences." *Games and Economic Behavior*, 1(2): 191–209.

Roth, Alvin E. 1991. "A Natural Experiment in the Organization of Entry-Level Labor Markets: Regional Markets for New Physicians and Surgeons in the United Kingdom." *The American Economic Review*, 81(3): 415–440.

Roth, Alvin E., and Uriel G. Rothblum. 1999. "Truncation Strategies in Matching Markets – in search of advice for participants." *Econometrica*, 67(1): 21–43.

Troyan, Peter. 2012. "Comparing school choice mechanisms by interim and ex-ante welfare." *Games and Economic Behavior*, 75(2): 936–947.

Ünver, M. Utku. 2001. "Backward unraveling over time: The evolution of strategic behavior in the entry level British medical labor markets." *Journal of Economic Dynamics and Control*, 25(6-7): 1039–1080.

A Best-response analysis for aligned environment under Boston

There are two cases present in the data which give Top students strict incentives to rank either Second or Third as their second choice, given that they ranked Best as their first choice

Case 1 (Skip the middle): All Top students rank Best as their first choice, at least one Average ranks Second as his first choice, and one of the Averages who might not be assigned in the first round ranks Third as his second choice. Tops are then constrained to rank Third as their second choice.

Case 2 (Do not skip the middle): All Top students rank Best as their first choice, at least one Average ranks Third as his first choice, and one of the Averages who might not be assigned in the first round ranks Second as his second choice. Tops are then constrained to rank Second second.

We restrict ourselves to studying sessions in which a large fraction (4 out of 5) or at least the majority (3 out of 5) of periods are either in Case 1 or Case 2 and in which the remaining periods do *not* fall into any of these two cases. That is, we focus on sessions in which Top students weakly best respond by submitting ($Best$, $Second$, x) or ($Best$, $Third$, x) in all five periods, and, in 4 (or 3) of those periods, this behavior is a *strict* best response. Table 6 shows the number of sessions that fall into these categories and whether Top students submit truthfully (i.e. submit ($Best$, $Second$, x)) or best respond as a primary strategy. Note that table below analyzes 4 distinct "Skip the middle" sessions.

Best response for Top students (BR)	# sessions where BR is weakly optimal in each of the last 5 rounds (# rounds where the BR is strict)	# of Tops who truth-tell at least 4 of the last 5 rounds	# of Tops who BR in at least 4 of the last 5 rounds
Skip the middle	2 (4/5)	5/6	0/6
Skip the middle	2 (3/5)	2/6	2/6
Don't skip the middle	1 (4/5)	3/3	3/3

Table 6: Unambiguously suboptimal behavior of Top students

B Theory appendix

B.1 The model

Let there be a set of students, $\mathcal{I}$, and a set of schools, $\mathcal{S}$. Also, let $\emptyset \notin \mathcal{S}$ represent the outcome of being unmatched. Each student i has a strict preference P_i over the elements of $\mathcal{S} \cup \{\emptyset\}$, and each school has a strict priority ordering $\succ_s$ over $\mathcal{I}$.[52] Denote the number of acceptable schools in a preference P_i by $|P_i| \equiv |\{s \in \mathcal{S} : s\, P_i\, \emptyset\}|$. As we are dealing with individually rational mechanisms, we will only consider the ordering of schools in P_i that are ranked above $\emptyset$, that is, if two preferences order their acceptable schools in the same way, then they are considered the same preference. A utility function u_i *rationalizes* a preference P_i if $u_i(P_i(a)) > u_i(P_i(b)) \Leftrightarrow a < b$, where $P_i(a)$ denotes the a^{th} ranked school in P_i. $q \equiv (q_s)_{s\in\mathcal{S}}$ is a vector of capacities for each school. The unmatched outcome, $\emptyset$, can be thought of as a special "school" with $q_\emptyset = \infty$. $P \equiv (P_i)_{i\in\mathcal{I}}$ and $\succ \equiv (\succ_s)_{s\in\mathcal{S}}$ are drawn from a joint probability measure λ.

A *school choice problem* is defined by $\langle \mathcal{I}, \mathcal{S}, q, \lambda \rangle$. A *matching* $\mu : \mathcal{I} \mapsto \mathcal{S} \cup \{\emptyset\}$ is a function such that $|\{i : \mu(i) = s\}| \leq q_s$ for all $s \in \mathcal{S}$. A *deterministic mechanism* is a function φ that maps reported student preferences $\widetilde{P}$ and school priorities $\succ$ to a matching. For a student, a strategy in the Bayesian game induced by a school choice problem and a deterministic mechanism is a mapping from true preferences to reported preferences, σ_i; let $\sigma \equiv (\sigma_i)_{i\in\mathcal{I}}$ denote the vector of strategies for all students. A set of strategies is a *Bayes-Nash equilibrium* if no student has a profitable unilateral deviation, that is, $\forall i, P_i, \widetilde{P}_i$

$$\mathbb{E}\left\{u_i\left[\varphi\left(\sigma\left(P\right);\succ\right)(i)\right]|P_i\right\} \geq \mathbb{E}\left\{u_i\left[\varphi\left(\widetilde{P}_i, \sigma_{-i}\left(P_{-i}\right);\succ\right)(i)\right]\middle|P_i\right\}$$

A set of strategies is an *ordinal Bayes-Nash equilibrium* if a student does weakly worse (in the first-order stochastic sense) when he deviates from equilibrium, that is, $\forall i, P_i, \widetilde{P}_i, |P_i| \geq b \geq 1$

$$\sum_{a=1}^{b} \Pr\left\{\varphi\left(\sigma\left(P\right);\succ\right)(i) = P_i\left(a\right)|P_i\right\} \geq \sum_{a=1}^{b} \Pr\left\{\varphi\left(\widetilde{P}_i, \sigma_{-i}\left(P_{-i}\right);\succ\right)(i) = P_i\left(a\right)\middle|P_i\right\}$$

Two equilibria are **outcome-equivalent** if, for any realization of $(P;\succ)$, they yield the same matching.

[52] Note that $\succ_s$ is not an ordering over $\mathcal{I} \cup \{\emptyset\}$, that is, schools are required to find all students acceptable.

B.2 The aligned environment

Proof of Proposition 1. Truth-telling is weakly dominant under DA. Since all Tops have a positive probability of having top priority at Best, failing to rank Best first is strictly dominated, so Tops must rank Best first. Given this, since all Tops have a positive probability of being the Top with lowest priority at Best, failing to rank Second second is strictly dominated, so Tops must also rank Second second. Given this, both Averages have positive probability of being the higest priority Average at Third; hence, failing to report Third acceptable is strictly dominated, so the Averages must report Third acceptable. □

Proof of Proposition 2. Note that at least two Tops must rank Best first in a pure-strategy Bayes-Nash equilibrium; otherwise, any Top not ranking Best first could increase his payoff.

If exactly two Tops rank Best first, the other must rank Second first. The Averages must respond to this by ranking Third first. Given this behavior, the Top ranking Second first prefers to rank Best first and Second second, contradicting equilibrium. Thus, there must be three Tops ranking Best first.

Three Tops ranking Best first means that the Averages must rank either Second or Third first. If they rank different schools, then one of the Tops can profitably deviate by ranking Second second, contradicting equilibrium. Hence, the Averages either both rank Second first or both rank Third first. Either way, the Tops respond by setting their second choices to get the last available school in the second round of the algorithm, limiting the Averages to whatever payoff they get in the first round of the algorithm. Since $\frac{1}{2} \cdot 67 = 33.5 > 25$, the Averages will rank Second second, which means that the Tops must rank Third second. □

B.3 The uncorrelated environment

B.3.1 Truth-telling under DA

Under DA, the existence of a truth-telling equilibrium is straightforward.

Proposition 4. *Under DA, for any school choice problem, there exists an ordinal Bayes-Nash equilibrium where all students are truth-telling.*

Proof. Roth (1989) proves that two-sided DA is strategy-proof for the proposing side, which implies that one-sided DA is strategy-proof. Hence, by reporting truthfully, a student guarantees himself a weakly better assignment for every realization of reported preferences and priorities, $\left(\widetilde{P_{-i}}; \succ\right)$. Hence, truth-telling weakly first-order stochastically dominates any other response to any profile of other students' strategies. All students truth-telling is thus an ordinal Bayes-Nash equilibrium. $\square$

Now, we turn to finding conditions that make truth-telling the unique equilibrium in weakly undominated strategies.

Definition 1. Let $I_i(s) \equiv \{i' \in \mathcal{I} : i' \succ_s i\}$ be the set of students who have higher priority than student i at school s under $\succ$, and let $m \geq 2$. We say that a set of school priorities $\succ$ **stably exposes preference $\boldsymbol{P_i}$ to the $\boldsymbol{m^{th}}$ choice** if $\forall J \in 2^{\{1,\ldots,m-1\}}$,

$$\left|\bigcup_{j \in J} I_i\left(P_i(j)\right)\right| \geq \sum_{j \in J} q_{P_i(j)}$$

A school choice problem stably exposes preference P_i to the m^{th} choice if its measure places positive probability on a set of school priorities $\succ$ that stably exposes P_i to the m^{th} choice.

Intuitively, the condition for preference exposure means that there are enough students with higher priority than student i at schools he ranks better than m^{th} that he could potentially be assigned to his m^{th} choice. That the condition is sufficient is obvious, but that it is necessary is a subtle result from bipartite graph theory known as Hall's Theorem.

Lemma 1. *Take $s \in \mathcal{S} \cup \{\emptyset\}$ such that $\widetilde{P}_i(m) = s$ and $\left|\widetilde{P}_i\right| + 1 \geq m \geq 2$. Under DA, there exist preference reports for the other students, $\widetilde{P_{-i}}$, such that DA matches i to s with positive probability if and only if the school choice problem stably exposes $\widetilde{P}_i$ to the m^{th} choice.*

Proof. Consider a vector of school priorities $\succ$, and construct the bipartite graph where there is a vertex for each student besides i, and $q_{s'}$ vertices for each school s' that i prefers to s. There is an edge between a student vertex and a school vertex if and only if the student has higher priority than i at that school. The existence of a matching that covers the school vertices is necessary for DA to assign i to s; otherwise, we would contradict the stability of DA. By Hall's Theorem, the necessary and sufficient condition for the existence of such a matching is that $\succ$ stably exposes $\widetilde{P}_i$ to the m^{th} choice. So, if the school choice problem fails to stably expose $\widetilde{P}_i$ to the m^{th} choice, then i can never match to his m^{th} choice, establishing necessity. Now, consider the $\succ$ that stably exposes $\widetilde{P_i}$ to the m^{th} choice. On the bipartite graph we mentioned above, Hall's Theorem guarantees us a matching that covers the school vertices. Construct a $\widetilde{P_{-i}}$ where every student ranks the school to which they are matched first and $\emptyset$ second (students that aren't matched simply rank $\emptyset$ first). By construction, i gets s under these preference reports, which establishes sufficiency. $\square$

Proposition 5. *Under DA, all students truth-telling is the unique equilibrium in weakly undominated strategies if and only if, for all P_i that occur with positive probability, the school choice problem stably exposes P_i to the* $\min\{|P_i|+1, |\mathcal{S}|\}^{th}$ *choice.*

Proof. Start by assuming that $\min\{|P_i|+1, |\mathcal{S}|\} = |P_i|+1$. If the school choice problem does expose P_i to the $(|P_i|+1)^{th}$ choice, then by the logic of the proof of Lemma 1, for any $2 \leq k \leq |P_i|+1$, there is a $\widetilde{P^k_{-i}}$ that makes the best school that i could match to his k^{th} choice. The possibility that all other students might rank $\emptyset$ first establishes that any report that isn't truthful about the first choice is weakly dominated. Of reports that are truthful about the first choice, $\widetilde{P^2_{-i}}$ establishes that any report that isn't truthful about the second choice is weakly dominated. Continuing with this logic, we can eliminate all strategies but truth-telling. Now, the only time that $\min\{|P_i|+1, |\mathcal{S}|\} \neq |P_i|+1$ can hold is if $|P_i| = |\mathcal{S}|$. If this is true, then we don't need eliminate strategies that lie past the $|P_i|^{th}$ choice, since there are no strategies that do so in the strategy space. Hence, Proposition 4 tells us that truth-telling is an equilibrium; we have established sufficiency. Now, say that the school choice problem stably exposes some P_i only to the k^{th} choice, where $k < \min\{|P_i|+1, |\mathcal{S}|\}$. By Lemma 1, DA will never match i to his $(k+1)^{th}$ choice; hence, he is free to lie about it, establishing necessity. $\square$

Now, we establish an easier to verify sufficient condition for the more complicated condition of the previous proposition.

Lemma 2. *Say for any priority ordering, there is a positive probability that all schools share that priority. Also, say $|\mathcal{I}| \geq \sum_{s\in\mathcal{S}} q_s$. Then, the school choice problem stably exposes any P_i to the* $\min\{|P_i|+1, |\mathcal{S}|\}^{th}$ *choice.*

Proof. If $|\mathcal{S}| \neq |P_i|$, then let all schools share a priority that puts $\sum_{sP_i\emptyset} q_s$ students ahead of i. If $|\mathcal{S}| = |P_i|$, then let all schools share a priority that puts $\sum_{s\neq P_i(|P_i|)} q_s$ students ahead of i. $\square$

It is worth mentioning two corollaries to the proof of Proposition 5 that are potentially of empirical importance. When we observe the preferences submitted to a

DA algorithm, some parts of that rank-order are constrained by equilibrium and some aren't. The constrained parts should obviously carry more weight in any analysis that tries to draw inferences about student preferences from their submitted rank-orders.

Corollary 1 (to the proof of Proposition 5)**.** *If a student's preference isn't stably exposed to the k^{th} choice, then his submitted rankings from k^{th} choice down are unconstrained in any equilibrium in weakly undominated strategies.*

Corollary 2 (to the proof of Proposition 5)**.** *In a submitted rank-order, the schools ranked higher than the school a student is ultimately assigned must be truthfully ranked in any equilibrium in weakly undominated strategies.*

Finally, we show that even when there are multiple equilibria in weakly undominated strategies, the multiplicity is degenerate. This tells us that whenever truth-telling fails to be the unique equilibrium, deviations from truth-telling won't affect the ultimate outcome.

Proposition 6. *Under DA, all equilibria in weakly undominated strategies are outcome-equivalent to the truth-telling equilibrium.*

Proof. If a student is matched to his k^{th} choice when he submits $\widetilde{P}_i$, then Lemma 1 tells us that the school choice problem exposes $\widetilde{P}_i$ to the k^{th} choice. By the logic of the proof to Proposition 5, failing to the tell the truth up to the k^{th} choice is weakly dominated. So all students are telling the truth, up to the school they are matched to. Since DA does not depend on the preferences below where a student ends up matched, any equilibrium in undominated strategies must yield the same matching. $\square$

B.3.2 Truth-telling under Boston

Clearly, similar theorems won't hold for arbitrary school choice problems under Boston; we will focus on school choice problems that have a certain symmetry property. Let $P_{-i}^{s\leftrightarrow s'}$ denote the preferences constructed by switching schools s and s' everywhere in all students' preferences, save for i. Similarly, let $\succ^{s\leftrightarrow s'}$ be the vector of schools' orderings where the orderings of schools s' and s are switched.

Definition 2. A probability measure λ is **school-symmetric with respect to *i*** if, for any P_i that occurs with positive probability and any s,$s' \in \mathcal{S}$, we have $\lambda\left[(P\,;\succ)|P_i\right] = \lambda\left[\left(P_i, P_{-i}^{s\leftrightarrow s'}\,;\succ^{s\leftrightarrow s'}\right)\middle|P_i\right]$. The measure is simply **school-symmetric** if it is school-symmetric with respect to i for all $i \in \mathcal{I}$. A school choice problem is school-symmetric if its measure is school-symmetric.

This definition does not allow for interchanges that involve $\emptyset$, as $\emptyset \notin \mathcal{S}$. Also, note that school-symmetry is not the same as drawing preferences from the uniform distribution; for instance, it allows for a student to know that it is unlikely that the other students find as many schools acceptable as he does. With school-symmetric reported preferences, we can show that truth-telling does at least as well as any other strategy a student might play. We do this via two lemmas. Let $B\left[P;\succ\right](i)$ denote student i's assignment under Boston when the reported preferences are P and the school priorities are $\succ$.

Lemma 3. *Let $s' \, P_i \, s$. Then*

- $B\left[P_i^{s'\leftrightarrow s}, P_{-i};\succ\right](i) \notin \{s', s\}$ *and* $B\left[P\,;\succ\right](i) \notin \{s', s\}$
  $\Rightarrow B\left[P_i^{s'\leftrightarrow s}, P_{-i}\,;\succ\right](i) = B\left[P\,;\succ\right](i)$
- $B[P\,;\succ](i) \neq s' \Rightarrow B[P_i^{s'\leftrightarrow s}, P_{-i}\,;\succ](i) \neq s'$
- $B[P\,;\succ](i) = s \Rightarrow B[P_i^{s'\leftrightarrow s}, P_{-i}\,;\succ](i) = s$
- *Reporting a school acceptable if and only if it is truly acceptable cannot hurt a student.*

Proof. The key insight for the first three items in the lemma is that what the Boston algorithm does in a given round depends solely on who isn't matched yet and which seats are left. Define j, k, and m by $P_i(j) = s'$, $P_i(k) = s$, and $P_i(m) = B\left[P\,;\succ\right](i)$, respectively. Now, we prove the lemma, bullet by bullet:

- If $m < j$, then i must also be assigned in round m when he reports $P_i^{s'\leftrightarrow s}$, as switching s and s' won't change what seats and students are still available at the start of the round m. If $m > j$, then by our assumptions that $B\left[P_i^{s'\leftrightarrow s}, P_{-i};\succ\right](i) \notin \{s', s\}$ and $B\left[P\,;\succ\right](i) \notin \{s', s\}$, the seats at both s and s' must be exhausted at the start of round j. Hence, the set of students and seats that remain available is the same for both reports going forward from round j. So i is assigned in round m under both preference reports.

- That i is rejected in round j when he reports P_i tells us that there are no seats available at s' at the end of round j. This must also be true when he reports $P_i^{s'\leftrightarrow s}$; hence he cannot get s' by ranking it $k > j$.

- P_i and $P_i^{s'\leftrightarrow s}$ are identical down to rank $j-1$, so i must be available at the start of round j, since he isn't matched until round $k > j$ when he reports P_i. This also tells us that there must be at least one unclaimed seat at s at the end of round j when i reports P_i. Reporting $P_i^{s'\leftrightarrow s}$ allows i to claim this seat.

- The key here is to realize that once Boston assigns a student, the assignment is permanent. Hence, if a student fails to report a truly acceptable school acceptable, it is costless and potentially profitable for the student to rank it between the lowest ranked acceptable school and $\emptyset$ in the original report. If a student reports a truly unacceptable school acceptable, it is costless and potentially profitable for the student to report it unacceptable and to move all schools originally declared inferior to the unacceptable school up one position in the reported rank order.

$\square$

Lemma 4. *Under the Boston mechanism, if the school choice problem is school-symmetric with respect to i, and $q_s = q'_s, \forall s \in \mathcal{S}$, then for i, truth-telling weakly first-order stochastically dominates any other preference report.*

Proof. Note that the proof of this theorem borrows heavily from Roth and Rothblum (1999), but while that paper focuses on strategies used by the proposed-to side of the market, this result focuses on the proposing side. To start, the final bullet in Lemma 3 tells us that there is no loss in limiting ourselves to reports that declare a school acceptable if and only if it is truly acceptable. Now, consider a student, i, who truthfully prefers school s' to s, that is $s' \, P_i \, s$. Table 7a lists the possibilities for what i gets with and without switching s and s'. The second and third bullets of Lemma 3 rule out the cells labeled "Impossible", while the first bullet tells us that u must be the same for both Truth and Lie in Case A.

By school-symmetry, the two columns under Truth in Table 7b are equally likely if i truthfully orders s and s', and the two columns under Lie are equally likely if i reverses their ordering. Row by row, truthfully ordering s and s' is at least as good as, and sometimes better than, switching their ordering. In other words, we can partition the states-of-the-world in a way that, for each element of the partition, ordering s and s' truthfully weakly first-order stochastically dominates reversing their order. This implies that truthfully ordering s and s' weakly first-order stochastically dominates any other report. Hence, truth-telling weakly first-order stochastically dominates any other strategy. $\square$

We have thus established conditions under which Boston admits a truth-telling equilibrium.

Proposition 7. *Under the Boston mechanism, if a school choice problem is school-symmetric and $q_s = q'_s, \forall s, s' \in \mathcal{S}$, then there exists an ordinal Bayes-Nash equilibrium where all students are truth-telling.*

		Lie: $B[P_i^{s'\leftrightarrow s}, P_{-i}\,;\succ](i)$		
		$= u \notin \{s, s'\}$	$= s$	$= s'$
	$= u \notin \{s, s'\}$	Case A	Case B	Impossible
Truth: $B[P_i, P_{-i}\,;\succ](i)$	$= s$	Impossible	Case C	Impossible
	$= s'$	Case D	Case E	Case F

(a) Table of cases

	Truth		Lie	
	$B[P_i, P_{-i};\succ](i)$	$B[P_i, P_{-i}^{s'\leftrightarrow s};\succ^{s'\leftrightarrow s}](i)$	$B[P_i^{s'\leftrightarrow s}, P_{-i};\succ](i)$	$B[P_i^{s'\leftrightarrow s}, P_{-i}^{s'\leftrightarrow s};\succ^{s'\leftrightarrow s}](i)$
Case A	u	u	u	u
Case B	u	s'	s	u
Case C	s	s'	s	s'
Case D	s'	u	u	s
Case E	s'	s'	s	s
Case F	s'	s	s'	s

(b) Outcomes for the cases

Table 7: Cases and outcomes for Lemma 4

Proof. If all students but i are truth-telling, then their reports are school-symmetric with respect to i. By Lemma 4, truth-telling weakly first-order stochastically dominates any other strategy for i. Hence, all students truth-telling is an ordinal Bayes-Nash equilibrium. $\square$

To get uniqueness, we will need to restrict ourselves to a class of strategies that is natural for symmetric environments.

Definition 3. A ***k*-permutation** is a bijection from $\{1, ..., k\}$ to itself. If $|P_i| = k$ and π is a k-permutation, then define $\pi(P_i)$ to be the preference of size k defined by $\pi(P_i)(j) \equiv P_i(\pi(j))$ for all $j \in \{1, ..., k\}$.

Definition 4. A strategy σ_i is ***anonymous*** if there exists a vector of permutations $(\pi_k)_{k=1}^{|\mathcal{S}|}$ such that, for each k, π_k is a k-permutation and $\sigma_i(P_i) = \pi_{|P_i|}(P_i)$ for all P_i.

Anonymous strategies must choose how to rank a school based only on its true rank and the number of truly acceptable schools. Note that we have ruled out strategies that change the number of acceptable schools, such as truncations. This will simplify the proof of Lemma 5 by keeping the mapping from true preferences to reported preferences one-to-one. Since we will always look at situations where changing the number of acceptable schools in the report is weakly dominated (see Lemma 6), this simplification is without loss of generality when considering equilibria in weakly undominated strategies.

Anonymous strategies are useful to us because they translate school-symmetric true preferences into school-symmetric reported preferences, allowing us to leverage Lemma 4.

Lemma 5. *In a school-symmetric school choice problem, when all students besides $i \in \mathcal{I}$ play anonymous strategies, the distribution of students' reported preferences and school orderings is school-symmetric with respect to i.*

Proof. Strategies for the students induce a measure over reported preferences and school priorities, $\widetilde{\lambda}$. Let $\overline{\pi}_{-i}$ denote the particular set of k-permutations that the anonymous strategy σ_i prescribes for the profile of true preferences P_{-i}. The following equation proves the lemma:

$$\widetilde{\lambda}(\sigma_{-i}(P_{-i})\,;\succ|P_i) = \widetilde{\lambda}(\overline{\pi}_{-i}(P_{-i})\,;\succ|P_i)$$
$$= \lambda(P_{-i}\,;\succ|P_i) = \lambda\left(P_{-i}^{s\leftrightarrow s'}\,;\succ^{s\leftrightarrow s'}\Big|P_i\right) = \widetilde{\lambda}\left(\overline{\pi}_{-i}\left(P_{-i}^{s\leftrightarrow s'}\right)\,;\succ^{s\leftrightarrow s'}\Big|P_i\right) =$$
$$\widetilde{\lambda}\left(\left(\overline{\pi}_{-i}(P_{-i})\right)^{s\leftrightarrow s'}\,;\succ^{s\leftrightarrow s'}\Big|P_i\right) = \widetilde{\lambda}\left(\left(\sigma_{-i}(P_{-i})\right)^{s\leftrightarrow s'}\,;\succ^{s\leftrightarrow s'}\Big|P_i\right)$$

The first and last lines are obvious; the meat is in the center line. The first equality comes from the definition of $\widetilde{\lambda}$. The second comes from student-symmetry of the true preferences. The third comes from the fact that $|P_i| = \left|P_i^{s\leftrightarrow s'}\right|$. The fourth comes from the fact that interchanging the names of two schools in the true preference will interchange the name of the same schools in the submitted preference.[53] $\square$

The final condition we need for our uniqueness result ensures that the distribution over preferences is such that a student is always strictly hurt by a deviation from truth-telling.

Definition 5. A school choice problem has **full support** if for any student i and any preference P_i that he draws with positive probability, any $(P_{-i}, \succ)$ such that all students have the same acceptable schools as P_i is also drawn with positive probability.

Finally, we use the full support assumption to get show that some deviations are weakly dominated.

Lemma 6. *Under Boston, if $|\mathcal{I}| \geq \sum_{s\in\mathcal{S}} q_s$ and the school choice problem has full support, then only strategies that report a acceptable school if and only if it is truly acceptable are weakly undominated.*

Proof. By Lemma 3, reporting a school acceptable if and only if it is truly acceptable can't hurt. First, we show that reporting a truly unacceptable school acceptable is weakly dominated. Let $\widetilde{P}_i(k) = s$ be the highest-ranked school in $\widetilde{P}_i$ such that $\emptyset P_i s$. For all schools s' such that $s' P_i \emptyset$, let $q_{s'}$ other students rank s' first, and let all other students report no school acceptable. i will be matched to s, which hurts him. Now, of strategies that remain potentially weakly undominated, we eliminate those that fail to rank all truly acceptable schools acceptable. To see this, say that $\emptyset \widetilde{P}_i s$ when really $s P_i \emptyset$. For all $s' \neq s$ such that $s' \widetilde{P}_i \emptyset$, let $q_{s'}$ students rank s' first, and let all other students report no schools acceptable. $\square$

Proposition 8. *Under the Boston mechanism, if a school choice problem is school-symmetric, has full support, $q_s = q_s', \forall s, s' \in \mathcal{S}$, and $|\mathcal{I}| \geq \sum_{s\in\mathcal{S}} q_s$, then the truth-telling equilibrium is the unique Bayes-Nash equilibrium in weakly undominated, anonymous strategies.*

Proof. Since the equilibrium is in weakly undominated strategies, by Lemma 6, all students declare a school acceptable if and only if it truly is. Now, consider an

[53] Note that we are not claiming that permutations commute: our interchange operator references school names and not positions in a rank-order list.

equilibrium in which student i reports $\widetilde{P}_i \neq P_i$. There must be two schools such that $s\,\widetilde{P}_i\,s'$, but $s' P_i s$. By Lemmas 4 and 5, truthfully ordering s and s' can't hurt i. So, if we find a profile $(P_{-i}, \succ)$ such that falsely ordering s and s' puts us in either Case B, Case D, or Case E of Lemma 4 (the cases in which reversing the true order of s and s' is *strictly* first-order stochastically dominated), then we establish that s and s' must be truthfully ordered in any equilibrium in anonymous strategies.

Let $\widetilde{P}_i(m) = s$. If $m = 1$, let i be ranked first in $\succ_s$ and $\succ_{s'}$, and let $q_{s'}$ students rank s' first. This is Case E from the proof to Lemma 4. If $m \geq 2$, then have $q_{\widetilde{P}_i(1)}$ students in $I_i\left(\widetilde{P}_i\,(1)\right)$ report $\widetilde{P}_i$, $q_{P_i(k)}$ other students rank $\widetilde{P}_i(k)$ first (for all $2 \leq k \leq m$), $q_{s'} - 1$ other students rank s' first, and all other students submitting $\widetilde{P}_i^{\,s \leftrightarrow s'}$. Let $\succ_{s'}$ rank i first. Under this profile of reports, i gets a seat at s' if he truthfully orders s and s', and he gets something that is neither s nor s' if he reports $\widetilde{P}_i$. This is Case D from the proof to Lemma 4. Since s and s' were arbitrary, any deviation from truthfully ordering the school reported acceptable strictly hurts i, so he must truth-tell in any equilibrium in anonymous, weakly undominated strategies. Hence, the truth-telling equilibrium must be the unique equilibrium in weakly undominated, anonymous strategies. $\square$

B.4 The arts and science schools example

Proof of Proposition 3. Conditional on the rules not changing a student's assigned school, he does better truth-telling, state-by-state. This is true by the strategy-proofness of DA. Now, we focus on states of the world in which the rules do change the student's outcome. Assume the other students truth-tell. If our student reports himself a scientist, then with probability $p_{\texttt{sSa},trigger} = \frac{1}{3} \cdot k \cdot (1-k) \cdot \frac{1-k}{k} = \frac{1}{3} \cdot (1-k)^2$ the `ssa` rule is triggered and changes his utility by $u(\emptyset) - u(a)$, and with probability $p_{\texttt{aaS}} = \frac{1}{3} \cdot (1-k)^2$ the `aas` rule is triggered and changes his utility by $u(s) - u(\emptyset)$. Since $p_{\texttt{sSa},trigger} = p_{\texttt{aaS}}$, conditional on the rule changing his outcome, he nets $u(s) - u(a)$ in utility above what he would have gotten under DA. Similarly, if he reports himself an artist, then with probability $p_{\texttt{ssA},trigger} = \frac{1}{3} \cdot k^2 \cdot \frac{1-k}{k} = \frac{1}{3} \cdot k \cdot (1-k)$ the `ssa` rule is triggered and changes his utility by $u(a) - u(\emptyset)$, and with probability $p_{\texttt{aAs}} = \frac{1}{3} \cdot k \cdot (1-k)$ the `aas` rule is triggered and his utility changes by $u(\emptyset) - u(s)$. Since $p_{\texttt{ssA},trigger} = p_{\texttt{aAs}}$, conditional on the rule changing his outcome, he nets $u(a) - u(s)$ in utility above what he would have gotten under DA. So, conditional on the rules changing a student's outcome, truthfully reporting his type first-order stochastically dominates a false report, as well as the outcome under plain DA. Hence, regardless of whether the rules change his outcome, when faced with other students truth-telling, he does first-order stochastically better to truth-tell. Thus, we have shown that truth-telling is an ordinal Bayes-Nash equilibrium that all students interim prefer their outcomes under this equilibrium to their outcomes under the truth-telling equilibrium of plain DA. □