清華大學

中 國 經 濟 研 究 中 心

學 術 論 文

Sorting Behavior and Matching Mechanisms: An Experimental Investigation

Bingbing Li, Stephanie W. Wang, Xiaohan Zhong

No.201906 June 2019

Working Paper

National Center for Economic Research

At

Tsinghua University, Beijing

Sorting Behavior and Matching Mechanisms: An Experimental Investigation

Bingbing Li[a], Stephanie W. Wang[b*], Xiaohan Zhong[a]

[a] Department of Economics, School of Economics and Management, Tsinghua University, Beijing, China

[b] Department of Economics, University of Pittsburgh, Pittsburgh, USA

Abstract: Many matching mechanisms induce non-truth-telling behaviors. In this paper we highlight one of them, the sorting behavior, by which agents put their most preferred achievable matching objective as the first choice. We examine sorting behaviors in a lab experiment under both truth-telling (SD) and non-truth-telling (Boston) mechanisms with different information. The sorting behavior is prevalent under all mechanisms, performs as well as truth-telling behaviors and better than other non-truth-telling behaviors. Promoting sorting behavior can improve the social welfare under Boston mechanisms, but not under SD mechanisms.

Keywords: Sorting Behavior, Boston, Serial dictatorship, incomplete information

JEL classification: C78, C91, D82

I. Introduction

Non-truth-telling behaviors are common in matching mechanisms. Roth and Sotomayor (1990, Chapter 4) prove, in general, "the impossibility of a 'strategy proof' stable mechanisms"[1]. In school choice or college admission, although mechanism *can be* strategy-proof from the student side, non-truth-telling can still be the equilibrium strategy in quite a few cases (Ergin and Sonmez, 2006; Haeringer and Klijn, 2009, etc.). Non-truth-telling behaviors are also widely observed in experimental lab (Chen and Sonmez, 2006; Klijn, Pais and Vorsatz, 2013, etc.), and in reality. For example, China's college admission is still non-strategy-proof, although the reform is directed toward strategy-proof (Chen and Kesten, 2017).

* Corresponding author at: University of Pittsburgh, Pittsburgh, USA. E-mail address: swwang@pitt.edu

[1] Conceptually, strategy-proof implies in a mechanism, all the players have dominant strategies and can avoid the complication of figuring out contingent equilibrium strategy. Truth-telling implies the dominant strategy is the truth-telling (report truly preference order). Therefore, a truth-telling mechanism must be strategy-proof, but not vice versa.

Several questions can be asked, given the prevalence of non-truth-telling behavior: Are there any general patterns of non-truth-telling behaviors? Under what environment people choose a (specific) non-truth-telling behavior? How do those behaviors perform compared with truth-telling and other behaviors? Should we design mechanisms promoting some specific non-truth-telling behaviors? Those questions are not fully explored in the literature and still wait for good answers.

In this paper we highlight one specific non-truth-telling behavior: the sorting behavior. *Sorting behavior is for a player to put his or her most preferred achievable matching objective as the first choice.* Sorting behavior is well (i.e., almost uniquely) defined in a matching mechanism where the stable matching is unique[2]: It is for players to put their unique stable-matched objective as the first choice.

Sorting behavior and truth-telling behavior are extremes on the spectrum of strategic behaviors. Taking the example of college admission, truth-telling behaviors need to list *all colleges* in the order of student's true preference, with the ignorance of preferences of other students and priorities of colleges. On the contrary, sorting behavior need information of all other student preferences and all school priorities, because the stable matching is derived from it.

Therefore, it is easy to argue that sorting behaviors are too sophisticated for players to figure out, let alone be willing to implement it. But first, we need to acknowledge that truth-telling can also be cognitively restrictive[3]. Consider Chinese college admission. There are over 2,000 colleges in China, and the choice set for each student can be as large as over 1,000. (Notice that we've already ignored choosing dozens of majors for any given college.) Obviously, students can only choose from a relatively small *chosen* set of colleges, a way of thinking leading to sorting behaviors. In addition, sorting behaviors sometimes can be simplified significantly due to mechanism features, as happened in Chinese college admission: college priorities are solely determined by rankings of college entrance exam (CEE) total scores. Therefore, students can figure out achievable colleges (or rule out unachievable ones), by attending mock exams, and by searching often publicly-announced score distribution after the exam.[4] In a smaller

[2] For example, when the school/college priorities are acyclic (Haeringer and Klijn, 2009, Theorem 7.3).

[3] Li (2017) argues that a strategy-proof mechanism may not be obviously strategy-proof (OSP), therefore, still subject to cognitive limitations of agents. Ashlagi and Gonczarowski (2017) and Troyan (2016) argues that stable matching mechanisms are in general not obviously strategy-proof, and require acyclic assumptions about preferences in order to be OSP-implemented. Note that even under acyclic assumption, mechanisms must be implemented in a specific way to be OSP.

[4] All the provinces now adopt "preference submission after score announcement". The two latest provinces making transformation are Beijing (in 2015) and Shanghai (in 2017). All other provinces completed the transformation in 2013.

matching game, sorting behavior can be even easier to implement given relatively small information burden.

So if sorting behavior is not difficult to implement, would agents be willing to choose it? The answer is yes for a broad class of matching mechanism. Back to Roth and Sotomayor (1990), sorting behavior has been proved to be equilibrium strategy for women in a men-optimal stable matching mechanism (Theorem 4.15). In Ergin and Sonmez (2006), sorting behavior is proved to be an equilibrium strategy under non-strategy-proof Boston mechanism (Theorem 1)[5]. Haeringer and Klijn (2009) consider the constrained school choice, where the number of schools a student can submit are binding constrains. They prove that under Top Trading Cycles (TTC) mechanism, any equilibrium given a shorter length of allowed preference list is also an equilibrium given a lengthier one. One immediate implication of it is that sorting behavior can be Nash equilibrium under the strategy-proof TTC mechanism.

However, sorting behavior has not been the focus of most studies until now. Literature has centered on designing strategy-proof mechanisms and promoting truth-telling behaviors. Non-strategy-proof mechanisms are usually regarded as undesirable, and various non-truth-telling behaviors regarded as nuisance. Empiricists have additional reasons to ignore non-truth-telling behaviors: they are difficult to classify and identify in a general way[6]. Non-truth-telling equilibrium is often not unique, and among them, sorting behavior is also not unique.

Our paper design a matching environment to examine sorting behavior. Sorting behavior is unique[7] due to acyclic school priority. We include both truth-telling and non-truth-telling mechanisms, while under truth-telling mechanisms, non-truth-telling equilibrium exists. We are also interested in how information would affect the choice of sorting behaviors versus other including truth-telling behaviors.

The experiment design is also motivated by China's college admissions system. In the system, students are uniformly ranked and matched to schools by their total score earned in the national College Entrance Examination (CEE). Still, there are disparities with this system across time and regions, mainly in two dimensions: preference submission timing and the matching mechanism. Preferences can be submitted before

---

[5] Wu and Zhong (2017) show that, under the Boston mechanism, if schools have acyclic priorities and students have some degree of preference homogeneity, sorting behavior is the only Nash equilibrium.

[6] A few experimental papers identify several patterns of non-truth-telling behaviors. Chen and Sonmez (2006) identify district school bias (DSB) and small school bias (SSB). Calsamiglia et al. (2010) consider modifications of truth-telling behavior, e.g., preservation of ranking and truncated truth-telling. Klijn et al. (2013) study protective strategy (i.e., max-min strategy). The only one focusing on sorting behavior is Pan (2017).

[7] More precisely, it is unique if we only consider the first choice. However, as we show later, in theory, admission by first choice is often equilibrium outcome, and in experimental results, first choice admission is very high under sorting behavior.

the exam, after the exam but before knowing the exam scores, or after knowing the exam scores. The two most common matching mechanisms used have been the Boston mechanism (known as mechanism "without parallel options" in China) and Serial Dictatorship mechanism (known as mechanism "with parallel options" in China). However, the SD mechanism nowadays is a constrained one, or half-way between Boston and SD (Chen and Kesten, 2017). Strategic plays have long been regarded as critical for success entrance into colleges.

Inspired by China's college admission, we experimentally investigate four matching environments: Boston under complete information (BOS_C for short) and incomplete information (BOS_I), and Serial Dictatorship under complete information (SD_C) and incomplete information (SD_I). Complete information corresponds to students submitting their preference ranking after the score realization, while incomplete information corresponds to submission before the score realization. College priorities are all determined uniformly by student score, the strongest case of acyclic priorities. Sorting behavior is defined as, for a student, the most preferred achievable school *given available information*.

We found, first, sorting behavior is prevalent in any of four environments. Under non-strategy-proof Boston mechanism, around two thirds of students play it. Under strategy-proof SD mechanism, the proportion is much lower, but still around 20 percent of students play it, while around two thirds play truth-telling. Information makes no difference on the choice of sorting behaviors. Second, sorting behavior leads to a high first choice admission rate (near 90 percent)., and makes its players as well off as truth-telling behaviors, under any mechanism. Third, an increase in sorting behavior improves the social welfare under Boston mechanism, but not under SD mechanism.

Here are some cautions. Our results do not support that Boston mechanisms perform better that SD mechanisms. In fact, we found in our experiment that SD mechanisms (where truth-telling dominates) always perform better than Boston mechanism (where sorting behavior dominates). Furthermore, when all players switch to truth-telling, social welfare is generally higher than when all players play sorting behavior. Sorting behavior is used to align the interests of players, when players are not willing to use truth-telling because they are afraid of being explored by other players.

The paper is organized as following: in Section II we summarize related literature and lay out theories on sorting behavior under various mechanisms. Section III is devoted to describe experimental design and measurements for gauging individual behavior as well as matching outcomes. Section IV presents various patterns of non-truth-telling behaviors and highlight sorting behavior. Section V focuses on how sorting behavior affect individual welfare. Section VI presents matching outcome under various environments and examine how varying behavior mixture would affect social

welfare. Section VII extends our measure of sorting behavior to include a broader class of misreporting behavior. Section VIII concludes the paper.

II. Related Literature

We summarize literature on sorting behavior under Boston and SD mechanisms with complete and incomplete information. We first highlight theoretical predictions which address sorting behavior under various mechanisms. We then discuss experimental literatures concerning non-truth-telling behaviors, including sorting behavior. Finally, we brief literature on matching outcomes under various mechanisms.

Theoretical Predictions

In the following we focus on matching environments where school priorities are acyclic (in the strongest sense). The immediate implication is that stable matching is unique (Haeringer and Klijn, 2009).

Ergin and Sonmez (2006, Theorem 1) prove that under Boston mechanism, sorting behavior is a Nash equilibrium which implements the stable matching. It immediately implies that under such an equilibrium, students are all admitted by their first choice. Haeringer and Klijn (2009) prove that under Top Trading Cycles (TTC) mechanism, any equilibrium under a shorter length of allowed preference list is also an equilibrium under a lengthier one. This immediately implies that sorting behavior, the Nash equilibrium with allowed preference list containing only one school, is also Nash equilibrium for TTC without any constrains. By Theorem 6.4 in their paper, if school priorities are acyclic, constrained (and unconstrained) TTC implement stable matching outcomes under Nash equilibrium. Therefore, students are also admitted by their first choice under sorting behaviors. Finally, note that under acyclic school priorities, TTC mechanism is mechanically equivalent to SD mechanism (Abodulkadiroglu and Sonmez, 2003).

We summarize those theories as the following:

Proposition 1: Under Boston mechanism and SD mechanism with acyclic school priorities and complete information, sorting behavior is Nash equilibrium and implements stable matching, and students are admitted by their first choice.

Furthermore, Wu and Zhong (2017, Proposition 2) prove that under Boston mechanism, if college slots are scarce resources, sorting behavior is the unique Nash equilibrium. College slots are scarce resources if the number of students preferring admission (by some colleges) to non-admission is larger than total number of slots.

Although our experimental design does not satisfy this condition, their proposition still highlights the important of sorting behavior in some real situations, e.g., China's college admission.

We also want to highlight one interesting theory result. Consider a mixture of truth-telling behavior with sorting behavior under SD with complete information (SD_C). That is, players either play truth-telling or sorting behavior. It is easy to see that such a strategy profile also forms a Nash equilibrium. We state the following proportion without proof:

Proposition 2: Any mixture of truth-telling and sorting behavior form a Nash equilibrium implementing stable matching outcome under SD mechanism with complete information.

SD_C mechanism is robust to any mixture of truth-telling and sorting behavior, while other mechanisms (e.g., Boston mechanism) may not. This is definitely a desirable property for SD_C.

The situation under incomplete information is more complicated. Usually, sorting behavior is not Nash equilibrium. Lien et al. (2017, Proposition 3.3) prove that under Boston mechanism with incomplete information (BOS_I), if students have homogeneous preference over schools, and each school has one slot, Boston mechanism implement ex-ante fair matching outcomes *only if* sorting behaviors are used by all students, except the one with the least score. Here ex-ante fairness is defined as stable matching outcome with regard to expected scores. Sorting behavior is accordingly defined as sorting to expected scores (more details in Section III.2&3). However, their paper's Theorem 3.2 states that ex-ante fairness can only be implemented under very restricted conditions, i.e., students have almost no competition relationship (i.e., any overlapping on realized scores) with each other[8]. The implication of these two theories is that sorting behavior are almost surely not Nash equilibrium under BOS_I. Under SD_I, since truth-telling is always dominant, it is easy to prove that sorting behavior may not be Nash equilibrium[9].

[8] Both Proposition 3.3 and Theorem 3.2 can be extended to more general cases, i.e., multi-slot school, and some degree of preference heterogeneity, see Section 5 in Lien et al. (2017).

[9] Consider one player (student A) deviates from sorting behavior to truth-telling, given other players still play sorting behavior. His realized score is larger than some students with higher expected score, among whom student B has the highest expected scores. Suppose for simplicity all students have the same ordinal preference. If everyone plays sorting behaviors, all get their ex-ante fair school. When player A deviates to truth-telling, he can switch to the ex-ante fair school belonging to student B, a more preferred school. Sorting behavior is not Nash equilibrium.

We summarize our findings as following:

Proposition 3: Under Boston mechanism and SD mechanism with incomplete information, sorting behavior may not be Nash equilibrium.

Although sorting behavior may not be a Nash equilibrium, it can still server as a focal point, especially when there is no obvious equilibrium strategy, e.g., under BOS_I. We will see what happens in the lab under mechanisms with incomplete information (i.e., BOS_I and SD_I).

Non-truth-telling in experiments

Experimental literature usually focuses on welfare consequences (efficiency and stability) of the whole matching mechanism, and often address truth-telling behavior. The study of non-truth-telling behavior is usually their byproduct, with only a few exceptions directly addressing non-truth-telling behaviors. Chen and Sonmez (2006) single out several patterns of non-truth-telling behavior under Boston, TTC and Gale-Shapley mechanisms, e.g., small school bias (SSB) and district school bias (DSB). Pais and Pinter (2008) consider SSB behavior as well as priority school bias (PSB, i.e., students rank schools where they have priority higher in the submitted rank) under Boston, TTC and GS mechanisms with various information. Calsamiglia et al. (2010) consider modifications of truth-telling behavior, e.g., preservation of ranking and truncated truth-telling, under constrained school choice. Klijn et al. (2013) study how an individual's risk preference influences his/her strategy, especially the choice of protective strategy (i.e., max-min strategy) under Gale–Shapley and Boston mechanism. Neither of them highlights sorting behavior as a general pattern of non-truth-telling behavior.

Featherstone and Niederle (2014) explore Boston mechanism (as well as others) with incomplete information, and argue that non-truth-telling equilibrium is difficult to achieve, because students fail to figure out equilibrium strategy. However, they focus on students' second choice - what they called the "skip the middle" bias. Students in fact work quite well in choosing their first choice (see their Table 2), supportive to our arguments. Second choice are important in their environment, because they assume homogenous ordinal student preference, which may intensify competitions among students. This is not the case in our environment, where we only assume partially aligned student preferences. Nevertheless, we never argue sorting behavior is equilibrium strategy under BOS_I as they did in their setup.

Pan (2017) is the closest paper to ours. She highlights the "sorting strategy" (Definition 5) and find that sorting strategy is skewed by overconfidence under BOS_I, resulting in more ex-ante unfair matchings than other matching environments. Our results agree with hers in that BOS_I in generally performs worse than other mechanisms. But our results do not support that the reason is that students usually over-report their first choice due to overconfidence. In fact, students play quite well in positioning their first choice. Under BOS_I, more than 60 percent of students fairly report their first choice, only 20 percent of students up-report, and 17 percent of students even down-report (See Section VII). There is no evidence overconfidence dominates. Furthermore, by increasing the proportion of sorting behavior, the whole matching outcome becomes better off. This could not happen when sorting behavior is driven by overconfidence. We believe in a more realistic environment as we designed, sorting behavior is not only driven by overconfidence, but also by other factors, e.g., risk attitude.[10]

Matching mechanisms and matching outcome

Among huge literature addressing this issue, we focus on those dealing with the effect of incomplete (or imperfect) information on matching mechanisms.

One important source of incomplete information is uncertain school priorities. For examples, schools may have indifferent priorities which need to be solved by random tie-breaking rules. Edril and Ergin (2008) found that when random tie-breaking is introduced, Gale-Shapley mechanism may generate *inefficient* stable outcomes. Abdulkadiroglu et al. (2011) prove that when students have the same ordinal preferences (but different cardinal preferences) and schools use random tie-breaking rule, Boston mechanism Pareto dominates Gale-Shapley mechanism. China's college admission provides another important example of uncertain school priorities: school priorities may be determined by CEE exam scores which can only be realized *after* preference submission. BOS_I can be more efficient than other mechanisms under conditions consistent with Abdulkadiroglu et al. (2011) (Lien et al., 2016; Chen, 2017). In addition to efficiency, Lien et al. (2017) raise the issue of ex-ante fairness. Ex-ante fairness is defined as stable matching as usual with the only exception that school priorities are determined by ranking students according to their *expected* scores (or, arguably, their true abilities), instead of score realization. They prove that BOS_I can be ex-ante fairness than other mechanisms, but the edge may be small. In lab

[10] Our experimental design differs from Pan (2017) in several ways: First, Pan's design assumes homogenous ordinal student preferences, and student preferences are (obviously) public information. Our design allows for partially aligned student preference; only the aligned part of student preference is publicly known. Second, our design contained 36 students, with 7 schools and multiple slots for each school, while Pan's design only contains 5 students and 5 unit-slot schools.

experiments, Lien et al. (2016) provide positive evidence for both ex-ante efficiency and fairness advantage of BOS_I, while Pan (2017) provide negative results for ex-ante fairness of BOS_I. Our environments do not fit into the model of either Abdulkadiroglu et al. (2011) and Lien et al. (2017), but rather a mixture of both: student scores are different ex ante, which is not in Abdulkadiroglu et al, (2011), while students have heterogeneous ordinal preferences, which is not in Lien et al. (2017). We are interested in how this more realistic setup would generate results concerning efficiency and ex ante fairness.

Another source of incomplete information is from private information of student preferences. Our setup also contains such elements: students only vaguely know others' preference. Pais and Pinter (2008) found that TTC mechanism works better than Boston mechanism, and less sensitive to information. However, Featherstone and Nierderle (2014) found that, Boston mechanism under incomplete information can have higher efficiency than Deferred Acceptance (i.e., Gale-Shapley) mechanism, and can even induce more truth-telling. However, their setup is extreme: they not only assume school priorities are purely random drawn, but also student ordinal preferences. In our paper we do not find more truth-telling and higher efficiency under BOS_I than other mechanisms.

III. Experimental Design and Measurement

In this section we first describe our experiment design, and then explain how to identify sorting behavior in our setup. We also describe our methods to measure matching outcomes.

III.1 Experimental Design

Our experiment is designed to compare different student behaviors (in particular sorting behavior and truth-telling) and matching outcomes under different mechanisms. In particular, we implement a 2x2 design, Boston and SD mechanism under either a complete or incomplete information environment. For the complete information environment, preference submission is done after the exam and all the students' scores are known. For the incomplete information environment, preference submission is done before the scores are known. For each treatment, we have 2 or 3 sessions, with 36 students in each session. We have 10 sessions in total and each student only participated in one session.

Ranking/Mechanism	Boston	Serial Dictatorship
Complete	2 (N=72)	2 (N=72)
Incomplete	3 (N=108)	3 (N=108)

In the complete information setting, the scores for the 36 students in each session are independently drawn without replacement from integers between 105 and 140. Each student is informed of his/her ranking. In the incomplete information environment, the *estimated* scores for the 36 students in each session are independently drawn without replacement from integers between 105 and 140. Each student is informed of his/her *estimated* ranking. The actual score for each student is drawn from a uniform distribution -10 to +10 around his/her estimated score. The actual ranking is not revealed until the preference rankings are already submitted.

In each session, there are 36 school slots across 7 schools A-G: 3 slots each at A and B, 5 slots at C and E, 6 slots at D and F, 8 slots in G. All students prefer A and B to the other five schools, which is common knowledge. Some students prefer A to B and some prefer B to A, but the proportion is not publicly known. Except the public information mentioned above, students only know his/her own preference ranking. The payoff structure is similar to Chen and Sönmez (2006) and the payoffs obtained are symmetric, i.e., each student gets the same payoff for the same preference ranking of choice. The actual monetary payment obtained is multiplied by 5 RMB from the payoff table. Schools always prefer students with higher (realized) scores. The payoff table is shown below.

	Preference Ranking						
Payoff	1	2	3	4	5	6	7
	16	13	11	9	7	5	2

We ran paper-and-pencil experiments at Tsinghua University in China on May 24th (2 sessions), June 1st (4 sessions) and June 2nd (4 sessions) of 2012. Each session lasted approximately an hour with a participation fee of 20 RMB. All sessions were conducted in Tsinghua University, School of Economics and Managements' Experimental Economics Laboratory (ESPEL).

We also collected information from the students in a post-experiment survey including their experience (origin province, whether or not they took the college entrance exam, and year of college entrance exam if they did), gender, age, and major.

We use the three series of paired lotteries in Tanaka et al. (2010) for an incentivized elicitation of the prospect theory parameters.

III.2 Identifying Sorting behavior

Sorting behavior is for a player to put his most preferred achievable school as the first choice. This definition does not put restrictions on second, third and any other choices. In our setup, because students do not have full information of other students' preference, the unique stable-matched (or fair) school is not fully observable for them. Another issue is what kind of stable matching we are talking about. There are two kinds of stable matching: stable matching with regard to realized scores, i.e., ex-post stable (or fair) matching; or stable matching with regard to estimated scores, i.e., ex-ante stable (or fair) matching.

For the uncertain scores (or school priorities) issue, there is an obvious solution to pin down sorting behavior. For mechanisms with incomplete information, because students only know their estimated scores, their judgement on score priorities are only determined by their ranking of estimated scores. For mechanism with complete information, their judgement on school priorities are determined by their ranking of realized scores.

For the issue of incomplete information on student preferences, we consider two alternative measures of sorting behavior, by relaxing the theoretical definition. First, we assume students only rely on public information to figure out their fair school, combined with information of their estimated or realized score ranking. Therefore, if a student is ranked (ex ante or ex post) among top 6, the total number of school A and B, which all students prefer to other schools, their achievable school set is {A, B}. Otherwise their achievable school set is {C, D, E, F, G}. Among their achievable school set, they choose their most preferred one according to the endowed school payoffs. For example, under an incomplete information setup, a student with an estimated score ranking as 5^{th}, payoff for B as 16, payoff for A as 13, would be regarded as playing a sorting behavior if he chooses school B as the first choice. Or, under a complete information setup, a student with a realized score ranking as 10^{th}, payoffs as C=9, D=11, E=7, F=5, G=2, is playing sorting behavior if he chooses school D as the first choice.

Second, for experimental designer, by a hindsight, we can identify whether a student put his (ex-ante/ex-post) stable-matched (or fair) school as his first choice. The shortcoming of this *de facto* sorting behavior is obvious: students cannot deliberately choose it due to lack of information. Yet it is still useful, because at least in some degree, it answers the "what if" question: i.e., how sorting behavior (or deviation from it) would

affect individual matching outcome, if fair school is fully observable (or calculable) by agents[11]?

In the later parts, we refer to the first measure of sorting behavior still as "sorting behavior", and the second measure as "*fair reporting*". Note that two measures converge to one when information on student preferences are fully revealed.

A final note. Sorting behavior and truth-telling can be overlapped. It happens when a student's true first choice is coincident with his fair school. When they overlap, truth-telling is a stricter concept that sort-behavior: it requires all schools are listed in the true preference order. So we give a higher priority to truth-telling. That is, if a behavior can be both sorting behavior and truth-telling, we statistically identify it as truth-telling, not sorting behavior.

III.3 Measuring Matching Outcomes

For our purpose, it is important not only to measure (or evaluate) matching outcomes for the whole matching, as common in literature, but also to measure matching outcomes at individual level. We first discuss the latter one. The measurements at the system level is somehow an aggregation of measurements at individual level, so we discuss it subsequently. The evaluation of the system matching outcomes involves a simulation procedure, which we will discuss briefly at the end of this subsection.

Measuring individual matching outcomes

We consider two measures: one is related to fairness and the other to efficiency.

The *degree of mismatch* is defined as the gap between the preference ranking of a student's matched school (in lab) and his/her fair school. Note that the sign of the degree of mismatch can be negative or positive, indicating down-matching or up-matching respectively. For example, a student's fair school is C, which in his payoff table is ranked 3th, yet in the lab experiment he is matched to school A, which is ranked first in his payoff table. Then his degree of mismatch is 3-1=2, and he is up-matched. The measure is intuitive to see how far (above or below) a student is from his fair matching outcome, but the shortcoming is that it is based on (somehow arbitrary) ordinal number of preference rankings.

We need to distinguish between environments with incomplete and complete information. For incomplete information, we use a student's ex ante fair school to

[11] No doubt, student behaviors would be different whether they are fully knowledgeable and choose fair school as the first choice cautiously, or maybe, they just try to guess the fair school and get it in a random way. Therefore, our results are only the best guess for the effect of this type of sorting behavior given our information setup.

calculate his/her degree of mismatch. Ex ante fair school can be derived, for example, by ranking students according to their estimated scores, and match them with schools by using SD mechanism. The matched school for each student is his/her ex ante fair school. For complete information, we consider a student's ex post fair school. The algorithm is still by using SD mechanism but ranking students according to their realized scores.

The *relative payoff* is defined as the proportion of a student's payoff in the lab over his payoff under stable matchings. This measure complements what degree of mismatch leaves out, i.e., the "real monetary gain" (normalized to fair gain) a student has. Note that those two measures are positively correlated.

Measuring system matching outcomes

The welfare consequences of the matching outcome under a mechanism can be evaluated by both efficiency and fairness. For fairness, we consider both ex ante fairness and ex post fairness, i.e., fairness with regard to students' estimated scores and to students' realized scores. Note that both ex ante and ex post fairness schools are applicable for either complete or incomplete information. For experimental designers, either students' estimated or realized scores are known after the experiments are done, for any environment. In fact, how each of those four mechanisms can achieve higher ex ante fairness as well as ex post fairness is a topic of a few previous studies (Lien et al., 2016; Pan, 2017; Wu and Zhong, 2014).

We discuss measures for each of them.

Efficiency. The (ex-ante) efficiency is calculated as the payoff per capita in a matching system.

Ex-post Fairness. In the literature, fairness is measured by the number of blocking pairs. A blocking pair is a pair of student-school in which the student prefers the school to his own matched school, and the school either has an empty seat or gives a higher priority to this student than another student it admits. Here the school priority is solely determined by the students' realized scores. In this paper, we adopt two measures of fairness in the same spirit as blocking pairs.

(i) *Ex-post Fairness by number of blocking pairs.* Define (i, S) as a blocking pair for student if student i prefers school S to his matched school and his score is above the minimum score of all the students matched to school S. We then count the number of blocking pairs in a matching system. The lower the number of blocking pairs, the more fair the matching outcome.

However, we also consider another measure, which is the aggregation of our measure of fairness at individual level (See Section III.2).

(ii) *Ex-post Fairness by average absolute value of degree of mismatch.* This measure is calculated by averaging *absolute value* of the degree of mismatch across all students. A higher degree of mismatch implies a lower level of fairness of the matching outcome.

Some discussions on relations between those two measures are deserved. Although the two measures share the same spirit, they may not always be consistent. For example, consider 3 students, 1, 2 and 3, who are matched to 3 schools, A, B, and C. All the students have the same preference: A>B>C, and all the schools have the same priority: 1>2>3. It is easy to calculate that among all the 6 possible matching outcomes, matching outcomes with the highest average degree of mismatch (i.e., highest unfairness by measure (ii)) are three: (3=A, 2=B, 1=C), (2=A, 3=B, 1=C), and (3=A, 1=B, 2=C), with the value of 4/3. While by measure (i), the number of blocking pairs, only (3=A, 2=B, 1=C) has the highest degree of unfairness, with the value of 1 (versus 2/3 for the other two). One may argue that (3=A, 2=B, 1=C) should be the only one most unfair, because it "reverses the whole matching order": it put the best student into the worse school, and the worst student into the best school, and so on. Yet another one may argue that (3=A, 2=B, 1=C) can be fairer than (2=A, 3=B, 1=C) and (3=A, 1=B, 2=C): In (3=A, 2=B, 1=C), at least student 2 is matched to his fair school, while in (2=A, 3=B, 1=C) and (3=A, 1=B, 2=C), all students are mismatched. We believe it is an important question to compare those two measures in general in the future study.

Ex-ante Fairness. Ex-ante fairness is defined similarly as ex-post fairness, except that fair matchings are defined based on expected student scores instead of realized scores. We still use two alternative measures, i.e., average degree of mismatch, and average number of blocking pairs as we measure ex-post fairness.

Simulation Issues. The initial condition of our game is that students only know the distribution of scores. Under the BOS and SD with incomplete information environments, students have to base their strategy on such score uncertainties. Yet under the BOS and SD with complete information environments, students know their realized scores, and then submit their preference rankings. To make the incomplete and complete information environments comparable, we need to compare their matching outcome under the same *prior* score realizations. For the mechanisms under incomplete information, we only need to realize the matching outcome many times according to the score distribution, each time using the strategies of all students we observe in the lab, which is obviously fixed across different score realizations. Yet for the mechanisms under complete information, we have to consider the counterfactuals, i.e., how the students (or subjects) would response to all the possible score realizations we could not observe. One conjecture would be that students only respond to their realized score ranking, so for any given realized score ranking, students would play the same strategy as the one with the same realized score ranking we really observe in the lab. However,

it would be naïve to assume that the realized score ranking is the only determinant of behavior and the observed behavior is the only "true" strategy. We therefore simulate student behavior using some empirical patterns we found in the experimental data. As a whole, we consider five scenarios in which we generate student behaviors based on observed behaviors (or their patterns) in our experiments. More details about simulation methods are in Appendix A.

IV. Patterns of Behaviors

We now present first group of results concerning patterns of behaviors. We want to identify as many as possible patterns of behaviors. But we try to convince you that sorting behavior is the most prominent one among all the non-truth-telling behaviors.

Table 1 shows the distributions of all behaviors patterns for each of four mechanisms (BOS_I/C, SD_I/C). Truth-telling plays a dominant role under two SD mechanisms, the strategy-proof mechanisms. There are 65% and 69% of all students playing truth-telling under SD_I and SD_C. Only 1% (1 person) play truth-telling under BOS_I, and the proportion is 11% (8 persons) for BOS_C.

Table 1: Distribution of reporting behavior

	BOS_I		BOS_C		SD_I		SD_C	
	#	%	#	%	#	%	#	%
All Students								
Truth-telling	1	0.93	8	11.11	70	64.81	50	69.44
Non-truth-telling	107	99.07	64	88.89	38	35.19	22	30.56
Sorting	68	62.96	47	65.28	19	17.59	17	23.61
Top 6 Students								
Truth-telling	1	5.56	7	58.33	15	83.33	10	83.33
Non-truth-telling	17	94.44	5	41.67	3	16.67	2	16.67
Sorting	15	83.33	3	25.00	3	16.67	2	16.67
Risk averse	2	11.11	2	16.67	0	0.00	0	0.00
Below top 6 Students								
Truth-telling	0	0.00	1	1.67	55	61.11	40	66.67
Non-truth-telling	90	100.00	59	98.33	35	38.89	20	33.33
Sorting	53	58.89	44	73.33	16	17.78	15	25.00
Risk seeking type-I	11	12.22	1	1.67	6	6.67	2	3.33
Risk seeking type-II	5	5.56	2	3.33	2	2.22	0	0.00
Safe choice	7	7.78	5	8.33	5	5.56	0	0.00
Equal slot switch	6	6.67	6	10.00	3	3.33	1	1.67
Less slot switch	8	8.89	1	1.67	3	3.33	2	3.33

What we focus is, however, non-truth-telling behaviors. Under BOS_I and BOS_C, sorting behavior dominates: 63% and 65% of all students play sorting behavior under BOS_I and BOS_C. Under SD_I and SD_C, there are still 18% and 24% of all students play sorting-behavior, dominating all other non-truth-telling behaviors.

We also divide students into two groups: Top 6 or Below Top 6, according to their expected/realized scores under incomplete/complete information. Unsurprisingly, truth-telling becomes more pronounced for Top 6 students, because they are more eligible for their true first choice school. Sorting behavior becomes more pronounced for below Top 6 students, with the only exception under BOS_I. Under BOS_I, Top 6 students tend to choose their first choice "right", i.e., to choose fair school (also their true first choice school), but choose other choices in a way inconsistent with their true preference order. This is understandable because BOS_I is not strategy-proof, and some students may apply the "skip the middle" strategy as suggested in Featherstone and Niederle (2014).

We also identify some other non-truth telling behaviors, for Top 6 students and below top 6 students separately. For Top 6 students, the only pattern beside truth-telling and sorting behavior is what we called "risk averse" behavior. That is, students put their less preferred one between school A and B as their first choice. Only 11% and 16% students under BOS_I and BOS_C play it, and none play it under and SD mechanism.

For below top 6 students, more non-truth-telling behavior patterns can be identified. Here "risk seeking type-I" is defined as choosing the most preferred school (between A and B) as their first choice. Under BOS_I, 11% students play this very risky strategy, while under other mechanisms, less than 7% student do that. "risk seeking type-II" is defined as choosing the second preferred school (between A and B) as their first choice. Surprisingly, students playing this strategy is less than those playing "risk seeking type-I", except under BOS_C, where it is 2% higher. Another choice we highlight is "safe choice", which means choosing a school which has more slots that the school sorting behavior should choose (i.e., sorting school). For example, when sorting behavior requires one to choose school C or E, he/she chooses among D, F, G. None plays it under SD_C, while under other mechanisms, players count on a proportion of 6-8%. For remained behaviors, we can categorize them into two: switching to schools with equal or less slots form their sorting school. Their proportions are roughly comparable with safe choice. These two patterns are hard to explain: they are either chosen randomly, or involve some high level thinking. Finally, note that there are no students who put their least preferred school first (not shown in Table 1).

In summary, by examining Table 1, we are fair to say that sorting behavior dominates under non-strategy-proof Boston mechanisms, as well as prevails under SD mechanisms.

In the following we focus on the two prominent behaviors: truth-telling and sorting behavior. We ask how frequent these two patterns are, although they are both prevalent, under various environments. Table A1-A2 use regression method to explore various factors affecting the emergence of those two behaviors, both across and within mechanisms. Table A1 is on determinants of truth-telling. Truth-telling is more prevalent under SD than Boston mechanisms. Within Boston mechanisms, it is more prevalent under complete information than incomplete information. Under complete information, students have certain estimation of their rankings. Therefore, for students with high rankings (e.g., the Top 6), they dare use truth-telling (see also Table 1). Within SD mechanisms, truth-telling does not differ between two information settings, reflecting that truth-telling is always a dominant strategy. Note also that the behavioral parameter σ has significantly positive effects on the choice of truth-telling, implying that more risk averse players (with lower σ) are less likely to be truth-telling.

Table A2 is on determinants of sorting behavior. Contrasting to truth-telling, sorting behavior is more prevalent under Boston mechanisms. Information has no influence on the emergence of sorting behavior, for either Boston or SD mechanism. Sorting behaviors are not influenced by any other factors we include in the regression.

V. Welfare Consequences of Behaviors

In this section we compare welfare consequences of truth-telling and sorting behaviors. Truth-telling is dominant under SD_I/C, while as we found, sorting behavior is dominant under BOS_I/C. It is tempting to think sorting behavior should perform better than truth-telling under BOS_I/C, while the opposite is true under SD_I/C. This may be not true. First, by Proposition 2, sorting behavior is not defeated by truth-telling under SD_C. Second, more generally, the performance of any behavior also depends on the mixture of various behaviors in the game, *in equilibrium* it is hard to see which strategy is better than others.

Sorting behavior is considered as a behavior for the player to achieve admission by their first choice, as stated in Proposition 1. Table 2 shows whether this is the case in the lab. For all students (or all behaviors), the first choice admission rate is higher under BOS_I/C than under SD_I/C, with the proportion of around 80% in the former and around 30% in the latter (Column (1)). Truth-telling (Column (2)) do not generate high first choice admission under SD mechanisms, understandably. It seems surprising that it does generate high first choice admission under Boston mechanisms, but note that very few players play this strategy and almost all of them are top students (Table 1). On the contrary, sorting behavior generates a high first choice admission, with a proportion of 85%-94% under every mechanism (Column (4)). This verifies that sorting behavior players indeed try to achieve first choice admission.

Table 2: Proportions of Top Choice Match in different mechanisms (%)

	All Students	Truth-telling	Sorting	Other Non-truth-telling	Fair-reporting
Treatment	(1)	(2)	(3)	(4)	(5)
BOS_I	81.48	57	92.59	62.74	93.56
BOS_C	88.89	87.5	93.62	76.47	97.78
SD_I	35.63	18.58	84.97	49.11	88.78
SD_C	31.94	16	88.24	0.00	100.00

Note: Percent of top choice match for BOS_I and SD_I is the average value after 200 simulations of score distribution.

Figure 1 compares the fairness consequences of truth-telling and sorting behavior for their players. (See also Table A3 for more details). The measurement is the degree of mismatch (can be positive or negative). For comparison, we also calculate the fairness of other non-truth-telling behavior excluding sorting behavior. A larger mean of degree of mismatch is more desirable. Truth-telling performs no better than sorting behavior under any environments, and even worse under BOS_I and SD_C (p-value of Wilcoxon rank-sum test is p=0.0001 and p=0.0020, by Table A3). On the contrary, other non-truth-telling plays worse than sorting behavior under any environment, significantly (by Table A3).[12] In addition, sorting behavior is also moderately safe, measures by the variance of degree of mismatch (in Table A3).

[12] We also compare variance (instead of mean) of degree of mismatch for various behaviors. Less variance is reasonably more desirable. Under BOS_I, truth-telling has a higher variance than sorting, whole under SD_I, it has a lower one. Under other environments there two are equally well (under SD_I) or non-comparable (under BOS_C, because the sample size for truth-telling is one.). See Table A3.

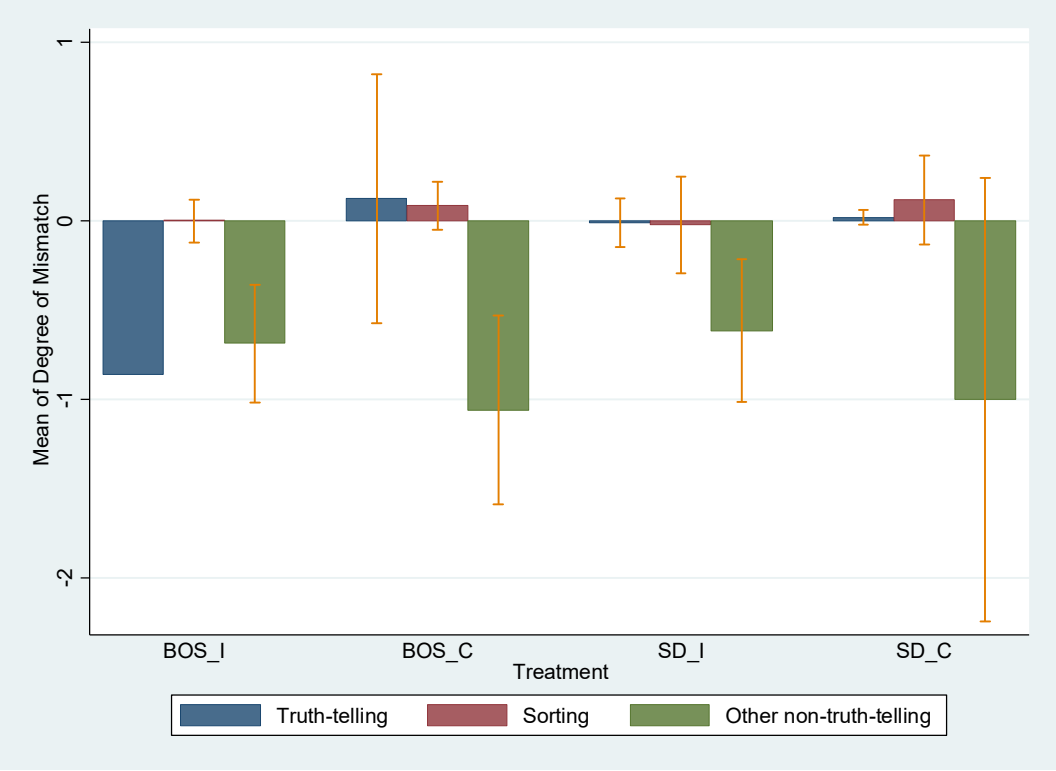


Figure1: Fairness Consequences of Different Behaviors

Figure 2 (and Table A4) compare the efficiency consequences of truth-telling and sorting behavior. The result is the same as for fairness. Sorting behavior performs equally well as truth-telling under BOS_C, SD_I/C, and better than truth-telling under BOS_I. Other non-truth-telling always performs the worst.

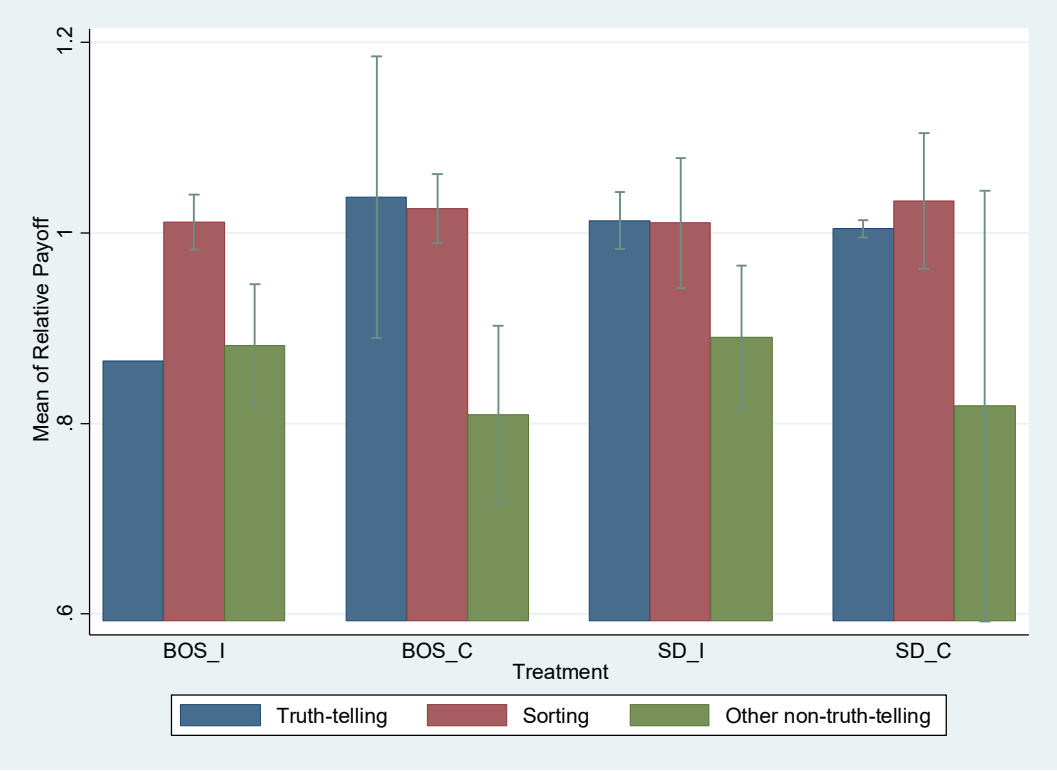


Figure 2: Efficiency Consequences of Different Behaviors

As a whole, our results suggest sorting behavior is indeed a "good" strategy under any environment, at least compared with truth-telling and other non-truth-telling behaviors as a whole. This may justify why it is prevalent under all environments, as we've already shown in Section IV.

VI. Promoting Sorting Behavior? A Social Welfare View

In Section V we show that sorting behavior performs as well as truth-telling or even better under various mechanisms. Does it imply that we should, as a policy maker, promote sorting behavior, compared to truth-telling, or other non-truth-telling behavior, to improve the social welfare (either efficiency or fairness)? Not necessary. Sorting behavior performs well in our lab only for the specific group of students who play it. It may cause negative (or positive) externality on other players playing either truth-telling (under BOS mechanisms) or other non-truth-telling (under all mechanisms). So it is naïve to extrapolate its welfare advantage for individual to welfare superiority for the whole society.

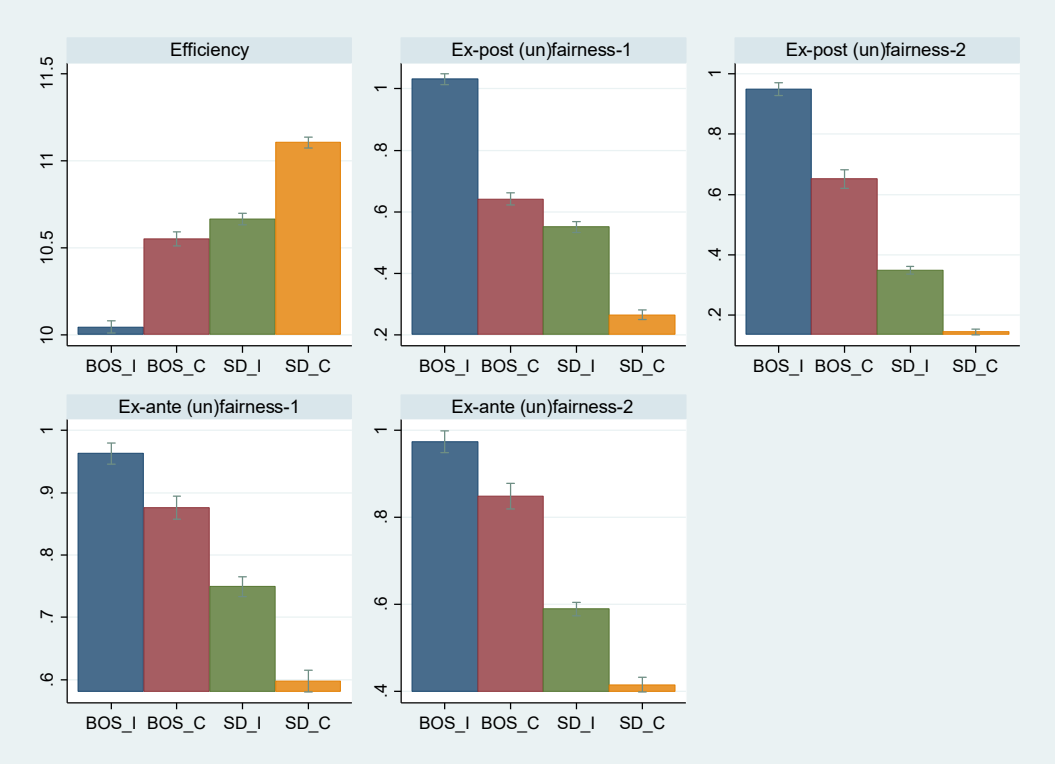
A useful starting point to consider the social welfare consequence of sorting behaviors (vs truth-telling or others) is to directly examine the welfare outcome of four mechanisms in our lab experiments. As we've seen in Section IV, sorting behavior is

as prevalent under BOS-I/C as truth-telling under SD_I/C. Therefore, comparing the welfare outcomes of those four mechanisms at least sheds some lights on how sorting behavior and truth-telling contribute to the social welfare.

Welfare Outcomes of Mechanisms

We use simulation methods to find welfare outcomes of four matching environments, as we explain in Section III.3 and in Appendix A. We consider three welfare criteria, (ex-ante) efficiency, ex-ante fairness and ex-post fairness, also explained in Section III.3. The results are shown in Figure 3 (and Table A5). Under all environments, and all measures, the outcome is consistent: SD_C is always the winner, followed by SD_I, while BOS_I is the loser, followed by BOS_C.

Therefore, the results do not favor sorting behavior (vs truth-telling) to promote social welfare. In particular, sorting behavior does not help BOS_I to achieve (ex-ante) efficiency or fairness, as claimed by Lien et al. (2016, 2017), but rather, aligned with the claim of Pan (2017), which suspects the "social value" of sorting behavior.



Note: Efficiency is measure by payoff per capita. Ex-ante or ex-post (un)fairness is measured by degree of mismatch (1) or the number of blocking pairs (2).

Figure 3: Social Welfare Comparison Among Four Mechanisms

Although higher frequency of sorting behavior is associate with lower social welfare (under BOS_I/C, vs SD_I/C), while higher truth-telling is associate with the opposite (under SD_I/C, vs BOS_I/C), it is still unfair to discard sorting behavior and promote truth-telling under *all* mechanisms. First, BOS_I/C is essentially non-truth-telling mechanisms. Unless you discard the whole mechanisms, you'd better not to "fool" players by suggesting them to play truth-telling. If you still try to use Boston mechanisms, for some theoretical arguments by Abdulkadiroglu et al. (2011) or Lien (2017), or some practical considerations (remember it is never possible to truth-tell under Chinese college admission system), you'd better consider some non-truth-telling behaviors.

Second, and more important, the aforementioned correlation between behavior frequency and welfare outcome does not necessarily mean that promoting sorting behavior cannot help any given *non-truth-telling* mechanisms (e.g., BOS_I/C) to achieve higher social welfare. It even does not suggest increasing the frequency of sorting behavior must decrease the welfare of *truth-telling* mechanisms. After all, if sorting behavior indeed incur negative externalities to *other* non-truth telling behaviors or truth-telling behavior, those externality effects may *diminish* when more players play sorting behaviors. In our next subsection, we ask, for any given environment, what happen when we increase the frequency of sorting behavior or replace truth-telling or other non-truth-telling behaviors.

Promoting Sorting Behavior: An Counterfactual Analysis

Since Boston mechanism are non-strategy-proof, our counterfactual test only considers the switch between sorting behavior and other non-truth-telling behavior. Yet we also include the all-truth-telling case (i.e., all players play truth-telling) as a benchmark. Under SD mechanisms, we instead consider the switch between sorting behavior and truth-telling. Our welfare measure and simulation method are described in Section III.3 and Appendix A. Because sorting behavior only characterizes students' first choice, we consider a simulation scenario applicable to this characterization, which is scenario 2'. (Other scenarios would be discussed in Section VII.)

Figure 4 (and Table A6) shows that when we increase the proportion of sorting behavior, all the welfare measures, including efficiency, ex-ante and ex-post fairness improve significantly (with p-value for testing difference all being zero, see Table A6).

For example, under BOS_I, when we increase the proportion of sorting behavior by 6 percent (roughly one tenth of its original proportion), efficiency increased by 2 percent (= (10.249-10.046)/10.046) (Table A6). Therefore, the elasticity, i.e., the percentage change of efficiency divided by the percentage change of sorting behavior proportion (i.e., 6 percent), is roughly 0.33. Ex-ante and ex-post fairness improve more,

by about 5-9 percent, with the elasticity close to or higher than 1. When the proportion of sorting behavior increases by 12 percent, the percentage change of welfare is roughly doubled, with the elasticity being roughly the same. A decrease in the proportion of sorting behavior results in the similar magnitude of changes, with an opposite direction.

The welfare improvement under BOS_C is even larger, especially for ex-ante and ex-post fairness. The sorting behavior elasticity is roughly between 1.5-2.5, larger than that under BOS_I. Since at the original case, BOS_C has already outperformed BOS_I, BOS_C will always outperform BOS_I along the changing of the proportion of sorting behavior. However, for ex-ante fairness, when the change is large enough so that the proportion of sorting behavior is close to 1, BOS_I outperforms BOS_C.

We also consider what happens when all the players play truth-telling. Compare with all-sorting case, all truth-telling case has a higher efficiency, ex-ante and ex-post fairness under BOS_I/C, except for one ex-ante fairness measure under BOS_I.

Note: Efficiency is measure by payoff per capita. Ex-ante or ex-post (un)fairness is measured by degree of mismatch (1) or the number of blocking pairs (2).

Figure 4: The Effect of Changing Sorting Behavior under Boston Mechanisms

Figure 5 (and Table A7) shows the welfare results of replacing sorting behavior by truth-telling under SD mechanisms. Increasing or decreasing sorting behavior around the origin case has almost no effects on all welfare measures. For those significant changes, the direction can be not monotonic. Even if we jump to the all sorting behavior case, for most measures, the effect for fairness is still insignificant,

although the effect for efficiency is positive. By the fairness measure of the number of blocking pairs, all sorting behavior performs even the worst among all the listed behavior mixtures. [13]However, all truth-telling generates the highest welfare level. As a whole, sorting behavior does not show up under SD mechanisms.

Note: Efficiency is measure by payoff per capita. Ex-ante or ex-post (un)fairness is measured by degree of mismatch (1) or the number of blocking pairs (2).

[13] The inconsistency of two fairness measures, i.e., number of blocking pairs and degree of mismatch, sometimes very sharp, may due to the methodology issue we briefly discuss in Section III.3.

Figure 5: The Effect of Changing Sorting Behavior under SD Mechanisms

As a whole, sorting behavior helps to approach desirable matching outcomes under Boston mechanism, but not under SD mechanisms. Under Boston mechanisms, it helps both at the margin and at the extreme. Under SD mechanisms, it does not help at the margin, and only has limited influence at the extreme where all players play sorting behavior. Although all truth-telling strategy profile seems the best under any environment, under Boston mechanism it is not an equilibrium. Therefore, promoting sorting behavior is still a good alternative under such mechanisms.

VII. Alternative Measure of Sorting Behaviors

In previous sections we focus on one measure of sorting behavior, i.e., choosing the most preferred one within the *set of possibly achievable schools*. In this section we focus on the other measure, fair-reporting, i.e., students choose a school as their first choice which *turns out to be* their fair school.

Furthermore, we can extent the definition of fair-reporting to include *all* non-truth-telling behaviors. We define *misreport* as a behavior where a student's first choice turns out *not* to be his/her fair school. The *degree of misreport* measures the difference between the preference ranking of the first choice and the fair school. For example, if a student's fair school is a school ranked No 6 in his/her preference list, while his/her first choice is a school ranked No 5, then the degree of mismatch is 5-6=-1. If the degree of mismatch is negative, then the student is said to down-report, otherwise he/she up-reports. Therefore, all non-truth-telling behaviors are divided into three types: up-report, down-report, and fair-report.[14]

Figure 6 shows the proportion of various misreport behaviors, as well as truth-telling behavior. Under BOS_C/I, fair-reporting behavior dominates other behavior patterns, with a proportion of over 60 percent. Under SD_C/I, although truth-telling is dominant, fair-reporting is still prevalent and counts on near 20 percent among all players. There are also significant proportion of up-reporting and down-reporting, with a proportion of 10-20 percent in general.

Table A8 further explores the determinants of fair-reporting, parallel to Table A2 on determinants of sorting behavior. The result is also similar as Table A2. Boston mechanisms generate more fair-reporting than SD mechanisms. Within Boston or SD

[14] It is not surprising that fair-reporting and sorting behavior (defined in Section 3.2) are highly overlapping. In fact, among all the non-truth-telling behavior, 49.35% is both, 19.9% is either, while other 30.7% is neither.

mechanism, information makes no difference. There are no other factors included in the table showing significance.

Table A9 explores the determinants of various misreport behaviors further. It shows how the degree of misreport changes across mechanisms and other school or student characteristics. To focus on non-truth-telling behavior, truth-telling behavior is excluded from the sample. Misreport behavior does not differ across four mechanisms. Among school characteristics, if a student has a larger size of fair school slots, he or she is encouraged to choose a first choice school ranked higher. Among student characteristics, a student with a higher expected/realized score ranking tends to up-report more. No other student characteristics has significant influence, except that student from Economics and Management school up-report more. Behavioral parameters keep quiet in all regressions.

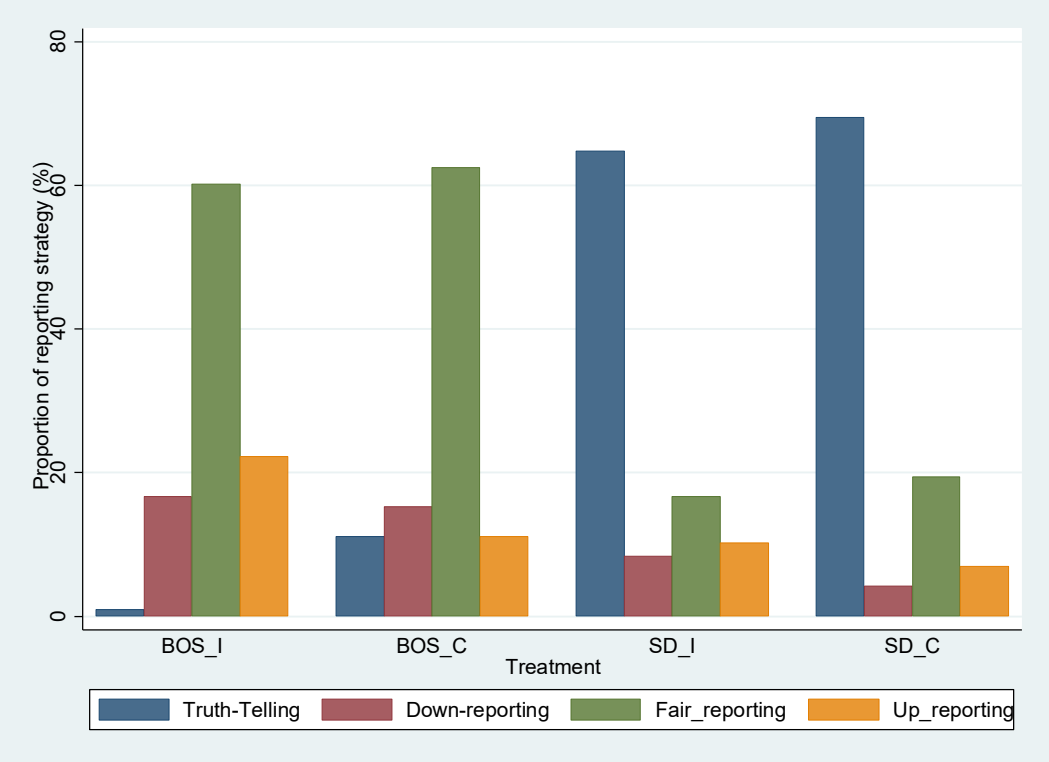


Figure 6: Proportions of Misreporting and Truth-telling

Column (5) in Table 2 shows that fair-reporting results a very high first choice admission under all mechanisms. Under SD_C it is 100 percent, while under other mechanisms it is close to or beyond 90 percent. Figure 7 compares fairness consequences of misreporting (including fair-reporting) under four mechanisms, through the measure of degree of mismatch. Table A10 use a regression method to

address the same issue. Fair-reporting almost always generates fair matching. Up-report results in a positive degree of mismatch, or an up-match, while down-reporting results in a negative degree of mismatch, or down-match. Truth-telling also generates fair matching under BOS_C and SD_I/C, but down-match under BOS_I.

Although up-match generate higher matching outcome on average, the benefit has its costs: the variance of degree of mismatch is also higher under up-report than under fair-report and other behaviors (Table A5). Therefore, although fair-report results in lower matching than up-report, it is safer.

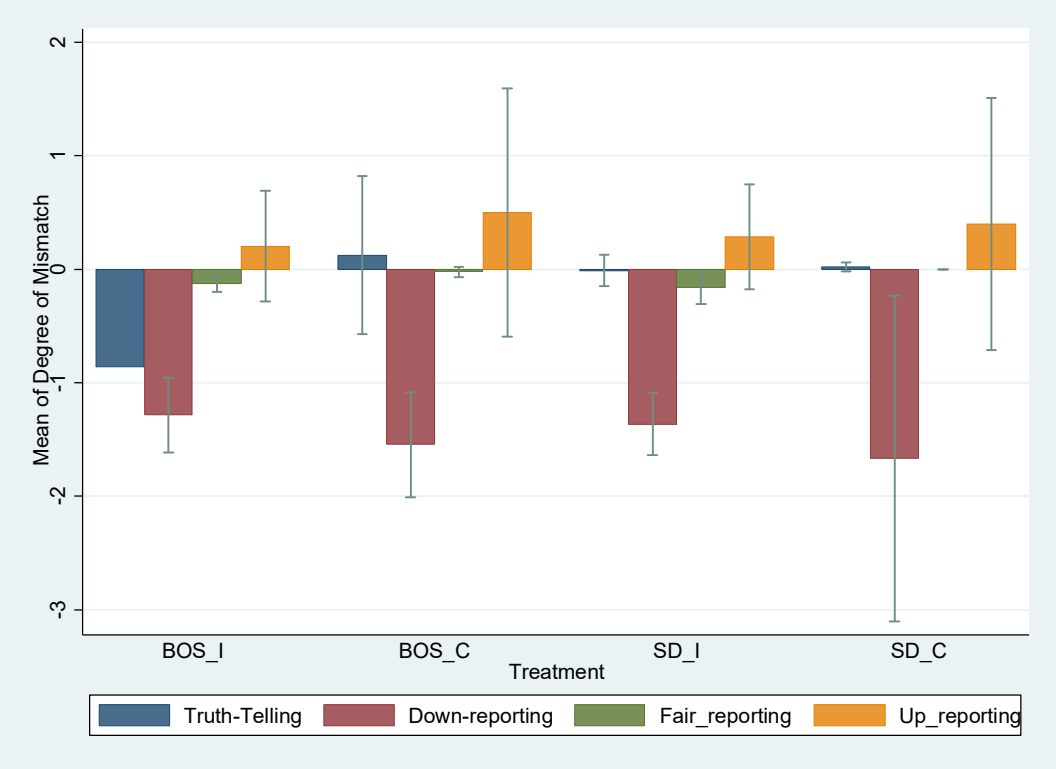


Figure 7: Degree of Mismatch under Misreporting

We can also simulate welfare outcomes of various mechanism by characterizing students' misreporting behavior. The measure of misreporting has an advantage over sorting behavior: it can not only characterize first choice, but any further choice. We can ask how far away a student's second choice is from his/her fair school, and so on. In scenario 1 of our simulations, we simply copy every student's behavior observed in the lab experiment data, only by his/her expected or realized ranking, as in Chen and Sonmez (2007). In scenario 2-4, we only draw information on distributions of students' misreporting behaviors by considering their first choice and second choices, and allow for a larger space for randomization. More details are given in Appendix A.

All the simulations deliver matching outcomes similar with what we have when we use sorting behavior as the "experimental" behavior (as in Table 1). SD_I/C always perform better than BOS_I/C under any welfare measure, either efficiency or ex-ante/ex-post fairness. The complete information case usually has a higher welfare also, with the only exception that BOS_I performs better than BOS_C under the ex-ante fairness measure. The results are shown in Figure A1-A5.

Our simulation results also support our simulation method by focusing on first choice. Although in different scenarios we always characterize the first choice, we randomize other choices in various degrees, with the full characterization of all choices in scenario 1. Although the absolute level of welfare differs across different scenarios, the relative rankings of welfare performance are almost unchanged.

VIII. Conclusions

Strategy-proof mechanisms are thought as good mechanisms because they induce good behaviors such as truth-telling. School choice and college admission literature, among others, often focus on strategy-proof mechanisms (such as SD or TTC mechanisms) and truth-telling behaviors, with non-strategy-proof mechanisms (such as Boston mechanism) as the control group, and non-truth-telling behavior as nuisances. However, non-strategy-proof mechanisms are frequently used in reality. Meanwhile, non-truth-telling behaviors emerge in both strategy-proof and non-strategy-proof mechanisms. Therefore, studying non-truth-telling behaviors have at least double significances. First, if we have to use non-strategy-proof mechanisms, what kind of non-truth-telling behavior is good for players and for the society? Second, under strategy-proof mechanisms, how often does some non-truth-telling behavior emerge and how they affect the matching outcomes?

In this paper we put non-truth-telling behavior as our "leading role". In particular, we highlight one non-truth-telling behavior: the sorting behavior, where plays list their most preferred achievable matching objective as they can perceive as their first choice, regardless of their further choices. Sorting behavior contrasts to truth-telling: it targets directly the matching objective a player think as the best under his/her ability or other external constraints (i.e., vacancies), instead of listing all his/her unconstrained matching objectives by his/her preference order.

In our lab experimental including both truth-telling mechanisms (i.e., SD mechanisms) and non-truth-telling mechanisms (i.e., Boston mechanisms), as well as with different information, we show that: First, sorting behavior prevalent under all mechanisms, and more so under non-strategy-proof mechanisms. Second, at individual level, sorting behavior performs as well as truth-telling behavior, under all mechanisms.

Third, at the society level, more sorting behavior under Boston mechanism benefit the society, while under SD mechanism, its influence is negligible.

The results explain why non-strategy-proof mechanisms are frequently used and why people frequently play non-truth-telling behavior in the reality. In particular, if a mechanism designer keeps in his mind that he can promote sorting behavior in the would-be designed non-strategy-proof mechanism, he might be confident that he can achieve a social welfare level as high as under a truth-telling mechanism. In fact, Chinese college admission system has been working under a non-truth-telling mechanism (i.e., Boston mechanism) for many years. High school teachers, parents, consulting firms, and even colleges have try their best to provide information or guidance to students, to help them figure out their "sorting" colleges. In dating service companies, client information is collected and analyzed so that each client is provided with a very limited list of dating candidates. The service is essentially one to help players to form a good sorting strategy. Another example would be that a PhD student may get advice or recommendation from his/her advisor to figure out his/her sorting employee(s) as he/she applies for a position in the job market.

We have already show that promoting behaviors can help individual or the society to achieve higher welfare. One remained question is how to do that. Information provision plays an important role, as we discussed above. Sorting behavior can also be induced by careful mechanism design. For example, constrained school choice mechanisms may help to promote sorting behavior, as implied in Haeringer and Klijn (2009) and Lien, et al., (2017). In general, it deserves more research to either examine the welfare consequence of sorting behavior or to design mechanisms to influence agents' choice of it.

References

Abdulkadiroğlu, A., Che, Y. K., Yasuda, Y., 2011. Resolving conflicting preferences in school choice: The "Boston mechanism" reconsidered. Amer. Econ. Rev. 101(1), 399-410.

Abdulkadiroğlu, A., Sönmez, T., 2003. School choice: A mechanism design approach. Amer. Econ. Rev. 93(3), 729-747.

Ashlagi, I., Gonczarowski, Y. A., 2017. Stable matching mechanisms are not obviously strategy-proof. Manuscript.

Calsamiglia, C., Haeringer G., and Klijn F., 2010. Constrained school choice: An experimental study. Amer. Econ. Rev. 100(1), 1860-1874.

Chen, Y., Jiang, M., Kesten, O., 2015. Chinese College Admissions reforms: experimental and empirical evaluations. Manuscript

Chen, Y., Kesten, O., 2017. Chinese college admissions and school choice reforms: A theoretical analysis. J. Polit. Economy. 125(1), 99-139.

Chen, Y., Sönmez, T., 2006. School choice: an experimental study. J. Econ. Theory. 127(1), 202-231.

Edril, A., Ergin, H., 2008. What's the matter with tie-breaking? Improving efficiency in school choice. Amer. Econ. Rev. 98(3), 669-689.

Ergin, H., Sönmez, T., 2006. Games of school choice under the Boston mechanism. J. Public Econ. 90(1), 215-237.

Featherstone, C., Niederle, M., 2014. Improving on strategy-proof school choice mechanisms: An experimental investigation. Unpublished paper, Stanford University.

Haeringer, G., Klijn, F., 2009. Constrained school choice. J. Econ. Theory. 144(5), 1921-1947.

Klijn, F., Pais, J., Vorsatz, M., 2013. Preference intensities and risk aversion in school choice: A laboratory experiment. Exper. Econ. 16(1), 1-22.

Li, S., 2017. Obviously strategy-proof mechanisms, Amer. Econ. Rev. 107(11): 3257-3287.

Lien, J. W., Zheng, J., Zhong, X., 2016. Preference submission timing in school choice matching: testing fairness and efficiency in the laboratory. Exper. Econ. 19(1), 116-150.

Lien, J. W., Zheng, J., Zhong, X., 2017. Ex-ante fairness in the Boston and Serial Dictatorship mechanisms under pre-exam and post-exam preference submission. Games Econ. Behav. 101, 98-120.

Pais, J., Pintér, Á., 2008. School choice and information: An experimental study on matching mechanisms. Games Econ. Behav. 64(1), 303-328.

Pan, S., 2016. The instability of matching with overconfident agents: laboratory and field investigations. Manuscript.

Roth, A.E., Sotomayor, M.A.O., Two Sided Matching: A Study in Game-Theoretic Modelling and Analysis, New York: Cambridge University Press, 1990.

Tanaka, T., Camerer, C. F., Nguyen, Q., 2010. Risk and time preferences: linking experimental and household survey data from Vietnam. Amer. Econ. Rev. 100(1), 557-571.

Troyan, P., 2016. Obviously Strategy-Proof Implementation of Top Trading Cycles. Manuscript.

Wu, B., Zhong, X., 2014. Matching mechanisms and matching quality: Evidence from a top university in China. Games Econ. Behav. 84, 196-215.

Wu, B., Zhong, X., 2017. Fairness of the Boston mechanism in China's centralized college admissions. Working paper.

Appendix A: Simulation Methods for Evaluating Matching Outcomes

We first describe our simulation process in general, and then we will describe five scenarios we use to simulate preference submission behaviors.

The simulation steps are as follows.

Step1. Randomly draw all the students' realized score rankings from the given score distribution.

Step2. Simulate students' preference submission based on each ex-ante/realized score distribution under the incomplete/complete-information mechanism, through each of the four scenarios we will describe below.

Step3. Match according to the realized scores and simulated preference submission behavior of all students under the Boston or SD mechanism.

Step4. Simulate step 1-3 200 times.

We now describe the four scenarios we propose to draw from the observed actions of students and capture behavioral patterns.

Scenario 1

For the incomplete information treatments, the matching is done according to each realized score ranking, each time under the same observed preference submission behavior of each student.

For the complete information treatments, we assume a student's behavior is only determined by his realized ranking, and randomly choose from one of the two sessions for each treatment the preference submission behavior (for all seven choices) for each score ranking.

Scenario 2

We use the distribution of truth-telling and up/down/fair reporting strategies for the first choice for simulating student choices. In particular, for each treatment, we calculate the proportion of truth-telling and the distribution of non-truth-telling reporting strategies by looking at the degree of misreport (w.r.t. ex-ante/realized scores for incomplete/complete-information treatments). We then assign students to be truth-telling according to the proportion of truth-telling for each treatment. For non-truth-telling students, we assign students to up/fair/down-reporting the first choice according to the distributions of non-truth-telling reporting strategies found in the data. We randomly assign the 2nd-7th choices for each non-truth-telling student.

Scenario 2'

We use the distribution of truth-telling, sorting behavior and other behavior patterns revealed in Table 1. In particular, for each treatment, we calculate the proportion of truth-telling, and the distribution of sorting and other non-truth-telling strategies by looking at students' first choice. We distinguish top 6 students and below top 6 students. We then assign students to be a specific behavior pattern according to the proportion of that pattern of behavior for each treatment found in the data (i.e., Table 1). We randomly assign the 2nd-7th choices for each non-truth-telling student.

Scenario 3

Scenario 3 further simulates the non-truth-telling students' second choice after simulating the first choice as in Scenario 2. We assign second choices to the students in each treatment to match the distribution of the degree of misreporting, that is, the gap between the preference ranking of the second choice and the calculated ex-ante/ex-post fair school according to the ex-ante/ex-post realized scores. If a student's simulated second choice is the same with the simulated first choice, then we re-generate the second choice randomly according to the distribution until they are different. The 3rd-7th choices for each student are randomly assigned.

Scenario 4

Scenario 4 simulates the non-truth-telling students' second choice by considering its relation with the first choice after simulating the first choice as in Scenario 2. In particular, the second choices are assigned to match the distribution of the gap between the preference ranking of the first and second choice in the data. The 3rd-7th choices for each student are randomly assigned.

Table and Figure Appendix

Table A1: Determinants of Truth-telling: Logit Model

Independent Variable	Dependent Variable: Truth-telling		
	(1)	(2)	(3)
Boston	-0.5833***	-0.5823***	-0.5949***
(BOS_C-SD_C)	(0.0646)	(0.0646)	(0.0626)
Incomplete	-0.0463	-0.03539	-0.0475
(SD_I-SD_C)	(0.0696)	(0.0730)	(0.0720)
Boston*Incomplete	-0.05556	-0.0629	-0.0471
(BOS_I-BOS_C)-(SD_I-SD_C)	(0.0792)	(0.0805)	(0.0793)
Rank	-0.00344	-0.00336	-0.00263
	(0.00237)	(0.00239)	(0.00240)
Fair School Slots	-0.0181	-0.0191	-0.0241
	(0.0156)	(0.0158)	(0.0162)
Female		-0.000912	0.00214
		(0.0405)	(0.0405)
Age		-0.00154	0.000496
		(0.0147)	(0.0145)
Econ		0.0153	0.0245
		(0.0550)	(0.0548)
Engineer		0.0350	0.0506
		(0.0572)	(0.0574)
Science		-0.00267	-0.00205
		(0.0734)	(0.0735)
σ			0.208***
			(0.0763)
α			-0.0926
			(0.0810)
λ			0.00997
			(0.0124)
Observations	360	360	360
Pseudo R-squared	0.4022	0.4034	0.4186
BOS_I-BOS_C	-0.1019***	-0.0982***	-0.0946***
	(0.0377)	(0.0374)	(0.0362)
SD_I-BOS_C	0.5370***	0.5469***	0.5475***
	(0.0579)	(0.0594)	(0.0583)

Note: *** p<0.01, ** p<0.05, * p<0.1. Coefficients report average marginal effects. λ takes the average value of its lower and upper bound. We also run regressions with the lower bound and upper bound of λ, and the results are similar.

Table A2: Determinants of Sorting: Logit Model

Independent Variable	Dependent Variable: Sorting		
	(1)	(2)	(3)
Boston	0.417***	0.421***	0.430***
(BOS_C-SD_C)	(0.0751)	(0.0748)	(0.0741)
Incomplete	-0.0602	-0.0557	-0.0496
(SD_I-SD_C)	(0.0620)	(0.0622)	(0.0623)
Boston*Incomplete	0.0370	0.0266	0.0116
(BOS_I-BOS_C)-(SD_I-SD_C)	(0.0956)	(0.0957)	(0.0957)
Rank	-0.00225	-0.00209	-0.00205
	(0.00299)	(0.00302)	(0.00302)
Fair School Slots	0.0133	0.0134	0.0111
	(0.0194)	(0.0194)	(0.0197)
Female		0.00702	0.0105
		(0.0510)	(0.0515)
Age		-0.0122	-0.0126
		(0.0171)	(0.0171)
Econ		-0.0213	-0.00773
		(0.0925)	(0.0941)
Engineer		-0.0391	-0.0402
		(0.0951)	(0.0963)
Science		-0.106	-0.108
		(0.113)	(0.115)
σ			-0.111
			(0.101)
α			0.154
			(0.104)
λ			-0.00967
			(0.0155)
Observations	360	360	360
Pseudo R-squared	0.155	0.159	0.164
BOS_I-BOS_C	-0.0231	-0.0291	-0.0380
	(0.0728)	(0.0731)	(0.0728)
SD_I-BOS_C	0.477***	0.477***	0.480***
	(0.0670)	(0.0674)	(0.0670)

Note: *** p<0.01, ** p<0.05, * p<0.1. Coefficients report average marginal effects. λ takes the average value of its lower and upper bound. We also run regressions with the lower bound and upper bound of λ, and the results are similar.

Table A3: Fairness Consequences of Different Behaviors

Mechanism	Report Strategy	# of Obs.	Mean	SD	Min	Max
	Panel A: Summary statistics of degree of mismatch					
BOS_I	Truth-telling	1	-0.86	.	-0.86	-0.86
	Sorting	68	-0.00044	0.5	-0.93	2
	Other non-truth-telling	39	-0.69	1	-3	2
BOS_C	Truth-telling	8	0.13	0.83	-1	2
	Sorting	47	0.085	0.46	-1	2
	Other non-truth-telling	17	-1.1	1	-3	1
SD_I	Truth-telling	70	-0.011	0.58	-0.95	1.6
	Sorting	19	-0.024	0.56	-0.84	1.9
	Other non-truth-telling	19	-0.61	0.83	-2	0.74
SD_C	Truth-telling	50	0.02	0.14	0	1
	Sorting	17	0.12	0.49	0	2
	Other non-truth-telling	5	-1	1	-2	0
	Panel B: Wilcoxon Rank-Sum Test of differences in degree of mismatch					
BOS_I	Sorting >Other≈ Truth-telling (p=0.0001) (p=0.8960)					
BOS_C	Truth-telling ≈ Sorting> Other (p=0.7743) (p=0.0000)					
SD_I	Truth-telling ≈ Sorting> Other (p=0.9680) (p=0.0411)					
SD_C	Sorting >Other>Truth-telling (p=0.0020) (p=0.0000)					
	Panel C: F test of variance of degree of mismatch					
BOS_I	Other> Sorting (p=0.0000)					
BOS_C	Other ≈ Truth-telling > Sorting (p=0.5923) (p=0.0061)					
SD_I	Other >Truth-telling ≈ Sorting (p=0.0157) (p=0. 9485)					
SD_C	Other ≈ Sorting >Truth-telling (p=0.156) (p=0.0000)					

Note: Degree of mismatch for BOS_I and SD_I is the average value after 200 simulations of score distribution.

Table A4: Efficiency Consequences of Different Behaviors

Mechanism	Report Strategy	# of Obs.	Mean	SD	Min	Max
	Panel A: Summary statistics of relative payoff					
BOS_I	Truth-telling	1	0.866	.	0.866	0.866
	Sorting	68	1.011	0.119	0.855	1.571
	Other non-truth-telling	39	0.882	0.199	0.455	1.455
BOS_C	Truth-telling	8	1.038	0.177	0.846	1.455
	Sorting	47	1.025	0.124	0.818	1.571
	Other non-truth-telling	17	0.809	0.182	0.556	1.182
SD_I	Truth-telling	70	1.013	0.124	0.826	1.451
	Sorting	19	1.01	0.141	0.869	1.554
	Other non-truth-telling	19	0.89	0.156	0.628	1.168
SD_C	Truth-telling	50	1.004	0.031	1	1.222
	Sorting	17	1.034	0.139	1	1.571
	Other non-truth-telling	5	0.818	0.182	0.636	1
	Panel B: Wilcoxon Rank-Sum Test of differences in relative payoff					
BOS_I	Sorting >Other ≈Truth-telling (p=0.0001) (p=0.8960)					
BOS_C	Truth-telling≈ Sorting>Other (p=0.7575) (p=0.0000)					
SD_I	Truth-telling ≈ Sorting>Other (p=0.7903) (p=0.0383)					
SD_C	Sorting ≈Truth-telling>Other (p=0.4061) (p=0.0000)					
	Panel C: F test of variance of relative payoff					
BOS_I	Other >Sorting (p=0.0001)					
BOS_C	Other ≈ Truth-telling > Sorting (p=0.9990) (p=0.0701)					
SD_I	Other ≈ Sorting ≈ Truth-telling (p=0.6743) (p=0.4444)					
SD_C	Other ≈ Sorting >Truth-telling (p=0.3891) (p=0.0000)					

Note: Relative payoff for BOS_I and SD_I is the average value after 200 simulations of score distribution.

Table A5: Social Welfare Comparison Among Four Mechanisms

Welfare Measures	Panel A: Mean of Welfare Measures			
	BOS_I	BOS_C	SD_I	SD_C
Efficiency (by payoff per capita)	10.046	10.551	10.664	11.105
Ex-post fairness (by degree of mismatch)	1.032	0.642	0.551	0.266
Ex-post fairness (by number of blocking pairs)	0.948	0.651	0.349	0.145
Ex-ante fairness (by degree of mismatch)	0.963	0.876	0.749	0.598
Ex-ante fairness (by number of blocking pairs)	0.973	0.849	0.589	0.416
	Panel B: Wilcoxon Rank-Sum Test Result			
Efficiency (by payoff per capita)	BOS_I < BOS_C < SD_I < SD_C (p=0.0000)(p=0.0000)(p=0.0000)			
Ex-post fairness (by degree of mismatch)	BOS_I > BOS_C > SD_I > SD_C (p=0.0000)(p=0.0000)(p=0.0000)			
Ex-post fairness (by number of blocking pairs)	BOS_I > BOS_C > SD_I > SD_C (p=0.0000)(p=0.0000)(p=0.0000)			
Ex-ante fairness (by degree of mismatch)	BOS_I > BOS_C > SD_I > SD_C (p=0.0000)(p=0.0000)(p=0.0000)			
Ex-ante fairness (by number of blocking pairs)	BOS_I > BOS_C > SD_I > SD_C (p=0.0000)(p=0.0000)(p=0.0000)			

Table A6: The Effects of Changing Sorting Behavior under Boston Mechanisms

	Origin	6%	-6%	12%	-12%	100% sorting	100% truth-telling
	（1）	（2）	（3）	（4）	（5）	（6）	（7）
BOS_I							
Payoff	10.046	10.249	9.897	10.429	9.74	11.431	11.611
Ex-post (un)fairness-1	1.032	0.949	1.09	0.869	1.165	0.435	0.157
Ex-post (un)fairness-2	0.948	0.894	1.001	0.84	1.066	0.638	0.071
Ex-ante (un)fairness-1	0.963	0.873	1.036	0.791	1.114	0.293	0.424
Ex-ante (un)fairness-2	0.973	0.898	1.022	0.825	1.103	0.541	0.262
Efficiency	All truth > all sorting > 12% > 6% > Origin > -6% > -12%						
	(0.000) (0.000) (0.000) (0.000) (0.000) (0.000)						
Ex-post (un)fairness-1	-12% > -6% > Origin > 6% > 12% > all sorting > all truth						
	(0.000) (0.000) (0.000) (0.000) (0.000) (0.000)						
Ex-post (un)fairness-2	-12% > -6% > Origin > 6% > 12% > all sorting > all truth						
	(0.0001) (0.0008) (0.0004) (0.0022) (0.000) (0.000)						
Ex-ante (un)fairness-1	-12% > -6% > Origin > 6% > 12% > all truth > all sorting						
	(0.000) (0.000) (0.000) (0.000) (0.000) (0.000)						
Ex-ante (un)fairness-2	-12% > -6% > Origin > 6% > 12% > all sorting > all truth						
	(0.000) (0.0033) (0.0001) (0.0001) (0.000) (0.000)						
BOS_C							
Efficiency	10.551	10.713	10.336	10.893	10.11	11.509	11.61
Ex-post (un)fairness-1	0.642	0.557	0.744	0.478	0.849	0.169	0.16
Ex-post (un)fairness-2	0.651	0.587	0.753	0.519	0.86	0.42	0.073
Ex-ante (un)fairness-1	0.876	0.797	0.969	0.722	1.065	0.468	0.417
Ex-ante (un)fairness-2	0.849	0.782	0.944	0.709	1.047	0.568	0.253
Efficiency	All truth > all sorting > 12% > 6% > Origin > -6% > -12%						
	(0.000) (0.000) (0.000) (0.000) (0.000) (0.000)						
Ex-post (un)fairness-1	-12% > -6% > Origin > 6% > 12% > all sorting ≈ all truth						
	(0.000) (0.000) (0.000) (0.000) (0.000) (0.3045)						
Ex-post (un)fairness-2	-12% > -6% > Origin > 6% > 12% > all sorting > all truth						
	(0.000) (0.000) (0.0015) (0.001) (0.000) (0.000)						
Ex-ante (un)fairness-1	-12% > -6% > Origin > 6% > 12% > all sorting > all truth						
	(0.000) (0.000) (0.000) (0.000) (0.000) (0.000)						
Ex-ante (un)fairness-2	-12% > -6% > Origin > 6% > 12% > all sorting > all truth						
	(0.000) (0.000) (0.0007) (0.0008) (0.000) (0.000)						

Note: Efficiency is measure by payoff per capita. Ex-ante or ex-post (un)fairness is measured by degree of mismatch (1) or the number of blocking pairs (2). Wilcoxon Rank-Sum Test is used to rank matching outcome under different behavior mixtures for each welfare measure.

Table A7: The Effects of Changing Sorting Behavior under SD Mechanisms

	Origin	6%	-6%	12%	-12%	100% sorting	100% Truth-telling
	（1）	（2）	（3）	（4）	（5）	（6）	（7）
SD_I							
Efficiency	10.664	10.636	10.691	10.643	10.657	11.171	11.528
Ex-post (un)fairness-1	0.551	0.583	0.541	0.585	0.53	0.511	0
Ex-post (un)fairness-2	0.349	0.364	0.336	0.374	0.333	0.742	0
Ex-ante (un)fairness-1	0.749	0.742	0.735	0.733	0.765	0.409	0.407
Ex-ante (un)fairness-2	0.589	0.586	0.576	0.597	0.594	0.667	0.258
Efficiency	All truth > all sorting > -6% ≈ Origin ≈ -12% ≈ +12% ≈ +6%						
		(0.000) (0.000)		(0.323) (0.611) (0.604) (0.867)			
Ex-post (un)fairness-1	12% ≈ 6% > Origin ≈ -6% ≈ -12% ≈ all sorting > all truth						
		(0.832) (0.017) (0.254) (0.522) (0.1225) (0.000)					
Ex-post (un)fairness-2	all sorting > 12% ≈ 6% > Origin ≈ -6% ≈ -12% > all truth						
		(0.000) (0.339) (0.090) (0.104) (0.939)(0.000)					
Ex-ante (un)fairness-1	-12% ≈ Origin ≈ 6% ≈ -6% ≈ 12% > all sorting ≈ all truth						
		(0.167) (0.659) (0.456) (0.755) (0.000)(0.306)					
Ex-ante (un)fairness-2	all sorting > 12% ≈ -12% ≈ Origin ≈ 6% ≈ -6% > all truth						
		(0.000) (0.617) (0.826) (0.586) (0.490)(0.000)					
SD_C							
Efficiency	11.105	11.115	11.096	11.091	11.153	11.318	11.532
Ex-post (un)fairness-1	0.266	0.261	0.269	0.27	0.239	0.231	0
Ex-post (un)fairness-2	0.145	0.144	0.148	0.152	0.129	0.51	0
Ex-ante (un)fairness-1	0.598	0.581	0.596	0.603	0.577	0.557	0.398
Ex-ante (un)fairness-2	0.416	0.411	0.429	0.428	0.392	0.662	0.254
Efficiency	All truth > all sorting > -12% ≈ 6% ≈ Origin ≈ -6% ≈ 12%						
		(0.000)		(0.000) (0.206) (0.503) (0.688) (0.928)			
Ex-post (un)fairness-1	12% ≈ -6% ≈ Origin ≈ 6% ≈ -12% ≈ all sorting > all truth						
		(0.887) (0.648) (0.718) (0.111) (0.294) (0.000)					
Ex-post (un)fairness-2	all sorting > 12% ≈ -6% ≈ Origin ≈ 6% > -12% > all truth						
		(0.000) (0.606) (0.508) (0.911) (0.078)(0.000)					
Ex-ante (un)fairness-1	12% ≈ Origin ≈ -6% ≈ 6% ≈ -12% ≈ all sorting > all truth						
		(0.899) (0.733) (0.164) (0.879) (0.121) (0.000)					
Ex-ante (un)fairness-2	all sorting > -6% ≈ 12% ≈ Origin ≈ 6% ≈ -12% > all truth						
		(0.000) (0.859) (0.286) (0.634) (0.151)(0.000)					

Note: Efficiency is measure by payoff per capita. Ex-ante or ex-post (un)fairness is measured by degree of mismatch (1) or the number of blocking pairs (2). Wilcoxon Rank-Sum Test is used to rank matching outcome under different behavior mixtures for each welfare measure.

Table A8: Determinants of Fair-Reporting: Logit Model

Independent Variable	Dependent Variable: Fair-Reporting		
	(1)	(2)	(3)
Boston	0.431***	0.439***	0.445***
(BOS_C-SD_C)	(0.0735)	(0.0730)	(0.0727)
Incomplete	-0.0278	-0.0280	-0.0258
(SD_I-SD_C)	(0.0587)	(0.0587)	(0.0589)
Boston*Incomplete	0.00463	-0.00726	-0.0174
(BOS_I-BOS_C)-(SD_I-SD_C)	(0.0942)	(0.0939)	(0.0941)
Rank	-0.00311	-0.00288	-0.00282
	(0.00296)	(0.00299)	(0.00300)
Fair School Slots	-0.000301	-0.00129	-0.00381
	(0.0192)	(0.0193)	(0.0195)
Female		0.0145	0.0190
		(0.0506)	(0.0510)
Age		0.00137	0.00154
		(0.0169)	(0.0169)
Econ		-0.0731	-0.0626
		(0.0903)	(0.0916)
Engineer		-0.108	-0.112
		(0.0928)	(0.0939)
Science		-0.0992	-0.106
		(0.111)	(0.112)
σ			-0.0631
			(0.0999)
α			0.134
			(0.104)
λ			-0.00981
			(0.0155)
Observations	360	360	360
Pseudo R-squared	0.158	0.162	0.166
BOS_I-BOS_C	-0.0231	-0.0353	-0.0432
	(0.0737)	(0.0737)	(0.0736)
SD_I-BOS_C	0.458***	0.467***	0.471***
	(0.0672)	(0.0669)	(0.0667)

Note: *** p<0.01, ** p<0.05, * p<0.1. Coefficients report average marginal effects. λ takes the average value of its lower and upper bound. We also run regressions with the lower bound and upper bound of λ, and the results are similar.

Table A9: Determinants of Misreport within Non-Truth-Telling

Independent Variable	Dependent Variable: Degree of Misreport					
	OLS			Ordered Probit		
	(1)	(2)	(3)	(4)	(5)	(6)
Boston	-0.0720	-0.0706	-0.153	-0.131	-0.130	-0.220
(BOS_C-SD_C)	(0.250)	(0.248)	(0.250)	(0.274)	(0.274)	(0.278)
Incomplete	0.116	0.128	0.0985	0.0354	0.0553	0.0114
(SD_I-SD_C)	(0.271)	(0.269)	(0.279)	(0.293)	(0.294)	(0.307)
Boston*Incomplete	0.0126	0.0115	0.101	0.0941	0.0864	0.176
(*BOS_I-BOS_C)-(SD_I-SD_C)*	(0.315)	(0.312)	(0.317)	(0.341)	(0.342)	(0.349)
Rank	-0.0171**	-0.0272***	-0.0264***	-0.0202***	-0.0331***	-0.0329***
	(0.00678)	(0.00838)	(0.00855)	(0.00754)	(0.00943)	(0.00968)
Fair School Slots		0.112**	0.104*		0.140**	0.134**
		(0.0550)	(0.0556)		(0.0608)	(0.0620)
Female			-0.0949			-0.0808
			(0.144)			(0.158)
Age			0.00718			0.00723
			(0.0480)			(0.0530)
Econ			0.425**			0.427**
			(0.189)			(0.210)
Engineer			0.326			0.330
			(0.198)			(0.219)
Science			0.0645			-0.0626
			(0.291)			(0.318)
σ			0.0227			0.103
			(0.300)			(0.330)
α			0.0209			-0.0181
			(0.321)			(0.353)
λ			-0.0663			-0.0737
			(0.0468)			(0.0514)
Observations	231	231	231	231	231	231
(Pseudo)R-squared	0.032	0.050	0.091	0.0141	0.0233	0.0394
BOS_I-BOS_C	0.1282	0.1394	0.1997	0.1295	0.1417	0.1870
	(0.1600)	(0.1590)	(0.1670)	(0.1743)	(0.1748)	(0.1852)
BOS_C-SD_I	-0.1875	-0.1985	-0.2513	-0.1665	-0.1854	-0.2317
	(0.2070)	(0.2056)	(0.2178)	(0.2231)	(0.2238)	(0.2395)

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses. λ is the average value of its lower and upper bound. We also run regression with the lower bound and upper bound of λ, and the results are similar.

Table A10: OLS Regression of the Degree of Mismatch on Misreport

Independent Variable	Dependent Variable							
	Degree of Mismatch				Variance of Degree of Mismatch			
	BOS_I	BOS_C	SD_I	SD_C	BOS_I	BOS_C	SD_I	SD_C
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Truth-telling	-0.416	0.146	0.246*	0.0133	0.423	0.585**	0.211	0.0127
	(0.671)	(0.265)	(0.139)	(0.0841)	(0.703)	(0.240)	(0.139)	(0.0557)
Up-report	0.278*	0.341	0.397**	0.390**	0.894***	0.806***	0.421**	0.379***
	(0.152)	(0.242)	(0.196)	(0.157)	(0.159)	(0.219)	(0.196)	(0.104)
Down-report	-1.351***	-1.446***	-1.105***	-1.798***	0.118	0.0812	-0.0652	0.112
	(0.179)	(0.206)	(0.217)	(0.194)	(0.188)	(0.187)	(0.217)	(0.129)
Rank	0.00111	0.0228**	-0.0191***	0.00915**	0.0230***	0.00731	0.0260***	0.00651**
	(0.00824)	(0.00944)	(0.00680)	(0.00424)	(0.00864)	(0.00856)	(0.00681)	(0.00281)
Fair School	0.0899*	-0.0964	0.194***	-0.0343	-0.109**	0.0683	-0.109**	-0.0228
Slots	(0.0515)	(0.0634)	(0.0420)	(0.0288)	(0.0540)	(0.0574)	(0.0420)	(0.0191)
Female	-0.300**	0.283*	-0.143	0.0826	0.136	-0.304**	-0.0881	0.0792
	(0.130)	(0.160)	(0.101)	(0.0795)	(0.136)	(0.145)	(0.101)	(0.0527)
Age	0.0515	0.0516	0.0118	-0.0439*	-0.0991**	0.00231	0.00644	-0.0395**
	(0.0463)	(0.0457)	(0.0417)	(0.0242)	(0.0485)	(0.0414)	(0.0417)	(0.0160)
Econ	0.225	0.195	0.0157	0.0600	-0.176	-0.165	-0.0611	0.0228
	(0.156)		(0.135)	(0.114)	(0.164)	(0.289)	(0.135)	(0.0756)
Engineer	0.00339	-0.00320	0.0549	0.194	0.282	-0.185	-0.0122	0.141
	(0.163)	(0.308)	(0.141)	(0.128)	(0.171)	(0.279)	(0.141)	(0.0850)
Science	0.660*	0.100	0.207	0.112	0.0133	-0.235	-0.207	0.0734
	(0.376)	(0.350)	(0.200)	(0.150)	(0.394)	(0.317)	(0.200)	(0.0991)
σ	-0.423	0.221	-0.328	-0.00749	0.0407	0.333	-0.147	0.00232
	(0.269)	(0.367)	(0.200)	(0.141)	(0.282)	(0.333)	(0.201)	(0.0936)
α	0.667**	0.737**	0.0102	-0.140	-0.121	-0.459	-0.0838	-0.0915
	(0.292)	(0.354)	(0.218)	(0.145)	(0.305)	(0.321)	(0.218)	(0.0963)
λ	0.0323	-0.0284	-0.0580	0.00658	-0.0287	0.0565	0.00229	0.00529
	(0.0389)	(0.0548)	(0.0361)	(0.0216)	(0.0408)	(0.0497)	(0.0361)	(0.0143)
Constant	-1.935*	-1.775*	-0.826	0.861	2.357**	-0.218	0.295	0.762**
	(1.071)	(0.985)	(0.928)	(0.523)	(1.122)	(0.892)	(0.929)	(0.347)
Observations	108	72	108	72	108	72	108	72
R-squared	0.498	0.641	0.516	0.713	0.371	0.339	0.199	0.418

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses. λ is the average value of its lower and upper bound. We also run regression with the lower bound and upper bound of λ, and the results are similar. Degree of mismatch for BOS_I and SD_I is the average value after 200 simulations of score distribution.

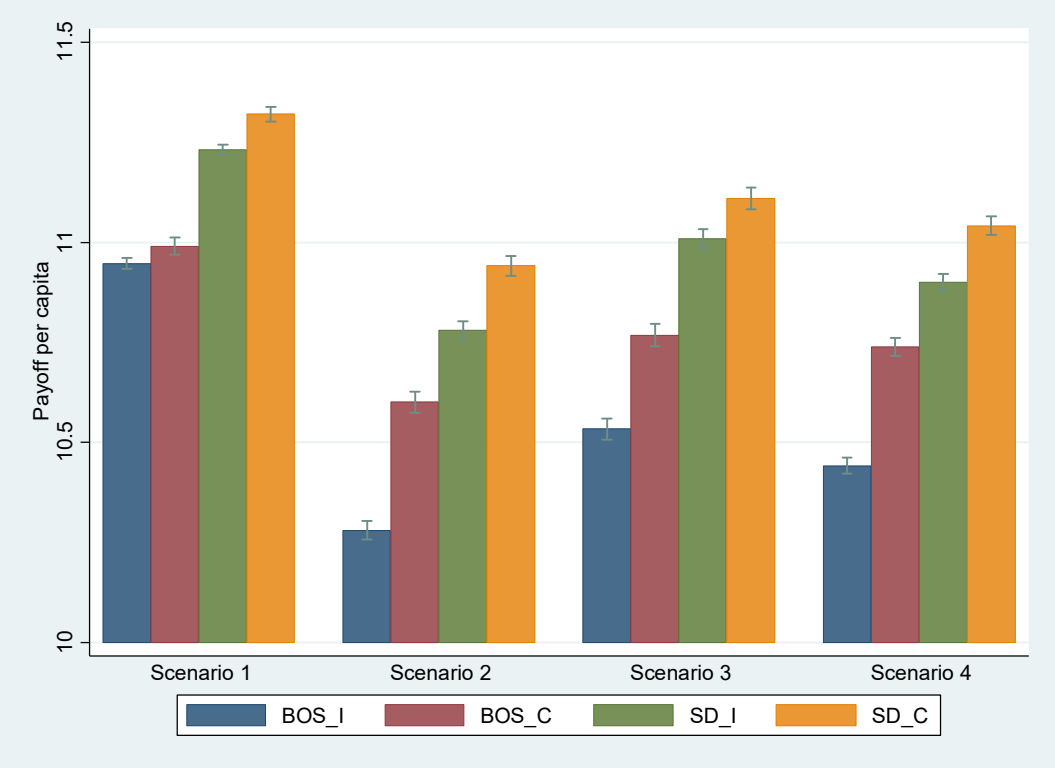


Figure A1: Efficiency result under different scenarios

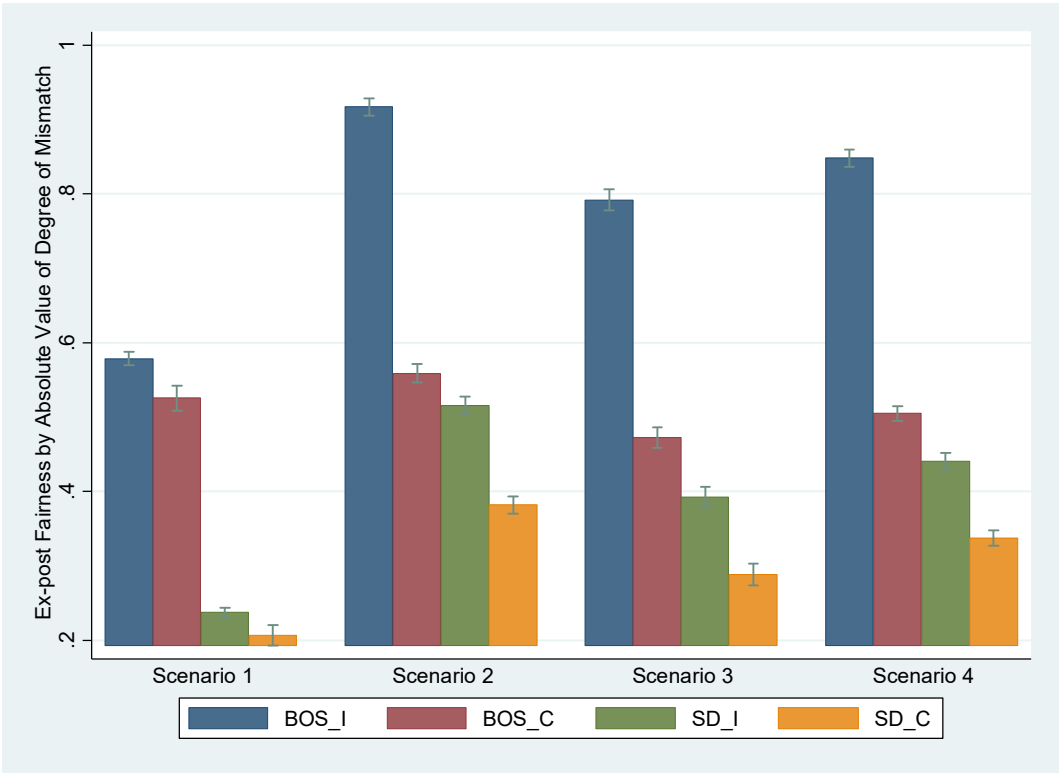


Figure A2: Ex-post fairness (by absolute value of degree of mismatch)

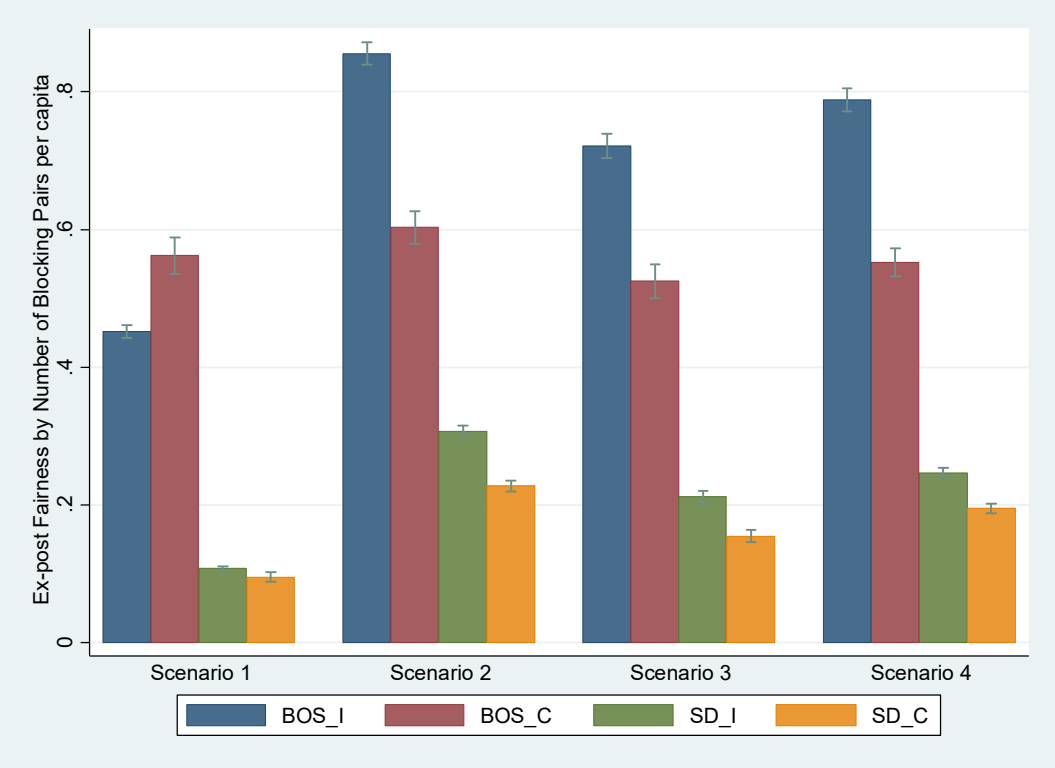


Figure A3: Ex-post fairness (by number of blocking pairs)

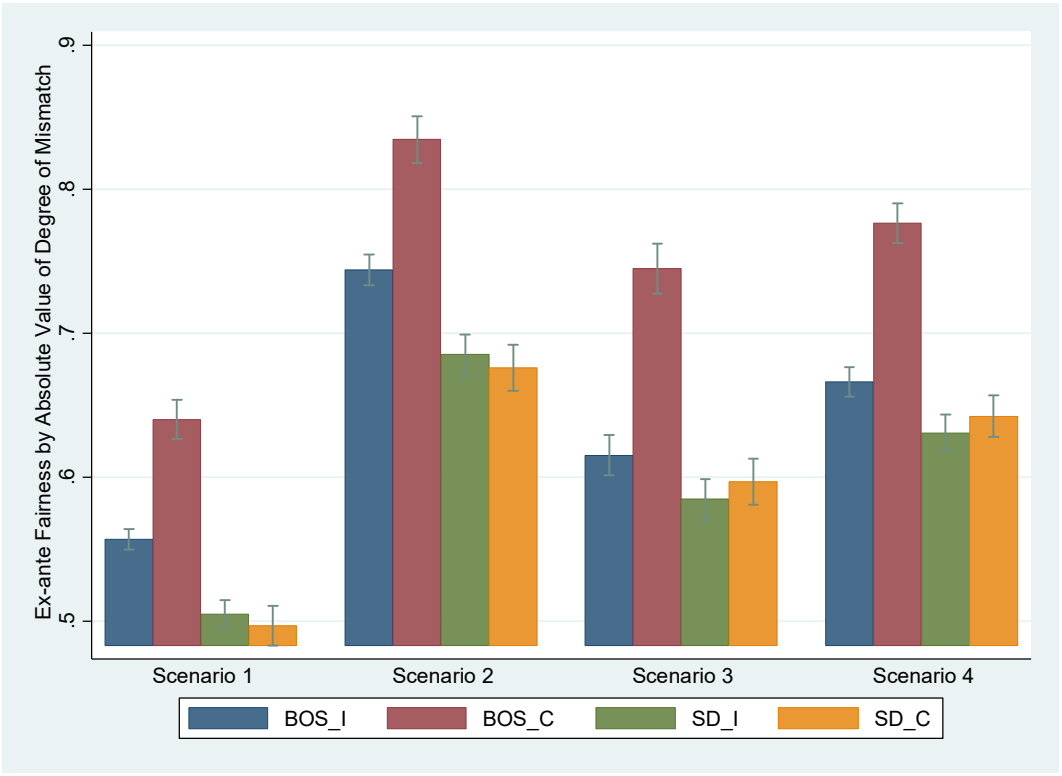


Figure A4: Ex-ante fairness (by absolute value of degree of mismatch)

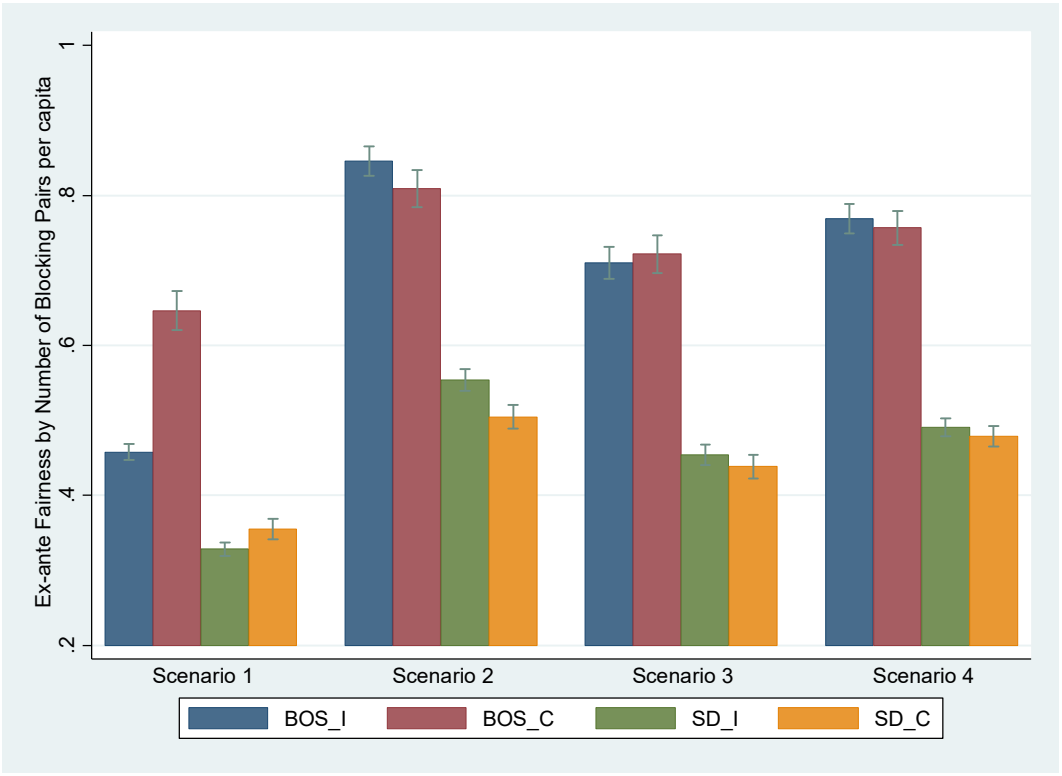


Figure A5: Ex-ante fairness (by number of blocking pairs)

通信地址：

北京　清华大学

中国经济研究中心

电话：86-10-62789695　　传真：86-10-62789697

邮编：100084

网址：http://www.ncer.tsinghua.edu.cn

E-mail: ncer@em.tsinghua.edu.cn

Adress:

National Center for Economic Research

Tsinghua University

Beijing 100084

China

Tel: 86-10-62789695　　Fax: 86-10-62789697

Web site: http://www.ncer.tsinghua.edu.cn

E-mail: ncer@em.tsinghua.edu.cn